Algebraic & Geometric Topology

Volume 24 (2024)

Issue 7 (pages 3571–4137)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

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

The subscription price for 2024 is US $705/year for the electronic version, and $1040/year (+$70, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

AGT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

Algebraic & Geometric Topology **24**:7 (2024) 3571–3604
DOI: 10.2140/agt.2024.24.3571
Published: 9 December 2024

Geography of bilinearized Legendrian contact homology

FRÉDÉRIC BOURGEOIS
DAMIEN GALANT

We study the geography of bilinearized Legendrian contact homology for closed connected Legendrian submanifolds with vanishing Maslov class in 1–jet spaces. We show that this invariant detects whether the two augmentations used to define it are DGA homotopic or not. We describe a collection of graded vector spaces containing all possible values for bilinearized Legendrian contact homology and then show that all these vector spaces can be realized.

53D42, 57R17

1 Introduction

Let Λ be a closed Legendrian submanifold of the 1–jet space $J^1(M)$ of a manifold M. Given a generic complex structure for the canonical contact structure on $J^1(M)$, one can associate to Λ its Chekanov–Eliashberg differential graded algebra $(\mathcal{A}(\Lambda), \partial)$; see Chekanov [3] and Ekholm, Etnyre and Sullivan [7; 9]. The homology of $(\mathcal{A}(\Lambda), \partial)$, called Legendrian contact homology, is an invariant of the Legendrian isotopy class of Λ, but it is often hard to compute. It is therefore useful to consider augmentations of $(\mathcal{A}(\Lambda), \partial)$, because such an augmentation ε can be used to define a linearized complex $(C(\Lambda), \partial^\varepsilon)$. The homology is denoted by $\mathrm{LCH}^\varepsilon(\Lambda)$ and called the linearized Legendrian contact homology of Λ with respect to ε. The collection of these homologies for all augmentations of $(\mathcal{A}(\Lambda), \partial)$ is also an invariant of the Legendrian isotopy class of Λ. The geography (i.e. the determination of all possible values) of a similar homological invariant defined using generating families was described by the first author with Sabloff and Traynor [2]. Using the work of Dimitroglou Rizell [4] on the effect of embedded surgeries on Legendrian contact homology, this geography can be shown to hold for linearized Legendrian contact homology as well. On the other hand, the first author and Chantraine [1] showed that it is possible to use two augmentations ε_1 and ε_2 of the Chekanov–Eliashberg DGA to define a bilinearized differential $\partial^{\varepsilon_1,\varepsilon_2}$ on $C(\Lambda)$. The corresponding homology is called bilinearized Legendrian contact homology and is denoted by $\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}(\Lambda)$. Our object is to describe the geography of bilinearized Legendrian contact homology. In other words, our goal is to describe a collection of Legendrian submanifolds equipped with two augmentations such that their bilinearized Legendrian contact homologies realize all possible values for this invariant.

When $\varepsilon_1 = \varepsilon_2$, bilinearized Legendrian contact homology coincides with linearized Legendrian contact homology. More generally, if the two augmentations are DGA homotopic, $\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}(\Lambda)$ is isomorphic to

$\mathrm{LCH}^{\varepsilon_1}(\Lambda)$. Our first result describes a crucial difference in the behavior of bilinearized Legendrian contact homology depending whether the two augmentations are DGA homotopic or not. More precisely, this different behavior is detected by a map $\tau_n : \mathrm{LCH}_n^{\varepsilon_1,\varepsilon_2}(\Lambda) \to H_n(\Lambda)$ appearing in the duality exact sequence for Legendrian contact homology (see Ekholm, Etnyre and Sabloff [6]) and described in Sections 2 and 3.

Theorem 1.1 *Let Λ be a closed connected Legendrian submanifold of $J^1(M)$ with $\dim M = n$. Let ε_1 and ε_2 be two augmentations of the Chekanov–Eliashberg DGA of Λ with coefficients in $\mathbb{Z}_2$. Then ε_1 and ε_2 are DGA homotopic if and only if the map $\tau_n : \mathrm{LCH}_n^{\varepsilon_1,\varepsilon_2}(\Lambda) \to H_n(\Lambda)$ is surjective.*

In other words, the fundamental class of Λ induces a class in linearized Legendrian contact homology, while the class of the point in Λ induces a class in bilinearized Legendrian contact homology with respect to non-DGA homotopic augmentations.

Corollary 1.2 *Bilinearized Legendrian contact homology is a complete invariant for DGA homotopy classes of augmentations of the Chekanov–Eliashberg DGA.*

The strength of this result will be illustrated in Section 3 by revisiting an important example of a Legendrian knot featuring only a partial study of its augmentations; see Melvin and Shrestha [14]. We complete the study of this Legendrian knot with a full description of its DGA homotopy classes of augmentations.

Our second result describes the geography of the Laurent polynomials that can be obtained as a Poincaré polynomial for bilinearized Legendrian contact homology. We will introduce in Definition 4.1 the explicit notion of a BLCH–admissible Laurent polynomial, and prove that only these polynomials can be obtained as the Poincaré polynomial of bilinearized Legendrian contact homology.

Theorem 1.3 *For any* BLCH*–admissible Laurent polynomial P, there exists a closed connected Legendrian submanifold Λ of $J^1(M)$ and there exist two non-DGA homotopic augmentations ε_1 and ε_2 of the Chekanov–Eliashberg DGA of Λ, with the property that the Poincaré polynomial of $\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}(\Lambda)$ with coefficients in $\mathbb{Z}_2$ is equal to P.*

We also will establish a similar result, namely Theorem 4.17, in the specific case of Legendrian spheres.

The collection of Poincaré polynomials that is realized by bilinearized Legendrian contact homology is considerably wider than the corresponding collection for the geography of linearized Legendrian contact homology [2, Theorem 1.1]. For this reason, the examples of Legendrian submanifolds that are constructed here in order to realize the geography of bilinearized Legendrian contact homology differ substantially from those considered in [2] and exhibit new interesting phenomena. In particular, while connected sums of Legendrian submanifolds played an important role in [2], such constructions cannot be used here because these tend to produce pairs of unwanted generators in bilinearized Legendrian contact homology. Moreover, we introduce a completely new construction in order to create pairs of generators in arbitrary degrees, instead of degrees summing to $n-1$ as in linearized Legendrian contact homology. We also introduce another completely new construction in order to obtain bilinearized Legendrian contact homologies of different ranks, depending on the ordering of the two non-DGA homotopic augmentations. Note that the examples we construct are convenient to work with, as they only have cusp singularities.

This paper is organized as follows. In Section 2 we review the definition of bilinearized Legendrian contact homology and state its main properties. In Section 3 we study fundamental classes in bilinearized Legendrian contact homology, prove Theorem 1.1 and Corollary 1.2 and study the effect of connected sums on bilinearized Legendrian contact homology. In Section 4 we study the geography of bilinearized Legendrian contact homology and prove Theorem 1.3 and its counterpart Theorem 4.17 for Legendrian spheres.

Acknowledgments We are indebted to Josh Sabloff for providing us computer code that computes linearized Legendrian contact homology of Legendrian knots in $\mathbb{R}^3$, using techniques of Henry and Rutherford [11; 12]. Although our exposition is independent from these sources, the generalization of this computer code by Galant to the calculation of bilinearized Legendrian contact homology played an essential role at the beginning of this work, before its generalization to higher dimensions. We thank Georgios Dimitroglou Rizell for a productive discussion of DGA homotopies of augmentations. An important refinement in the constructions from Section 4 emerged after an interesting conversation with Sylvain Courte. Special thanks go to Filip Strakoš for spotting a mistake in the proof of Proposition 4.2 impacting other parts of an earlier version of the paper. We also thank Cyril Falcon for his remarks on the original manuscript. Bourgeois was partially supported by the Institut Universitaire de France and by the ANR projects Quantact (16-CE40-0017), Microlocal (15-CE40-0007) and COSY (21-CE40-0002). Galant is an FRS-FNRS research fellow.

2 Bilinearized Legendrian contact homology

The 1–jet space $J^1(M) = T^*M \times \mathbb{R}$ of a smooth n–dimensional manifold M is equipped with a canonical contact structure $\xi = \ker(dz - \lambda)$, where λ is the Liouville 1–form on T^*M and z is the coordinate along $\mathbb{R}$. Let Λ be a closed Legendrian submanifold of this contact manifold, i.e. a closed embedded submanifold of dimension n such that $T_p\Lambda \subset \xi_p$ for any $p \in \Lambda$.

We first describe the definition of a differential graded algebra associated to Λ, following its construction by Ekholm, Etnyre and Sullivan [7]. The Reeb vector field associated to the contact form $\alpha = dz - \lambda$ for ξ is simply $R_\alpha = \partial/\partial z$. A Reeb chord of Λ is a finite nontrivial piece of integral curve for R_α with endpoints on Λ. After performing a Legendrian isotopy, we can assume that all Reeb chords of Λ are nondegenerate, i.e. the canonical projections to the tangent space of T^*M of the tangent spaces to Λ at the endpoints of each chord intersect transversally. Let us assume that the Maslov class $\mu(\Lambda)$ of Λ vanishes; see [7, section 2.2].

We denote by $\mathcal{A}(\Lambda)$ the unital noncommutative algebra freely generated over $\mathbb{Z}_2$ by the Reeb chords of Λ. Each Reeb chord c is graded by its Conley–Zehnder $\nu(c) \in \mathbb{Z}$; when Λ is connected, this does not depend on any additional choice since $\mu(\Lambda) = 0$. The grading of c is defined as $|c| = \nu(c) - 1$. Hence, in this case, the algebra $\mathcal{A}(\Lambda)$ is naturally graded.

Let J be a complex structure on ξ which is compatible with its conformal symplectic structure. This complex structure naturally extends to an almost complex structure, which we still denote by J, on the symplectization $(\mathbb{R} \times J^1(M), d(e^t\alpha))$ by $J\partial/\partial t = R_\alpha$. We consider the moduli space $\widetilde{\mathcal{M}}(a; b_1, \ldots, b_k)$ of J–holomorphic disks in $\mathbb{R} \times J^1(M)$ with boundary on $\mathbb{R} \times \Lambda$ and with $k+1$ punctures on the boundary that are asymptotic at the first puncture to the Reeb chord a at $t = +\infty$ and at the other punctures to the Reeb chords $b_1, \ldots, b_k$ at $t = -\infty$. For a generic choice of J, this moduli space is a smooth manifold of dimension $|a| - \sum_{i=1}^k |b_i|$; see [7, Proposition 2.2]. This moduli space carries a natural $\mathbb{R}$–action corresponding to the translation of J–holomorphic disks along the t–coordinate. When $\{b_1, \ldots, b_k\} \neq \{a\}$, let us denote by $\mathcal{M}(a; b_1, \ldots, b_k)$ the quotient of this moduli space by this free action.

We define a differential ∂ on $\mathcal{A}(\Lambda)$ by

$$\partial a = \sum_{\substack{b_1,\ldots,b_k \\ \dim \mathcal{M}(a;b_1,\ldots,b_k)=0}} \#_2\mathcal{M}(a; b_1, \ldots, b_k) b_1 \cdots b_k,$$

where $\#_2\mathcal{M}(a; b_1, \ldots, b_k)$ is the number of elements in the corresponding moduli space, modulo 2. This differential has degree -1 and satisfies $\partial \circ \partial = 0$.

The resulting differential graded algebra $(\mathcal{A}(\Lambda), \partial)$ is called the Chekanov–Eliashberg DGA, and its homology is called Legendrian contact homology and denoted by $\mathrm{LCH}(\Lambda)$. This graded algebra over $\mathbb{Z}_2$ depends only on the Legendrian isotopy class of Λ.

Let us now turn to the definition of a linearized version of Legendrian contact homology. An augmentation of $(\mathcal{A}(\Lambda), \partial)$ is a unital DGA map $\varepsilon\colon (\mathcal{A}(\Lambda), \partial) \to (\mathbb{Z}_2, 0)$. In other words, it is a choice of $\varepsilon(c) \in \mathbb{Z}_2$ for all Reeb chords c of Λ, it satisfies $\varepsilon(1) = 1$, it extends to $\mathcal{A}(\Lambda)$ multiplicatively and additively, and it satisfies $\varepsilon \circ \partial = 0$.

Such an augmentation can be used to define a linearization of $(\mathcal{A}(\Lambda), \partial)$. Let $C(\Lambda)$ be the vector space over $\mathbb{Z}_2$ freely generated by all Reeb chords of Λ. We also define the linearized differential ∂^ε on $C(\Lambda)$ by

$$\partial^\varepsilon a = \sum_{\substack{b_1,\ldots,b_k \\ \dim \mathcal{M}(a;b_1,\ldots,b_k)=0}} \#_2\mathcal{M}(a; b_1, \ldots, b_k) \sum_{i=1}^k \varepsilon(b_1) \cdots \varepsilon(b_{i-1}) b_i \varepsilon(b_{i+1}) \cdots \varepsilon(b_k).$$

This differential has degree -1 and satisfies $\partial^\varepsilon \circ \partial^\varepsilon = 0$. The homology of the resulting linearized complex $(C(\Lambda), \partial^\varepsilon)$ is called linearized Legendrian contact homology (with respect to ε) and denoted by $\mathrm{LCH}^\varepsilon(\Lambda)$. The collection of these graded modules over $\mathbb{Z}_2$ for all augmentations of Λ depends only on the Legendrian isotopy class of Λ.

Linearized Legendrian contact homology fits into a duality long exact sequence [6] together with its cohomological version $\mathrm{LCH}_\varepsilon(\Lambda)$ and with the singular homology $H(\Lambda)$ of the underlying n–dimensional manifold Λ:

$$\cdots \to \mathrm{LCH}_\varepsilon^{n-k-1}(\Lambda) \to \mathrm{LCH}_k^\varepsilon(\Lambda) \xrightarrow{\tau_k} H_k(\Lambda) \to \mathrm{LCH}_\varepsilon^{n-k}(\Lambda) \to \cdots.$$

Moreover, the map τ_n in the above exact sequence does not vanish. These properties induce constraints on the graded modules over $\mathbb{Z}_2$ that can be realized as the linearized Legendrian contact homology of some Legendrian submanifold, with respect to some augmentation. These constraints can be formulated in terms of the Poincaré polynomial of $\mathrm{LCH}^\varepsilon(\Lambda)$, which is the Laurent polynomial defined by

$$P_{\Lambda,\varepsilon}(t) = \sum_{k\in\mathbb{Z}} \dim_{\mathbb{Z}_2} \mathrm{LCH}^\varepsilon_k(\Lambda) t^k.$$

When Λ is connected, the duality exact sequence and the nonvanishing of τ_n imply that the above Poincaré polynomial has the form

$$P_{\Lambda,\varepsilon}(t) = q(t) + p(t) + t^{n-1} p(t^{-1}), \tag{2-1}$$

where q is a monic polynomial of degree n with integral coefficients (corresponding to the image of the maps τ_k) and p is a Laurent polynomial with integral coefficients (corresponding to the kernel of the maps τ_k). We shall say that a Laurent polynomial of this form is LCH–admissible.

The first author together with Sabloff and Traynor [2] studied generating family homology $\mathrm{GH}(f)$, an invariant for isotopy classes of Legendrian submanifolds $\Lambda \subset (J^1(M), \xi)$ admitting a generating family f. This invariant is also a graded module over $\mathbb{Z}_2$ and satisfies the same duality exact sequence as above. In this study, the effect of Legendrian ambient surgeries on this invariant was determined and these operations were used to produce many interesting examples of Legendrian submanifolds admitting generating families. More precisely, for any LCH–admissible Laurent polynomial P, a connected Legendrian submanifold Λ_P admitting a generating family f_P realizing P as the Poincaré polynomial of $\mathrm{GH}(f_P)$ was constructed using these operations. On the other hand, Dimitroglou Rizell [4] showed in particular that Legendrian ambient surgeries have the same effect as above on linearized Legendrian contact homology (for more details in the case of the connected sum, see the proof of Proposition 3.5). This result can be used step by step in the constructions of [2] to show that, for any LCH–admissible Laurent polynomial P, there exists an augmentation ε_P for Λ_P such that $\mathrm{LCH}^{\varepsilon_P}(\Lambda_P) \cong \mathrm{GH}(f_P)$. Therefore, the geography question for linearized Legendrian contact homology is completely determined by the above LCH–admissible Laurent polynomials.

Finally, we turn to a generalization of linearized LCH introduced by the first author together with Chantraine [1]. Using two augmentations ε_1 and ε_2 of $(\mathcal{A}(\Lambda), \partial)$, we can define another differential $\partial^{\varepsilon_1,\varepsilon_2}$ on $C(\Lambda)$, called the bilinearized differential:

$$\partial^{\varepsilon_1,\varepsilon_2} a = \sum_{\substack{b_1,\dots,b_k \\ \dim \mathcal{M}(a;b_1,\dots,b_k)=0}} \#_2\mathcal{M}(a;b_1,\dots,b_k) \sum_{i=1}^k \varepsilon_1(b_1)\cdots\varepsilon_1(b_{i-1}) b_i \varepsilon_2(b_{i+1})\cdots\varepsilon_2(b_k).$$

As above, this differential has degree -1 and satisfies $\partial^{\varepsilon_1,\varepsilon_2} \circ \partial^{\varepsilon_1,\varepsilon_2} = 0$. The homology of the resulting bilinearized complex $(C(\Lambda), \partial^{\varepsilon_1,\varepsilon_2})$ is called bilinearized Legendrian contact homology (with respect to ε_1 and ε_2) and denoted by $\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}(\Lambda)$. The collection of these graded modules over $\mathbb{Z}_2$ for all pairs of augmentations of Λ depends only on the Legendrian isotopy class of Λ.

Bilinearized Legendrian contact homology also satisfies a duality exact sequence [1], but one has to take care of the ordering of the augmentations:

$$\cdots \to \mathrm{LCH}^{n-k-1}_{\varepsilon_2,\varepsilon_1}(\Lambda) \to \mathrm{LCH}^{\varepsilon_1,\varepsilon_2}_k(\Lambda) \xrightarrow{\tau_k} H_k(\Lambda) \xrightarrow{\sigma_{n-k}} \mathrm{LCH}^{n-k}_{\varepsilon_2,\varepsilon_1}(\Lambda) \to \cdots. \tag{2-2}$$

Moreover, unlike in the linearized case, there exist [1, Section 5] connected Legendrian submanifolds Λ with augmentations ε_1 and ε_2 such that the map τ_n vanishes. Our goal here is to understand when the map τ_n vanishes, and to study the geography of the Poincaré polynomials

$$P_{\Lambda,\varepsilon_1,\varepsilon_2}(t) = \sum_{k\in\mathbb{Z}} \dim_{\mathbb{Z}_2} \mathrm{LCH}^{\varepsilon_1,\varepsilon_2}_k(\Lambda) t^k$$

for bilinearized Legendrian contact homology.

3 Fundamental classes in bilinearized Legendrian contact homology

There are several notions of equivalence for augmentations of DGAs that were introduced in the literature and used in the context of the Chekanov–Eliashberg DGA. As the results of this section will show, it turns out that the equivalence relation among augmentations that best controls the behavior of bilinearized LCH is the notion of DGA homotopic augmentations [16, Definition 5.13]. Let ε_1 and ε_2 be two augmentations of the DGA $(\mathcal{A},\partial)$ over $\mathbb{Z}_2$. Recall that a linear map $K\colon \mathcal{A}\to\mathbb{Z}_2$ is said to be an $(\varepsilon_1,\varepsilon_2)$–derivation if $K(ab)=\varepsilon_1(a)K(b)+K(a)\varepsilon_2(b)$ for any $a,b\in\mathcal{A}$. We say that ε_1 is DGA homotopic to ε_2, and we write $\varepsilon_1\sim\varepsilon_2$, if there exists an $(\varepsilon_1,\varepsilon_2)$–derivation $K\colon\mathcal{A}\to\mathbb{Z}_2$ of degree $+1$ such that $\varepsilon_1-\varepsilon_2=K\circ\partial$. It is a standard fact that DGA homotopy is an equivalence relation [10, Lemma 26.3].

Note that the defining condition for a DGA homotopy admits a beautiful and convenient reformulation in terms of the bilinearized complex.

Lemma 3.1 *Two augmentations ε_1 and ε_2 are DGA homotopic if and only if there exists a linear map $\overline{K}\colon C(\Lambda)\to\mathbb{Z}_2$ of degree $+1$ such that $\varepsilon_1-\varepsilon_2=\overline{K}\circ\partial^{\varepsilon_1,\varepsilon_2}$ on $C(\Lambda)$.*

Proof Suppose first that ε_1 is DGA homotopic to ε_2. This implies in particular that $\varepsilon_1(c)-\varepsilon_2(c)=K\circ\partial c$ for any $c\in C(\Lambda)$. Since K is an $(\varepsilon_1,\varepsilon_2)$–derivation, it directly follows from the definition of the bilinearized differential that $K\circ\partial c = K\circ\partial^{\varepsilon_1,\varepsilon_2}c$. It then suffices to take $\overline{K}$ to be the restriction of K to $C(\Lambda)$.

Suppose now that there exists a linear map $\overline{K}\colon C(\Lambda)\to\mathbb{Z}_2$ of degree $+1$ such that $\varepsilon_1-\varepsilon_2=\overline{K}\circ\partial^{\varepsilon_1,\varepsilon_2}$ on $C(\Lambda)$. The map $\overline{K}$ determines a unique $(\varepsilon_1,\varepsilon_2)$–derivation $K\colon\mathcal{A}\to\mathbb{Z}_2$ via the relation

$$K(a_1\cdots a_n)=\sum_{i=1}^{k}\varepsilon_1(a_1\cdots a_{i-1})\overline{K}(a_i)\varepsilon_2(a_{i+1}\cdots a_n)$$

for all $a_1,\ldots,a_n\in\mathcal{A}$. As above, these maps satisfy $K\circ\partial c=\overline{K}\circ\partial^{\varepsilon_1,\varepsilon_2}c$, so that $\varepsilon_1-\varepsilon_2=\overline{K}\circ\partial$ on $C(\Lambda)$. Now observe that $\varepsilon_1(ab)-\varepsilon_2(ab)=\varepsilon_1(a)(\varepsilon_1(b)-\varepsilon_2(b))+(\varepsilon_1(a)-\varepsilon_2(a))\varepsilon_2(b)$, and on the other hand $K\circ\partial(ab)=\varepsilon_1(\partial a)K(b)+\varepsilon_1(a)K(\partial b)+K(\partial a)\varepsilon_2(b)+K(a)\varepsilon_2(\partial b)=\varepsilon_1(a)K(\partial b)+K(\partial a)\varepsilon_2(b)$.

Hence if a and b satisfy the DGA homotopy relation, then ab satisfies it as well. Since this relation holds on $C(\Lambda)$, it follows that it is also satisfied on $\mathcal{A}$. □

Note that, in the above proof, the extension of the linear map $\bar{K}$ to a unique $(\varepsilon_1, \varepsilon_2)$–derivation on $\mathcal{A}$ as well as the extension of the homotopy relation from $C(\Lambda)$ to $\mathcal{A}$ were first established in a more general setup by Kálmán in [13, Lemma 2.18].

With this suitable notion of equivalence for augmentations, we can now turn to the study of the fundamental class in bilinearized LCH, via the maps τ_0 and τ_n from the duality long exact sequence. The following proposition generalizes [6, Theorem 5.5].

Proposition 3.2 *Let ε_1 and ε_2 be augmentations of the Chekanov–Eliashberg DGA $(\mathcal{A}, \partial)$ of a closed connected n–dimensional Legendrian submanifold Λ in $(J^1(M), \xi)$. The map $\tau_0\colon \mathrm{LCH}_0^{\varepsilon_1,\varepsilon_2}(\Lambda) \to H_0(\Lambda)$ from the duality long exact sequence vanishes if and only if ε_1 and ε_2 are DGA homotopic.*

Proof Let f be a Morse function on Λ with a unique minimum at point m, and g be a Riemannian metric on Λ. Since the stable manifold of m is open and dense in Λ, for a generic choice of the Morse–Smale pair (f, g), the endpoints of all Reeb chords of Λ are in this stable manifold. The vector space $H_0(\Lambda)$ is generated by m and we identify it with $\mathbb{Z}_2$. By the results of [6], the map τ_0 counts rigid J–holomorphic disks with boundary on Λ, with a positive puncture on the boundary and with a marked point on the boundary mapping to the stable manifold of m. This disk can have extra negative punctures on the boundary; these are augmented by ε_1 if they sit between the positive puncture and the marked point, and by ε_2 if they sit between the marked point and the positive puncture. Since mapping to m is an open condition on Λ, such rigid configurations can only occur when the image of the disk boundary is discrete in Λ. In other words, the holomorphic disk maps to the symplectization of a Reeb chord c of Λ. Since there is a unique positive puncture, this map is not a covering, and there is a unique negative puncture at c. There is a unique such J–holomorphic disk for any chord c of Λ. The marked point maps to the starting point or to the ending point of the chord c in Λ. If the marked point maps to the starting point of c, the negative puncture sits between the positive puncture and the marked point on the boundary of the disk, which therefore contributes $\varepsilon_2(c)$ to $\tau_0(c)$ at chain level. If the marked point maps to the ending point of c, the negative puncture sits between the marked point and the positive puncture on the boundary of the disk, which therefore contributes $\varepsilon_1(c)$ to $\tau_0(c)$. We conclude that the map τ_0 is given at chain level by $\varepsilon_1 - \varepsilon_2$.

If ε_1 and ε_2 are DGA homotopic, then by Lemma 3.1 the map τ_0 is nullhomotopic and therefore vanishes in homology. On the other hand, if ε_1 and ε_2 are not DGA homotopic, Lemma 3.1 implies that the map $\varepsilon_1 - \varepsilon_2\colon C_0(\Lambda) \to \mathbb{Z}_2$ does not factor through the bilinearized differential $\partial^{\varepsilon_1,\varepsilon_2}$. In other words, there exists $a \in C_0(\Lambda)$ such that $\partial^{\varepsilon_1,\varepsilon_2} a = 0$ but $\varepsilon_1(a) - \varepsilon_2(a) \neq 0$. But then the homology class $[a] \in \mathrm{LCH}_0^{\varepsilon_1,\varepsilon_2}(\Lambda)$ satisfies $\tau_0([a]) \neq 0$, so that τ_0 does not vanish in homology. □

We are now in position to prove our first main result.

Proof of Theorem 1.1 In the duality long exact sequence (2-2) for bilinearized LCH, the maps τ_k and σ_k are adjoint in the sense of [6, Proposition 3.9] as in the linearized case. The proof of this fact is essentially identical in the bilinearized case: the holomorphic disks counted by τ_k are still in bijective correspondence with those counted by σ_k. In the bilinearized case, it is also necessary to use the fact that the extra negative punctures on corresponding disks are augmented with the same augmentations in order to reach the conclusion.

In particular, τ_n vanishes if and only if σ_n vanishes. Since $H_0(\Lambda) \cong \mathbb{Z}_2$, the exactness of the duality sequence (2-2) implies that σ_n vanishes if and only if τ_0 does not vanish. By Proposition 3.2, this means that τ_n vanishes if and only if the augmentations ε_1 and ε_2 are not DGA homotopic. □

This difference in the behavior of bilinearized LCH can be used to determine DGA homotopy classes of augmentations. More precisely, the next proposition shows that bilinearized LCH provides an explicit criterion to decide whether two augmentations are DGA homotopic or not.

Proposition 3.3 *Let ε_1 and ε_2 be augmentations of the Chekanov–Eliashberg DGA $(\mathcal{A}, \partial)$ of a closed connected n–dimensional Legendrian submanifold Λ in $(J^1(M), \xi)$. Then*

$$\dim_{\mathbb{Z}_2} \mathrm{LCH}_n^{\varepsilon_2,\varepsilon_1}(\Lambda) - \dim_{\mathbb{Z}_2} \mathrm{LCH}^{\varepsilon_1,\varepsilon_2}_{-1}(\Lambda) = \begin{cases} 0 & \text{if } \varepsilon_1 \nsim \varepsilon_2, \\ 1 & \text{if } \varepsilon_1 \sim \varepsilon_2. \end{cases}$$

Proof By the duality exact sequence (2-2), we have

$$H_0(\Lambda) \cong \mathbb{Z}_2 \xrightarrow{\sigma_n} \mathrm{LCH}^n_{\varepsilon_2,\varepsilon_1}(\Lambda) \to \mathrm{LCH}^{\varepsilon_1,\varepsilon_2}_{-1}(\Lambda) \to H_{-1}(\Lambda) = 0.$$

In other words, $\mathrm{LCH}^n_{\varepsilon_2,\varepsilon_1}(\Lambda)/\operatorname{im}\sigma_n \cong \mathrm{LCH}^{\varepsilon_1,\varepsilon_2}_{-1}(\Lambda)$. Taking into account that

$$\dim_{\mathbb{Z}_2} \mathrm{LCH}^n_{\varepsilon_2,\varepsilon_1}(\Lambda) = \dim_{\mathbb{Z}_2} \mathrm{LCH}_n^{\varepsilon_2,\varepsilon_1}(\Lambda),$$

we obtain the desired result since, as in the proof of Theorem 1.1, the rank of σ_n is 1 when $\varepsilon_1 \sim \varepsilon_2$ and vanishes when $\varepsilon_1 \nsim \varepsilon_2$. □

Corollary 1.2 follows immediately from the above proposition.

Example 3.4 Let us consider the Legendrian knot K_2 studied by Melvin and Shrestha in [14, Section 3], which is topologically the mirror image of the knot 8_{21}, and illustrated in Figure 1.

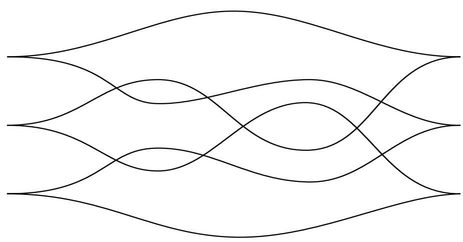


Figure 1: Front projection of the Legendrian knot K_2.

It is shown in [14, Section 3] that the Chekanov–Eliashberg DGA of this Legendrian knot K_2 has exactly 16 augmentations, which split into a set A of 4 augmentations and a set B of 12 augmentations, such that $P_{K_2,\varepsilon}(t) = 2t+4+t^{-1}$ if $\varepsilon \in A$ and $P_{K_2,\varepsilon}(t) = t+2$ if $\varepsilon \in B$. This implies that augmentations in A are not DGA homotopic to augmentations in B. However, the number of DGA homotopy classes of augmentations for K_2 was not determined in [14], as linearized LCH does not suffice to obtain this information.

Using Proposition 3.3, these DGA homotopy classes can be determined systematically. It turns out that the augmentations in A are pairwise not DGA homotopic, because the Poincaré polynomial of any such pair of augmentations is $t+3+t^{-1}$. On the other hand, the set B splits into six DGA homotopy classes $\mathcal{C}_1, \dots, \mathcal{C}_6$ of augmentations. The BLCH Poincaré polynomials are given by $t+2$ for two DGA homotopic augmentations in B, by 1 for two non-DGA homotopic augmentations both in $\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$ or in $\mathcal{C}_4 \cup \mathcal{C}_5 \cup \mathcal{C}_6$, and by $t+2$ and $2+t^{-1}$ otherwise.

These calculations are straightforward but tedious. Suitable Python code gives the above answer instantly.

We conclude our study of the fundamental classes in bilinearized LCH with a useful description of their behavior when performing a connected sum. To this end, it is convenient to introduce some additional notation about the map τ_n in the duality exact sequence (2-2). Its target space $H_n(\Lambda)$ is spanned by the fundamental classes $[\Lambda_i]$ of the connected components Λ_i of the Legendrian submanifold Λ. We can therefore decompose τ_n as $\sum_i \tau_{n,i}[\Lambda_i]$, where the maps $\tau_{n,i}$ take their values in $\mathbb{Z}_2$.

Proposition 3.5 *Let Λ be a Legendrian link in $J^1(M)$ equipped with two augmentations ε_1 and ε_2. Let $\overline{\Lambda}$ be the Legendrian submanifold obtained by performing a connected sum between two connected components Λ_1 and Λ_2 of Λ, and let $\bar\varepsilon_1$ and $\bar\varepsilon_2$ be the augmentations induced by ε_1 and ε_2.*

If the map $\tau_{n,1} - \tau_{n,2}$ constructed from the map τ_n in the duality exact sequence (2-2) vanishes, then $P_{\overline{\Lambda},\bar\varepsilon_1,\bar\varepsilon_2}(t) = P_{\Lambda,\varepsilon_1,\varepsilon_2}(t) + t^{n-1}$. Otherwise, $P_{\overline{\Lambda},\bar\varepsilon_1,\bar\varepsilon_2}(t) = P_{\Lambda,\varepsilon_1,\varepsilon_2}(t) - t^n$.

Proof As explained in [1, Section 3.2.5], the map τ_n in the duality exact sequence (2-2) for Λ counts holomorphic disks in the symplectization of $J^1(M)$ with boundary on the symplectization of Λ, having a positive puncture asymptotic to a chord c of Λ and a marked point on the boundary mapped to a fixed generic point p_j of a connected component Λ_j of Λ. This disk can also carry negative punctures on the boundary; let us say that those located between the positive puncture and the chord (with respect to the natural orientation of the boundary) are asymptotic to chords $c_1^-, \dots, c_r^-$, while those between the marked point and the positive puncture are asymptotic to $c_{r+1}^-, \dots, c_{r+s}^-$. Let us denote by $\mathcal{M}(c; c_1^-, \dots, c_r^-, p_j, c_{r+1}^-, \dots, c_{r+s}^-)$ the moduli space of such holomorphic disks, modulo translation in the $\mathbb{R}$ direction of the symplectization. The map τ_n is then given by

$$\tau_n(c) = \sum_j \#_2\mathcal{M}(c; c_1^-, \dots, c_r^-, p_j, c_{r+1}^-, \dots, c_{r+s}^-)\varepsilon_1(c_1^-)\cdots\varepsilon_1(c_r^-)\varepsilon_2(c_{r+1}^-)\cdots\varepsilon_2(c_{r+s}^-)[\Lambda_j].$$

On the other hand, the effect of a connected sum on bilinearized LCH can be deduced from the results of Dimitroglou Rizell on the full Chekanov–Eliashberg DGA [4, Theorem 1.6]. There is an isomorphism of DGAs $\Psi\colon (\mathcal{A}(\overline{\Lambda}), \partial_{\overline{\Lambda}}) \to (\mathcal{A}(\Lambda; S), \partial_S)$ between the Chekanov–Eliashberg DGA of $\overline{\Lambda}$ and the DGA $(\mathcal{A}(\Lambda; S), \partial_S)$ generated by the Reeb chords of Λ as well as a formal generator s of degree $n-1$, equipped with a differential ∂_S satisfying in particular $\partial_S s = 0$. In this notation, S stands for the pair of points $\{p_1 \in \Lambda_1, p_2 \in \Lambda_2\}$ in a neighborhood of which the connected sum is performed. Any augmentation ε of the Chekanov–Eliashberg DGA of Λ can be extended to an augmentation of $(\mathcal{A}(\Lambda; S), \partial_S)$ by setting $\varepsilon(s) = 0$. Moreover, the pullback $\Psi^*\varepsilon$ of this extension to the Chekanov–Eliashberg DGA of $\overline{\Lambda}$ coincides with the augmentation induced on $\overline{\Lambda}$ from the original augmentation ε for Λ via the surgery Lagrangian cobordism between $\overline{\Lambda}$ and Λ. In particular, we have $\bar{\varepsilon}_1 = \Psi^*\varepsilon_1$ and $\bar{\varepsilon}_2 = \Psi^*\varepsilon_2$. Applying the bilinearization procedure to the map Ψ, we obtain a chain complex isomorphism $\Psi^{\varepsilon_1,\varepsilon_2}$ between the bilinearized chain complex for $\overline{\Lambda}$ and the chain complex $(C(\Lambda, S), \partial_S^{\varepsilon_1,\varepsilon_2})$ generated by Reeb chords of Λ and the formal generator s. Since $\partial_S^{\varepsilon_1,\varepsilon_2} s = 0$, the line spanned by s forms a subcomplex of $(C(\Lambda, S), \partial_S^{\varepsilon_1,\varepsilon_2})$. Moreover, the quotient complex is exactly the bilinearized chain complex for Λ. We therefore obtain a long exact sequence in homology

$$\cdots \to \mathrm{LCH}_k^{\bar{\varepsilon}_1,\bar{\varepsilon}_2}(\overline{\Lambda}) \to \mathrm{LCH}_k^{\varepsilon_1,\varepsilon_2}(\Lambda) \xrightarrow{\rho_k} \mathbb{Z}_2[s]_{k-1} \to \mathrm{LCH}_{k-1}^{\bar{\varepsilon}_1,\bar{\varepsilon}_2}(\overline{\Lambda}) \to \cdots$$

that corresponds to the long exact sequence obtained in [2, Theorem 2.1] for generating family homology. This exact sequence implies that bilinearized LCH remains unchanged by a connected sum, except possibly in degrees $n-1$ and n. The map ρ_n is the part of the bilinearized differential $\partial_S^{\varepsilon_1,\varepsilon_2}$ from the bilinearized complex for Λ to the line spanned by s. According to the definition [4, Section 1.1.3] of ∂_S and the above description of τ_n, this map is given by $\rho_n = (\tau_{n,1} - \tau_{n,2})s$.

If $\rho_n = 0$, the generator s injects into $\mathrm{LCH}_{n-1}^{\bar{\varepsilon}_1,\bar{\varepsilon}_2}(\overline{\Lambda})$, resulting in an exact term t^{n-1} in the Poincaré polynomial. If $\rho_n \neq 0$, the map $\mathrm{LCH}_n^{\bar{\varepsilon}_1,\bar{\varepsilon}_2}(\overline{\Lambda}) \to \mathrm{LCH}_n^{\varepsilon_1,\varepsilon_2}(\Lambda)$ has a 1–dimensional cokernel, resulting in the loss of a term t^n in the Poincaré polynomial. □

4 Geography of bilinearized Legendrian contact homology

In this section we study the possible values for the Poincaré polynomial $P_{\Lambda,\varepsilon_1,\varepsilon_2}$ of the bilinearized LCH for a closed connected Legendrian submanifold Λ in $J^1(M)$ with $\dim M = n$, equipped with two augmentations ε_1 and ε_2 of its Chekanov–Eliashberg DGA.

When $\varepsilon_1 = \varepsilon_2$, this geography question was completely answered in [2] for generating family homology. As explained in Section 2, this result extends to linearized LCH via the work of Dimitroglou Rizell [4]. Moreover, bilinearized LCH is invariant under changes of augmentations within their DGA homotopy classes [16, Section 5.3]. Therefore, the geography of bilinearized LCH is already known when $\varepsilon_1 \sim \varepsilon_2$.

4.1 Basic properties of BLCH Poincaré polynomials

We now turn to the case $\varepsilon_1 \nsim \varepsilon_2$, and describe the possible Poincaré polynomials for bilinearized LCH.

Definition 4.1 A BLCH–admissible polynomial is the data of a Laurent polynomial P with nonnegative integral coefficients together with a splitting $P = q + p$ involving two Laurent polynomials with nonnegative integral coefficients p and q such that

(i) q is a polynomial of degree at most $n-1$ with $q(0) = 1$, and

(ii) $p(-1)$ is even if $n = 1$ and $p(-1) \leq \frac{1}{2}(1 - q(-1))$ if $n = 2$.

We first show that the Poincaré polynomial of bilinearized LCH always has this form.

Proposition 4.2 *Let ε_1 and ε_2 be augmentations of the Chekanov–Eliashberg DGA $(\mathcal{A}, \partial)$ of a closed connected n–dimensional Legendrian submanifold Λ with vanishing Maslov class in $(J^1(M), \xi)$. If ε_1 and ε_2 are not DGA homotopic, then the Poincaré polynomial $P_{\Lambda,\varepsilon_1,\varepsilon_2}$ corresponding to* $\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}(\Lambda)$ *is* BLCH*–admissible.*

Proof Considering the map τ_k from the duality exact sequence (2-2), we have $\dim_{\mathbb{Z}_2} \mathrm{LCH}_k^{\varepsilon_1,\varepsilon_2}(\Lambda) = \dim_{\mathbb{Z}_2} \ker \tau_k + \dim_{\mathbb{Z}_2} \operatorname{im} \tau_k$. Let p and q be the Poincaré polynomials constructed using the terms in the right-hand side of this relation: $p(t) = \sum_{k\in\mathbb{Z}} \dim_{\mathbb{Z}_2} \ker \tau_k t^k$ and $q(t) = \sum_{k\in\mathbb{Z}} \dim_{\mathbb{Z}_2} \operatorname{im} \tau_k t^k$. This provides the desired splitting $P_{\Lambda,\varepsilon_1,\varepsilon_2} = q + p$.

Let us prove (i). Since $\operatorname{im} \tau_k \subset H_k(\Lambda)$, q is a polynomial of degree at most n. By Proposition 3.2, since $\varepsilon_1 \nsim \varepsilon_2$, $\operatorname{im} \tau_0 \neq 0$. But $H_0(\Lambda) = \mathbb{Z}_2$ as Λ is connected, so that $q(0) = 1$. On the other hand, by Theorem 1.1, since $\varepsilon_1 \nsim \varepsilon_2$ we have that $\tau_n = 0$. Therefore the term of degree n in q vanishes and q is a polynomial of degree at most $n-1$.

Let us now prove (ii). Assume first that n is odd. Since the generators of the chain complex $C(\Lambda)$ do not depend on the augmentations, the Euler characteristic $P_{\Lambda,\varepsilon_1,\varepsilon_2}(-1)$ does not depend on the augmentations either. Equation (2-1) then implies that $P_{\Lambda,\varepsilon_1,\varepsilon_2}(-1)$ has the same parity as $\frac{1}{2}\sum_{k\in\mathbb{Z}} \dim_{\mathbb{Z}_2} H_k(\Lambda)$, since $(-1)^{n-1} = 1$ when n is odd. If $n = 1$, then condition (i) sets $q(t) = 1$ so that $q(-1) = 1$ while $\frac{1}{2}\sum_{k\in\mathbb{Z}} \dim_{\mathbb{Z}_2} H_k(\Lambda) = 1$. By subtraction, we deduce that $p(-1)$ must be even. Note that if $n \geq 3$, this does not impose any condition on $p(-1)$ since $q(-1)$ can take arbitrary integer values.

Assume now that n is even. By [8, Proposition 3.3], the Thurston–Bennequin invariant of Λ is given by $\mathrm{tb}(\Lambda) = (-1)^{(n-1)(n-2)/2} P_{\Lambda,\varepsilon_1,\varepsilon_2}(-1)$. On the other hand, $\mathrm{tb}(\Lambda) = (-1)^{n/2+1}\frac{1}{2}\mathcal{X}(\Lambda)$ when n is even by [8, Proposition 3.2]. Hence $P_{\Lambda,\varepsilon_1,\varepsilon_2}(-1) = \frac{1}{2}\mathcal{X}(\Lambda)$. When $n = 2$, we have that $\frac{1}{2}\mathcal{X}(\Lambda) = \frac{1}{2}(1 - \dim_{\mathbb{Z}_2} H_1(\Lambda) + 1) \leq \frac{1}{2}(1 + q(-1))$. By subtraction, we get that $p(-1) \leq \frac{1}{2}(1 - q(-1))$. Note that if $n \geq 4$, this does not impose any condition on $p(-1)$ since $\frac{1}{2}\mathcal{X}(\Lambda)$ can take arbitrary integer values. □

Remark 4.3 If we restrict ourselves to Legendrian spheres Λ, the Laurent polynomials $P = q + p$ that can arise as the Poincaré polynomial of bilinearized LCH can also be characterized. More precisely, revisiting the proof of Proposition 4.2 shows that in this case q and p satisfy the more restrictive conditions

(i′) $q(t) = 1$, and

(ii′) $p(-1)$ is even if n is odd and $p(-1) = 0$ if n is even.

The duality exact sequence imposes fewer restrictions on $\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}(\Lambda)$ than in the case of linearized LCH because it mainly relates this invariant to $\mathrm{LCH}^{\varepsilon_2,\varepsilon_1}(\Lambda)$ with exchanged augmentations. This fact, however, means that one of these invariants determines the other one. In order to formulate this more precisely, let us consider the duality exact sequence obtained from (2-2) after reversing the ordering of the augmentations:

$$\cdots \to \mathrm{LCH}^{n-k-1}_{\varepsilon_1,\varepsilon_2}(\Lambda) \to \mathrm{LCH}^{\varepsilon_2,\varepsilon_1}_k(\Lambda) \xrightarrow{\tilde{\tau}_k} H_k(\Lambda) \xrightarrow{\tilde{\sigma}_{n-k}} \mathrm{LCH}^{n-k}_{\varepsilon_1,\varepsilon_2}(\Lambda) \to \cdots. \tag{4-1}$$

In the next proposition, we denote by $P_\Lambda(t)$ the Poincaré polynomial for the singular homology of Λ with coefficients in $\mathbb{Z}_2$.

Proposition 4.4 *Let ε_1 and ε_2 be non-DGA homotopic augmentations of the Chekanov–Eliashberg DGA $(\mathcal{A},\partial)$ of a closed connected n–dimensional Legendrian submanifold Λ with vanishing Maslov class in $(J^1(M),\xi)$. If $P_{\Lambda,\varepsilon_1,\varepsilon_2}$ decomposes as $q+p$ as in Definition 4.1, then $P_{\Lambda,\varepsilon_2,\varepsilon_1}$ decomposes as $\tilde{q}+\tilde{p}$ with $\tilde{q}(t)=P_\Lambda(t)-t^n q(t^{-1})$ and $\tilde{p}(t)=t^{n-1}p(t^{-1})$.*

Proof Let us decompose $P_{\Lambda,\varepsilon_2,\varepsilon_1}(t)=\tilde{q}(t)+\tilde{p}(t)$ as in Definition 4.1. The polynomial p was defined as $p(t)=\sum_{k\in\mathbb{Z}}\dim_{\mathbb{Z}_2}\ker\tau_k t^k$ in the proof of Proposition 4.2. But $\ker\tau_k$ is the image of the map $\mathrm{LCH}^{n-k-1}_{\varepsilon_2,\varepsilon_1}(\Lambda)\to\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}_k(\Lambda)$, which is isomorphic to a supplementary subspace of $\operatorname{im}\sigma_{n-k-1}$ in $\mathrm{LCH}^{n-k-1}_{\varepsilon_2,\varepsilon_1}(\Lambda)$. Since σ_{n-k-1} is the adjoint in the sense of [6, Proposition 3.9] of the map $\tilde{\tau}_{n-k-1}\colon \mathrm{LCH}^{\varepsilon_2,\varepsilon_1}_{n-k-1}(\Lambda)\to H_{n-k-1}(\Lambda)$, the spaces $\ker\tau_k$ and $\ker\tilde{\tau}_{n-k-1}$ are isomorphic. Therefore, the polynomial $\tilde{p}$ is given by $\tilde{p}(t)=\sum_{k\in\mathbb{Z}}\dim_{\mathbb{Z}_2}\ker\tau_k t^{n-k-1}=t^{n-1}p(t^{-1})$.

On the other hand, we have $\tilde{q}(t)=\sum_{k\in\mathbb{Z}}\dim_{\mathbb{Z}_2}\operatorname{im}\tilde{\tau}_k t^k$ as in the proof of Proposition 4.2. But $\operatorname{im}\tilde{\tau}_k=\ker\tilde{\sigma}_{n-k}$ and since $\tau\sigma_{n-k}$ is the adjoint in the sense of [6, Proposition 3.9] of the map τ_{n-k}, we have that $\ker\tilde{\sigma}_{n-k}$ is isomorphic to a supplementary subspace of $\operatorname{im}\tau_{n-k}$ in $H_{n-k}(\Lambda)$. Hence

$$\tilde{q}(t)=\sum_{k\in\mathbb{Z}}(\dim_{\mathbb{Z}_2}H_{n-k}(\lambda)-\dim_{\mathbb{Z}_2}\operatorname{im}\tau_{n-k})t^k=P_\Lambda(t)-t^nq(t^{-1}),$$

as announced. □

Note that, since the data of $P_{\Lambda,\varepsilon_1,\varepsilon_2}$ and $P_{\Lambda,\varepsilon_2,\varepsilon_1}$ determine P_Λ, the question of finding Λ, ε_1 and ε_2 with prescribed polynomials $P_{\Lambda,\varepsilon_1,\varepsilon_2}$ and $P_{\Lambda,\varepsilon_2,\varepsilon_1}$ is more complicated than our geography question. We will not address this more complicated question.

4.2 Motivating example

We now describe a fundamental example in view of the construction of Legendrian submanifolds and augmentations realizing BLCH–admissible polynomials.

Example 4.5 With $n=1$, consider the right-handed trefoil knot Λ with maximal Thurston–Bennequin invariant, depicted in its front projection in Figure 2. The same Legendrian knot was studied in Section 5.1 of [1]. We consider it this time in the front projection, after applying Ng's resolution procedure [15].

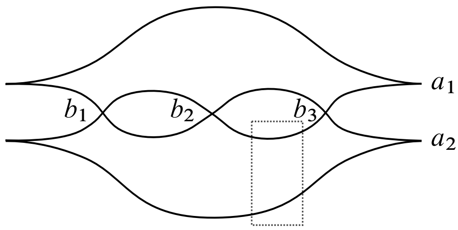


Figure 2: Front projection of the maximal tb right-handed trefoil.

The Chekanov–Eliashberg DGA has five generators: a_1 and a_2 correspond to right cusps and have grading 1, while b_1, b_2 and b_3 correspond to crossings and have grading 0. The differential is given by

$$\partial a_1 = 1 + b_1 + b_3 + b_1 b_2 b_3 \quad \text{and} \quad \partial a_2 = 1 + b_1 + b_3 + b_3 b_2 b_1.$$

This DGA admits 5 augmentations $\varepsilon_1, \dots, \varepsilon_5$ given by

	b_1	b_2	b_3
ε_1	1	1	1
ε_2	1	0	0
ε_3	1	1	0
ε_4	0	0	1
ε_5	0	1	1

A straightforward calculation shows that $P_{\Lambda,\varepsilon_i,\varepsilon_j}(t) = 1$ for all $i \neq j$. In view of Definition 4.1 and Proposition 4.2, this is the simplest Poincaré polynomial that can be obtained using bilinearized LCH.

In order to produce other terms in this Poincaré polynomial, let us replace the portion of Λ contained in the dotted rectangle in Figure 2 by the fragment represented in Figure 3. This produces a Legendrian link Λ'.

The additional generator a_3 corresponds to a right cusp and has grading 1. The four mixed chords between the unknot and the trefoil have a grading that depends on a shift $k \in \mathbb{Z}$ between the Maslov potentials of the trefoil and of the unknot. These gradings are given by

$$|c_1| = k - 1, \quad |c_2| = k, \quad |d_1| = 1 - k \quad \text{and} \quad |d_2| = -k.$$

The augmentations $\varepsilon_1, \dots, \varepsilon_5$ can be extended to this enlarged DGA by sending all new generators to 0. The bilinearized differential of the original generators is therefore unchanged. The differential of the new generators is, on the other hand, given by

$$\partial c_1 = 0, \quad \partial c_2 = (1 + b_2 b_1) c_1, \quad \partial d_1 = d_2 (1 + b_2 b_1), \quad \partial d_2 = 0 \quad \text{and} \quad \partial a_3 = d_1 c_1 + d_2 c_2.$$

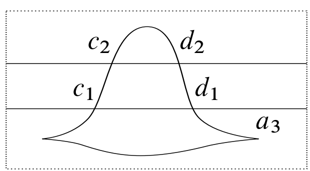


Figure 3: Replacement for the dotted rectangle in Figure 2.

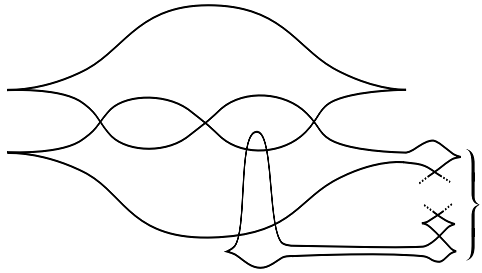


Figure 4: Front projection of the Legendrian knot Λ''.

If we choose $\varepsilon_L = \varepsilon_1$ or ε_3 and $\varepsilon_R = \varepsilon_2$, ε_4 or ε_5, then the bilinearized differential is

$$\partial^{\varepsilon_L,\varepsilon_R} c_1 = 0, \quad \partial^{\varepsilon_L,\varepsilon_R} c_2 = 0, \quad \partial^{\varepsilon_L,\varepsilon_R} d_1 = d_2, \quad \partial^{\varepsilon_L,\varepsilon_R} d_2 = 0 \quad \text{and} \quad \partial^{\varepsilon_L,\varepsilon_R} a_3 = 0.$$

The Poincaré polynomial of the resulting homology is therefore $P_{\Lambda',\varepsilon_L,\varepsilon_R}(t) = t^k + t^{k-1} + t + 1$. We now perform a connected sum between the right cusps corresponding to a_2 and a_3 in order to obtain the connected Legendrian submanifold Λ'' represented by Figure 4. A Legendrian isotopy involving a number of first Reidemeister moves is performed before the connected sum in order to ensure that the Maslov potentials agree on the cusps to be merged. This connected sum induces a Lagrangian cobordism L from Λ'' to Λ', and we can use this cobordism to pull back the augmentations ε_L and ε_R to the Chekanov–Eliashberg DGA of Λ''.

By Proposition 3.5, since $[a_3] \in \mathrm{LCH}_1^{\varepsilon_L,\varepsilon_R}(\Lambda')$ corresponds to the fundamental class of the Legendrian unknot depicted in Figure 4, we obtain the Poincaré polynomial $P_{\Lambda'',\varepsilon_L,\varepsilon_R}(t) = t^k + t^{k-1} + 1$. This corresponds to $q(t) = 1$ and $p(t) = t^k + t^{k-1}$ in Definition 4.1.

4.3 A family of Legendrian spheres with a basic BLCH Poincaré polynomial

In order to generalize Example 4.5 to higher dimensions, let us consider the standard Legendrian Hopf link, or in other words the 2–copy of the standard Legendrian unknot $\Lambda^{(2)} \subset J^1(\mathbb{R}^n)$. This will lead to a generalization of the trefoil knot from Figure 2, since it can be obtained from the standard Legendrian Hopf link in $\mathbb{R}^3$ via a connected sum. Let us denote by l the length of the unique Reeb chord of the standard Legendrian unknot and by ε the positive shift (much smaller than l) in the Reeb direction between the two components Λ_1 and Λ_2 of $\Lambda^{(2)}$. We assume that the top component is perturbed by a Morse function of amplitude δ much smaller than ε with exactly one maximum M and one minimum m. In particular, among the continuum of Reeb chords of length ε between the two components, only two chords corresponding to these extrema persist after perturbation. We also assume that thanks to this perturbation, all Reeb chords of $\Lambda^{(2)}$ lie above distinct points of $\mathbb{R}^n$. In order to define the grading of mixed Reeb chords in this link, we choose the Maslov potential of the upper component Λ_2 to be given by the Maslov potential of the lower component Λ_1 plus k.

Proposition 4.6 *The Chekanov–Eliashberg DGA of $\Lambda^{(2)} \subset J^1(\mathbb{R}^n)$ has the following six generators*:

	grading	*length*
c_{11}	n	l
c_{22}	n	l
c_{12}	$n+k$	$l+\varepsilon$
c_{21}	$n-k$	$l-\varepsilon$
m_{12}	$k-1$	$\varepsilon-\delta$
M_{12}	$n+k-1$	$\varepsilon+\delta$

Its differential is given by

$$\partial c_{12} = M_{12} + m_{12}c_{11} + c_{22}m_{12}, \quad \partial c_{11} = c_{21}m_{12} \quad \text{and} \quad \partial c_{22} = m_{12}c_{21},$$

and $\partial M_{12} = \partial m_{12} = \partial c_{21} = 0$.

Proof The front projection of each component in $\Lambda^{(2)}$ consists of two sheets, having parallel tangent hyperplanes above a single point of $\mathbb{R}^n$ before the perturbation by the Morse function. The number of Reeb chords above that point is the number of pairs of sheets, which is $\frac{1}{2}4(4-1) = 6$. The chords between the two highest or the two lowest sheets belong to a continuum of chords of length ε between the two components, which is replaced by two chords M_{12} for the maximum M and m_{12} for the minimum m after the perturbation by the Morse function. Their lengths are therefore $\varepsilon \pm \delta$. Their gradings are given by the Morse index of the corresponding critical point plus the difference of Maslov potentials minus one, so that we obtain $n+k-1$ and $k-1$.

The four other chords will be denoted by c_{ij}, where i numbers the component of origin for the chord and j numbers the component of the endpoint of the chord. Each of these chords corresponds to a maximum of the local difference function between the heights of the sheets it joins. We therefore obtain the announced gradings and lengths.

The link $\Lambda^{(2)}$ and its Reeb chords determine a quiver represented in Figure 5, in which each component of the link corresponds to a vertex and each Reeb chord corresponds to an oriented edge. When computing the differential of a generator, the terms to be considered correspond to paths formed by a sequence of edges in this quiver with the same origin and endpoint as the generator, with total grading one less than the grading of the generator and with total length strictly smaller than the length of the generator.

For ∂c_{12}, the only possible terms are M_{12}, $m_{12}c_{11}$ and $c_{22}m_{12}$. Such terms cannot contain c_{21} because two other chords from Λ_1 to Λ_2 would be needed as well. The resulting total length would be smaller than

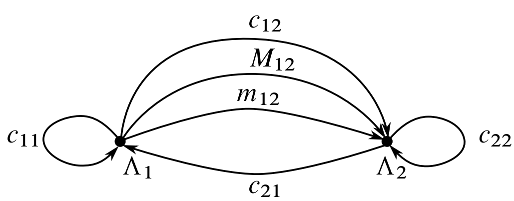


Figure 5: Quiver corresponding to the standard Hopf link.

the length of c_{12} only in the case of $m_{12}c_{21}m_{12}$, but this term is of grading 2 lower than c_{12}. The generators c_{11} and c_{22} can appear at most once due to their length, and due to total length constraint, only m_{12} can appear (only once) as a factor, leading to the possibilities $m_{12}c_{11}$ and $c_{22}m_{12}$. Finally, if M_{12} appears, then no other chord can appear as a factor by the previous discussion, leading to the possibility M_{12}.

Let us show that each possible term in ∂c_{12} is realized by exactly one Morse flow tree [5], which in turn corresponds to a unique holomorphic curve. To obtain M_{12}, we start at the chord c_{12} and follow the negative gradient of the local height difference function in the unique direction leading to the chord M_{12}. At this chord, we have a 2–valent puncture of the Morse flow tree and we continue by following the negative gradient of the local height difference function corresponding to one of the components Λ_1 or Λ_2 (depending on which hemisphere the maximum M is located on). This gradient trajectory will generically not hit any other Reeb chord and will finally hit the cusp equator of that component, which is the end of the Morse flow tree. To obtain $m_{12}c_{11}$, we start at the chord c_{12} and follow the negative gradient of the local height difference function in the unique direction leading to the chord c_{11}. At this chord, we have a 2–valent puncture of the Morse flow tree and we continue by following the negative gradient of the local height difference function corresponding to the highest two sheets, which is the Morse function used to perturb the Hopf link. Generically, this gradient trajectory will reach the minimum m so that we obtain a 1–valent puncture of the Morse flow tree at m_{12}. The term $c_{22}m_{12}$ is obtained similarly.

For ∂c_{11}, the only possible term is $c_{21}m_{12}$. Indeed, when $n > 1$, the chord c_{21} is the only one available to start an admissible path from Λ_1 to itself, because the empty path is not admissible. When $n = 1$, the empty path is admissible but there are two holomorphic disks having c_{11} as a positive puncture and no negative puncture, which cancel each other. Due to its length, the only chord we can still use is m_{12}, and after this no other chord can be added. Let us show that this possible term for ∂c_{11} is realized by exactly one Morse flow tree. We start at the chord c_{11} and follow the negative gradient of the local height difference function in the unique direction leading to the chord c_{21}. At this chord, we have a 2–valent puncture of the Morse flow tree and we continue by following the negative gradient of the local height difference function corresponding to the lowest two sheets, which is the Morse function used to perturb the Hopf link. Generically, this gradient trajectory will reach the minimum m so that we obtain a 1–valent puncture of the Morse flow tree at m_{12}. The calculation of ∂c_{22} is analogous.

For ∂c_{12}, there are no possible terms because no other chord can lead from Λ_1 to Λ_2. For ∂M_{12}, the only chord which is short enough to appear is m_{12} but its grading $k-1$ is strictly smaller when $n > 1$ than the necessary grading $n+k-2$. When $n = 1$, there are two gradient trajectories from the maximum to the minimum of a Morse function on the circle, which cancel each other. Finally, $\partial m_{12} = 0$ because it is the shortest chord and it joins different components. □

Corollary 4.7 *If $k = 1$, the Chekanov–Eliashberg DGA of $\Lambda^{(2)} \subset J^1(\mathbb{R}^n)$ has two augmentations ε_L and ε_R such that $\varepsilon_L(m_{12}) = 0$ and $\varepsilon_R(m_{12}) = 1$, and that vanish on the other Reeb chords. When $n > 1$, there are no other augmentations.*

Proof When $n > 1$, m_{12} is the only generator of degree 0, so that the maps ε_L and ε_R are the only two degree-preserving algebra morphisms $\mathcal{A} \to \mathbb{Z}_2$. In order to show that these are augmentations, we need to check that $1, m_{12} \notin \operatorname{im} \partial$. This follows from the fact that there is no term 1 and that m_{12} always appears as a factor of another generator in the expression of ∂ in Proposition 4.6. □

The above augmentations ε_L and ε_R can be used in order to obtain a bilinearized differential associated to the differential from Proposition 4.6. We obtain $\partial^{\varepsilon_L,\varepsilon_R} c_{12} = M_{12} + c_{22}$ and $\partial^{\varepsilon_L,\varepsilon_R} c_{11} = c_{21}$, while the differential of the other four generators vanishes. The corresponding homology is therefore generated by $[M_{12}] = -[c_{22}]$ in degree n and by $[m_{12}]$ in degree 0. Hence, the Poincaré polynomial $P_{\Lambda^{(2)},\varepsilon_L,\varepsilon_R}(t)$ is given by $1 + t^n$.

After this preliminary calculation, let us consider a combination of several such links in view of obtaining more general Poincaré polynomials than those in Example 4.5. To this end, consider the $2N$–copy of the standard Legendrian unknot $\Lambda^{(2N)} \subset J^1(\mathbb{R}^n)$ for $N \geq 1$. This link contains the components $\Lambda_1, \dots, \Lambda_{2N}$ numbered from bottom to top. If l denotes the length of the unique Reeb chord of Λ_i and ε denotes the positive shift between any two consecutive components, we require that $2N\varepsilon$ is much smaller than l. We perturb the component Λ_i for $i = 2, \dots, 2N$ by a Morse function f_i with two critical points and of amplitude δ much smaller than ε such that all differences $f_i - f_j$ with $i \neq j$ are Morse functions with two critical points. In order to define the gradings of mixed Reeb chords in this link, we choose the Maslov potential of the component Λ_i to be given by the Maslov potential of the lowest component Λ_1 plus $i - 1$.

A direct application of Proposition 4.6 to each pair of components Λ_i and Λ_j gives the chords of $\Lambda^{(2N)}$:

	grading	length
$c_{i,i}$	n	l
$c_{i,j}$	$n+j-i$	$l+\varepsilon(j-i)$
$c_{j,i}$	$n-j+i$	$l-\varepsilon(j-i)$
$m_{i,j}$	$j-i-1$	$\varepsilon(j-i)-\delta$
$M_{i,j}$	$n+j-i-1$	$\varepsilon(j-i)+\delta$

Here the indices i and j take all possible values between 1 and $2N$ such that $i < j$.

Proposition 4.8 *The algebra morphisms ε_L and ε_R defined by $\varepsilon_L(m_{i,i+1}) = 1$ when i is even, $\varepsilon_R(m_{i,i+1}) = 1$ when i is odd and that vanish on all other chords are augmentations of the Chekanov–Eliashberg DGA of $\Lambda^{(2N)}$.*

Proof Let us show that $m_{i,i+1} \notin \operatorname{im} \partial$ for all $i = 1, \dots, 2N-1$. If $m_{i,i+1}$ was a term in ∂a for some a in the Chekanov–Eliashberg of $\Lambda^{(2N)}$, then a would have to be a linear combination of chords from Λ_i to Λ_{i+1}. Indeed, ∂c does not contain the term 1 for any chord c of $\Lambda^{(2N)}$, say from Λ_i to Λ_j, because it would give rise to a term 1 in Proposition 4.6 for the Legendrian Hopf link composed of Λ_i and Λ_j. Therefore ∂ does not decrease the number of factors in terms it acts on. Since a must be a single chord from Λ_i to Λ_{i+1}, if there were a term $m_{i,i+1}$ in ∂a, then there would already be such a term in Proposition 4.6 for the Legendrian Hopf link composed of Λ_i and Λ_{i+1}. Hence $m_{i,i+1} \notin \operatorname{im} \partial$, as announced.

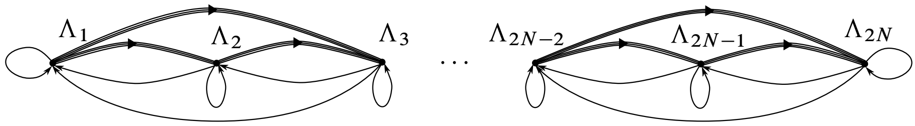


Figure 6: Quiver corresponding to the $2N$–copy of the standard Legendrian unknot.

This implies that ε_L and ε_R are augmentations, because any element of $\operatorname{im}\partial$ is composed of monomials having at least one factor which is not of the form $m_{i,i+1}$, and in particular not augmented, so that ε_L and ε_R vanish on $\operatorname{im}\partial$. ☐

Proposition 4.9 *The bilinearized differential $\partial^{\varepsilon_L,\varepsilon_R}$ of $\Lambda^{(2N)}$ is given by*

$$\partial^{\varepsilon_L,\varepsilon_R} c_{i,i} = \bar{i}\, c_{i,i-1} + \bar{i}\, c_{i+1,i}, \qquad \partial^{\varepsilon_L,\varepsilon_R} c_{i,j} = M_{i,j} + \bar{j}\, c_{i,j-1} + \bar{i}\, c_{i+1,j},$$
$$\partial^{\varepsilon_L,\varepsilon_R} c_{j,i} = \bar{i}\, c_{j,i-1} + \bar{j}\, c_{j+1,i}, \qquad \partial^{\varepsilon_L,\varepsilon_R} m_{i,j} = \bar{j}\, m_{i,j-1} + \bar{i}\, m_{i+1,j},$$
$$\partial^{\varepsilon_L,\varepsilon_R} M_{i,j} = \bar{j}\, M_{i,j-1} + \bar{i}\, M_{i+1,j},$$

with $i < j$ and where $\bar{i}$ and $\bar{j}$ are the modulo-2 reductions of i and j. In the above formulas, any generator with one of its indices equal to 0 or $2N+1$ or of the form $m_{i,i}$ or $M_{i,i}$ should be replaced by zero.

Proof The link $\Lambda^{(2N)}$ and its Reeb chords determine a quiver represented in Figure 6, and as in the proof of Proposition 4.6, the terms in the differential of a chord from Λ_i to Λ_j must form a path from vertex i to vertex j.

Let us compute $\partial^{\varepsilon_L,\varepsilon_R} c_{i,i}$. The only possible terms in $\partial c_{i,i}$ that could lead to a nonzero contribution to $\partial^{\varepsilon_L,\varepsilon_R} c_{i,i}$ are $c_{i+1,i} m_{i,i+1}$ and $m_{i-1,i} c_{i,i-1}$. Indeed, there are no other chords of Λ_i, so a change of component is needed. Since only chords of the form $m_{i,i+1}$ are augmented by ε_L and ε_R, there must be exactly one chord from Λ_j to Λ_k with $j > k$. Moreover, since neither ε_L nor ε_R augment consecutive chords in the quiver determined by $\Lambda^{(2N)}$, we must have $|j-k| = 1$ and $j = i$ or $k = i$. Considering the Legendrian Hopf link composed of Λ_i and Λ_{i+1}, Proposition 4.6 gives the term $c_{i+1,i} m_{i,i+1}$, while considering the Legendrian Hopf link composed of Λ_{i-1} and Λ_i, it gives the term $m_{i-1,i} c_{i,i-1}$. With the first term, since $m_{i,i+1}$ has to be augmented by ε_R, we obtain the contribution $c_{i+1,i}$ when i is odd. With the second term, since $m_{i-1,i}$ has to be augmented by ε_L, we obtain the contribution $c_{i,i-1}$ when $i-1$ is even. In other words, we obtain $\partial^{\varepsilon_L,\varepsilon_R} c_{i,i} = \bar{i}\, c_{i,i-1} + \bar{i}\, c_{i+1,i}$, as announced.

Let us compute $\partial^{\varepsilon_L,\varepsilon_R} c_{i,j}$ with $i < j$. All terms in $\partial c_{i,j}$ involving a single chord from Λ_i to Λ_j correspond to terms with a single factor in the expression for ∂c_{12} in Proposition 4.6. We therefore obtain the term $M_{i,j}$. The other terms must involve augmented chords; since ε_L and ε_R do not have consecutive augmented chords, these other terms could come from $m_{j-1,j} c_{i,j-1}$, $c_{i+1,j} m_{i,i+1}$, $m_{j-1,j} c_{i+1,j-1} m_{i,i+1}$ or analogous terms with $c_{k,l}$ replaced with $m_{k,l}$ or $M_{k,l}$. The last two possibilities lead to elements with a too small grading, so that the unaugmented chord is of the type $c_{k,l}$. The possibilities $m_{j-1,j} c_{i,j-1}$ and $c_{i+1,j} m_{i,i+1}$ are each realized by a single holomorphic disk, corresponding to the contribution $m_{12} c_{11} + c_{22} m_{12}$ in the expression for ∂c_{12} in Proposition 4.6. The remaining possibility

$m_{j-1,j}c_{i+1,j-1}m_{i,i+1}$ has a too small grading. Summing up, the possibility $m_{j-1,j}c_{i,j-1}$ leads to the term $c_{i,j-1}$ when j is odd and the possibility $c_{i+1,j}m_{i,i+1}$ leads to the term $c_{i+1,j}$ when i is odd, so that we obtain $\partial^{\varepsilon_L,\varepsilon_R}c_{i,j} = M_{i,j} + \bar{j}c_{i,j-1} + \bar{i}c_{i+1,j}$, as announced.

The computation of $\partial^{\varepsilon_L,\varepsilon_R}c_{j,i}$ with $i < j$ is similar. Since there are no other chords from Λ_i to Λ_j, the only contributions involve augmented chords and come from $m_{i-1,i}c_{j+1,i-1}m_{j,j+1}$, $m_{i-1,i}c_{j,i-1}$ or $c_{j+1,i}m_{j,j+1}$, The first possibility has a too small grading, while the last two possibilities are each realized by a single holomorphic disk, corresponding to the contributions $c_{21}m_{12}$ and $m_{12}c_{21}$ in the expressions for ∂c_{11} and ∂c_{22} in Proposition 4.6. The possibility $m_{i-1,i}c_{j,i-1}$ leads to the term $c_{j,i-1}$ when i is odd and the possibility $c_{j+1,i}m_{j,j+1}$ leads to the term $c_{j+1,i}$ when j is odd, so that we obtain $\partial^{\varepsilon_L,\varepsilon_R}c_{j,i} = \bar{i}c_{j,i-1} + \bar{j}c_{j+1,i}$, as announced.

The computation of $\partial^{\varepsilon_L,\varepsilon_R}m_{i,j}$ and $\partial^{\varepsilon_L,\varepsilon_R}M_{i,j}$ with $i < j-1$ involves only chords of the type $m_{k,l}$ and $M_{k,l}$ since all other chords are much longer. Let us start with $\partial^{\varepsilon_L,\varepsilon_R}m_{i,j}$. Arguing as above, since $m_{i,j}$ is the shortest chord from Λ_i to Λ_j, the only contributions involve augmented chords and come from $m_{i-1,i}m_{j,i-1}$, $m_{j+1,i}m_{j,j+1}$ or $m_{i-1,i}m_{j+1,i-1}m_{j,j+1}$. The last possibility has a too small grading, and the first two possibilities are each realized by a unique Morse flow tree [5], which in turn corresponds to a unique holomorphic curve. Both Morse flow trees start with a constant gradient trajectory at $m_{i,j}$, which is the minimum of the difference function $f_j - f_i$. The only possibility to leave $m_{i,j}$ is to have a 3–valent vertex, corresponding to the splitting of the gradient trajectory into two gradient trajectories, for $f_j - f_k$ and for $f_k - f_i$, for some k strictly between i and j. These trajectories converge to the corresponding minima $m_{k,j}$ and to $m_{i,k}$, so we obtain the desired trees for $k = i+1$ and $k = j-1$. Summing up, we obtain as above $\partial^{\varepsilon_L,\varepsilon_R}m_{j,i} = \bar{i}m_{j,i-1} + \bar{j}m_{j+1,i}$, as announced. The computation of $\partial^{\varepsilon_L,\varepsilon_R}M_{i,j}$ is completely analogous, except for the description of the Morse flow trees. Both Morse flow trees start with a gradient trajectory from $M_{i,j}$ to a priori any point of the sphere. In order to reach $M_{i+1,j}$ or $M_{i,j-1}$ it is necessary for the gradient trajectory to end exactly at the maximum of the corresponding height difference function. There, we have a 2–valent puncture of the Morse flow tree and we continue with a gradient trajectory converging to the minimum $m_{i,i+1}$ or $m_{j-1,j}$. Again, $\partial^{\varepsilon_L,\varepsilon_R}M_{j,i} = \bar{i}M_{j,i-1} + \bar{j}M_{j+1,i}$, as announced. □

Proposition 4.10 *The Poincaré polynomial of $\Lambda^{(2N)}$ with respect to the augmentations ε_L and ε_R is given by $P_{\Lambda^{(2N)},\varepsilon_L,\varepsilon_R}(t) = N(1+t^n)$.*

Proof We need to compute the homology of the complex described in Proposition 4.9.

Let us first consider the subcomplex spanned by the chords $m_{i,j}$ with $i < j$. For any $k, l = 1, \dots, N$ with $k < l-1$, the generators $m_{2k-1,2l-1}$, $m_{2k,2l-1}$, $m_{2k-1,2l-2}$ and $m_{2k,2l-2}$ form an acyclic subcomplex. When $k = l-1$, we just have a subcomplex with the three generators $m_{2l-3,2l-1}, m_{2l-2,2l-1}$ and $m_{2l-3,2l-2}$, which has homology spanned by $[m_{2l-2,2l-1}] = [m_{2l-3,2l-2}]$ in degree 0. We therefore obtain $N-1$ such generators. For any $k = 1, \dots, N-1$, the generators $m_{2k-1,2N}$ and $m_{2k,2N}$ form an acyclic subcomplex. Finally, the generator $m_{2N-1,2N}$ survives in homology and has degree 0. The total contribution of the chords $m_{i,j}$ to the polynomial $P_{\Lambda^{(2N)},\varepsilon_L,\varepsilon_R}$ is therefore the term N.

Consider now the subcomplex spanned by the chords $M_{i,j}$ with $i<j$ and $c_{i,j}$ for all $i,j=1,\dots,2N$. For any $k,l=1,\dots,N$ with $k<l-1$, the generators $c_{2k-1,2l-1}$, $c_{2k,2l-1}$, $c_{2k-1,2l-2}$, $c_{2k,2l-2}$, $M_{2k-1,2l-1}$, $M_{2k,2l-1}$, $M_{2k-1,2l-2}$ and $M_{2k,2l-2}$ form an acyclic subcomplex. When $k=l-1$, we just have a subcomplex with the seven generators $c_{2k-1,2l-1}$, $c_{2k,2l-1}$, $c_{2k-1,2l-2}$, $c_{2k,2l-2}$, $M_{2k-1,2l-1}$, $M_{2k,2l-1}$ and $M_{2k-1,2l-2}$, which has homology spanned by $c_{2l-2,2l-2}$ in degree n. We therefore obtain $N-1$ such generators. For any $k=1,\dots,N-1$, the generators $c_{2k-1,2N}$,$c_{2k,2N}$, $M_{2k-1,2N}$ and $M_{2k,2N}$ form an acyclic subcomplex. But the subcomplex spanned by the three generators $c_{2N-1,2N}$, $c_{2N,2N}$, $M_{2N-1,2N}$ has homology generated by $[c_{2N,2N}]=[M_{2N-1,2N}]$ in degree n. For any $k,l=1,\dots,N$ with $k\le l$ and $k>1$, the generators $c_{2l-1,2k-1}$, $c_{2l,2k-1}$, $c_{2l-1,2k-2}$ and $c_{2l,2k-2}$ form an acyclic subcomplex. When $k=1$, we just have an acyclic subcomplex with the 2 generators $c_{2l-1,1}$ and $c_{2l,1}$. The total contribution of the chords $M_{i,j}$ with $i<j$ and $c_{i,j}$ to the polynomial $P_{\Lambda^{(2N)},\varepsilon_L,\varepsilon_R}$ is therefore the term Nt^n.

The sum of the above two contributions therefore gives $P_{\Lambda^{(2N)},\varepsilon_L,\varepsilon_R}(t)=N(1+t^n)$, as announced. □

The next step is to perform some type of connected sum on the Legendrian link $\Lambda^{(2N)}$ in order to obtain a Legendrian sphere $\widetilde{\Lambda}^{(2N)}\subset J^1(\mathbb{R}^n)$. More precisely, for each $i=1,\dots,N-1$, we consider the Legendrian link formed by Λ_{2i-1}, Λ_{2i}, Λ_{2i+1} and Λ_{2i+2} as the 2–copy of the Legendrian link formed by Λ_{2i-1} and Λ_{2i+1}, and we perform the 2–copy of the connected sum of Λ_{2i-1} and Λ_{2i+1} as follows:

We deform Λ_{2i-1} by a Legendrian isotopy corresponding to the spinning of two iterated first Reidemeister moves on one half of the standard Legendrian unknot in $J^1(\mathbb{R})$. Since this front in $J^0(\mathbb{R})$ has a vertical symmetry axis, we can spin it around this axis to produce a Legendrian surface in $J^1(\mathbb{R}^2)$ as in [2, Section 3.2]. The resulting front has vertical symmetry planes, and hence is spinnable around such a plane; iterating the spinning construction, we obtain the desired 2–component Legendrian link in $J^1(\mathbb{R}^n)$ with cusp edges from (the deformation of) Λ_{2i-1} and Λ_{2i+1} facing each other and having the same Maslov potentials. This is illustrated by Figure 7.

In this figure, we consider the rectangular area limited by a dashed line: its image in $J^0(\mathbb{R}^+)\subset J^0(\mathbb{R}^n)$, i.e. with all spinning angles set to zero, is a rectangular area intersecting Λ_{2i-1}, Λ_{2i}, Λ_{2i+1} and Λ_{2i+2} in the 2–copy of two cusps facing each other. We then replace a neighborhood of this rectangular area with the 2–copy of a connecting tube, as shown in Figure 8. This operation is equivalent to the 2–copy of the connected sum operation described in [2, Section 4].

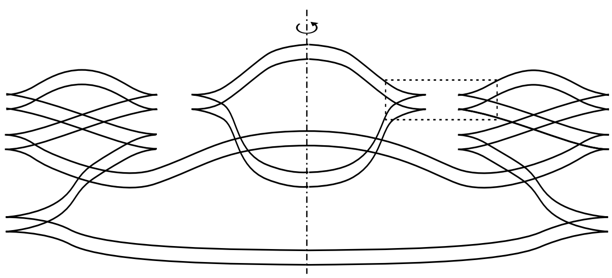


Figure 7: Isotopy of Λ_{2i-1}, Λ_{2i}, Λ_{2i+1} and Λ_{2i+2}.

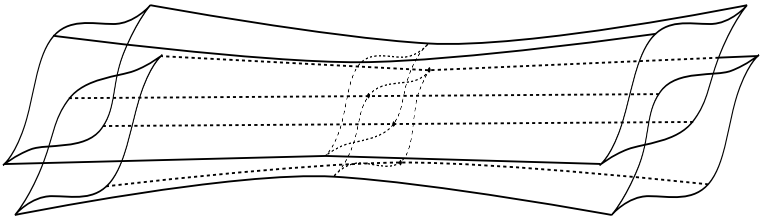


Figure 8: Double tube.

Finally, after performing $N-1$ times these 2–copies of connected sums, we are left with a Legendrian link composed of two connected components: Λ_{odd}, resulting from the connected sum of Λ_{2i-1} for $i=1,\dots,N$, and Λ_{even}, resulting from the connected sum of Λ_{2i} for $i=1,\dots,N$. We then perform an (ordinary) connected sum between these components in order to obtain the Legendrian sphere $\widetilde{\Lambda}^{(2N)}$.

Proposition 4.11 *The augmentations ε_L and ε_R of $\Lambda^{(2N)}$ induce augmentations $\tilde{\varepsilon}_L$ and $\tilde{\varepsilon}_R$ of $\widetilde{\Lambda}^{(2N)}$.*

Proof It suffices to show that an augmentation induces another augmentation after a single 2–copy of a connected sum. To this end, we describe this operation differently, in order to gain a better control on the Reeb chords during this process. Before performing the 2–copy connected sum connecting Λ_{2i-1} and Λ_{2i} to Λ_{2i+1} and Λ_{2i+2}, respectively, we deform these components by a Legendrian isotopy in order to create a pair of canceling critical points $m'_{2i-1,2i}$ of index 0 and $s_{2i-1,2i}$ of index 1 for the Morse function $f_{2i}-f_{2i-1}$, and a similar pair $m'_{2i+1,2i+2}$ and $s_{2i+1,2i+2}$ for $f_{2i+2}-f_{2i+1}$ near the attaching locus of the connecting double tube. More precisely, the chords $m'_{2i-1,2i}$ and $m'_{2i+1,2i+2}$ are contained in the small balls that are removed during the connected sums, while the chords $s_{2i-1,2i}$ and $s_{2i+1,2i+2}$ are just outside these balls. The connecting double tube is the thickening of an $(n-1)$–dimensional standard Legendrian Hopf link, and we shape each tube so that its thickness in the z–direction is minimal in the middle. We extend the Morse functions $f_{2i}-f_{2i-1}$ and $f_{2i+2}-f_{2i+1}$ by a Morse function on the connecting tube decreasing towards its middle and having exactly two critical points (of index 0 and $n-1$) in its middle slice. All Reeb chords for the connecting double tube are contained in this middle slice and correspond to the generators described in Proposition 4.6 with $k=1$ and n replaced with $n-1$:

	grading	length
$c^h_{2i-1,2i-1}$	$n-1$	$l'<l$
$c^h_{2i,2i}$	$n-1$	l'
$c^h_{2i-1,2i}$	n	$l'+\varepsilon$
$c^h_{2i,2i-1}$	$n-2$	$l'-\varepsilon$
$m^h_{2i-1,2i}$	0	$\varepsilon-\delta$
$M^h_{2i-1,2i}$	$n-1$	$\varepsilon+\delta$

The last two generators correspond to the critical points of the Morse function on the connecting tube mentioned above. The unital subalgebra $\mathcal{A}^h$ generated by these six generators is a subcomplex of the Chekanov–Eliashberg DGA, because Morse flow trees are pushed towards the middle of the double connecting tube due to its shape. By Corollary 4.7, this subcomplex has two augmentations such that only $m^h_{2i-1,2i}$ is possibly augmented. On the other hand, we have $\partial s_{2i-1,2i} = m_{2i-1,2i} + m'_{2i-1,2i}$ with no other terms because the length of $s_{2i-1,2i}$ is very short. Hence, for any augmentation ε, we must have $\varepsilon(m'_{2i-1,2i}) = \varepsilon(m_{2i-1,2i})$ and this forces the choice of the augmentation for $\mathcal{A}^h$. More precisely, the map $\tilde{\varepsilon}$ induced by ε must satisfy $\tilde{\varepsilon}(m^h_{2i-1,2i}) = \varepsilon(m_{2i-1,2i})$. Similarly, arguing with $s_{2i+1,2i+2}$, we also have $\tilde{\varepsilon}(m^h_{2i-1,2i}) = \varepsilon(m_{2i+1,2i+2})$. Note that these relations are compatible since each of ε_L and ε_R have the same value on $m_{2i-1,2i}$ and $m_{2i+1,2i+2}$.

Let us check that the resulting maps $\tilde{\varepsilon}_L, \tilde{\varepsilon}_R \colon \mathcal{A}(\widetilde{\Lambda}^{(2N)}) \to \mathbb{Z}_2$ satisfy $\tilde{\varepsilon}_L \circ \partial = 0 = \tilde{\varepsilon}_R \circ \partial$. We already saw that these relations are satisfied on $\mathcal{A}^h$ as well as on $s_{2i-1,2i}$ and $s_{2i+1,2i+2}$. On any other chord c, the relation was satisfied before the 2–copy of connected sum. We claim that the augmented terms in ∂c are modified by the 2–copy of connected sum in the following way: all occurrences of $m'_{2i-1,2i}$ and $m'_{2i+1,2i+2}$ are replaced with $m^h_{2i-1,2i}$. In particular, the maps $\tilde{\varepsilon}_L$ and $\tilde{\varepsilon}_R$ keep the same value on these terms and the augmentation relation continues to hold after the 2–copy of connected sum.

To verify the claim, note that the region in which the 2–copy of connected sum is taking place is a trap for Morse flow trees: any portion of such a tree entering this region cannot leave it, because all relevant gradient vector fields are pointing inwards. We only have to consider augmented terms, since these are the only ones that could harm the augmentation relation. We first consider an augmented term that contains neither $m'_{2i-1,2i}$ nor $m'_{2i+1,2i+2}$. If the corresponding Morse flow tree enters the region in which the 2–copy of connected sum is taking place, it must end at a cusp edge. Moreover, it cannot contain any trivalent vertex, otherwise it would not be rigid. Hence, it is a single gradient trajectory ending at a cusp edge. After the 2–copy of connected sum, it becomes another gradient trajectory, also ending at a cusp edge. Hence the corresponding term is not affected by the 2–copy of connected sum. Consider now an augmented term containing $m'_{2i-1,2i}$ or $m'_{2i+1,2i+2}$. A rigid Morse flow tree cannot have a 2–valent negative puncture at such a chord, since it is a minimum of the Morse function $f_{2i} - f_{2i-1}$ or $f_{2i+2} - f_{2i+1}$ [5, Lemma 3.7], so that these chords are 1–valent negative punctures. The only other way a fragment of Morse flow tree contained in the region in which the 2–copy of connected sum is taking place can end is at a cusp edge. As above, it cannot contain any trivalent vertex, otherwise it would not be rigid. Hence, it is a single gradient trajectory ending at a minimum $m'_{2i-1,2i}$ or $m'_{2i+1,2i+2}$. After the 2–copy of connected sum, it becomes another gradient trajectory, also ending at a minimum $m^h_{2i-1,2i}$. Conversely, consider an augmented term containing $m^h_{2i-1,2i}$ after the 2–copy of connected sum. In particular, the corresponding Morse flow tree can only end at the chord $m^h_{2i-1,2i}$ (at a 1–valent negative puncture, as above) or at a cusp edge. For the same reason as above, such a rigid tree cannot contain a trivalent vertex in the 2–copy of the connecting tube. Hence, it is just a single gradient trajectory ending at $m^h_{2i-1,2i}$. If we remove the 2–copy of the connecting tube and replace it with the regions containing the minima $m'_{2i-1,2i}$ and $m'_{2i+1,2i+2}$, this gradient trajectory is replaced with a single gradient trajectory ending at one of these minima. In other words, such an augmented term involving $m^h_{2i-1,2i}$ always comes from the substitution of $m'_{2i-1,2i}$ and $m'_{2i+1,2i+2}$ with $m^h_{2i-1,2i}$, proving the claim. □

We are now in position to show that these 2–copies of connected sums destroy almost all terms in the Poincaré polynomial for bilinearized LCH.

Proposition 4.12 *The Poincaré polynomial* $P_{\widetilde{\Lambda}^{(2N)},\tilde{\varepsilon}_L,\tilde{\varepsilon}_R}$ *is equal to* 1.

Proof Let us show by induction that, after applying i successive 2–copies of connected sums on $\Lambda^{(2N)}$, its Poincaré polynomial is given by $(N-k)(1+t^n)$. Proposition 4.10 corresponds to the case $i=0$. As shorthand, we denote by C_* the BLCH chain complex after $i-1$ successive 2–copies of connected sums, and by $\widetilde{C}_*$ the BLCH chain complex after i successive 2–copies of connected sums. Using the description of the i^{th} 2–copy of connected sum in the proof of Proposition 4.11, we see that this operation has two effects on the complex C_*. First, the generators $m'_{2i-1,2i}$ and $m'_{2i+1,2i+2}$ are removed. Second, we add generators of the BLCH complex C_*^h of the $(n-1)$–dimensional standard Legendrian Hopf link with distinct augmentations. Recall that C_*^h forms a subcomplex of $\widetilde{C}_*$ (see the proof of Proposition 4.11).

Since the 2–copy of connected sum is performed away from rigid holomorphic disks connecting generators of $\widetilde{C}_*/C_*^h$, the differential on this quotient complex is directly induced from that of C_*. In particular, we have $\partial s_{2i-1,2i}=m_{2i-1,2i}$ and $\partial s_{2i+1,2i+2}=m_{2i+1,2i+2}$ in $\widetilde{C}_*/C_*^h$. Hence, its homology coincides with the homology of C_*, except in degree 0, where it has two fewer generators. Hence, its Poincaré polynomial is $(N-i-1)+(N-i+1)t^n$. On the other hand, the homology of C_*^h is given by Proposition 4.10 with $N=1$ and n replaced with $n-1$. Hence its Poincaré polynomial is $1+t^{n-1}$.

In order to deduce the homology of $\widetilde{C}_*$, we consider the long exact sequence

$$\cdots\to H_{k+1}(\widetilde{C}_*/C_*^h)\to H_k(C_*^h)\to H_k(\widetilde{C}_*)\to H_k(\widetilde{C}_*/C_*^h)\to H_{k-1}(C_*^h)\to\cdots.$$

When $k=0$, we see that the generator $[m^h_{2i-1,2i}]$ of $H_0(C_*^h)$ injects into $H_0(\widetilde{C}_*)$, as it can only be hit by $s_{2i-1,2i}$ and $s_{2i+1,2i+2}$, but these do not survive in the homology of the quotient complex. Hence the rank of $H_0(\widetilde{C}_*)$ is $N-i$.

When $k=n$, we see that the generator $[c_{2i+2,2i+2}]$ in $H_n(\widetilde{C}_*/C_*^h)$, which was not affected by the $i-1$ first 2–copies of connected sums, hits the generator $[c^h_{2i,2i}]$ of $H_{n-1}(C_*^h)$, because there exists a single Morse flow tree connecting them. Indeed, in Figure 7 the chord $c_{2i+2,2i+2}$ is in the middle of the uppermost connected component, and the Morse flow tree starts from there to the right in the plane of the figure (corresponding to all spinning angles set to zero), then enters the dotted rectangle (hence the upper tube in Figure 8), until it reaches the chord $c^h_{2i,2i}$ sitting in the middle of that tube. Hence, the rank of $H_n(\widetilde{C}_*)$ is $N-i$. The Poincaré polynomial for the homology of $\widetilde{C}_*$ is therefore $(N-i)(1+t^n)$, as announced.

After these $N-1$ operations, we are therefore left with the Poincaré polynomial $1+t^n$. The last step in the construction of $\widetilde{\Lambda}^{(2N)}$ is an ordinary connected sum between the remaining two connected components Λ_{even} (the connected sum of Λ_{2i} for $i=1,\dots,N$) and Λ_{odd} (the connected sum of Λ_{2i-1} for $i=1,\dots,N$). Let us denote the corresponding 2–component Legendrian link by Λ'.

As in the proof of Proposition 3.2, the map $\tilde{\tau}_0$ from the duality exact sequence (2-2) with $\varepsilon_1=\tilde{\varepsilon}_R$ and $\varepsilon_2=\tilde{\varepsilon}_L$ is given at chain level by $\tilde{\varepsilon}_R-\tilde{\varepsilon}_L$, except that we must refine according to the connected component

Λ_{even} or Λ_{odd} which is hit. Note that all chords augmented by $\tilde{\varepsilon}_L$ start on Λ_{odd} and all chords augmented by $\tilde{\varepsilon}_R$ end on Λ_{odd}. This means that $\tilde{\tau}_0$ necessarily takes its values in $H_0(\Lambda_{\text{odd}})$. By Proposition 4.4, since $P_{\Lambda',\tilde{\varepsilon}_L,\tilde{\varepsilon}_R}(t) = 1 + t^n$ and $H_*(\Lambda')$ has rank 4, we must have $p = 0$, and hence $P_{\Lambda',\tilde{\varepsilon}_R,\tilde{\varepsilon}_L}(t) = 1 + t^n$ as well. Therefore, the image of the map $\tilde{\tau}_0 \colon \mathrm{LCH}_0^{\tilde{\varepsilon}_R,\tilde{\varepsilon}_L}(\Lambda') \to H_0(\Lambda')$ is equal to $H_0(\Lambda_{\text{odd}})$.

We deduce that $\ker \tilde{\sigma}_n = H_0(\Lambda_{\text{odd}})$ in the duality exact sequence (2-2) with $\varepsilon_1 = \tilde{\varepsilon}_R$ and $\varepsilon_2 = \tilde{\varepsilon}_L$. Consider now the map τ_n in the duality exact sequence (2-2) with $\varepsilon_1 = \tilde{\varepsilon}_L$ and $\varepsilon_2 = \tilde{\varepsilon}_R$. Since $\tilde{\sigma}_n$ and τ_n are adjoint in the sense of [6, Proposition 3.9], $\operatorname{im} \tau_n$ is the annihilator of $H_0(\Lambda_{\text{odd}})$ for the intersection pairing, which is $H_n(\Lambda_{\text{even}})$. In particular, the map $\tau_{n,1} - \tau_{n,2} = \tau_{n,\text{odd}} - \tau_{n,\text{even}}$ from Proposition 3.5 does not vanish, so that this last connected sum modifies the Poincaré polynomial by $-t^n$. We are therefore left with $P_{\widetilde{\Lambda}^{(2N)},\tilde{\varepsilon}_L,\tilde{\varepsilon}_R}(t) = 1$, as announced. □

4.4 Geography of BLCH for Legendrian spheres

The next step in our construction is to add to $\widetilde{\Lambda}^{(2N)}$ a standard Legendrian unknot Λ_0 which forms with the bottom k components $\Lambda_1, \dots, \Lambda_k$ a Legendrian link isotopic to the $(k+1)$–copy of the standard Legendrian unknot, but which is unlinked with the $2N - k$ top components $\Lambda_{k+1}, \dots, \Lambda_{2N}$. We fix the Maslov potential of the component Λ_0 to be given by the Maslov potential of Λ_1 plus $m - 1$, for some integer m. We can deform this link by a Legendrian isotopy in order to widen the components $\Lambda_1, \dots, \Lambda_k \subset J^1(\mathbb{R}^n)$ so that their projection to $\mathbb{R}^n$ becomes much larger than the projection of the components $\Lambda_{k+1}, \dots, \Lambda_{2N}$. We further narrow the component Λ_0 so that its projection to $\mathbb{R}^n$ does not intersect the projection of the components $\Lambda_{k+1}, \dots, \Lambda_{2N}$. We denote the resulting Legendrian link by $\widetilde{\Lambda}^{(2N)}_{(k,m)}$.

The addition of Λ_0 to $\widetilde{\Lambda}^{(2N)}$ is illustrated by Figure 9 in the case $k = 4$, where the picture zooms in on the bottom strata of the k components $\Lambda_1, \dots, \Lambda_k$, which are represented as portions of horizontal planes.

This Legendrian link $\widetilde{\Lambda}^{(2N)}_{(k,m)}$ has several additional Reeb chords compared to $\widetilde{\Lambda}^{(2N)}$. These are easily identified within the $(k+1)$–copy of the standard Legendrian unknot formed by $\Lambda_0, \Lambda_1, \dots, \Lambda_k$ and are given by

	grading	length
$c_{0,0}$	n	l
$c_{0,j}$	$n + j - m$	$l + \varepsilon j$
$c_{j,0}$	$n - j + m$	$l - \varepsilon j$
$m_{0,j}$	$j - m - 1$	$\varepsilon j - \delta$
$M_{0,j}$	$n + j - m - 1$	$\varepsilon j + \delta$

where the index j takes all possible values between 1 and k.

We extend the augmentations $\tilde{\varepsilon}_L$ and $\tilde{\varepsilon}_R$ by zero on these additional chords in order to define augmentations, still denoted by $\tilde{\varepsilon}_L$ and $\tilde{\varepsilon}_R$, on the Chekanov–Eliashberg DGA of $\widetilde{\Lambda}^{(2N)}_{(k,m)}$. Since the mixed chords involving Λ_0 are not augmented, it follows that the vector space generated by the above chords is a direct summand of the bilinearized complex with respect to the differential $\partial^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R}$.

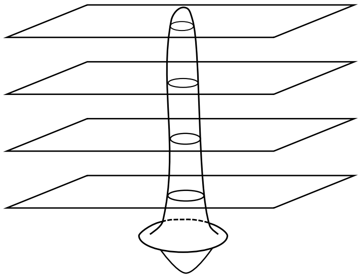


Figure 9: Additional component Λ_0 with $k=4$.

Proposition 4.13 *The bilinearized differential $\partial^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R}$ of $\widetilde{\Lambda}^{(2N)}_{(k,m)}$ on the subcomplex generated by the chords involving the component Λ_0 is given by*

$$\partial^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R} c_{0,0} = 0, \qquad \partial^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R} c_{0,j} = M_{0,j} + \bar{j} c_{0,j-1}, \quad \partial^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R} c_{j,0} = \bar{j} c_{j+1,0},$$
$$\partial^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R} m_{0,j} = \bar{j} m_{0,j-1}, \quad \partial^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R} M_{0,j} = \bar{j} M_{0,j-1},$$

for $j=1,\dots,k$, where $\bar{j}$ is the modulo-2 reduction of j and where in the right-hand sides $c_{k+1,0}$, $c_{0,0}$, $m_{0,0}$ and $M_{0,0}$ should be replaced by zero.

Proof This result follows from the same computations as in Proposition 4.9, in which we replace $2N$ with k, i with 0 and where all terms obtained by changing the index i are omitted since the mixed Reeb chords involving Λ_0 are not augmented. □

Proposition 4.14 *Consider the Legendrian link $\widetilde{\Lambda}^{(2N)}_{(k,m)} \subset J^1(\mathbb{R}^n)$. Its Poincaré polynomial with respect to the augmentations $\tilde{\varepsilon}_L$ and $\tilde{\varepsilon}_R$ is given by*

$$P_{\widetilde{\Lambda}^{(2N)}_{(k,m)},\tilde{\varepsilon}_L,\tilde{\varepsilon}_R}(t) = 1 + t^n + t^{-m} + t^a,$$

where

$$a = \begin{cases} k-m-1 & \text{if } k \text{ is even,} \\ n-k+m & \text{if } k \text{ is odd.} \end{cases} \tag{4-2}$$

Proof Let us compute the homology of the subcomplex generated by all Reeb chords involving the component Λ_0. First note that $c_{0,0}$ is always a generator in homology, leading to the term t^n in the Poincaré polynomial. Moreover, the complex generated by the chords $c_{0,1},\dots,c_{0,k}$ and $M_{0,1},\dots,M_{0,k}$ is acyclic.

If k is even, the complex generated by the chords $c_{1,0},\dots,c_{k,0}$ is acyclic. On the other hand, the complex generated by the chords $m_{0,1},\dots,m_{0,k}$ has its homology generated by $m_{0,1}$ and $m_{0,k}$. These lead to the terms t^{-m} and t^{k-m-1} in the Poincaré polynomial.

If k is odd, the complex generated by the chords $c_{1,0},\dots,c_{k,0}$ has its homology generated by $c_{k,0}$. This leads to the term t^{n-k+m} in the Poincaré polynomial. On the other hand, the complex generated by the chords $m_{0,1},\dots,m_{0,k}$ has its homology generated by $m_{0,1}$. This leads to the term t^{-m} in the Poincaré polynomial.

Adding these contributions to the Poincaré polynomial of $\widetilde{\Lambda}^{(2N)}$ from Proposition 4.12, we obtain the announced result. □

Remark 4.15 As a variant of the above construction, if we choose Λ_0 to be unlinked with Λ_1 in addition to $\Lambda_{k+1}, \dots, \Lambda_{2N}$, then we obtain instead the Poincaré polynomial $1+t^n+t^{n+m-2}+t^a$ with the same a as in Proposition 4.14. This is because the subcomplex generated by all Reeb chords involving the component Λ_0 considered in the above proof does not contain the generators $c_{1,0}$ and $m_{0,1}$ anymore. Therefore, when k is even its homology is generated by $c_{2,0}$ and $m_{0,k}$, and when k is odd it is generated by $c_{2,0}$ and $c_{k,0}$. Hence, in the Poincaré polynomial the exponent $-m = |m_{0,1}|$ is replaced with $n+m-2 = |c_{2,0}|$.

The next step in our construction is to perform a connected sum between the component Λ_0 and the original knot $\widetilde{\Lambda}^{(2N)}$. This can be done after a Legendrian isotopy of Λ_0 similar to the one depicted in Figure 7, so that a piece of cusp in the deformed Λ_0 faces a piece of cusp from the component Λ_1. In this case, it will be necessary to use a different number of first Reidemeister moves as in Figure 4 before spinning the resulting front, so that the Maslov potentials near the facing cusps agree. We denote by $\overline{\Lambda}^{(2N)}_{(k,m)}$ the resulting Legendrian knot in $J^1(\mathbb{R}^n)$. We denote by $\bar{\varepsilon}_L$ and $\bar{\varepsilon}_R$ the augmentations induced from $\tilde{\varepsilon}_L$ and $\tilde{\varepsilon}_R$ via the exact Lagrangian cobordism between $\overline{\Lambda}^{(2N)}_{(k,m)}$ and $\widetilde{\Lambda}^{(2N)}_{(k,m)}$.

Proposition 4.16 *Consider the Legendrian knot $\overline{\Lambda}^{(2N)}_{(k,m)} \subset J^1(\mathbb{R}^n)$. We have*

$$P_{\overline{\Lambda}^{(2N)}_{(k,m)},\bar{\varepsilon}_L,\bar{\varepsilon}_R}(t) = 1 + t^{-m} + t^a,$$

where a is given by (4-2).

Proof By Proposition 4.14, the generator $[c_{0,0}] \in \mathrm{LCH}_n^{\tilde{\varepsilon}_L,\tilde{\varepsilon}_R}(\widetilde{\Lambda}^{(2N)}_{(k,m)})$ corresponds to the fundamental class $[\Lambda_0]$ of the component Λ_0 of the Legendrian link $\widetilde{\Lambda}^{(2N)}_{(k,m)}$. By Proposition 3.5, the effect of the connected sum with this component is to remove the term t^n from the Poincaré polynomial, so that we obtain the announced result. □

Note that, instead of adding a single component Λ_0 to the Legendrian knot $\widetilde{\Lambda}^{(2N)}$, we can add a collection of components $\Lambda_{0,1}, \dots, \Lambda_{0,r} \subset J^1(\mathbb{R}^n)$ with similar properties. More precisely, for all $i = 1, \dots, r$, $\Lambda_{0,i}$ forms with the bottom k_i components $\Lambda_1, \dots, \Lambda_{k_i}$ a Legendrian link isotopic to the (k_i+1)–copy of the standard Legendrian unknot, but the projection of $\Lambda_{0,i}$ to $\mathbb{R}^n$ is disjoint from the projection of the other components $\Lambda_{k_i+1}, \dots, \Lambda_{2N}$. The Maslov potential of $\Lambda_{0,i}$ is fixed as the Maslov potential of Λ_1 plus $m_i - 1$, for some integer m_i. With $\bar{k} = (k_1, \dots, k_r)$ and $\bar{m} = (m_1, \dots, m_r)$, we denote the resulting Legendrian link by $\widetilde{\Lambda}^{(2N)}_{(\bar{k},\bar{m})}$.

The addition of $\Lambda_{0,1}, \dots, \Lambda_{0,r}$ to $\widetilde{\Lambda}^{(2N)}$ is illustrated by Figure 10 in the case $r = 3$ and $\{k_1, k_2, k_3\} = \{1, 3, 4\}$, where the picture zooms in on the bottom strata of the k components $\Lambda_1, \dots, \Lambda_k$, which are represented as portions of horizontal planes.

Each additional component $\Lambda_{0,i}$ gives rise to an additional subcomplex in the bilinearized complex as in Proposition 4.13, and hence to additional terms in the Poincaré polynomial of the form $t^n + t^{-m_i} + t^{a_i}$ with a_i given by (4-2). After the connected sum of these components with $\widetilde{\Lambda}^{(N)}$, we obtain a Legendrian

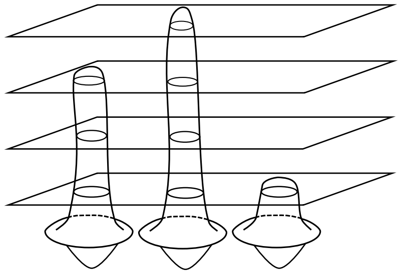


Figure 10: Additional components $\Lambda_{0,i}$ with $r=3$ and $\{k_1,k_2,k_3\}=\{1,3,4\}$.

knot $\overline{\Lambda}^{(2N)}_{(k,m)}$ and, arguing as in Proposition 4.16, its Poincaré polynomial is given by

(4-3) $$P_{\overline{\Lambda}^{(2N)}_{(\bar k,\bar m)},\bar\varepsilon_L,\bar\varepsilon_R}(t)=1+\sum_{i=1}^{r}(t^{-m_i}+t^{a_i}).$$

At this point of our constructions we have realized the geography of BLCH for Legendrian spheres Λ.

Theorem 4.17 *Let $P=q+p$ be the sum of Laurent polynomials with nonnegative integral coefficients satisfying conditions (i′) and (ii′) from Remark 4.3. Then there exists a Legendrian sphere Λ in $J^1(\mathbb{R}^n)$ and two non-DGA homotopic augmentations ε_1 and ε_2 of the Chekanov–Eliashberg DGA of Λ, with the property that the Poincaré polynomial of $\mathrm{LCH}^{\varepsilon_1,\varepsilon_2}(\Lambda)$ with coefficients in $\mathbb{Z}_2$ is equal to P.*

Proof Let us show that the Poincaré polynomials obtained in (4-3) realize all polynomials $P=q+p$ satisfying conditions (i′) and (ii′).

Indeed, let $q(t)=1$ and p be a Laurent polynomial satisfying (ii′). If n is even, $p(-1)=0$, so the polynomial p can be expressed as a sum of polynomials of the form $\sum_{i=1}^r(t^{u_i}+t^{v_i})$, where $u_i<v_i$ have different parities. If n is odd, $p(-1)$ is even, so the polynomial p can be expressed as the sum of polynomials of the form $\sum_{i=1}^r(t^{u_i}+t^{v_i})$, with no parity conditions on u_i and v_i.

In order to realize the polynomial $t^{u_i}+t^{v_i}$ when u_i and v_i have different parities, we can choose $m_i=-u_i$ and $k_i=v_i-u_i+1$, which is even. When u_i and v_i have the same parity, which can happen only if n is odd, we proceed as follows. If $u_i+v_i\le n-1$, we can choose $m_i=-u_i$ and $k_i=n-u_i-v_i$, which is odd. If $u_i+v_i\ge n-1$, we use the variant of the construction with Λ_0 described in Remark 4.15 with $m_i=u_i+2-n$ and $k_i=u_i+v_i+3-n$, which is even.

Let us define $\bar k=(k_1,\dots,k_r)$ and $\bar m=(m_1,\dots,m_r)$, and let N be the smallest even integer such that $k_i\le 2N$ for all $i=1,\dots,r$. Then, in view of (4-3), the Legendrian sphere $\overline{\Lambda}^{(2N)}_{(\bar k,\bar m)}$ satisfies

$$P_{\overline{\Lambda}^{(2N)}_{(\bar k,\bar m)},\bar\varepsilon_L,\bar\varepsilon_R}(t)=1+p(t)=q(t)+p(t),$$

as desired. □

4.5 Geography of BLCH for general Legendrian submanifolds

In order to obtain Poincaré polynomials with all possible polynomials q satisfying condition (i) from Definition 4.1, we use the following construction from [2, Corollary 6.7]:

Proposition 4.18 *For any monic polynomial $\bar{q}$ of degree n satisfying $\bar{q}(0)=0$, there exists a connected Legendrian submanifold $\Lambda_{\bar{q}} \subset J^1(\mathbb{R}^n)$ equipped with an augmentation ε such that $P_{\Lambda_{\bar{q}},\varepsilon} = \bar{q}$.*

If q is a polynomial satisfying condition (i) from Definition 4.1, then the polynomial $\bar{q}$ given by $\bar{q}(t) = q(t) + t^n - 1$ satisfies the assumptions of Proposition 4.18.

Let $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$ be the disjoint union of the Legendrian knots $\overline{\Lambda}^{(2N)}_{(\bar{k},\bar{m})}$ and $\Lambda_{\bar{q}}$ such that the projections of these components to $\mathbb{R}^n$ are disjoint. We denote by $\hat{\varepsilon}_L$ and $\hat{\varepsilon}_R$ the augmentations for $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$ induced by the augmentation ε for $\Lambda_{\bar{q}}$ and the augmentations $\bar{\varepsilon}_L$ and $\bar{\varepsilon}_R$ for $\overline{\Lambda}^{(2N)}_{(\bar{k},\bar{m})}$. The Poincaré polynomial of $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$ is given by the sum of the Poincaré polynomials of its components:

$$P_{\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})},\hat{\varepsilon}_L,\hat{\varepsilon}_R}(t) = t^n + q(t) + \sum_{i=1}^{r} (t^{-m_i} + t^{a_i}).$$

We then perform a connected sum on the Legendrian link $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$ in order to obtain a Legendrian knot $\widetilde{\Lambda}^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$, equipped with two augmentations still denoted by $\hat{\varepsilon}_L$ and $\hat{\varepsilon}_R$. Since the augmentations $\bar{\varepsilon}_L$ and $\bar{\varepsilon}_R$ coincide (with ε) on the component $\Lambda_{\bar{q}}$, by Proposition 3.2 the fundamental class $[\Lambda_{\bar{q}}]$ of this component is in the image of the map τ_n in the duality exact sequence (2-2). By Proposition 3.5, the effect of the connected sum with $\Lambda_{\bar{q}}$ is to remove a term t^n from the Poincaré polynomial. We therefore obtain

$$P_{\widetilde{\Lambda}^{(2N)}_{\bar{q},(\bar{k},\bar{m})},\hat{\varepsilon}_L,\hat{\varepsilon}_R}(t) = q(t) + \sum_{i=1}^{r} (t^{-m_i} + t^{a_i}).$$

Although these Poincaré polynomials realize all polynomials q satisfying condition (i) from Definition 4.1, we are still missing some Laurent polynomials p, since these can be arbitrary when $n > 2$. In order to realize these more general Laurent polynomials p, we describe a generalization of the embedded surgery construction on which Proposition 4.18 and its proof in [2, Corollary 6.7] are based.

From now on, assume that $n \geq 2$. Consider a point on the cusp locus of the component Λ_1 of the $2N$–copy of the standard Legendrian unknot $\Lambda^{(2N)} \subset J^1(\mathbb{R}^n)$. By a Legendrian isotopy, it is always possible to arrange so that, in a neighborhood of this point, the front of $\Lambda^{(2N)}$ in $J^0(\mathbb{R}^n)$ with local coordinates $(x_1, \dots, x_n, z)$ is locally described as follows: the fragment of Λ_1 in this neighborhood is composed of a bottom stratum $z=0$ and of a top stratum satisfying $z^2 = x_n^3$, both for $x_n \geq 0$. Moreover, the fragments of the bottom strata of the components Λ_i in this neighborhood satisfy $z = (i-1)\varepsilon$ for $i = 2, \dots, 2N$, and no other parts of the front of $\Lambda^{(2N)}$ lie in this neighborhood. Note that it is possible to arrange so that this local model still holds for the more sophisticated Legendrian $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$ after our above constructions.

For a given $m' \in \{0, \dots, n-2\}$, we consider an embedded sphere $S^{m'}$ of dimension m' in the cusp locus $\{x_n = z = 0\}$ of Λ_1. In view of our assumptions on the front of $\Lambda^{(2N)}$, this sphere bounds an embedded disk of dimension $m'+1$ with its interior disjoint from the front of $\Lambda^{(2N)}$. For a given $k' \in \{2, \dots, 2N\}$, we define a function f on the cusp locus of Λ_1, equal to $\big((k'+\tfrac{2}{3})\varepsilon\big)^{2/3}$ along $S^{m'}$, given by $\big(\big((k'+\tfrac{2}{3})\varepsilon\big)^{2/3}/r_0^{1/2}\big)\sqrt{r_0 - r}$ at distance $r \in (0, r_0]$ from $S^{m'}$ and extended by 0 everywhere else. We

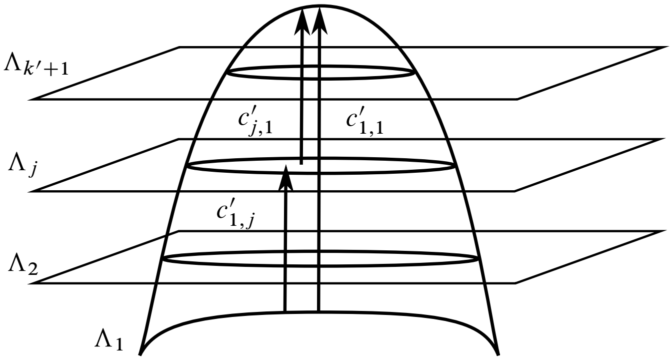


Figure 11: Center of a generalized handle.

remove from the front of Λ_1 the region satisfying $x_n < f(x_1, \dots, x_{n-1})$; the resulting front has boundary diffeomorphic to the cartesian product of $S^{m'}$ with a standard Legendrian sphere of dimension $n-m'-1$, with a flat bottom stratum. We now perform an m'–surgery on $\Lambda^{(2N)}$ by attaching a standard Legendrian handle diffeomorphic to $D^{m'+1} \times S^{n-m'-1}$ to the above front along its boundary. By construction, along the boundary of this handle, the standard Legendrian sphere of dimension $n-m'-1$ has height $\left(k'+\frac{2}{3}\right)\varepsilon$. We shape the handle so that this height decreases monotonically from the boundary of $D^{m'+1}$ to its center, where it takes the minimal value $\left(k'+\frac{1}{3}\right)\varepsilon$. This is a standard Legendrian surgery on Λ_1, but it is of a more general nature if we consider the whole $\Lambda^{(2N)}$, since the front of the attached handle intersects the front of the components $\Lambda_2, \dots, \Lambda_{k'+1}$ (but not of the components $\Lambda_{k'+2}, \dots, \Lambda_{2N}$). When this operation is performed on the Legendrian submanifold $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$, we denote the resulting Legendrian submanifold by $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')}$.

In order to minimize the number of Reeb chords created by this operation, we shape the standard Legendrian sphere of dimension $n-m'-1$ as shown in Figure 11, with both top and bottom strata being the graphs of concave functions. Assuming for simplicity that the minima of the perturbing Morse functions $f_i - f_j$ for $i \neq j$ are located in the bottom strata and that the corresponding maxima are located in the top strata, the bottom strata of the Λ_i are slightly moving away from each other in the z–direction as x_n decreases to 0. Hence, the bottom stratum of the standard Legendrian sphere of dimension $n-m'-1$ is slightly moving down from the boundary of $D^{m'+1}$ to its center. In particular, all new Reeb chords are located very close to the center of the handle: $c'_{1,1}$ with endpoints on the handle, $c'_{1,j}$ from the handle to Λ_j and $c'_{j,1}$ from Λ_j to the handle, for $j = 2, \dots, k'+1$, as shown in Figure 11. On the other hand, we can perturb the resulting Legendrian submanifold so that there are no Reeb chords between the attached handle and the components $\Lambda_{k'+2}, \dots, \Lambda_{2N}$. Summarizing, the gradings and lengths of the new Reeb chords are given by

	grading	length
$c'_{1,1}$	$n-m'-1$	$\left(k'+\frac{1}{3}\right)\varepsilon$
$c'_{j,1}$	$n-m'-j$	$\left(k'-j+\frac{4}{3}\right)\varepsilon$
$c'_{1,j}$	$m'+j-1$	$(j-1)\varepsilon$

Proposition 4.19 *The augmentations $\hat\varepsilon_L$ and $\hat\varepsilon_R$ can be extended by zero on the new chords to augmentations of $\Lambda^{(2N)}_{\bar q,(\bar k,\bar m),(k',m')}$. The vector space spanned by the new chords $c'_{1,1}$, $c'_{1,j}$ and $c'_{j,1}$ is a subcomplex with respect to the bilinearized differentials $\partial^{\hat\varepsilon_L,\hat\varepsilon_R}$ and $\partial^{\hat\varepsilon_R,\hat\varepsilon_L}$. These differentials are given by*

$$\partial^{\hat\varepsilon_L,\hat\varepsilon_R} c'_{1,j+1} = \overline{j+1}\, c'_{1,j}, \quad \partial^{\hat\varepsilon_L,\hat\varepsilon_R} c'_{j,1} = \bar j\, c'_{j+1,1} \quad \text{and} \quad \partial^{\hat\varepsilon_L,\hat\varepsilon_R} c'_{k'+1,1} = 0,$$

and respectively by

$$\partial^{\hat\varepsilon_R,\hat\varepsilon_L} c'_{1,j+1} = \begin{cases} \bar j\, c'_{1,j} & \text{if } j \neq 1,\\ 0 & \text{if } j = 1,\end{cases} \qquad \partial^{\hat\varepsilon_R,\hat\varepsilon_L} c'_{j,1} = \overline{j+1}\, c'_{j+1,1} \qquad \text{and} \qquad \partial^{\hat\varepsilon_R,\hat\varepsilon_L} c'_{k'+1,1} = 0$$

for $j = 1, \dots, k'$, where $\bar j$ is the modulo-2 reduction of j.

Proof We first show that $\hat\varepsilon_L \circ \partial c = \hat\varepsilon_R \circ \partial c = 0$ for any Reeb chord c of $\Lambda^{(2N)}_{\bar q,(\bar k,\bar m),(k',m')}$. If c is a Reeb chord of $\Lambda^{(2N)}_{\bar q,(\bar k,\bar m)}$, then ∂c consists of terms from the differential for $\Lambda^{(2N)}_{\bar q,(\bar k,\bar m)}$, and hence in the kernel of $\hat\varepsilon_L$ and $\hat\varepsilon_R$, and of terms involving at least one new chord of $\Lambda^{(2N)}_{\bar q,(\bar k,\bar m),(k',m')}$. Since $\hat\varepsilon_L$ and $\hat\varepsilon_R$ vanish on these new chords, we obtain the desired relations.

If c is a new chord of $\Lambda^{(2N)}_{\bar q,(\bar k,\bar m),(k',m')}$, we claim that any term in ∂c contains an unaugmented chord as a factor, and hence is in the kernel of $\hat\varepsilon_L$ and $\hat\varepsilon_R$. Indeed, the only augmented chords go from Λ_j to Λ_{j+1}, with a parity condition on j depending on the augmentation. Moreover, Morse flow trees cannot entirely go across a connecting tube (since they are attracted to its center), so chords are the only way to jump from Λ_i to Λ_j with $i \neq j$. Since the new chords have at least one endpoint on Λ_1, if a Morse flow tree has all negative ends at augmented chords, it must start at $c'_{1,1}$ or at $c'_{1,2}$. But $|c'_{1,1}| = n - m' - 1$ equals 1 if and only if $m' = n-2$, and in that case a Morse flow tree with endpoints remaining on Λ_1 must remain in the center of the handle, which is a 1–dimensional standard Legendrian knot, so that there are 2 such Morse flow trees with no negative end, canceling each other. On the other hand, $|c'_{1,2}| = m'+1$ equals 1 if and only if $m' = 0$, and in that case a Morse flow tree with endpoints remaining on Λ_1 and Λ_2 must connect the critical point $c'_{1,2}$ of $f_2 - f_1$ of index 1 to the critical point $m_{1,2}$ of $f_2 - f_1$ of index 0. There are two such Morse flow trees, corresponding to the two sides of the 1–dimensional unstable manifold of $c'_{1,2}$, and these cancel each other.

Let us now compute the bilinearized differentials. If a rigid Morse flow tree starting at $c'_{1,j}$ with $j = 1, \dots, k'+1$, has only one negative end, it will leave the handle radially and then flow to the minimum $m_{1,j}$ of $f_j - f_1$. Such a configuration is rigid if and only if $|m_{1,j}| = j-2 = |c'_{1,j}| - 1 = m' + j - 2$, but when $m' = 0$ there are two such Morse flow trees as above, canceling each other. If it has more negative ends and contributes to the bilinearized differential of $c'_{1,j}$, it can only have a negative end at $m_{j-1,j}$, and the other one must then be at $c'_{1,j-1}$. There is a unique such Morse flow tree, flowing from $c'_{1,j}$ to the position of $c'_{1,j-1}$ in the $D^{m'+1}$–factor of the handle, then splitting at the bottom stratum of Λ_{j-1}, so that one part flows in the $S^{n-m'-1}$–factor of the handle to $c'_{1,j-1}$ and the other part flows to the minimum $m_{j-1,j}$ of $f_j - f_{j-1}$. This term $m_{j-1,j} c'_{1,j-1}$ gives rise to the term $c'_{1,j-1}$ in $\partial^{\hat\varepsilon_L,\hat\varepsilon_R} c'_{1,j}$ if and only if $\hat\varepsilon_L(m_{j-1,j}) = 1$, i.e. when j is odd and > 1. It gives rise to the term $c'_{1,j-1}$ in $\partial^{\hat\varepsilon_R,\hat\varepsilon_L} c'_{1,j}$ if and only if $\hat\varepsilon_R(m_{j-1,j}) = 1$, i.e. when j is even.

Let us now consider a rigid Morse flow tree starting at $c'_{j,1}$ with $j = 2, \dots, k'+1$. Such a tree cannot have only one negative end, and if it contributes to the bilinearized differential of $c'_{j,1}$, it must have two negative ends, one at $m_{j,j+1}$ and the other one at $c'_{j+1,1}$. There is a unique such Morse flow tree, flowing from $c'_{j,1}$ to the position of $c'_{j+1,1}$ in the $S^{n-m'-1}$–factor of the handle, then splitting at the bottom stratum of Λ_{j+1}, so that one part flows in the $D^{m'+1}$–factor of the handle to $c'_{j+1,1}$ and the other part flows to the minimum $m_{j,j+1}$ of $f_{j+1} - f_j$. This term $c'_{j+1,1} m_{j,j+1}$ gives rise to the term $c'_{j+1,1}$ in $\partial^{\hat{\varepsilon}_L,\hat{\varepsilon}_R} c'_{j,1}$ if and only if $\hat{\varepsilon}_R(m_{j,j+1}) = 1$, i.e. when j is odd and $< k'+1$. It gives rise to the term $c'_{j+1,1}$ in $\partial^{\hat{\varepsilon}_R,\hat{\varepsilon}_L} c'_{j,1}$ if and only if $\hat{\varepsilon}_L(m_{j,j+1}) = 1$, i.e. when j is even and $< k'+1$. □

As an immediate consequence of Proposition 4.19, the homology with respect to $\partial^{\hat{\varepsilon}_L,\hat{\varepsilon}_R}$ of the subcomplex generated by the new Reeb chords is generated by $[c'_{k'+1,1}]$ in degree $n - m' - k' - 1$ if k' is even, and by $[c'_{1,k'+1}]$ in degree $m' + k'$ if k' is odd. Similarly, the homology with respect to $\partial^{\hat{\varepsilon}_R,\hat{\varepsilon}_L}$ of this subcomplex is generated by $[c'_{1,1}]$ in degree $n - m' - 1$, $[c'_{1,2}]$ in degree $m' + 1$, and by $[c'_{1,k'+1}]$ in degree $m' + k'$ if k' is even, and by $[c'_{k'+1,1}]$ in degree $n - m' - k' - 1$ if k' is odd.

Proposition 4.20 *The* BLCH *Poincaré polynomials of* $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')}$ *are given by*

$$P_{\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')},\hat{\varepsilon}_L,\hat{\varepsilon}_R}(t) = P_{\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})},\hat{\varepsilon}_L,\hat{\varepsilon}_R}(t) + t^b$$

and by

$$P_{\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')},\hat{\varepsilon}_R,\hat{\varepsilon}_L}(t) = P_{\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})},\hat{\varepsilon}_R,\hat{\varepsilon}_L}(t) + t^{n-m'-1} + t^{m'+1} + t^{n-1-b},$$

where $b = n - m' - k' - 1$ if k' is even and $b = m' + k'$ if k' is odd.

Proof Observe that the image of $[c'_{1,1}]$ by the map

$$\tilde{\tau}_{n-m'-1} \colon \mathrm{LCH}^{\hat{\varepsilon}_R,\hat{\varepsilon}_L}_{n-m'-1}(\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')}) \to H_{n-m'-1}(\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')})$$

from the duality exact sequence (4-1) is the homology class of the cocore sphere of the attached handle. Indeed, all Morse flow trees starting at $c'_{1,1}$ and with no negative end must remain in the cocore sphere of the handle, since it is narrowest there. The resulting Morse flow trees start at $c'_{1,1}$ in any direction and finish at the cusp of the cocore sphere. The boundary of the corresponding holomorphic disks foliate the cocore sphere minus the endpoints of $c'_{1,1}$ so that the image of the cycle $c'_{1,1}$ in the bilinearized complex is the cycle corresponding to the cocore sphere in the singular complex of $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')}$. Since the corresponding homology class does not vanish in $H_{n-m'-1}(\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')})$, it follows that $[c'_{1,1}]$ does not vanish in bilinearized homology either.

Similarly, observe that the image of $[c'_{1,2}]$ by the map

$$\tilde{\tau}_{m'+1} \colon \mathrm{LCH}^{\hat{\varepsilon}_R,\hat{\varepsilon}_L}_{m'+1}(\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')}) \to H_{m'+1}(\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')})$$

from the duality exact sequence (4-1) is the Poincaré dual of the homology class of the cocore sphere of the attached handle. Indeed, all Morse flow trees starting at $c'_{1,2}$ and with no negative end must follow radii of the disk factor $D^{m'+1}$ for the handle. Once such a Morse flow tree exits the handle, it will flow down the chord $m_{1,2}$ corresponding to the minimum of the perturbing Morse function $f_2 - f_1$. The chord $m_{1,2}$ is augmented for $\hat{\varepsilon}_R$ so that the image by $\tilde{\tau}_{m'+1}$ is obtained by considering the part of the boundary of the corresponding holomorphic disks lying in Λ_1. This is a sphere of dimension $m'+1$, intersecting the cocore sphere at the endpoint of $c'_{1,2}$ in Λ_1. Since the corresponding homology class does not vanish in $H_{m'+1}(\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')})$, it follows that $[c'_{1,2}]$ does not vanish in bilinearized homology either.

In view of the long exact sequence relating the bilinearized homology of our subcomplex with the bilinearized homologies of our Legendrian submanifold before and after the generalized handle attachment, the effect of $[c'_{k'+1,1}]$ or $[c'_{1,k'+1}]$ could either be to add a term in the BLCH Poincaré polynomial in the degree of this generator, or to remove a term in the degree of this generator, plus one.

In terms of Proposition 4.4, we have just shown that the polynomial $\tilde{q}$ gains the terms $t^{n-m'-1} + t^{m'+1}$ as an effect of this generalized handle attachment. Since the dimension of the singular homology of the Legendrian submanifold increased by 2, it follows that the modifications due to $[c'_{k'+1,1}]$ and $[c'_{1,k'+1}]$ are affecting the polynomials p and $\tilde{p}$. Since the relation $\tilde{p}(t) = t^{n-1}p(t^{-1})$ must hold at all times, it follows that the changes to both BLCH Poincaré polynomials must occur in degrees that add up to $n-1$. But since the sum of the gradings of $[c'_{k'+1,1}]$ and of $[c'_{1,k'+1}]$ is $n-1$, it follows that the effect of these generators is necessarily to add a term in their corresponding BLCH Poincaré polynomial.

Since the four generators

$$[c'_{1,1}], \quad [c'_{1,2}], \quad [c'_{k'+1,1}] \quad \text{and} \quad [c'_{1,k'+1}]$$

each give rise to an additional term in one of the BLCH Poincaré polynomials of $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(k',m')}$, the announced relations follow. □

We can repeat the above generalized handle attachment as many times as we want, with different values of k' and m'. Repeating it s times with parameters k'_i and m'_i, let us define $\bar{k}' = (k'_1, \dots, k'_s)$ and $\bar{m}' = (m'_1, \dots, m'_s)$, and after choosing N so that $k'_i + 1 \le 2N$ for all $i = 1, \dots, s$. Applying these operations on $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m})}$, we denote the resulting Legendrian submanifold by $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(\bar{k}',\bar{m}')}$.

Corollary 4.21 *The* BLCH *Poincaré polynomial of* $\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(\bar{k}',\bar{m}')}$ *is given by*

$$P_{\Lambda^{(2N)}_{\bar{q},(\bar{k},\bar{m}),(\bar{k}',\bar{m}')},\hat{\varepsilon}_L,\hat{\varepsilon}_R}(t) = q(t) + \sum_{i=1}^{r}(t^{-m_i} + t^{a_i}) + \sum_{i=1}^{s} t^{b_i},$$

where

$$a_i = \begin{cases} k_i - m_i - 1 & \text{if } k_i \text{ is even,} \\ n - k_i + m_i & \text{if } k_i \text{ is odd,} \end{cases} \quad \text{and} \quad b_i = \begin{cases} n - k'_i - m'_i - 1 & \text{if } k'_i \text{ is even,} \\ k'_i + m'_i & \text{if } k'_i \text{ is odd.} \end{cases}$$

Proof of Theorem 1.3 Note that if $n = 1$, any connected Legendrian submanifold Λ is a circle. Since we already showed that the BLCH geography for spheres is realized by the submanifolds $\overline{\Lambda}^{(2N)}_{(\bar{k},\bar{m})}$ with Poincaré polynomial given by (4-3) with $q(t) = 1$ and $p(-1)$ even, we can assume that $n \ge 2$.

Assume first that $n > 2$. Let $q + p$ be a BLCH–admissible polynomial in the sense of Definition 4.1. Writing $p(t) = \sum_{i=1}^{s} t^{w_i}$, for any term $i = 1, \dots, s$ we can find $k_i' \geq 1$ and $0 \leq m_i' \leq n-2$ such that $b_i = w_i$ as in Corollary 4.21: if $w_i > 0$ is odd we can choose $m_i' = 0$ and $k_i' = w_i$, if $w_i > 0$ is even we can choose $m_i' = 1$ and $k_i' = w_i - 1$, if $w_i \leq 0$ has the same parity as n we can choose $m_i' = 1$ and $k_i' = n - 2 - w_i$, and if $w_i \leq 0$ has the same parity as $n-1$ we can choose $m_i' = 0$ and $k_i' = n - 1 - w_i$. Then the Legendrian submanifold $\Lambda^{(2N)}_{\bar{q},(\bar{k}',\bar{m}')}$ has the desired BLCH Poincaré polynomial $q + p$.

Finally, in the case $n = 2$, we cannot use the above choices of parameters since we must have $m_i' = 0$ for all $i = 1, \dots, s$. Let $q + p$ be a BLCH–admissible polynomial in the sense of Definition 4.1. Let us decompose p as $p_0 + p_1$ where p_0 and p_1 are Laurent polynomials with nonnegative integral coefficients, $p_0(-1) = 0$ and $p_1(1)$ is minimal with respect to these properties. We have already showed that there exists a Legendrian sphere $\bar{\Lambda}^{(2N)}_{(\bar{k},\bar{m})}$ with BLCH Poincaré polynomial given by $1 + p_0$ in view of (4-3). Since $p_1(1)$ is minimal, it follows that all terms in p_1 have degrees of the same parity.

If this parity is odd, all terms in p_1 are of the form t^{w_i} with w_i odd. If $w_i \geq 1$, we choose $k_i' = w_i$ odd, and if $w_i \leq -1$, we choose $k_i' = 1 - w_i$ even, as in Corollary 4.21. Therefore, using as many generalized handle attachments as needed, we can realize the BLCH Poincaré polynomial $1 + p_0 + p_1$, regardless of the value of $p_1(-1) \leq 0$. Then, by a connected sum with the Legendrian submanifold $\Lambda_{\bar{q}}$ from Proposition 4.18, we realize the BLCH Poincaré polynomial $q + p$ as desired.

If the terms in p_1 have degrees of even parity, we use generalized handle attachments on Λ_2 instead of Λ_1: the effect of this modified operation will be as described by Proposition 4.20, with the ordering of the augmentations reversed. In other words, each such generalized handle attachment will add $2t + t^{1-b_i}$ to the BLCH Poincaré polynomial of our Legendrian submanifold, with $1 - b_i = 1 - k_i'$ even, as in Corollary 4.21. If $q(t) = 1 + at$ then we can perform up to $\lfloor \frac{1}{2}a \rfloor$ such attachments. Therefore, for any polynomial p_1 such that $p_1(-1) = p_1(1) \leq \frac{1}{2}a = \frac{1}{2}(1 - q(-1))$, we can realize the BLCH Poincaré polynomial $1 + 2p_1(1)t + p_0 + p_1$. Setting $q_0(t) = q(t) - 2p_1(1)t$, we then perform a connected sum with the Legendrian submanifold $\Lambda_{\bar{q}_0}$ from Proposition 4.18 in order to realize the BLCH Poincaré polynomial $q + p$, as desired. □

References

[1] **F Bourgeois**, **B Chantraine**, *Bilinearized Legendrian contact homology and the augmentation category*, J. Symplectic Geom. 12 (2014) 553–583 MR Zbl

[2] **F Bourgeois**, **J M Sabloff**, **L Traynor**, *Lagrangian cobordisms via generating families: construction and geography*, Algebr. Geom. Topol. 15 (2015) 2439–2477 MR Zbl

[3] **Y Chekanov**, *Differential algebra of Legendrian links*, Invent. Math. 150 (2002) 441–483 MR Zbl

[4] **G Dimitroglou Rizell**, *Legendrian ambient surgery and Legendrian contact homology*, J. Symplectic Geom. 14 (2016) 811–901 MR Zbl

[5] **T Ekholm**, *Morse flow trees and Legendrian contact homology in 1–jet spaces*, Geom. Topol. 11 (2007) 1083–1224 MR Zbl

[6] **T Ekholm**, **J B Etnyre**, **J M Sabloff**, *A duality exact sequence for Legendrian contact homology*, Duke Math. J. 150 (2009) 1–75 MR Zbl

[7] **T Ekholm**, **J Etnyre**, **M Sullivan**, *The contact homology of Legendrian submanifolds in* $\mathbb{R}^{2n+1}$, J. Differential Geom. 71 (2005) 177–305 MR Zbl

[8] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Non-isotopic Legendrian submanifolds in* $\mathbb{R}^{2n+1}$, J. Differential Geom. 71 (2005) 85–128 MR Zbl

[9] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Legendrian contact homology in* $P \times \mathbb{R}$, Trans. Amer. Math. Soc. 359 (2007) 3301–3335 MR Zbl

[10] **Y Félix**, **S Halperin**, **J-C Thomas**, *Rational homotopy theory*, Graduate Texts in Math. 205, Springer (2001) MR Zbl

[11] **M B Henry**, **D Rutherford**, *A combinatorial DGA for Legendrian knots from generating families*, Commun. Contemp. Math. 15 (2013) art. id. 1250059 MR Zbl

[12] **M B Henry**, **D Rutherford**, *Equivalence classes of augmentations and Morse complex sequences of Legendrian knots*, Algebr. Geom. Topol. 15 (2015) 3323–3353 MR Zbl

[13] **T Kálmán**, *Contact homology and one parameter families of Legendrian knots*, Geom. Topol. 9 (2005) 2013–2078 MR Zbl

[14] **P Melvin**, **S Shrestha**, *The nonuniqueness of Chekanov polynomials of Legendrian knots*, Geom. Topol. 9 (2005) 1221–1252 MR Zbl

[15] **L L Ng**, *Computable Legendrian invariants*, Topology 42 (2003) 55–82 MR Zbl

[16] **L Ng**, **D Rutherford**, **V Shende**, **S Sivek**, **E Zaslow**, *Augmentations are sheaves*, Geom. Topol. 24 (2020) 2149–2286 MR Zbl

FB: *Université Paris-Saclay, CNRS, Laboratoire de Mathématiques d'Orsay*
91405 Orsay, France

DG: *INSA Hauts-de-France, CERAMATHS – Laboratoire de Matériaux Céramiques et de Mathématiques, Université Polytechnique Hauts-de-France*
F-59313 Valenciennes, France

DG: *F.R.S.–FNRS, Département de Mathématiques, Université de Mons*
Place du Parc, 20, B-7000 Mons, Belgium

frederic.bourgeois@universite-paris-saclay.fr, damien.galant@umons.ac.be

https://www.imo.universite-paris-saclay.fr/~frederic.bourgeois/,
https://damien-gal.github.io/

Received: 7 August 2020 Revised: 25 August 2022

Algebraic & Geometric Topology **24**:7 (2024) 3605–3620
DOI: 10.2140/agt.2024.24.3605
Published: 9 December 2024

The deformation spaces of geodesic triangulations of flat tori

Yanwen Luo
Tianqi Wu
Xiaoping Zhu

We prove that the deformation space of geodesic triangulations of a flat torus is homotopically equivalent to a torus. This solves an open problem proposed by Connelly et al. in 1983 in the case of flat tori. A key tool of the proof is a generalization of Tutte's embedding theorem for flat tori. While this paper was under preparation, Erickson and Lin proved a similar result, which works for all convex drawings.

55Q52, 57N65, 57R19, 57S05, 58D10

1 Introduction

This paper is a continuation of the previous work [Luo et al. 2023], where we proved that the deformation space of geodesic triangulations of a surface with negative curvature is contractible. The purpose of this paper is to identify the homotopy type of the deformation space of geodesic triangulations of a flat torus. This solves an open question proposed in [Connelly et al. 1983]. The main result of this paper is:

Theorem 1.1 *The deformation space of geodesic triangulations of a flat torus is homotopically equivalent to a torus.*

It is conjectured in [Connelly et al. 1983] that the space of geodesic triangulations of a closed orientable surface S with constant curvature deformation retracts to the group of orientation-preserving isometries of S homotopic to the identity. This paper affirms this conjecture in the case of flat tori. The case of hyperbolic surfaces was proved in [Luo et al. 2023]. In a very recent work, Erickson and Lin [2021] proved independently a generalized version of our Theorem 1.1 for general graph drawings on a flat torus.

The study of the homotopy types of spaces of geodesic triangulations stemmed from [Cairns 1944]. A brief history of this problem can be found in [Luo et al. 2023]. These spaces are closely related to diffeomorphism groups of surfaces. Bloch, Connelly and Henderson [Bloch et al. 1984] proved that the space of geodesic triangulations of a convex polygon is contractible. The space of geodesic triangulations of a planar polygon is equivalent to the space of simplexwise linear homeomorphisms. Hence, the Bloch–Connelly–Henderson theorem can be viewed as a discrete analogue of Smale's theorem, which states that the diffeomorphism group of the closed 2–disk fixing the boundary pointwise is contractible. Earle and Eells [1969] proved that the group of orientation-preserving diffeomorphisms of a torus isotopic to the identity is homotopically equivalent to a torus. Theorem 1.1 can be regarded as a discrete version of this theorem.

Similar to the previous work [Luo et al. 2023], our key idea to prove Theorem 1.1 originates from Tutte's embedding theorem.

1.1 Set up and the main theorem

Let $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2 = [0,1]^2/\sim$ be the flat torus constructed by gluing the opposite sides of the unit square in $\mathbb{R}^2$.

A topological triangulation of $\mathbb{T}^2$ can be identified as a homeomorphism ψ from $|\mathcal{T}|$ to $\mathbb{T}^2$, where $|\mathcal{T}|$ is the carrier of a 2–dimensional simplicial complex $\mathcal{T} = (V, E, F)$ with the vertex set V, and the edge set E, and the face set F. For convenience, we label the vertices as $v_1, \dots, v_n$, where $n = |V|$ is the number of the vertices. Denote by $|E|$ the number of edges, by $\vec{e}_{ij}$ the directed edge from the vertex i to its neighbor j, by $\vec{E} = \{\vec{e}_{ij} \mid ij \in E\}$ the set of directed edges, and by $N(i)$ the indices of neighboring vertices of $v_i \in V$.

A *geodesic triangulation* with the combinatorial type $(\mathcal{T}, \psi)$ is an embedding φ from the one-skeleton $\mathcal{T}^{(1)}$ to $\mathbb{T}^2$ satisfying that

(a) the restriction φ_{ij} of φ on each edge e_{ij}, identified with a unit interval $[0,1]$, is a geodesic of constant speed, and

(b) φ is homotopic to the restriction of ψ on $\mathcal{T}^{(1)}$.

Let $X = X(\mathbb{T}^2, \mathcal{T}, \psi)$ denote the set of all such geodesic triangulations, which is called a *deformation space of geodesic triangulations of* $\mathbb{T}^2$. This space can be defined for other flat tori in a similar fashion. Perturbing each vertex locally, we can construct a family of geodesic triangulations from an initial geodesic triangulation. Therefore, the space X is naturally a $2n$–dimensional manifold.

For any geodesic triangulation $\varphi \in X$, we can always translate φ on $\mathbb{T}^2$ to make the image $\varphi(v_1)$ of the first vertex v_1 be at the (quotient of the) origin $(0,0)$. By this normalization, we can decompose X as $X = X_0 \times \mathbb{T}^2$, where

$$X_0 = X_0(\mathbb{T}^2, \mathcal{T}, \psi) = \{\varphi \in X \mid \varphi(v_1) = (0,0)\}.$$

Since there are affine transformations between any two flat tori, and an affine transformation always preserves the geodesic triangulations, Theorem 1.1 reduces to the following.

Theorem 1.2 *Given a topological triangulation $(\mathcal{T}, \psi)$ of $\mathbb{T}^2$, the space $X_0 = X_0(\mathbb{T}^2, \mathcal{T}, \psi)$ is contractible.*

1.2 Key tool: generalized Tutte's embedding theorem

Let φ be a map from $\mathcal{T}^{(1)}$ to $\mathbb{T}^2$. Assume φ maps every edge in E to a geodesic arc parametrized by $[0,1]$ with constant speed on $\mathbb{T}^2$. A positive assignment $w \in \mathbb{R}_+^{\vec{E}}$ on the set of directed edges is called a *weight* of $\mathcal{T}$. We say φ is *w–balanced* at v_i if

$$\sum_{j\in N(i)} w_{ij}\dot{\varphi}_{ij} = 0,$$

where $\dot{\varphi}_{ij} = \dot{\varphi}_{ij}(0) \in T_{\varphi(v_i)}\mathbb{T}^2 \cong \mathbb{R}^2$. Then $\dot{\varphi}_{ij}$ indicates the direction of the edge $\varphi(e_{ij})$ and $\|\dot{\varphi}_{ij}\|$ equals the length of $\varphi(e_{ij})$. A map φ is called *w–balanced* if it is w–balanced at each vertex in V. We have the following version of Tutte's embedding theorem, which is a special case of Gortler, Gotsman and Thurston's embedding result in [Gortler et al. 2006] and Theorem 1.6 in [Luo et al. 2023].

Theorem 1.3 *Assume $(\mathcal{T},\psi)$ is a topological triangulation of $\mathbb{T}^2$, and φ is a map from $\mathcal{T}^{(1)}$ to $\mathbb{T}^2$ such that φ is homotopic to $\psi|_{\mathcal{T}^{(1)}}$ and the restriction φ_{ij} of φ on each edge e_{ij} is a geodesic parametrized by constant speed. If φ is w–balanced for some weight $w\in\mathbb{R}^{\vec{E}}_{+}$, then φ is an embedding, or equivalently φ is a geodesic triangulation.*

To be self-contained, we will give a simple proof for Theorem 1.3, which is adapted from the argument in [Gortler et al. 2006].

The classical Tutte's embedding theorem [1963] states that a straight-line embedding of a simple 3–vertex-connected planar graph can be constructed by fixing an outer face as a convex polygon and solving interior vertices on the condition that each vertex is in the convex hull of its neighbors. Various new proofs of Tutte's embedding theorem have been proposed by Floater [2003b], Gortler, Gotsman and Thurston [Gortler et al. 2006], et al.

Tutte's embedding theorem has been generalized by Colin de Verdière [1991], Delgado and Friedrichs [2005], and Hass and Scott [2015] to surfaces with nonpositive Gaussian curvatures. They showed that the minimizer of a discrete Dirichlet energy is a geodesic triangulation. Here the fact that φ is a minimizer of a discrete Dirichlet energy means that φ is w–balanced for some symmetric weight w in $\mathbb{R}^{\vec{E}}_{+}$ with $w_{ij} = w_{ji}$. Their result also implies that $X = X(\mathbb{T}^2,\mathcal{T},\psi)$ is not an empty set for any topological triangulation $(\mathcal{T},\psi)$. Recently, Luo, Wu and Zhu [Luo et al. 2023] proved a new version of Tutte's embedding theorem for nonsymmetric weights and triangulations of orientable closed surfaces with nonpositive Gaussian curvature.

Gortler, Gotsman and Thurston [Gortler et al. 2006] generalized Tutte's embedding theorem to flat tori. In contrast to the case of convex polygons and surfaces of negative curvatures, it is not always possible to construct a geodesic triangulation of $\mathbb{T}^2$ such that it is w–balanced with respect to a given nonsymmetric weight w. See [Chambers et al. 2021, Section 1.1] for a detailed discussion.

1.3 Outline of the proof

Fix a lifting $(x_i,y_i)\in\mathbb{R}^2$ of $\psi(v_i)\in\mathbb{T}^2$ for each $i=1,\dots,n$. Then for any $\vec{e}_{ij}\in\vec{E}$, there exists a unique lifting $\tilde{\psi}_{ij}:[0,1]\to\mathbb{R}^2$ of $\psi_{ij}=\psi|_{e_{ij}}:[0,1]\to\mathbb{T}^2$ such that $\tilde{\psi}_{ij}(0)=(x_i,y_i)$. Then

$$\tilde{\psi}_{ij}(1) = (x_j,y_j) + (b^x_{ij}, b^y_{ij})$$

for some lattice point $(b_{ij}^x, b_{ij}^y) \in \mathbb{Z}^2$. Notice that for any $j \in N(i)$, the liftings $\tilde{\psi}_{ij}$ have the same base point (x_i, y_i), and b_{ij}^x and b_{ij}^y are determined by the liftings $\tilde{\psi}_i$, (x_i, y_i), and (x_j, y_j).

A geodesic triangulation φ in X_0 can be represented by $(x_i, y_i) \in \mathbb{R}^2$ for $i = 1, \dots, n$ with $(x_1, y_1) = (0, 0)$. Under this representation, $\varphi_{ij} : [0, 1] \to \mathbb{T}^2$ is the quotient of the linear map

$$\varphi_{ij}(t) = t(x_j + b_{ij}^x, y_j + b_{ij}^y) + (1-t)(x_i, y_i),$$

and the equations for w–balanced conditions at all the vertices can be written as

$$\sum_{j \in N(i)} w_{ij}(x_j - x_i + b_{ij}^x) = 0 \quad \text{and} \quad \sum_{j \in N(i)} w_{ij}(y_j - y_i + b_{ij}^y) = 0.$$

In a closed matrix form, we can write

$$A(w)\boldsymbol{x} = \boldsymbol{b}(w), \tag{1}$$

where the weight matrix $A(w)$ is

$$A(w) = \begin{pmatrix} -\sum_{j=1}^n w_{1j} & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & -\sum_{j=1}^n w_{2j} & w_{23} & \cdots & w_{2n} \\ w_{31} & w_{32} & -\sum_{j=1}^n w_{3j} & \cdots & w_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \cdots & -\sum_{j=1}^n w_{nj} \end{pmatrix},$$

and

$$\boldsymbol{x} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix}, \quad \boldsymbol{b}(w) = \begin{pmatrix} -\sum_{j=1}^n w_{1j} b_{1j}^x & -\sum_{j=1}^n w_{1j} b_{1j}^y \\ -\sum_{j=1}^n w_{2j} b_{2j}^x & -\sum_{j=1}^n w_{2j} b_{2j}^y \\ \vdots & \vdots \\ -\sum_{j=1}^n w_{nj} b_{nj}^x & -\sum_{j=1}^n w_{nj} b_{nj}^y \end{pmatrix}.$$

Here we write $w_{ij} = 0$ if $e_{ij} \notin E$.

A weight w in $\mathbb{R}_+^{\vec{E}}$ is called *admissible* if (1) is solvable. Let W be the space of admissible weights. For any $w \in W$, a solution $\boldsymbol{x}$ to (1) uniquely determines the coordinates of the vertices and a w–balanced map φ that is homotopic to $\psi|_{\mathcal{T}^{(1)}}$. By Theorem 1.3, such a φ is an embedding, and $\varphi \in X$. Noticing that $A(w)$ is weakly diagonally dominant and the graph $\mathcal{T}^{(1)}$ is connected, the solution to (1) is unique up to a 2–dimensional translation, and is unique if we require $(x_1, y_1) = (0, 0)$. Define the *Tutte map* as

$$\Psi : W \to X_0$$

sending an admissible weight w to the unique w–balanced geodesic triangulation in X_0. The Tutte map is continuous by the continuous dependence of the solutions to the coefficients in a linear system.

The Tutte map is also surjective, since there exists a smooth map σ from X_0 to W by the "mean value coordinates"

$$w_{ij} = \frac{\tan(\alpha_i^j/2) + \tan(\beta_i^j/2)}{l_{ij}}$$

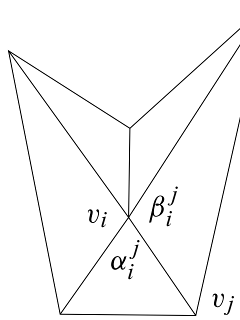


Figure 1: The mean value coordinate.

introduced in [Floater 2003a]. Here α_i^j, β_i^j are two inner angles adjacent to e_{ij} at the vertex v_i, and l_{ij} is the edge length of φ_{ij}. See Figure 1 for an illustration. Floater [2003a] showed that any geodesic triangulation φ is $\sigma(\varphi)$–balanced, ie $\Psi \circ \sigma = \mathrm{Id}_{X_0}$.

Having the knowledge of the Tutte map and the mean value coordinates, Theorem 1.2 reduces to the following proposition.

Proposition 1.4 *There exists a continuous map* $\Phi \colon \mathbb{R}_+^{\vec{E}} \to W$ *such that*

$$\Phi|_W = \mathrm{Id}_W.$$

Proof of Theorem 1.2 assuming Proposition 1.4 By Proposition 1.4, W is contractible since there exists a retraction Φ from the contractible space $\mathbb{R}_+^{\vec{E}}$ to W. So $\sigma \circ \Psi$ is homotopic to the identity map on W. On the other hand $\Psi \circ \sigma = \mathrm{Id}_{X_0}$, and thus X_0 is homotopic to W and contractible. □

1.4 Organization of this paper

In Section 2, we will prove Proposition 1.4 by constructing a flow. In Section 3, we prove Theorem 1.3 following the idea in [Gortler et al. 2006].

Acknowledgements

We really appreciate Professor Jeff Erickson for providing valuable insights and references about this problem. The authors were supported in part by the NSF grants 1737876, 1760471, DMS-FRG-1760527 and DMS-1811878.

2 Proof of Proposition 1.4

Set an energy function on the weight space $\mathbb{R}_+^{\vec{E}}$ as

$$\mathcal{E}(w) = \min_{\boldsymbol{x} \in \mathbb{R}^{n\times 2}} \|A(w)\boldsymbol{x} - \boldsymbol{b}(w)\|^2 = \min_{\boldsymbol{x} \in \mathbb{R}^{n\times 2} | x_1 = y_1 = 0} \|A(w)\boldsymbol{x} - \boldsymbol{b}(w)\|^2,$$

where the norm is the Frobenius norm of a matrix. The second minimization problem above is a least square problem with $2(n-1)$ real variables and a nondegenerate coefficient matrix. By the standard formula in linear least squares (LLS) or quadratic programming (QP), the minimizer, denoted by $\boldsymbol{x}(w)$, is a smooth function of w, and thus $\mathcal{E}(w)$ is also a smooth function of w. Note that $\mathcal{E}(w)=0$ if and only if w is admissible, and intuitively $\mathcal{E}(w)$ measures the deviation of w from being admissible. The key idea of the proof is to construct a flow on $\mathbb{R}_+^{\vec{E}}\backslash W$ to minimize $\mathcal{E}(w)$, as in the following lemma.

Lemma 2.1 *There exists a smooth function $\Theta\colon \mathbb{R}_+^{\vec{E}}\backslash W \to \mathbb{R}^{\vec{E}}$ and a positive continuous function $C(w)$ on $\mathbb{R}_+^{\vec{E}}$ such that, for any initial value $w^0 \in \mathbb{R}_+^{\vec{E}}$, the flow $w(t)$ defined by*

$$\begin{cases} \dot{w}(t)=\Theta(w(t)), \\ w(0)=w^0 \end{cases} \tag{2}$$

satisfies that, for any t in the maximum existing interval $[0,T)$,

(a) $0 \le \dot{w}_{ij}(t) \le w_{ij}(t)$, *and*

(b) $\dfrac{d\mathcal{E}(w(t))}{dt} \le -C(w^0)\sqrt{\mathcal{E}(w(t))}$.

Proposition 1.4 is proved in Section 2.1, assuming Lemma 2.1. Then we construct a flow in Section 2.2, and in Section 2.3 show that this flow is satisfactory for Lemma 2.1.

2.1 Proof of Proposition 1.4 assuming Lemma 2.1

Assume $\Theta(w)$ and $C=C(w^0)$ are as in Lemma 2.1. Given $w^0 \in \mathbb{R}_+^{\vec{E}}\backslash W$, assume $w(t)$ is the flow defined by equation (2), and $[0,T)$ is the maximum existing interval.

We claim that $T=T(w^0)<\infty$ and $w(t)$ converges to some $\bar{w}$ as $t \to T$. Since

$$\frac{d\mathcal{E}}{dt} \le -C\sqrt{\mathcal{E}},$$

we have

$$\frac{d(\sqrt{\mathcal{E}})}{dt}(t) \le -\frac{C}{2}$$

and

$$\sqrt{\mathcal{E}(w(t))} \le \sqrt{\mathcal{E}(w(0))}-\frac{Ct}{2},$$

which implies

$$T \le \frac{2\sqrt{\mathcal{E}(w^0)}}{C(w^0)} < \infty.$$

Since

$$0 \le \dot{w}_{ij}(t) \le w_{ij}(t),$$

we have

$$w_{ij}(t) \le w_{ij}^0 e^t \le w_{ij}^0 e^T. \tag{3}$$

Then by the monotone convergence theorem, $w(t)$ converges to some $\bar{w}$. By the maximality of T, $\bar{w}$ has to be in W. Let $\Phi\colon \mathbb{R}_+^{\vec{E}} \to W$ be such that $\Phi(w^0) = \bar{w}$ if $w^0 \notin W$, and $\Phi(w) = w$ if $w \in W$.

Now we prove that Φ is continuous, ie for any $w \in \mathbb{R}_+^{\vec{E}}$ and $\epsilon > 0$ there exists $\delta > 0$ such that $|\Phi(w') - \Phi(w)|_\infty \le \epsilon$ for any w' with $|w' - w|_\infty < \delta$. We consider the two cases $w \in W$ and $w \notin W$.

2.1.1 $\boldsymbol{w \in W}$ Since $C(w)$ is continuous, there exist $C_1 > 0$ and $\delta_1 > 0$ such that $C(w') \ge C_1$ for any w' with $|w - w'|_\infty \le \delta_1$. Since $\mathcal{E}$ is continuous, there exists $\delta_2 \in (0, \delta_1)$ such that

$$\mathcal{E}(w') \le \left[\frac{C_1}{2}\log\left(1 + \frac{\epsilon}{2|w|_\infty + \epsilon}\right)\right]^2$$

if $|w' - w|_\infty < \delta_2$. Then we will show that $\delta = \min\{\delta_2, \epsilon/2\}$ is satisfactory. Assume w' satisfies that $|w' - w|_\infty < \delta$. If $w' \in W$, then $|\Phi(w') - \Phi(w)|_\infty = |w' - w|_\infty < \delta \le \epsilon$. If $w' \notin W$,

$$T(w') \le \frac{2\sqrt{\mathcal{E}(w')}}{C(w')} \le \frac{C_1 \log(1 + \epsilon/(2|w|_\infty + \epsilon))}{C_1} = \log\left(1 + \frac{\epsilon}{2|w|_\infty + \epsilon}\right) \le \log\left(1 + \frac{\epsilon}{2|w'|_\infty}\right).$$

So by inequality (3),

$$|\Phi(w')_{ij} - \Phi(w)_{ij}| \le |\Phi(w')_{ij} - w'_{ij}| + |w'_{ij} - w_{ij}| < w'_{ij}(e^{T(w')} - 1) + \frac{\epsilon}{2} \le \epsilon.$$

2.1.2 $\boldsymbol{w \notin W}$ Assume $\bar{w} = \Phi(w) \in W$. Then by the result of the previous case, there exists $\delta_1 > 0$ such that $|\Phi(w') - \Phi(w)|_\infty < \epsilon$ for any w' with $|w' - \bar{w}|_\infty < \delta_1$. Assume $w(t)$ is the flow determined by (2) with the initial value $w^0 = w$. Then there exists some t_0, such that $|w(t_0) - \bar{w}|_\infty < \delta_1/2$. By the continuous dependence of the solutions of ODEs on the initial values, there exists $\delta > 0$ such that if $|w' - w|_\infty < \delta$, then $|w'(t_0) - w(t_0)| < \delta_1/2$, where $w'(t)$ is the flow determined by (2) with the initial value $w^0 = w'$. So if $|w - w'|_\infty < \delta$, we have $|w'(t_0) - \bar{w}| < \delta_1$ and

$$|\Phi(w') - \Phi(w)|_\infty = |\Phi(w'(t_0)) - \Phi(w)|_\infty < \epsilon.$$

2.2 Construction of the flow

Denote by $\boldsymbol{x}(w)$ the minimizer of the second minimization problem in the definition of the energy function $\mathcal{E}(w)$. Define the residual $\boldsymbol{r}(w)$ as

$$\boldsymbol{r}(w) = A(w)\boldsymbol{x}(w) - \boldsymbol{b}(w),$$

where

$$\boldsymbol{r}(w) = \begin{pmatrix} r_1^x(w) & r_1^y(w) \\ r_2^x(w) & r_2^y(w) \\ \vdots & \vdots \\ r_n^x(w) & r_n^y(w) \end{pmatrix} = \begin{pmatrix} \boldsymbol{r}^x(w) & \boldsymbol{r}^y(w) \end{pmatrix} = \begin{pmatrix} \boldsymbol{r}_1(w) \\ \boldsymbol{r}_2(w) \\ \vdots \\ \boldsymbol{r}_n(w) \end{pmatrix} \in \mathbb{R}^{n\times 2}.$$

The vector $\boldsymbol{r}_i$ is the residual at the vertex $v_i \in V$, and $\boldsymbol{r}^x$, $\boldsymbol{r}^y$ are the projections of the total residual in the directions of the x–axis and the y–axis respectively. Then by the minimality of $\boldsymbol{x}(w)$,

$$A^T(w)\boldsymbol{r}^x(w) = 0 \quad \text{and} \quad A^T(w)\boldsymbol{r}^y(w) = 0.$$

Equivalently,

$$\boldsymbol{r}^x(w) \perp A(w)(\mathbb{R}^n) \quad \text{and} \quad \boldsymbol{r}^y(w) \perp A(w)(\mathbb{R}^n).$$

Since $\operatorname{rank}(A(w)) = n-1$, we have $\boldsymbol{r}^x \parallel \boldsymbol{r}^y$. So $\operatorname{rank}(\boldsymbol{r}) \le 1$, and $\boldsymbol{r}_i \parallel \boldsymbol{r}_j$ for any $1 \le i, j \le n$. Here $u \parallel v$ means that vectors u and v are parallel, ie linearly dependent.

Lemma 2.2 *Assume $\boldsymbol{r} \neq 0$ and the following properties hold for $\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_n$:*

(a) *The vectors have the same direction, namely,*

$$\langle \boldsymbol{r}_i, \boldsymbol{r}_j \rangle > 0 \quad \text{for all } 1 \le i, j \le n.$$

(b) *If*

$$C = \max_{\vec{e}_{ij} \in \vec{E}} \frac{w_{ij}}{w_{ji}},$$

then

$$\frac{\|\boldsymbol{r}_i\|_2}{\|\boldsymbol{r}_j\|_2} \le C^{n-1} \quad \text{for all } 1 \le i, j \le n.$$

Proof Without loss of generality, after a rotation we can assume that all the vectors $\boldsymbol{r}_i$ are parallel to the x–axis, namely $\boldsymbol{r}^y = 0$.

To prove part (a), assume that $\langle \boldsymbol{r}_i, \boldsymbol{r}_j \rangle = r_i^x \cdot r_j^x \le 0$ for some $1 \le i, j \le n$. Then one can find a nonzero vector $\boldsymbol{p} = (p_1, \dots, p_n)^T \in \mathbb{R}^n_{\ge 0}$ so that $\boldsymbol{p} \perp \boldsymbol{r}^x$. Then $\boldsymbol{p} \in A(w)(\mathbb{R}^n)$ and there exists $\boldsymbol{q} = (q_1, \dots, q_n)^T \in \mathbb{R}^n$ with $\boldsymbol{p} = A(w)\boldsymbol{q}$. Then if $q_i = \max_j q_j$ for some i,

$$0 \le p_i = \sum_{j=1}^n w_{ij}(q_j - q_i) \le 0,$$

and thus $q_j = q_i$ if $j \in N(i)$. By this maximum principle and the connectedness of the graph, $q_j = q_i$ for any $j \in V$, and $\boldsymbol{p} = A(w)\boldsymbol{q} = 0$. This contradicts that $\boldsymbol{p}$ is nonzero.

If part (b) is not true, ordering the set V based on the values of r_i^x monotonically, one can find a nonempty proper subset $V_0 \subsetneq V$ such that

$$\frac{\min_{i \in V_0}\{r_i^x\}}{\max_{i \in V - V_0}\{r_i^x\}} > C.$$

Choose a vector $\boldsymbol{p} \in \mathbb{R}^n$ such that $p_i = 1$ if $i \in V_0$, and $p_i = 0$ otherwise. Then the contradiction follows from

$$\begin{aligned} 0 = \langle \boldsymbol{r}^x, A(w)\boldsymbol{p} \rangle &= \sum_{i \in V} r_i^x \sum_{j \in N(i)} w_{ij}(p_j - p_i) = \sum_{\vec{e}_{ij} \in \vec{E}} w_{ij} r_i^x p_j - \sum_{i \in V} r_i^x p_i \sum_{j \in N(i)} w_{ij} \\ &= \sum_{\substack{\vec{e}_{ij} \in \vec{E} \\ j \in V_0}} w_{ij} r_i^x - \sum_{i \in V_0} r_i^x \sum_{j \in N(i)} w_{ij} = \sum_{\substack{i \in V_0, j \in V - V_0 \\ j \in N(i)}} (r_j^x w_{ji} - r_i^x w_{ij}) < 0. \qquad \square \end{aligned}$$

Assume $\boldsymbol{n} = \boldsymbol{n}(w) \in \mathbb{R}^2$ is the unit vector that is parallel to $\boldsymbol{r}_1$ and $\langle \boldsymbol{n}, \boldsymbol{r}_1 \rangle > 0$. Define for each directed edge

$$u_{ij} = u_{ij}(w) = \boldsymbol{n} \cdot (\boldsymbol{x}_j - \boldsymbol{x}_i + (b_{ij}^x, b_{ij}^y)) \quad \text{for all } \vec{e}_{ij} \in \vec{E},$$

where $\boldsymbol{x}_i = (x_i, y_i)$ is given by the minimizer $\boldsymbol{x}(w)$. Note that

$$\|\boldsymbol{r}_i\|_2 = \boldsymbol{n} \cdot \boldsymbol{r}_i = \sum_{j \in N(i)} w_{ij} u_{ij}.$$

Lemma 2.3 *There exists a constant $\beta = \beta(T, \psi) > 0$ such that for any $w \in \mathbb{R}_+^{\vec{E}}$, there exists $\vec{e}_{ij} \in \vec{E}$ such that $u_{ij}(w) \le -\beta$.*

Proof Since $u_{ij} = -u_{ji}$ for any $ij \in E$, it suffices to find $\vec{e}_{ij}$ such that $|u_{ij}(w)| \ge \beta$. Assume $\boldsymbol{n} = (n_1, n_2) \in \mathbb{R}^2$. Then $|n_1| \ge 1/\sqrt{2}$ or $|n_2| \ge 1\sqrt{2}$.

If $|n_1| \ge 1/\sqrt{2}$, let $\gamma_1 = \mathbb{T} \times \{0\}$ be a horizontal simple loop in $\mathbb{T}^2$. Then $\psi^{-1}(\gamma_1)$ is a simple loop in the carrier of $\mathcal{T}$, and it is not difficult to show that there exists a sequence of vertices $v(1), \dots, v(k) = v(0)$ such that $v(i) \sim v(i+1)$ for any $i = 0, \dots, k-1$, and the union $\bigcup_{i=0}^{k-1} e_{v(i)v(i+1)}$ is a piecewise linear loop in $|\mathcal{T}|$, which is homotopic to $\psi^{-1}(\gamma_1)$. By choosing an appropriate orientation, we have

$$\sum_{i=0}^{k-1} (\boldsymbol{x}_{v(i)+1} - \boldsymbol{x}_{v(i)} + (b^x_{v(i+1)v(i)}, b^y_{v(i+1)v(i)})) = (1, 0).$$

So

$$\sum_{i=0}^{k-1} u_{v(i)v(i+1)} = \boldsymbol{n} \cdot (1, 0) = n_1,$$

and there exists some i such that $|u_{v(i)v(i+1)}| \ge |n_1|/k \ge 1/(\sqrt{2}k)$. Notice that here k is a constant depending only on $\mathcal{T}$ and ψ.

Similarly, if $|n_2| \ge 1/\sqrt{2}$, there exists some $\vec{e}_{ij} \in \vec{E}$ such that $|u_{ij}| \ge 1/(\sqrt{2}k')$ for some constant $k' = k'(\mathcal{T}, \psi)$. □

We define the smooth flow Θ on the domain $\mathbb{R}_+^{\vec{E}} \backslash W$ on each edge as

$$\begin{cases} \dot{w}_{ij} = w_{ij} \cdot g\Big(\dfrac{1}{\alpha}(w_{ij} + w_{ji}) u_{ij}\Big) \cdot h(w_{ij} - w_{ji}), \\ w_{ij}(0) = w^0_{ij}, \end{cases} \tag{4}$$

where g and h are smooth nonincreasing functions such that

(a) $g \equiv 1$ on $(-\infty, -1)$ and $g \equiv 0$ on $[0, +\infty)$, and
(b) $h \equiv 1$ on $(-\infty, 1)$ and $h \equiv 0$ on $[2, +\infty)$, and
(c) $\alpha = \alpha(w) = \beta \cdot \Big(2|E| + \sum_{\vec{e}_{ij} \in \vec{E}} w_{ij}^{-1}\Big)^{-1}$.

Roughly speaking, the function g tends be positive if $u_{ij} < 0$, meaning that $w_{ij} u_{ij}$ will decrease so as to reduce the residual $\|\boldsymbol{r}_i\|_2$. The function h controls the difference between w_{ij} and w_{ji}, and note that $\alpha(w)$ is smooth and very small. Specifically we have

$$\alpha(w) \le \frac{\beta}{2|E|} \quad \text{and} \quad \alpha(w) \le \beta L(w), \tag{5}$$

where $L(w) = \min_{\vec{e}_{ij} \in \vec{E}} w_{ij}$ is a continuous function on $w \in \mathbb{R}_+^{\vec{E}}$. Define

$$M(w) = \max\Big\{2, \max_{\vec{e}_{ij} \in \vec{E}} |w_{ij} - w_{ji}|\Big\},$$

which is another continous function on $\mathbb{R}_+^{\vec{E}}$.

Lemma 2.4 *Assume the flow $w(t)$ satisfies (4). Then we have the following*:

(a) $0 \le \dot{w}_{ij} \le w_{ij}$.

(b) $u_{ij} \ge 0$ *implies* $\dot{w}_{ij} = 0$. *Then* $\dot{w}_{ij} u_{ij} \le 0$ *for all directed edges.*

(c) $w_{ij} - w_{ji} \ge 2$ *implies* $\dot{w}_{ij} = 0$.

(d) $L(w(t))$ *is nondecreasing and* $M(w(t))$ *is nonincreasing.*

(e) *For any edge* ij,

$$\frac{w_{ij}(t)}{w_{ji}(t)} \le 1 + \frac{M}{L}.$$

(f) *The residual vectors* $\boldsymbol{r}_i(t)$ *satisfy*

$$\frac{\max_{i \in V} \|\boldsymbol{r}_i(t)\|_2}{\min_{i \in V} \|\boldsymbol{r}_i(t)\|_2} \le (1 + \frac{M}{L})^{n-1} \quad \textit{for all } 1 \le i, j \le n.$$

(f) *The residual vectors* $\boldsymbol{r}_i(t)$ *satisfy*

$$\frac{\sqrt{\mathcal{E}(w(t))}}{\sqrt{n}(1 + M/L)^{n-1}} \le \|\boldsymbol{r}_i(t)\|_2 \quad \textit{for all } 1 \le i \le n. \tag{6}$$

Proof Parts (a)–(d) are straightforward from (4) and the defining properties of smooth functions g and h. Part (e) follows from

$$\frac{w_{ij}(t)}{w_{ji}(t)} = 1 + \frac{w_{ij}(t) - w_{ji}(t)}{w_{ji}(t)} \le 1 + \frac{M}{L}.$$

Part (f) follows from part (e) and Lemma 2.2. For part (g), by definition

$$\mathcal{E}(t) = \sum_{j=1}^{n} \|\boldsymbol{r}_j(t)\|_2^2.$$

Part (f) implies that

$$\mathcal{E}(t) \le n\left(1 + \frac{M}{L}\right)^{2n-2} \|\boldsymbol{r}_i(t)\|_2^2 \quad \text{and} \quad \frac{\sqrt{\mathcal{E}(t)}}{\sqrt{n}(1 + M/L)^{n-1}} \le \|\boldsymbol{r}_i(t)\|_2 \quad \text{for all } 1 \le i \le n. \qquad \square$$

2.3 Proof of Lemma 2.1

Proof Let

$$C(w) = \frac{\beta L/M}{2\sqrt{n}(1 + M/L)^{n-1}},$$

where $L = L(w)$ and $M = M(w)$ are continous functions on $\mathbb{R}_+^{\vec{E}}$, on which $C(w)$ is also continuous. We claim that such a function $C(w)$ and the flow Θ defined as (4) are satisfactory. Assume $w^0 \in \mathbb{R}_+^{\vec{E}} \backslash W$ and $w(t)$ is a flow defined by (4). By part (a) of Lemma 2.4 we only need to prove part (b) of Lemma 2.1. By part (d) of Lemma 2.4, it is easy to see that $C(w(t))$ is nondecreasing on t. So we only need to prove that

$$\frac{d\mathcal{E}(w(t))}{dt} \leq -C(w(t))\sqrt{\mathcal{E}(w(t))}.$$

Given $w \in R_+^{\vec{E}}$ and $\boldsymbol{x} \in \mathbb{R}^{n\times 2}$, define

$$\widetilde{\mathcal{E}}(w, \boldsymbol{x}) = \|A(w)\boldsymbol{x} - \boldsymbol{b}(w)\|^2,$$

and then

$$\begin{aligned}\frac{d\mathcal{E}(w(\cdot))}{dt}(t) &= \lim_{\epsilon\to 0} \frac{\mathcal{E}(w(t+\epsilon)) - \mathcal{E}(w(t))}{\epsilon} \\ &\leq \lim_{\epsilon\to 0} \frac{\widetilde{\mathcal{E}}(w(t+\epsilon), \boldsymbol{x}(w(t))) - \widetilde{\mathcal{E}}(w(t), \boldsymbol{x}(w(t)))}{\epsilon} = \frac{\partial \widetilde{\mathcal{E}}}{\partial w}(w(t), \boldsymbol{x}(w(t)))\cdot \dot{w}.\end{aligned}$$

So it suffices to show

$$\frac{\partial \widetilde{\mathcal{E}}}{\partial w}(w(t), \boldsymbol{x}(w(t)))\cdot \dot{w} \leq -C\sqrt{\mathcal{E}(w(t))}.$$

Notice that

$$\begin{aligned}\frac{\partial \widetilde{\mathcal{E}}}{\partial w_{ij}}(w(t), \boldsymbol{x}(w(t))) &= \left(\frac{\partial}{\partial w_{ij}} \sum_{i=1}^{n} \Big\| \sum_{j\in N(i)} w_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_i + (b_{ij}^x, b_{ij}^y))\Big\|_2^2\right)\Big|_{(w,\boldsymbol{x}(w))} \\ &= 2\boldsymbol{r}_i \cdot (\boldsymbol{x}_j - \boldsymbol{x}_i + (b_{ij}^x, b_{ij}^y)) \\ &= 2\|\boldsymbol{r}_i\|_2 \boldsymbol{n}\cdot(\boldsymbol{x}_j - \boldsymbol{x}_i + (b_{ij}^x, b_{ij}^y)) = 2\|\boldsymbol{r}_i\|_2 \cdot u_{ij}\end{aligned}$$

and then,

$$\frac{\partial \widetilde{\mathcal{E}}}{\partial w}(w(t), \boldsymbol{x}(w(t)))\cdot \dot{w} = 2\sum_{\vec{e}_{ij}\in\vec{E}} \|\boldsymbol{r}_i\|_2 u_{ij}\dot{w}_{ij} \leq \frac{2\sqrt{\mathcal{E}(w(t))}}{\sqrt{n}(1+M/L)^{n-1}} \sum_{\vec{e}_{ij}\in\vec{E}} u_{ij}\dot{w}_{ij}. \tag{7}$$

Here we use the fact that $\dot{w}_{ij}u_{ij} \leq 0$ for all directed edges (part (b) of Lemma 2.4), and inequality (6). It remains to show that

$$\sum_{\vec{e}_{ij}\in\vec{E}} u_{ij}\dot{w}_{ij} \leq -\frac{\beta L}{2M}.$$

By Lemma 2.3 there exists a directed edge $\vec{e}_{i'j'}$ with $u_{i'j'} \leq -\beta$. Then we will consider the following two cases:

Case 1 $(w_{i'j'} - w_{j'i'} \leq 1)$ By the definition of the function h,

$$h(w_{i'j'} - w_{j'i'}) = 1.$$

We also have

$$g\Big(\frac{1}{\alpha}(w_{i'j'} + w_{j'i'})u_{i'j'}\Big) = 1,$$

since

$$(w_{i'j'}+w_{j'i'})u_{i'j'} \leq -2\beta L \leq -\alpha.$$

By (4), $\dot{w}_{i'j'} = w_{i'j'}$. Notice that $\dot{w}_{ij}u_{ij} \leq 0$ for any $\vec{e}_{ij} \in \vec{E}$ by part (b) of Lemma 2.4, so

$$\sum_{\vec{e}_{ij}\in\vec{E}} u_{ij}\dot{w}_{ij} \leq u_{i'j'}\dot{w}_{i'j'} = u_{i'j'}w_{i'j'} \leq -\beta L \leq -\frac{\beta L}{2M}.$$

Case 2 $(w_{i'j'}-w_{j'i'} \geq 1)$ Define

$$\vec{E}_0 = \{\vec{e}_{ij} \in \vec{E} \mid u_{ij} < 0,\ (w_{ij}-w_{ji})u_{ij} \geq \alpha\}.$$

If $\vec{e}_{ij} \in \vec{E}_0$, then obviously $w_{ij}-w_{ji} < 0$ and

$$h(w_{ij}-w_{ji}) = 1.$$

Also,

$$g\Big(\frac{1}{\alpha}(w_{ij}+w_{ji})u_{ij}\Big) = 1$$

since

$$(w_{ij}+w_{ji})u_{ij} \leq (w_{ji}-w_{ij})u_{ij} \leq -\alpha.$$

By (4), $\dot{w}_{ij} = w_{ij}$ and

$$\sum_{\vec{e}_{ij}\in\vec{E}} \dot{w}_{ij}u_{ij} \leq \sum_{\vec{e}_{ij}\in\vec{E}_0} \dot{w}_{ij}u_{ij} = \sum_{\vec{e}_{ij}\in\vec{E}_0} w_{ij}u_{ij} \leq -\frac{L}{M}\sum_{e_{ij}\in E_0}(w_{ij}-w_{ji})u_{ij}. \tag{8}$$

The last inequality uses the fact that $w_{ij} \geq -L(w_{ij}-w_{ji})/M$, which is equivalent to $w_{ji}/w_{ij} \leq 1+M/L$.

By the fact that $u_{i'j'} \leq -\beta$, and the assumption $w_{i'j'}-w_{j'i'} \geq 1$,

$$(w_{i'j'}-w_{j'i'})u_{i'j'} \leq -\beta < 0 < \alpha,$$

and thus $\vec{e}_{i'j'} \notin \vec{E}_0$. Notice that

$$\sum_{\vec{e}_{ij}\in\vec{E}} w_{ij}u_{ij} = \sum_{i=1}^{n}\sum_{j\in N(i)} w_{ij}u_{ij} = \sum_{i=1}^{n}\|\boldsymbol{r}_i\|_2 \geq 0,$$

and

$$\begin{aligned}\sum_{\vec{e}_{ij}\in\vec{E}} w_{ij}u_{ij} &= \sum_{\vec{e}_{ij}\in\vec{E}:u_{ij}<0}(w_{ij}-w_{ji})u_{ij}\\ &= \sum_{\vec{e}_{ij}\in\vec{E}_0}(w_{ij}-w_{ji})u_{ij} + \sum_{e_{ij}\in\vec{E}-\vec{E}_0-\{e_{i'j'}\}:u_{ij}<0}(w_{ij}-w_{ji})u_{ij} + (w_{i'j'}-w_{j'i'})u_{i'j'}\\ &\leq \sum_{\vec{e}_{ij}\in\vec{E}_0}(w_{ij}-w_{ji})u_{ij} + |E|\alpha-\beta.\end{aligned}$$

Then

$$\sum_{\vec{e}_{ij} \in \vec{E}_0} (w_{ij} - w_{ji}) u_{ij} \geq \beta - |E|\alpha \geq \frac{\beta}{2},$$

and

$$\sum_{\vec{e}_{ij} \in \vec{E}} u_{ij} \dot{w}_{ij} \leq -\frac{\beta L}{2M}. \qquad \square$$

3 Proof of Theorem 1.3

We will first introduce the concept of discrete one-forms, and the index theorem proposed in [Gortler et al. 2006].

3.1 Discrete one-forms and the index theorem

A *discrete one-form* is a real-valued function η on the set of directed edges such that it is antisymmetric on each undirected edge. Specifically, let $\eta_{ij} = \eta(\vec{e}_{ij})$ be the value of η on the directed edge from v_i to v_j; then we have $\eta_{ij} = -\eta_{ji}$.

For a discrete one-form, an edge is degenerate (resp nonvanishing) if the one-form is zero (resp nonzero) on it. A vertex is degenerate (resp nonvanishing) if all of edges connected to it are degenerate (resp nonvanishing). A face is degenerate (resp nonvanishing) if all of its three edges are degenerate (resp nonvanishing). A one-form is degenerate (resp nonvanishing) if all the edges are degenerate (resp nonvanishing). Each edge is either degenerate or nonvanishing. However, vertices or faces can be degenerate, nondegenerate but vanishing on some edges, or nonvanishing.

Assume η is a discrete one-form. Denote by $\mathrm{sc}(\eta, v)$ the number of *sign changes* of the nonzero values of η on the directed edges starting from v, counted in counterclockwise order. For a vertex $v \in V$, define the *index* of v as $\mathrm{Ind}(\eta, v) = (2 - \mathrm{sc}(\eta, v))/2$. Similarly, for a nondegenerate face $t \in F$, the index of t is $\mathrm{Ind}(\eta, t) = (2 - \mathrm{sc}(\eta, t))/2$, where $\mathrm{sc}(\eta, t)$ is the number of sign changes of the nonzero values of η on the three edges of t, counted in counterclockwise order.

The following theorem is a special case of the index theorem from [Gortler et al. 2006], which is a discrete version of the Poincaré–Hopf theorem for discrete one-forms.

Theorem 3.1 *Let η be a nonvanishing discrete one-form on a triangulation of a torus. Then*

$$\sum_{v_i \in V} \mathrm{Ind}(\eta, v_i) + \sum_{t_{ijk} \in F} \mathrm{Ind}(\eta, t_{ijk}) = 0.$$

Assume φ satisfies the assumption in Theorem 1.3; then for any unit vector $\boldsymbol{n} \in \mathbb{R}^2$ we can naturally construct a discrete one-form η by letting $\eta_{ij} = \dot{\varphi}_{ij} \cdot \boldsymbol{n}$. If $\varphi \in X$, a generic unit vector determines a nonvanishing discrete one-form η. Further, if $\varphi \in X$ and such a constructed η is nonvanishing, it is

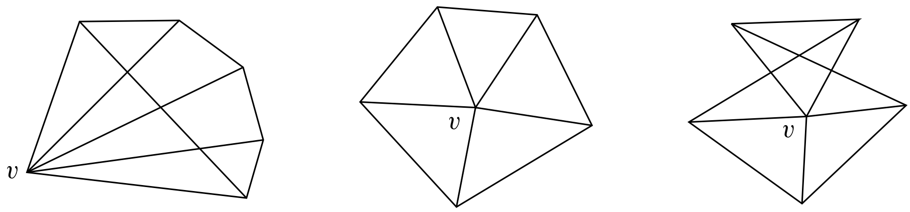


Figure 2: Typical vertex with positive (left), zero (middle), and negative (right) index.

not difficult to show that all the indices of the vertices and faces are zero. Figure 2 illustrates how the neighborhood of v looks if it has positive, or zero, or negative index for the case $\boldsymbol{n} = (1, 0)$.

Based on this construction, we have:

Lemma 3.2 *Given a triangulation $(\mathcal{T}, \psi)$ of $\mathbb{T}^2$, denote by $t_{ijk} \in F$ the triangle with three vertices v_i, v_j, and v_k. There exists a nonvanishing discrete one-form η such that $\eta_{ij} > 0$ and $\eta_{jk} > 0$. Moreover, all the indices of the vertices and faces of η are zero.*

Proof By the result of [Colin de Verdière 1991] and [Hass and Scott 2015], the space $X(\mathcal{T}, \psi)$ is not empty for any $(\mathcal{T}, \psi)$. Let φ be a geodesic triangulation in X. Then it is not difficult to find a unit vector $\boldsymbol{n}$ such that $\dot{\varphi}_{ij} \cdot \boldsymbol{n} > 0$ and $\dot{\varphi}_{jk} \cdot \boldsymbol{n} > 0$. Define the discrete one-form η as $\eta_{ij} = \dot{\varphi}_{ij} \cdot \boldsymbol{n}$. We can perturb the unit vector $\boldsymbol{n}$ a little bit to make η nonvanishing, and then such an η is satisfactory. □

3.2 The proof of Theorem 1.3

Assume $\varphi\colon \mathcal{T}^{(1)} \to \mathbb{T}^2$ satisfies the assumption of Theorem 1.3; then there exists a unique extension $\bar{\varphi}\colon |\mathcal{T}| \to \mathbb{T}^2$ such that the restriction of $\bar{\varphi}$ to every face is linear. Such a $\bar{\varphi}$ is homotopic to ψ, and φ is a geodesic triangulation in X if and only if $\bar{\varphi}$ is a homeomorphism.

For any triangle $t_{ijk} \in X$, we say that $\bar{\varphi}(t_{ijk})$ is *degenerate* if $\bar{\varphi}(t_{ijk})$ is contained in some geodesic λ. If $\bar{\varphi}(t_{ijk})$ is not degenerate, we can naturally define its inner angle θ^i_{jk} at $\varphi(v_i)$. We claim that

(a) $\bar{\varphi}(t_{ijk})$ is not degenerate for any $t_{ijk} \in F$, and

(b) $\bar{\varphi}$ is locally a homeomorphism.

Then $\bar{\varphi}$ is a proper local homeomorphism, and thus is a covering map. Since $\bar{\varphi}$ is homotopic to the homeomorphism ψ, we know $\bar{\varphi}$ is indeed a degree-1 covering map, ie a homeomorphism.

3.2.1 Proof of claim (a) Assume there is some triangle $t \in F$ such that $\bar{\varphi}(t)$ is degenerate and hence contained in a geodesic λ. Here λ is assumed to be a closed geodesic, or a densely immersed complete geodesic. Let $\mathcal{C}$ be the union of all triangles t such that $\bar{\varphi}(t) \subset \lambda$. Then $\mathcal{C}$ is not the whole complex $\mathcal{T}$; otherwise $\bar{\varphi}$ is not homotopic to the homeomorphism ψ. So, we can find a vertex $v_0 \in \partial\mathcal{C}$. Denote by star$(v_0)$ the star-neighborhood of v_0 in $\mathcal{T}$. Then $\bar{\varphi}(\text{star}(v_0))$ is not in λ, but $\bar{\varphi}(t_0) \subset \lambda$ for some triangle t_0 in star(v_0).

Let $\boldsymbol{n}$ be a unit vector that is orthogonal to the geodesic λ, and define η as $\eta_{ij} = \dot{\varphi}_{ij} \cdot \boldsymbol{n}$. Then the vertex v_0 is nondegenerate with respect to η, but the face t_0 is degenerate. Let ξ be a discrete one-form in Lemma 3.2 with the triangle t_0 and $v_j = v_0$. Scale ξ to make it very small so that $\eta + \xi$ has the same signs with η on the nondegenerate edges of η.

Notice that $\mathrm{sc}(\eta, v_0) \neq 0$, or equivalently $\mathrm{sc}(\eta, v_0) \geq 2$; otherwise all the edges connecting v_0 lie on a half-space, which contradicts the assumption that φ is balanced. Since t_0 is degenerate in η, taking the opposite $-\xi$ instead of ξ if necessary, we can assume that $\mathrm{sc}(\eta, v_0) < \mathrm{sc}(\eta + \xi, v_0)$, and thus $\mathrm{Ind}(\eta + \xi, v_0) < 0$.

Noticing that $\eta + \xi$ is nonvanishing, we will derive a contradiction with the index theorem (Theorem 3.1), by showing that the index of $\eta + \xi$ is nonpositive for any vertex and face.

If a face t is degenerate in η, then $\mathrm{Ind}(\eta + \xi, t) = \mathrm{Ind}(\xi, t) = 0$. If a face t is nondegenerate in η, then $\mathrm{Ind}(\eta + \xi, t) = \mathrm{Ind}(\eta) = 0$. In fact, the index of any nondegenerate face is zero.

If a vertex v is degenerate in η, then $\mathrm{Ind}(\eta + \xi, v) = \mathrm{Ind}(\xi, v) = 0$. If a vertex v is nonvanishing, then $\mathrm{Ind}(\eta + \xi, v) = \mathrm{Ind}(\eta, v)$. Since φ is balanced, $\mathrm{Ind}(\eta, v) \leq 0$.

If a vertex v is nondegenerate but vanishing at some edges in η, then

$$\mathrm{Ind}(\eta + \xi, v) \leq \mathrm{Ind}(\eta, v) \leq 0,$$

since adding ξ can only introduce more sign changes.

3.2.2 Proof of claim (b) Since φ is w–balanced, it is not difficult to show that for any vertex i,

$$\sum_{jk : t_{ijk} \in F} \theta^i_{jk} \geq 2\pi,$$

and the equality holds if and only if all the edges around v_i do not "fold" under the map $\bar{\varphi}$. By the Gauss–Bonnet theorem,

$$\sum_{i=1}^{n} \Bigl(2\pi - \sum_{jk : t_{ijk} \in F} \theta^i_{jk} \Bigr) = 0.$$

So

$$\sum_{jk : t_{ijk} \in F} \theta^i_{jk} = 2\pi$$

for any vertex v_i, and all the edges in E do not fold. Thus, $\bar{\varphi}$ is a local homeomorphism.

References

[Bloch et al. 1984] **E D Bloch**, **R Connelly**, **D W Henderson**, *The space of simplexwise linear homeomorphisms of a convex 2–disk*, Topology 23 (1984) 161–175 MR Zbl

[Cairns 1944] **S S Cairns**, *Isotopic deformations of geodesic complexes on the 2–sphere and on the plane*, Ann. of Math. 45 (1944) 207–217 MR Zbl

[Chambers et al. 2021] **E W Chambers**, **J Erickson**, **P Lin**, **S Parsa**, *How to morph graphs on the torus*, from "Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms", SIAM, Philadelphia, PA (2021) 2759–2778 MR Zbl

[Colin de Verdière 1991] **Y Colin de Verdière**, *Comment rendre géodésique une triangulation d'une surface?*, Enseign. Math. 37 (1991) 201–212 MR Zbl

[Connelly et al. 1983] **R Connelly**, **D W Henderson**, **C W Ho**, **M Starbird**, *On the problems related to linear homeomorphisms, embeddings, and isotopies*, from "Continua, decompositions, manifolds", Univ. Texas Press, Austin, TX (1983) 229–239 MR Zbl

[Delgado-Friedrichs 2005] **O Delgado-Friedrichs**, *Equilibrium placement of periodic graphs and convexity of plane tilings*, Discrete Comput. Geom. 33 (2005) 67–81 MR Zbl

[Earle and Eells 1969] **C J Earle**, **J Eells**, *A fibre bundle description of Teichmüller theory*, J. Differential Geom. 3 (1969) 19–43 MR Zbl

[Erickson and Lin 2021] **J Erickson**, **P Lin**, *Planar and toroidal morphs made easier*, from "Graph drawing and network visualization", Lecture Notes in Comput. Sci. 12868, Springer (2021) 123–137 MR Zbl

[Floater 2003a] **M S Floater**, *Mean value coordinates*, Comput. Aided Geom. Design 20 (2003) 19–27 MR Zbl

[Floater 2003b] **M S Floater**, *One-to-one piecewise linear mappings over triangulations*, Math. Comp. 72 (2003) 685–696 MR Zbl

[Gortler et al. 2006] **S J Gortler**, **C Gotsman**, **D Thurston**, *Discrete one-forms on meshes and applications to 3D mesh parameterization*, Comput. Aided Geom. Design 23 (2006) 83–112 MR Zbl

[Hass and Scott 2015] **J Hass**, **P Scott**, *Simplicial energy and simplicial harmonic maps*, Asian J. Math. 19 (2015) 593–636 MR Zbl

[Luo et al. 2023] **Y Luo**, **T Wu**, **X Zhu**, *The deformation space of geodesic triangulations and generalized Tutte's embedding theorem*, Geom. Topol. 27 (2023) 3361–3385 MR Zbl

[Tutte 1963] **W T Tutte**, *How to draw a graph*, Proc. Lond. Math. Soc. 13 (1963) 743–767 MR Zbl

Department of Mathematics, Rutgers University
New Brunswick, NJ, United States

Center of Mathematical Sciences and Applications, Harvard University
Cambridge, MA, United States

Department of Mathematics, Rutgers University
New Brunswick, NJ, United States

yl1594@rutgers.edu, tianqi@cmsa.fas.harvard.edu, xz349@rutgers.edu

Received: 11 July 2021 Revised: 11 August 2023

Algebraic & Geometric Topology 24:7 (2024) 3621–3639
DOI: 10.2140/agt.2024.24.3621
Published: 9 December 2024

Finite presentations of the mapping class groups of once-stabilized Heegaard splittings

DAIKI IGUCHI

Let $g \geq 2$ and assume that we are given a genus g Heegaard splitting of a closed orientable 3–manifold with distance greater than $2g+2$. We prove that the mapping class group of the once-stabilization of such a Heegaard splitting is finitely presented.

57K30, 57M60

1 Introduction

Let (M, Σ) be a Heegaard splitting of a compact orientable 3–manifold M. The *mapping class group* $\mathrm{MCG}(M, \Sigma)$ of the Heegaard splitting (M, Σ) is defined to be the group $\pi_0(\mathrm{Diff}(M, \Sigma))$ of path-connected components of the group $\mathrm{Diff}(M, \Sigma)$, where we denote by $\mathrm{Diff}(M, \Sigma)$ the group of diffeomorphisms of M that preserve Σ setwise. There is a natural homomorphism from $\mathrm{MCG}(M, \Sigma)$ to the mapping class group $\mathrm{MCG}(M)$ of M. Following Johnson [2011], we call the kernel of this natural homomorphism the *isotopy subgroup* of $\mathrm{MCG}(M, \Sigma)$, and denote it by $\mathrm{Isot}(M, \Sigma)$.

In this paper, we are interested in the isotopy subgroup of the mapping class group of a once-stabilized Heegaard splitting. Let (M, Σ') be a genus $g(\Sigma') \geq 2$ Heegaard splitting of a closed orientable 3–manifold M. We say that a Heegaard splitting (M, Σ) is a (once-)*stabilization* of (M, Σ') if it is obtained from (M, Σ') by adding a 1–handle whose core is parallel into Σ'. Corresponding to two handlebodies $V^-_{\Sigma'}$ and $V^+_{\Sigma'}$ in M with $\partial V^-_{\Sigma'} = \partial V^+_{\Sigma'} = \Sigma'$, there are two obvious subgroups of $\mathrm{Isot}(M, \Sigma)$: one is $\mathrm{Isot}(V^-_{\Sigma'}, \Sigma^-)$ and the other is $\mathrm{Isot}(V^+_{\Sigma'}, \Sigma^+)$, where Σ^- (resp. Σ^+) is the Heegaard surface obtained by pushing Σ into $V^-_{\Sigma'}$ (resp. $V^+_{\Sigma'}$) slightly. It is natural to ask when these subgroups generate $\mathrm{Isot}(M, \Sigma)$. Johnson [2011] proved that if the distance $d(\Sigma')$ of the Heegaard splitting (M, Σ') is greater than $2g(\Sigma') + 2$, then the two subgroups defined above generate $\mathrm{Isot}(M, \Sigma)$. As a consequence of this fact, together with a result of Scharlemann [2013] that says $\mathrm{Isot}(V^\pm_{\Sigma'}, \Sigma^\pm)$ are finitely generated, it follows that $\mathrm{Isot}(M, \Sigma)$ and $\mathrm{MCG}(M, \Sigma)$ are finitely generated. In that paper, Johnson conjectured that $\mathrm{Isot}(M, \Sigma)$ is an amalgamation of the two groups $\mathrm{Isot}(V^-_{\Sigma'}, \Sigma^-)$ and $\mathrm{Isot}(V^+_{\Sigma'}, \Sigma^+)$. This is the main result of the paper:

Theorem 1.1 *Suppose that (M, Σ') is Heegaard splitting of a closed orientable 3–manifold M with $d(\Sigma') > 2g(\Sigma') + 2$, and that (M, Σ) is a once-stabilization of (M, Σ'). Suppose that $(V^-_{\Sigma'}, \Sigma^-)$ (resp.*

$(V^+_{\Sigma'}, \Sigma^+)$) *is the Heegaard splitting of* $V^-_{\Sigma'}$ (*resp.* $V^+_{\Sigma'}$) *obtained by pushing* Σ *into* $V^-_{\Sigma'}$ (*resp.* $V^+_{\Sigma'}$) *slightly, where* $V^-_{\Sigma'}$ *and* $V^+_{\Sigma'}$ *are handlebodies in* M *bounded by* Σ'. *Then* $\mathrm{Isot}(M, \Sigma)$ *is isomorphic to an amalgamation of the two groups* $\mathrm{Isot}(V^-_{\Sigma'}, \Sigma^-)$ *and* $\mathrm{Isot}(V^+_{\Sigma'}, \Sigma^+)$.

One might expect that the above theorem has something to do with van Kampen's theorem. This idea can be justified as follows. Following Johnson and McCullough [2013], we define the space $\mathcal{H}(M, \Sigma)$ to be $\mathrm{Diff}(M)/\mathrm{Diff}(M, \Sigma)$ and call it the *space of Heegaard splittings* equivalent to (M, Σ). Let $\mathcal{H}$ denote the path-connected component of $\mathcal{H}(M, \Sigma)$ containing the left coset $\mathrm{id}_M \cdot \mathrm{Diff}(M, \Sigma)$. It is known that if a 3–manifold admits a Heegaard splitting with the distance greater than two, then such a 3–manifold must be hyperbolic. By a result in [Johnson and McCullough 2013] (see Theorem 2.1 below for more details) together with this fact, it follows that $\mathrm{Isot}(M, \Sigma)$ is isomorphic to $\pi_1(\mathcal{H})$.

Now fix a *spine* $K = K^- \cup K^+$ of the Heegaard splitting (M, Σ'), that is, K^- and K^+ are finite graphs embedded in M such that the complement $M \setminus K$ is diffeomorphic to $\Sigma' \times (-1, 1)$ and Σ' is a slice of this product structure. Denote by $\mathcal{H}^-$ (resp. $\mathcal{H}^+$) the subspace of $\mathcal{H}$ consisting of those elements represented by a Heegaard surface T such that T is a genus $g(\Sigma') + 1$ Heegaard surface of the genus $g(\Sigma')$ handlebody $M \setminus \mathrm{Int}(N(K^+))$ (resp. $M \setminus \mathrm{Int}(N(K^-))$), where $N(K^+)$ (resp. $N(K^-)$) is a small neighborhood of K^+ (resp. K^-). By the similar reason as above (see Theorem 2.2 below), we can identify $\mathrm{Isot}(V^-_{\Sigma'}, \Sigma^-)$ and $\mathrm{Isot}(V^+_{\Sigma'}, \Sigma^+)$ with the fundamental groups $\pi_1(\mathcal{H}^-)$ and $\pi_1(\mathcal{H}^+)$ respectively. Set $\mathcal{H}^\cup := \mathcal{H}^- \cup \mathcal{H}^+$. Theorem 1.1 is a corollary of the following.

Theorem 1.2 *The inclusion* $\mathcal{H}^\cup \to \mathcal{H}$ *is a homotopy equivalence.*

It is well known that a genus $g+1$ Heegaard splitting of a genus g handlebody is unique up to isotopy. Similarly, a genus $g+1$ Heegaard splitting of the space $F_g \times [-1, 1]$ is unique up to isotopy, where we denote by F_g a closed genus g surface. In other words, $\mathcal{H}^+$, $\mathcal{H}^-$ and $\mathcal{H}^- \cap \mathcal{H}^+$ are all connected, and hence van Kampen's theorem applies to the triple $(\mathcal{H}^-, \mathcal{H}^+, \mathcal{H}^- \cap \mathcal{H}^+)$.

The proof of Theorem 1.2 is based on the concept of *graphics*, which was first introduced by Cerf [1968] and then successfully applied to the study of Heegaard splittings by Rubinstein and Scharlemann [1996]. More precisely, we prove Theorem 1.2 by generalizing the method developed by Johnson [2010; 2011]. We also use an argument due to Hatcher [1976] crucially, which is a parametrized version of the innermost disk argument.

In Section 5, we confirm that the isotopy subgroup of a genus $g+1$ Heegaard splitting of a genus g handlebody is finitely presented:

Theorem 1.3 *Let* V *be a handlebody of genus* $g(V) \geq 2$, *and let* (V, Σ) *be a genus* $g(V)+1$ *Heegaard splitting of* V. *Then* $\mathrm{Isot}(V, \Sigma)$ *is finitely presented.*

It follows from Theorem 1.3 that $\pi_1(\mathcal{H}^-)$ and $\pi_1(\mathcal{H}^+)$ are finitely presented. As a consequence, we have:

Corollary 1.4 *Let (M, Σ') be a Heegaard splitting of a closed orientable 3–manifold M with*

$$d(\Sigma') > 2g(\Sigma') + 2.$$

Let (M, Σ) be a once-stabilization of (M, Σ'). Then $\mathrm{Isot}(M, \Sigma)$ *and* $\mathrm{MCG}(M, \Sigma)$ *are finitely presented.*

We remark that a problem related to this work was treated by Koda and Sakuma [2023]. In that paper, the concept of the "homotopy motion group" was introduced, and they considered the question that asks when the homotopy motion group $\Pi(M, \Sigma)$ of a Heegaard surface in a 3–manifold M can be written as an amalgamation of the two homotopy motion groups $\Pi(U_\Sigma^-, \Sigma)$ and $\Pi(U_\Sigma^+, \Sigma)$ corresponding to the two handlebodies U_Σ^- and U_Σ^+ with $\partial U_\Sigma^- = \partial U_\Sigma^+ = \Sigma$.

The paper is organized as follows. In Section 2, we recall from [Johnson and McCullough 2013] some facts about the space of Heegaard splittings. We also recall the definition of the distance of a Heegaard splitting. To prove Theorem 1.2, we will need to deal with the graphic determined by a 4–parameter family of Heegaard surfaces. In Section 3, we give a quick review of the theory of graphics, and then we see that some ideas in [Johnson 2010] can be adapted to our setting. In Section 4, we prove Theorem 1.2. Finally, we give the proof of Theorem 1.3 in Section 5.

Acknowledgements The author would like to thank his advisor Yuya Koda for much advice and sharing his insight. He is also grateful to the referees for their valuable comments that improved the manuscript. This work was supported by JSPS KAKENHI grant JP21J10249.

2 Preliminaries

Throughout the paper, we will use the following notation. For a topological space X, we denote by $|X|$ the number of path-connected components of X. For a subspace Y of X, $\mathrm{Int}(Y)$ and $\mathrm{Cl}(Y)$ denote the interior and the closure of Y in X, respectively. We will denote by J the closed interval $[-1, 1]$.

2.1 The space of Heegaard splittings

Let M be a compact orientable 3–manifold (possibly with boundary). Let (M, Σ) be a Heegaard splitting of M. This means that $\Sigma \subset M$ is a closed orientable embedded surface cutting M into the two compression bodies. Here, a *compression body* is a 3–manifold with nonempty boundary admitting a Morse function without critical points of index 2 and 3. A handlebody is a typical example of a compression body. The space $\mathcal{H}(M, \Sigma) = \mathrm{Diff}(M)/\mathrm{Diff}(M, \Sigma)$ is called the *space of Heegaard splittings* equivalent to (M, Σ). Note that there is a one-to-one correspondence between $\mathcal{H}(M, \Sigma)$ and the set of images of Σ under diffeomorphisms of M. We often identify an element of $\mathcal{H}(M, \Sigma)$ with the corresponding Heegaard surface. We always take the surface Σ as the basepoint of $\mathcal{H}(M, \Sigma)$, which corresponds to the left coset $\mathrm{id}_M \cdot \mathrm{Diff}(M, \Sigma)$. The space $\mathcal{H}(M, \Sigma)$ admits a structure of a Fréchet manifold, and this implies that $\mathcal{H}(M, \Sigma)$ has the homotopy type of a CW complex.

Theorem 2.1 [Johnson and McCullough 2013, Corollary 1] *Suppose that M is closed, orientable, irreducible and $\pi_1(M)$ is infinite, and that M is not a non-Haken infranilmanifold. Then $\pi_k(\mathcal{H}(M,\Sigma)) = 0$ for $k \geq 2$, and there is an exact sequence*

$$1 \to Z(\pi_1(M)) \to \pi_1(\mathcal{H}(M,\Sigma)) \to \mathrm{Isot}(M,\Sigma) \to 1.$$

A similar statement as above holds for handlebodies and the space $F_g \times J$:

Theorem 2.2 *Let $g' \geq g \geq 2$. Suppose that M is a genus g handlebody or the space $F_g \times J$, where F_g denotes a closed orientable surface of genus g. Suppose that (M,Σ) is a genus g' Heegaard splitting of M. Then $\pi_1(\mathcal{H}(M,\Sigma)) \cong \mathrm{Isot}(M,\Sigma)$ and $\pi_k(\mathcal{H}(M,\Sigma)) = 0$ for $k \geq 2$.*

Proof By [Johnson and McCullough 2013, Theorem 1], $\pi_k(\mathcal{H}(M,\Sigma)) = \pi_k(\mathrm{Diff}(M))$ for $k \geq 2$, and there is an exact sequence

$$1 \to \pi_1(\mathrm{Diff}(M)) \to \pi_1(\mathcal{H}(M,\Sigma)) \to \mathrm{Isot}(M,\Sigma) \to 1.$$

By Earle and Eells [1969] and Hatcher [1976], $\pi_k(\mathrm{Diff}(M)) = 0$ for $k \geq 1$. □

2.2 The distance of a Heegaard splitting

Let (M,Σ') be a genus $g(\Sigma') \geq 2$ Heegaard splitting of a closed orientable 3–manifold M. Denote by $V^-_{\Sigma'}$ and $V^+_{\Sigma'}$ the handlebodies in M with $V^-_{\Sigma'} \cap V^+_{\Sigma'} = \partial V^-_{\Sigma'} = \partial V^+_{\Sigma'} = \Sigma'$. The *curve graph* $\mathscr{C}(\Sigma')$ is the graph defined as follows. The vertices of $\mathscr{C}(\Sigma')$ are isotopy classes of nontrivial simple closed curves in Σ', and the edges are pairs of vertices that admit disjoint representatives. We denote by $d_{\mathscr{C}(\Sigma')}$ the simplicial metric on $\mathscr{C}(\Sigma')$.

Let $\mathscr{D}^-$ (resp. $\mathscr{D}^+$) denote the set of vertices in $\mathscr{C}(\Sigma')$ that are represented by simple closed curves bounding disks in $V^-_{\Sigma'}$ (resp. $V^+_{\Sigma'}$). Then the (Hempel) *distance* $d(\Sigma')$ of the Heegaard splitting (M,Σ') is defined to be

$$d(\Sigma') := d_{\mathscr{C}(\Sigma')}(\mathscr{D}^-,\mathscr{D}^+).$$

For example, if M contains an essential sphere, then any Heegaard splitting of M has distance zero (see Haken [1968]). If M contains an essential torus, then any Heegaard splitting of M has distance at most two. Furthermore, any Heegaard splitting of a Seifert manifold has distance at most two. See Hempel [2001] for these two facts. As a consequence of the geometrization theorem and these facts, we have:

Theorem 2.3 *Suppose that (M,Σ') is a Heegaard splitting of a closed orientable 3–manifold M. If $d(\Sigma') > 2$, then M admits a hyperbolic structure.*

3 Sweep-outs and graphics

In this section, we recall the definition of graphics and summarize their properties. In what follows, let M denote a closed orientable 3–manifold.

3.1 Graphics

Let (M, Σ) be a Heegaard splitting of M. A *sweep-out* associated with (M, Σ) is a function

$$h\colon M \to J = [-1, 1]$$

such that the level set $h^{-1}(t)$ is a Heegaard surface isotopic to Σ if $t \in \mathrm{Int}(J)$, and $h^{-1}(t)$ is a finite graph in M if $t \in \partial J$. The preimage $h^{-1}(\partial J)$ is called the *spine* of h.

Lemma 3.1 *Let $n > 0$ and (M, Σ) be a Heegaard splitting of a closed orientable 3–manifold M. Let $\varphi\colon D^n \to \mathcal{H}(M, \Sigma)$. Then there exists a family $\{h_u\colon M \to J \mid u \in D^n\}$ of sweep-outs such that $h_u^{-1}(0) = \varphi(u)$ for $u \in D^n$.*

Proof Take a sweep-out $h\colon M \to J$ with $h^{-1}(0) = \Sigma$. We note that

$$\mathrm{Diff}(M) \to \mathrm{Diff}(M)/\mathrm{Diff}(M, \Sigma) = \mathcal{H}(M, \Sigma)$$

is a fibration [Johnson and McCullough 2013]. So, the map φ lifts to a map $\tilde{\varphi}\colon D^n \to \mathrm{Diff}(M)$. Now define $h_u := h \circ \tilde{\varphi}(u)^{-1}$ for $u \in D^n$. □

Let (M, Σ) and (M, Σ') be Heegaard splittings of M. Let $f\colon M \to J$ be a sweep-out with $f^{-1}(0) = \Sigma'$. Furthermore, let $\{h_u\colon M \to J \mid u \in D^2\}$ be a family of sweep-outs associated with (M, Σ). We define the map $\Phi\colon M \times D^2 \to J^2 \times D^2$ by $\Phi(x, u) = (f(x), h_u(x), u)$.

Set $L := \Phi^{-1}(\partial J^2 \times D^2)$, and $W := (M \times D^2) \setminus L$. Define $S = S(\Phi|_W)$ to be the set of all points $w \in W$ such that $\operatorname{rank} d(\Phi|_W)_w < 4$. The image Γ of S in $J^2 \times D^2$ is called the *graphic* defined by f and $\{h_u\}$.

After a small perturbation, we may assume that the map Φ is generic in the following sense. First, for $u \in D^2$, the spine $h_u^{-1}(\partial J)$ intersects each level set of f at finitely many points. Similarly, for $u \in D^2$, the spine $f^{-1}(\partial J)$ intersects each level set of h_u at finitely many points. Furthermore, Φ is "excellent" on W. This means that the set S of singular points of $\Phi|_W$ is a 3–dimensional submanifold in W, and S is divided into four parts, S_2, S_3, S_4 and S_5, where S_k consists of singular points of codimension k. (In the notation of [Boardman 1967], we can write $S_2 = \Sigma^{2,0}$, $S_3 = \Sigma^{2,1,0}$, $S_4 = \Sigma^{2,1,1,0}$ and $S_5 = \Sigma^{2,1,1,1,0} \cup \Sigma^{2,2,0}$.) For $k \neq 5$, Φ has one of the following canonical forms around a point $w \in S_k$:[1] there exist local coordinates (a, b, c, x, y) centered at w and (A, B, X, Y) centered at $\Phi(w)$ such that

$$(A \circ \Phi, B \circ \Phi, X \circ \Phi, Y \circ \Phi) = \begin{cases} (a, b, c, x^2 + y^2) & \text{definite fold } (w \in S_2), \\ (a, b, c, x^2 - y^2) & \text{indefinite fold } (w \in S_2), \\ (a, b, c, x^3 + ax - y^2) & \text{cusp } (w \in S_3), \\ (a, b, c, x^4 + ax^2 + bx + y^2) & \text{definite swallowtail } (w \in S_4), \\ (a, b, c, x^4 + ax^2 + bx - y^2) & \text{indefinite swallowtail } (w \in S_4). \end{cases}$$

[1] We do not know if there exist canonical forms for the singularities of type $\Sigma^{2,2,0}$. However, the singularities in S_5 are not important for our present purpose.

Furthermore, for $2 \le k \le 5$, $\Phi|_{S_k}$ is an immersion with normal crossings, and the images of the S_k are in general position. The main reference about these materials is the book by Golubitsky and Guillemin [1973]. Hatcher and Wagoner [1973] also contains a helpful review for our present purpose.

In the remaining part of the paper, we always assume that the map Φ has the property described above. Under this assumption, Γ has the natural stratification: we can write $\Gamma = F_3 \cup F_2 \cup F_1 \cup F_0$, where $\dim F_k = k$ for $0 \le k \le 3$ and each F_k has the following description.

F_3 This consists of those points $y \in \Gamma$ such that $(\Phi|_S)^{-1}(y) \subset S_2$ and $|(\Phi|_S)^{-1}(y)| = 1$.

F_2 This consists of those points $y \in \Gamma$ such that
- $(\Phi|_S)^{-1}(y) \subset S_2$ and $|(\Phi|_S)^{-1}(y)| = 2$, or
- $(\Phi|_S)^{-1}(y) \subset S_3$ and $|(\Phi|_S)^{-1}(y)| = 1$.

F_1 This consists of those points $y \in \Gamma$ such that
- $(\Phi|_S)^{-1}(y) \subset S_2$ and $|(\Phi|_S)^{-1}(y)| = 3$,
- $(\Phi|_S)^{-1}(y) \subset S_2 \cup S_3$ and $|(\Phi|_S)^{-1}(y)| = 2$, or
- $(\Phi|_S)^{-1}(y) \subset S_4$ and $|(\Phi|_S)^{-1}(y)| = 1$.

F_0 This consists of those points $y \in \Gamma$ such that
- $(\Phi|_S)^{-1}(y) \subset S_2$ and $|(\Phi|_S)^{-1}(y)| = 4$,
- $(\Phi|_S)^{-1}(y) \subset S_2 \cup S_3$ and $|(\Phi|_S)^{-1}(y)| = 3$,
- $(\Phi|_S)^{-1}(y) \subset S_2 \cup S_4$ and $|(\Phi|_S)^{-1}(y)| = 2$,
- $(\Phi|_S)^{-1}(y) \subset S_3$ and $|(\Phi|_S)^{-1}(y)| = 2$, or
- $(\Phi|_S)^{-1}(y) \subset S_5$ and $|(\Phi|_S)^{-1}(y)| = 1$.

3.2 Labeling the regions of $J^2 \times D^2$

In this subsection, we will see that some definitions in [Johnson 2010] can be modified slightly and adapted to our setting.

Let (M, Σ) be a Heegaard splitting. We assume that one component of $M \setminus \Sigma$ is assigned the label $-$ and the other is assigned the label $+$ in some way. We denote by U_Σ^- and U_Σ^+ the components of $M \setminus \Sigma$ labeled by $-$ and $+$ respectively. (Typically, such a labeling is determined by a given sweep-out h with $h^{-1}(0) = \Sigma$. In this case, we can define $U_\Sigma^- = h^{-1}([-1, 0])$ and $U_\Sigma^+ = h^{-1}([0, 1])$.) Such an assignment of the labels $-$ or $+$ to the components of $M \setminus \Sigma$ is called a *transverse orientation* of Σ.

Definition Let (M, Σ) and $U_\Sigma^\pm$ be as above. Suppose $\Sigma' \subset M$ is a closed embedded surface. Then we say that Σ' is *mostly above* Σ if Σ' is transverse to Σ, and if every component of $\Sigma' \cap U_\Sigma^-$ is contained in a disk subset of Σ'. Similarly, we say that Σ' is *mostly below* Σ if Σ' is transverse to Σ, and if every component of $\Sigma' \cap U_\Sigma^+$ is contained in a disk subset of Σ'.

Suppose that $f : M \to J$ is a sweep-out, and that Σ is a transversely oriented Heegaard surface of M. We say that Σ is a *spanning surface* for f if there exist values $a, b \in \mathrm{Int}(J)$ such that $f^{-1}(a)$ is mostly above Σ and $f^{-1}(b)$ is mostly below Σ. We say that Σ is a *splitting surface* for f if it satisfies the following.

First, there does not exist value $s \in \mathrm{Int}(J)$ such that $f^{-1}(s)$ is mostly above or mostly below Σ. Second, $f|_\Sigma$ is *almost Morse*, that is, $f|_\Sigma$ has only nondegenerate critical points and $f|_\Sigma$ is Morse away from -1 and 1, but there may be more than one minima and maxima at the levels -1 and 1 respectively. We note that these definitions are coming from that in [Johnson 2010, Definitions 11 and 12].

Proposition 27 in [Johnson 2010], which will be used in the proof of Theorem 1.2, can be stated in our term as follows:

Lemma 3.2 [Johnson 2010, Proposition 27] *Let $f: M \to J$ be a sweep-out associated with a Heegaard splitting (M, Σ'). If f admits a splitting surface Σ, then $d(\Sigma') \le 2g(\Sigma)$.*

Let (M, Σ) and (M, Σ') be Heegaard splittings of M. Assume that $f: M \to J$ is a sweep-out with $f^{-1}(0) = \Sigma'$, and that $\{h_u: M \to J \mid u \in D^2\}$ is a family of sweep-outs associated with (M, Σ). Let $\Phi: M \times D^2 \to J^2 \times D^2$ be as in the previous subsection. Following [Johnson 2010], let us consider the two subsets $\mathcal{R}_a$ and $\mathcal{R}_b$ of $J^2 \times D^2$ defined as

$$\mathcal{R}_a := \{(s,t,u) \in J^2 \times D^2 \mid f^{-1}(s) \text{ is mostly above } h_u^{-1}(t)\},$$
$$\mathcal{R}_b := \{(s,t,u) \in J^2 \times D^2 \mid f^{-1}(s) \text{ is mostly below } h_u^{-1}(t)\}.$$

Here, for each $u \in D^2$ and $t \in J$, the transverse orientation of $h_u^{-1}(t)$ is determined by the sweep-out h_u. For example, if t is sufficiently close to -1, then the point (s,t,u) is in $\mathcal{R}_a$ because $f^{-1}(s) \cap h_u^{-1}([-1,t])$ consists of finitely many properly embedded disks in the handlebody $h_u^{-1}([-1,t])$. Similarly, if t is sufficiently close to 1, then the point (s,t,u) is in $\mathcal{R}_b$. The regions $\mathcal{R}_a$ and $\mathcal{R}_b$ are nonempty open subsets in $J^2 \times D^2$. The next proposition follows directly from the definition.

Proposition 3.3 *The following hold*:

1. *$\mathcal{R}_a$ and $\mathcal{R}_b$ are disjoint as long as $g(\Sigma') \neq 0$.*
2. *$\mathcal{R}_a$ and $\mathcal{R}_b$ are bounded by Γ.*
3. *The regions $\mathcal{R}_a$ and $\mathcal{R}_b$ are convex in the t–direction, that is, if (s,t,u) is in $\mathcal{R}_a$ (resp. $\mathcal{R}_b$), then so is (s,t',u) for any $t' \le t$ (resp. $t' \ge t$).*

Set $J_u^2 := J^2 \times \{u\} \subset J^2 \times D^2$ for $u \in D^2$. Then, for $u \in D^2$, the intersection $\Gamma \cap J_u^2 \subset J_u^2$ can be viewed as the (2D) graphic defined by sweep-outs f and h_u.

Definition Let f and h_u be as above.

(i) We say that h_u *spans* f if there exists $t \in J$ such that $h_u^{-1}(t)$ is a spanning surface for f.

(ii) We say that h_u *splits* f if there exists $t \in J$ such that $h_u^{-1}(t)$ is a splitting surface for f.

We also say that the graphic defined by f and h_u is *spanned* if h_u spans f. Similarly, we say that the graphic defined by f and h_u is *split* if h_u splits f.

Remark 3.4 By Lemma 3.2, the graphic defined by f and h_u cannot be split if $d(\Sigma') > 2g(\Sigma)$.

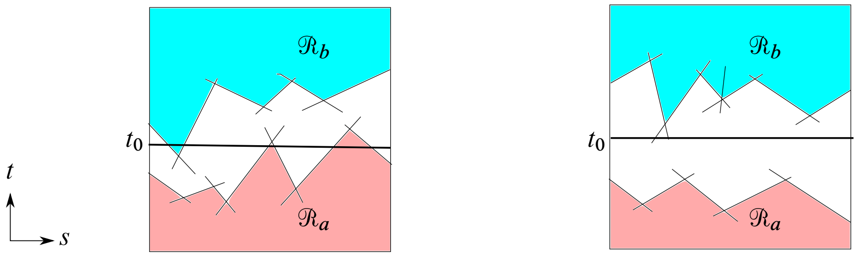


Figure 1: The graphic defined by f and h_u is spanned if there exists a horizontal segment in J_u^2 intersecting both $\mathcal{R}_a$ and $\mathcal{R}_b$ (left). On the other hand, the graphic is split if there exists a horizontal segment disjoint from both $\mathrm{Cl}(\mathcal{R}_a)$ and $\mathrm{Cl}(\mathcal{R}_b)$ (right).

Here are further remarks on the above definition. First, we remark that the condition (i) is equivalent to the following: there exists $t_0 \in J$ such that the horizontal segment $\{t = t_0\}$ in J_u^2 intersects both $\mathcal{R}_a$ and $\mathcal{R}_b$ (the left in Figure 1). We also note that J_u^2 intersects F_3 transversely for $u \in D^2$, and hence $J_u^2 \cap F_3$ consists of finitely many open arcs. This is because $d(p_3 \circ \Phi)_w$ has maximal rank for $w \in W$, where $p_3\colon J^2 \times D^2 \to D^2$ denotes the projection onto the third coordinate. Furthermore, after perturbing Φ if necessary, $J_u^2 \cap F_k$ consists of finitely many points for $0 \le k \le 2$ and $u \in D^2$. Under this assumption, condition (ii) is equivalent to the following: there exists $t_0 \in J$ such that the horizontal segment $\{t = t_0\}$ in J_u^2 is disjoint from both $\mathrm{Cl}(\mathcal{R}_a)$ and $\mathrm{Cl}(\mathcal{R}_b)$ (the right in Figure 1).

Proposition 3.5 *If $g(\Sigma') \ge 2$, then $\mathrm{Cl}(\mathcal{R}_a)$ and $\mathrm{Cl}(\mathcal{R}_b)$ intersect only at points of F_0.*

Proof We first note that by Proposition 3.3(1) and (2), the intersection between $\mathrm{Cl}(\mathcal{R}_a)$ and $\mathrm{Cl}(\mathcal{R}_b)$ is contained in Γ. Suppose that $\mathrm{Cl}(\mathcal{R}_a) \cap \mathrm{Cl}(\mathcal{R}_b) \neq \varnothing$ and $y \in \mathrm{Cl}(\mathcal{R}_a) \cap \mathrm{Cl}(\mathcal{R}_b) \subset \Gamma$. Let (s_0, t_0, u_0) be the coordinate of y. Let l be the segment in $J^2 \times D^2$ defined by $l := \{s = s_0\} \cap \{u = u_0\}$. Note that $y \in l$ by definition. Furthermore, it follows from Proposition 3.3(3) that $l \subset \mathrm{Cl}(\mathcal{R}_a) \cup \mathrm{Cl}(\mathcal{R}_b)$. Consider a point (s_0', t_0, u_0') obtained by perturbing the point (s_0, t_0, u_0) in the s– and u–directions. We may assume that the segment $\tilde{l} := \{s = s_0'\} \cap \{u = u_0'\}$ is transverse to each stratum of Γ. The preimage $\Phi^{-1}(\tilde{l})$ of $\tilde{l}$ is the genus $g(\Sigma')$ Heegaard surface in $M \times \{u_0'\} \subset M \times D^2$, which can be naturally identified with Σ'. Let $h\colon \Sigma' \to J$ denote the function defined to be the restriction of $h_{u_0'}$ on $\Phi^{-1}(\tilde{l}) \cong \Sigma'$. Then h is almost Morse.

Now suppose, for the sake of contradiction, that y is in F_k with $k \ge 1$. Let (s_0', t_-, u_0') denote the coordinate of the intersection point between $\tilde{l}$ and the boundary of $\mathrm{Cl}(\mathcal{R}_a)$. Note that such a point is unique by Proposition 3.3(3). Similarly, let (s_0', t_+, u_0') denote the coordinate of the intersection point between $\tilde{l}$ and the boundary of $\mathrm{Cl}(\mathcal{R}_b)$. Then, h satisfies the following.

- For any regular value $t \in J \setminus [t_-, t_+]$, every loop of $h^{-1}(t)$ is trivial in Σ'.
- The interval $[t_-, t_+]$ contains at most three critical values of h.

It is easily seen that the Euler characteristic of such Σ' must be at least -1, but this is impossible because $g(\Sigma') \ge 2$ by assumption. □

4 Proof of Theorem 1.2

Suppose that M is a closed orientable 3–manifold, and that (M, Σ') is a genus $g(\Sigma') \geq 2$ Heegaard splitting with $d(\Sigma') > 2g(\Sigma') + 2$. Suppose that (M, Σ) is a once-stabilization of (M, Σ'). Let $\mathcal{H}$ denote the path-connected component of $\mathcal{H}(M, \Sigma)$ containing Σ. Let $K^{\pm}$ and $\mathcal{H}^{\pm} \subset \mathcal{H}$ be as in Section 1. Set $\mathcal{H}^{\cup} := \mathcal{H}^{-} \cup \mathcal{H}^{+}$ and $\mathcal{H}^{\cap} := \mathcal{H}^{-} \cap \mathcal{H}^{+}$.

By Theorem 2.2, $\pi_k(\mathcal{H}^{-}) = \pi_k(\mathcal{H}^{+}) = \pi_k(\mathcal{H}^{\cap}) = 0$ for $k \geq 2$. By the Mayer–Vietoris exact sequence, it follows that $H_k(\mathcal{H}^{\cup}; \mathbb{Z}) = 0$ for $k \geq 2$. Applying Hurewicz's theorem, we have $\pi_k(\mathcal{H}^{\cup}) = 0$ for $k \geq 2$. On the other hand, by Theorems 2.1 and 2.3, $\pi_k(\mathcal{H}) = 0$ for $k \geq 2$. So, to prove Theorem 1.2, it is enough to show the following.

Lemma 4.1 *The inclusion $\mathcal{H}^{\cup} \to \mathcal{H}$ induces the isomorphism $\pi_1(\mathcal{H}^{\cup}) \to \pi_1(\mathcal{H})$.*

Johnson [2011] proved that $\pi_1(\mathcal{H})$ is generated by $\pi_1(\mathcal{H}^{-})$ and $\pi_1(\mathcal{H}^{+})$, and hence the induced map is a surjection. (In fact, using the notations in this paper, what he proved in [Johnson 2011] can be written as

$$\pi_1(\mathcal{H}, \mathcal{H}^{\cup}) = 1.$$

See [Johnson 2011, Lemmas 2 and 3]. The following argument is motivated by this observation.) So in this paper, we focus on the proof of the injectivity of the induced map. In other words, we will show the following:

Lemma 4.2 *The second homotopy group $\pi_2(\mathcal{H}, \mathcal{H}^{\cup})$ of the pair $(\mathcal{H}, \mathcal{H}^{\cup})$ vanishes.*

Let $e_0 \in \partial D^2$ be the basepoint. Let $\varphi\colon (D^2, \partial D^2, e_0) \to (\mathcal{H}, \mathcal{H}^{\cup}, \Sigma)$. We will show that $[\varphi] = 0 \in \pi_2(\mathcal{H}, \mathcal{H}^{\cup})$. Let $f\colon M \to J$ be a sweep-out with $f^{-1}(0) = \Sigma'$ and $f^{-1}(\pm 1) = K^{\pm}$. By Lemma 3.1, there exists a family $\{h_u\colon M \to J \mid u \in D^2\}$ of sweep-outs such that $h_u^{-1}(0) = \varphi(u)$ for $u \in D^2$. The key of the proof is the following.

Lemma 4.3 *For any $u \in D^2$, the graphic defined by f and h_u is spanned.*

Proof Suppose, contrary to our claim, there exists $u_0 \in D^2$ such that the graphic defined by f and h_{u_0} is not spanned. Put $J^2_{u_0} := \{(s, t, u) \in J^2 \times D^2 \mid u = u_0\}$. For brevity, we denote the restriction of Φ on $W = (M \times D^2) \setminus L$ by the same symbol Φ in the following. Set $\Gamma := \Phi(S(\Phi))$. The intersection $\Gamma \cap J^2_{u_0} \subset J^2_{u_0}$ is precisely the graphic defined by f and h_{u_0}.

As noted in Remark 3.4, this graphic cannot be split. Then, there exists $t_0 \in J$ such that the horizontal segment $l := \{t = t_0\} \subset J^2_{u_0}$ intersects both $\mathrm{Cl}(\mathcal{R}_a)$ and $\mathrm{Cl}(\mathcal{R}_b)$ at their boundaries (Figure 2). By Proposition 3.5, $\mathrm{Cl}(\mathcal{R}_a)$ and $\mathrm{Cl}(\mathcal{R}_b)$ intersect only at points of F_0. So, pushing l out of $J^2_{u_0}$ slightly, we get an arc $\tilde{l} \subset J^2 \times D^2$ such that

- $\tilde{l}$ is disjoint from both $\mathcal{R}_a$ and $\mathcal{R}_b$, and
- $\tilde{l}$ is transverse to each stratum of Γ.

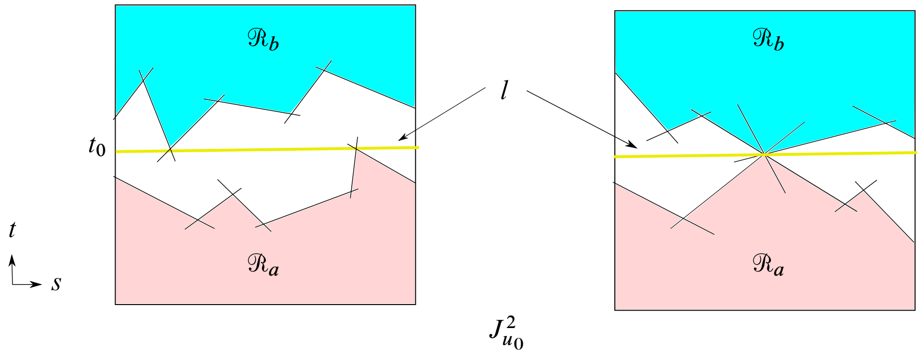


Figure 2: If the graphic defined by f and h_{u_0} is not spanned, then there exists a horizontal segment $l \subset J^2_{u_0}$ intersecting $\mathrm{Cl}(\mathcal{R}_a)$ and $\mathrm{Cl}(\mathcal{R}_b)$ at their boundaries; the intersection is either separate vertices (left) or a single common vertex (right). In either case l can be perturbed in $J^2 \times D^2$ so that l does not meet $\mathrm{Cl}(\mathcal{R}_a) \cup \mathrm{Cl}(\mathcal{R}_b)$.

Furthermore, there is a family $\{l_t \mid t \in I\}$ of arcs with $l_0 = l$ and $l_1 = \tilde{l}$ such that for any $t \in (0,1]$, l_t is transverse to each stratum of Γ. Note that $d(p_2 \circ \Phi)$ and $d(p_3 \circ \Phi)$ have maximal ranks, where $p_2 \colon J^2 \times D^2 \to J$ is the projection onto the second coordinate and $p_3 \colon J^2 \times D^2 \to D^2$ is the projection onto the third coordinate. This means that Φ is transverse to $l = l_0$. As a consequence, Φ is transverse to l_t for all $t \in [0,1]$, and hence $\tilde{\Sigma} := \Phi^{-1}(\tilde{l})$ is a closed embedded surface in $M \times D^2$ isotopic to $\Sigma_{t_0,u_0} := \Phi^{-1}(l)$ $(= h^{-1}_{u_0}(t_0) \times \{u_0\})$. In particular, $g(\tilde{\Sigma}) = g(\Sigma') + 1$.

Let $q_1 \colon M \times D^2 \to M$ denote the projection onto the first coordinate. Since the restriction

$$q_1|_{\Sigma_{t_0,u_0}} \colon \Sigma_{t_0,u_0} \to M$$

is an embedding, so is $q_1|_{\tilde{\Sigma}} \colon \tilde{\Sigma} \to M$. We see that $q_1(\tilde{\Sigma})$ is a splitting surface for f. Consider the restriction $f \circ q_1|_{\tilde{\Sigma}}$ on $\tilde{\Sigma}$. The arc $\tilde{l}$ intersects Γ only at points in F_3, which correspond to fold points of Φ. Thus, $f \circ q_1|_{\tilde{\Sigma}}$ is almost Morse. Let $s \in J$ be any regular value of $f \circ q_1|_{\tilde{\Sigma}}$. By definition, we can write

$$(f \circ q_1|_{\tilde{\Sigma}})^{-1}(s) = (h_u^{-1}(t) \times \{u\}) \cap (f^{-1}(s) \times \{u\}) \subset M \times D^2$$

for some $t \in J$ and $u \in D^2$. Since $\tilde{l}$ is disjoint from both $\mathcal{R}_a$ and $\mathcal{R}_b$, the preimage $(f \circ q_1|_{\tilde{\Sigma}})^{-1}(s)$ contains at least one loop that is nontrivial in the surface $f^{-1}(s) \times \{u\}$. This implies that $q_1(\tilde{\Sigma})$ is a splitting surface for f. But it follows from Lemma 3.2 that $d(\Sigma') \le 2g(\tilde{\Sigma}) = 2g(\Sigma') + 2$, and this contradicts the assumption. □

We now return to the proof of Lemma 4.2.

Proof of Lemma 4.2 Let $p_2 \colon J^2 \times D^2 \to J$ denote the projection onto the second coordinate. For $u \in D^2$, set $I_u := p_2(\mathrm{Cl}(\mathcal{R}_a)) \cap p_2(\mathrm{Cl}(\mathcal{R}_b))$. Then $t \in J$ is in $\mathrm{Int}(I_u)$ if and only if $h_u^{-1}(t)$ is a spanning surface for f. By Lemma 4.3, each I_u is a nonempty subset of J. Furthermore, it follows from Proposition 3.3(3) that each I_u is a closed interval in J. So $\bigsqcup_{u \in D^2} I_u$ is an (trivial) I–bundle over D^2.

Let $\sigma\colon D^2 \to \bigsqcup_{u\in D^2} I_u$ be a section of this I–bundle. Define $\tilde\sigma\colon D^2 \to \mathcal{H}$ by $\tilde\sigma(u) := h_u^{-1}(\sigma(u))$. Recall that $\varphi(u) = h_u^{-1}(0)$ for $u \in D^2$. The straight line homotopy connecting the 0–section of $J \times D^2 \to D^2$ to σ induces the homotopy $\{\varphi_r\colon D^2 \to \mathcal{H} \mid r \in [0,1]\}$ with $\varphi_0 = \varphi$ and $\varphi_1 = \tilde\sigma$. By Proposition 3.3(3), we may choose σ such that for $u \in \partial D^2$, $\varphi(u)$ is isotopic to $\tilde\sigma(u)$ through surfaces disjoint from K^- or K^+, depending on if $\varphi(u) \in \mathcal{H}^+$ or $\varphi(u) \in \mathcal{H}^-$ holds. This means that $\varphi_r(u) \in \mathcal{H}^\cup$ for $u \in \partial D^2$ and $r \in [0,1]$. Clearly, $\{\varphi_r\}$ can be chosen so that it preserves the basepoint. Thus, φ and $\tilde\sigma$ represent the same element of $\pi_2(\mathcal{H}, \mathcal{H}^\cup)$. Applying the homotopy described above, from now on, we may assume that the map φ satisfies the following: for any $u \in D^2$, $\Sigma_u := \varphi(u)$ is a spanning surface for f.

We think about a fixed $u \in D^2$ for a moment. By assumption, there exist values $a, b \in J$ such that $\Sigma'^+ := f^{-1}(a)$ is mostly above Σ_u, and $\Sigma'^- := f^{-1}(b)$ is mostly below Σ_u. By definition, every loop of $\Sigma_u \cap (\Sigma'^+ \cup \Sigma'^-)$ bounds a disk in $\Sigma'^+ \cup \Sigma'^-$. The following observation is due to Johnson [2011].

Claim *One of the two (possibly both) holds:*

1. *Every loop in $\Sigma_u \cap \Sigma'^+$ bounds a disk in Σ_u.*
2. *Every loop in $\Sigma_u \cap \Sigma'^-$ bounds a disk in Σ_u.*

Proof If we compress the surface Σ_u along innermost loops in $\Sigma'^+ \cup \Sigma'^-$ repeatedly, we have a collection of surfaces disjoint from both Σ'^+ and Σ'^-. The point is that there is a surface S in the collection that separates Σ'^+ from Σ'^-, and so S is in the product region between Σ'^+ and Σ'^-. Note that such S must have genus at least $g(\Sigma')$. This means that at most one of the two surfaces Σ'^+ and Σ'^- contains an actual compression for Σ_u (ie a loop in $\Sigma_u \cap \Sigma'^+$ or $\Sigma_u \cap \Sigma'^-$ that is nontrivial in Σ_u) because $g(\Sigma_u) = g(\Sigma') + 1$. Therefore, either (1) or (2) holds. □

Put $T' = \Sigma'^+ \cup \Sigma'^-$. Take a loop ℓ in $\Sigma_u \cap T'$ satisfying the following condition:

(∗) ℓ is trivial in Σ_u and ℓ is innermost in T' among all the loops of $\Sigma_u \cap T'$.

If we compress Σ_u along ℓ and discard the sphere component, then the loop ℓ (and possibly some other loops in $\Sigma_u \cap T'$) is removed. Since M is irreducible, this process can actually be achieved by an isotopy. Repeating this process as long as possible, all the loops in $\Sigma_u \cap T'$ satisfying the condition (∗) are finally removed. In particular, the resulting surface is disjoint from Σ'^+ or Σ'^- depending on if (1) or (2) holds.

We wish to do the above process simultaneously for $u \in D^2$. In fact, it is always possible using an argument of Hatcher [1976]. The following is a sketch of the argument in [Hatcher 1976].

We will construct a smooth family $\{\Theta_{u,r}\colon \Sigma_u \to M \mid u \in D^2, r \in [0,1]\}$ of isotopies such that for any $u \in D^2$, $\Theta_{u,0}(\Sigma_u) = \Sigma_u$ and $\Theta_{u,1}(\Sigma_u)$ is disjoint from either K^- or K^+. By the above argument, we can see that there exist a finite cover $\{B_i\}$ of D^2 with $B_i \cong D^2$ and a family $\{T_i'\}$ of (disconnected) surfaces with the following properties:

- T_i' is the union of the two level surfaces $\Sigma_i'^+$ and $\Sigma_i'^-$ of f.
- If $u \in B_i$, then $\Sigma_i'^+$ is mostly above Σ_u, and $\Sigma_i'^-$ is mostly below Σ_u.

For $u \in D^2$, let $\widetilde{\mathscr{C}}_u$ be the set of intersection loops between Σ_u and $\bigcup_i T_i'$, where the union is taken over all i such that $u \in B_i$. We denote by $D'(\ell)$ the disk in T_i' bounded by ℓ for $\ell \in \widetilde{\mathscr{C}}_u$. (Note that such a disk is unique because each component of T_i' is not homeomorphic to S^2.) For $u \in D^2$, let $\mathscr{C}_u$ be the subset of $\widetilde{\mathscr{C}}_u$ consisting of those loops ℓ such that ℓ is trivial in Σ_u, and that $D'(\ell)$ contains no other intersection loop that is nontrivial in Σ_u. Furthermore, for $\ell \in \mathscr{C}_u$, we denote by $D(\ell)$ the disk in Σ_u bounded by ℓ. (Again, note that such a disk is unique because $\Sigma_u \neq S^2$.) For $u \in D^2$, we define the partial order $<'$ on $\mathscr{C}_u$ by

$$\ell <' m \iff D'(\ell) \subset D'(m).$$

Let $\{B_i'\}$ be a finite cover of D^2 obtained by shrinking each B_i slightly so that $B_i' \subset \mathrm{Int}(B_i)$ for i. Take a family $\{\alpha_u \colon \mathscr{C}_u \to (0,2) \mid u \in D^2\}$ of functions with the following properties:

- If $\ell, m \in \mathscr{C}_u$ and $\ell <' m$, then $\alpha_u(\ell) < \alpha_u(m)$.
- If $u \in B_i'$ and $\ell \subset \Sigma_u \cap T_i'$, then $\alpha_u(\ell) < 1$.
- If $u \in \partial B_i$ and $\ell \subset \Sigma_u \cap T_i'$, then $\alpha_u(\ell) > 1$.

The function α_u shows the times when intersection loops that belong to $\mathscr{C}_u$ are eliminated by compressing. Let G denote the union of the images of the α_u in $D^2 \times [0,2]$. Note that for each intersection loop ℓ, the images of the loops corresponding to ℓ form a 2D sheet over some B_i, and so G can be written as the union of these sheets. We can view G as a "chart" to compress the surface Σ_u; if we compress Σ_u following this chart upward from $r = 0$ to $r = 2$, then we get the sequence of surfaces. Note that the following subtle case may occur: if ℓ and m are loops of $\Sigma_u \cap T_i'$ with $D(\ell) \subset D(m)$ and $\alpha_u(m) < \alpha_u(\ell)$, then the loop ℓ is eliminated automatically before the time $\alpha_u(\ell)$. This example shows that we should use the "reduced" chart $\widehat{G}$ rather than G, which is obtained from G by removing the parts of the sheets corresponding to any such ℓ.

For every u, we will define the isotopy $\Theta_{u,r}$ as follows. Let $N(\widehat{G})$ denote a small fibered neighborhood of $\widehat{G}$. The interval $\{u\} \times [0,2]$ intersects $N(\widehat{G})$ at its subintervals $J_u^{(k)}$, where $1 \le k \le n = n(u)$. Define $\widetilde{\Theta}_{u,r}$ to be the isotopy obtained by piecing together the isotopies $\theta_{u,r}^{(1)}, \dots, \theta_{u,r}^{(n)}$ in the way suggested by $\widehat{G}$. Here each $\theta_{u,r}^{(k)}$ is an isotopy with its r–support in $J_u^{(k)}$, and corresponds to the compression along a loop in $\mathscr{C}_u$. See Figure 3. Now we define $\Theta_{u,r}$ as the restriction of $\widetilde{\Theta}_{u,r}$ on $[0,1]$.

It remains to see that we can modify the above construction to get the isotopy $\{\Theta_{u,r}\}$ to be smooth for $u \in D^2$. It is enough to show that each factor $\theta_{u,r}^{(k)}$ of $\Theta_{u,r}$ can be chosen so that it varies smoothly for u. For simplicity, we will think about the isotopy $\theta_{u,r}^{(1)}$ in the following although the same argument applies to any $\theta_{u,r}^{(k)}$. The isotopy $\theta_{u,r}^{(1)}$ corresponds to the compression along a loop $\ell_u \in \mathscr{C}_u$ for each u. Assume that $\ell_u \subset T_i'$ for any u. Denote by $D^3(\ell_u)$ the 3–ball in M bounded by the 2–sphere $D(\ell_u) \cup D'(\ell_u)$. (Note that such a 3–ball is unique because $M \neq S^3$.) Let (D^3, D, D') be the standard triple of disks, that is, D and D' are the upper and lower hemispheres in the boundary ∂D^3 of the standard 3–ball D^3, respectively. There is an identification $\phi_u \colon (D^3(\ell_u), D(\ell_u), D'(\ell_u)) \to (D^3, D, D')$ for every u. Then the arguments

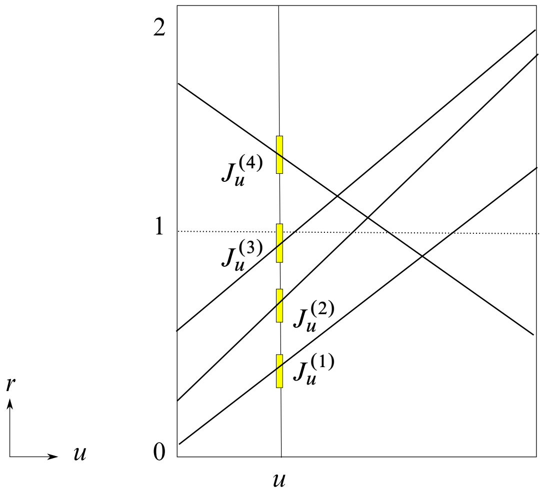


Figure 3: The isotopy $\widetilde{\Theta}_{u,r}$ is obtained by piecing together the isotopies $\theta_{u,r}^{(1)}$, $\theta_{u,r}^{(2)}$ $\theta_{u,r}^{(3)}$ and $\theta_{u,r}^{(4)}$. (In this example, $\widetilde{\Theta}_{u,r}$ can be written as the concatenation $\theta_{u,r}^{(1)} * \theta_{u,r}^{(2)} * \theta_{u,r}^{(3)} * \theta_{u,r}^{(4)}$ of the small isotopies.) The r–support of $\theta_{u,r}^{(k)}$ is contained in $J_u^{(k)}$.

in [Hatcher 1976] together with the Smale conjecture (the space $\mathrm{Diff}(D^3 \operatorname{rel} \partial D^3)$ is contractible), which is proved in [Hatcher 1983], show that ϕ_u can be chosen such that it varies smoothly for u.[2] Now we can define $\theta_{u,r}^{(1)} \equiv \phi_u^{-1} \circ F_r \circ \phi_u$ on $D(\ell_u) \subset \Sigma_u$ and $\theta_{u,r}^{(1)} \equiv \Theta_{u,0}$ on the complement of a small neighborhood of $D(\ell_u)$ in Σ_u, where $\{F_r : D^3 \to D^3 \mid r \in [0,1]\}$ is an isotopy that carries D to D' across D^3. Therefore, it follows that $\{\Theta_{u,r}\}$ is smooth for $u \in D^2$.

Finally, we see that $\Theta_{u,1}(\Sigma_u) \in \mathcal{H}^{\cup}$ for $u \in D^2$. Let u be any point in D^2. Take a path $\rho : [0,1] \to D^2$ with $\rho(0) = e_0$ and $\rho(1) = u$. It suffices to show that the path $\tilde{\rho} : [0,1] \to \mathcal{H}$ defined by $\tilde{\rho}(t) := \Theta_{\rho(t),1}(\Sigma_{\rho(t)})$ is wholly contained in $\mathcal{H}^{\cup}$.

For brevity, we denote by Σ_t the surface $\Theta_{\rho(t),1}(\Sigma_{\rho(t)})$ for $t \in [0,1]$ in the following. The cover $\{B_i\}$ of D^2 induces the cover $\{I_k \mid 0 \le k \le n\}$ of $[0,1]$ by finitely many closed intervals. By passing to a subcover if necessary, we may assume that $I_k \cap I_j = \varnothing$ if $|k - j| > 1$. As we have seen above, there exists a family $\{\Sigma_k'^{+} \cup \Sigma_k'^{-}\}$ $(= \{T_k'\})$ of level surfaces of f and the following hold:

- $\Sigma_k'^{+}$ is mostly above Σ_t if $t \in I_k$. Similarly, $\Sigma_k'^{-}$ is mostly below Σ_t if $t \in I_k$.
- For each k, one of the two surfaces $\Sigma_k'^{+}$ and $\Sigma_k'^{-}$ is disjoint from Σ_t if $t \in I_k$.

As is naturally expected, the following holds:

Claim *Suppose that $t \in I_k$. If $\Sigma_t \cap \Sigma_k'^{+} = \varnothing$ and $\Sigma_t \cap \Sigma_k'^{-} \neq \varnothing$, then $\Sigma_t \in \mathcal{H}^{-}$. Similarly, if $\Sigma_t \cap \Sigma_k'^{-} = \varnothing$ and $\Sigma_t \cap \Sigma_k'^{+} \neq \varnothing$, then $\Sigma_t \in \mathcal{H}^{+}$.*

[2]More specifically, we need the arguments at the end of Section 1 in [Hatcher 1976], where the sought isotopy, denoted by h_{tu} in that paper, is constructed. It starts by taking a suitable triangulation of D^n and then proceeds by extending the isotopy over the k–skeleton inductively. The homotopy group $\pi_k(\mathrm{Diff}(D^3 \operatorname{rel} D))$ appears as an obstruction to extending a map. (As we work in the smooth category, we use the Smale conjecture instead of the Alexander trick.)

Proof The proof is by induction on k. The following argument is based on the idea in [Johnson 2011]. By definition, $\Sigma_t \in \mathcal{H}^\cap$ for $t \in I_0$. Thus our claim holds on I_0. So, in what follows, we assume that $k > 0$ and that our claim holds on any interval I_j with $0 \le j < k$.

Let $t \in I_k$. Without loss of generality, we may assume that $\Sigma_t \cap \Sigma_k'^+ = \varnothing$ and $\Sigma_t \cap \Sigma_k'^- \neq \varnothing$. Fix $t_0 \in I_k \cap I_{k-1}$. Note that Σ_{t_0} and Σ_t are isotopic through surfaces disjoint from K^+. Thus it is enough to show that $\Sigma_{t_0} \in \mathcal{H}^-$. There are three cases to consider.

Case 1 $\Sigma_{t_0} \cap \Sigma_{k-1}'^+ = \varnothing$ and $\Sigma_{t_0} \cap \Sigma_{k-1}'^- \neq \varnothing$.

By the assumption of induction, this implies that $\Sigma_{t_0} \in \mathcal{H}^-$ and our claim holds in this case.

Case 2 $\Sigma_{t_0} \cap \Sigma_{k-1}'^+ \neq \varnothing$ and $\Sigma_{t_0} \cap \Sigma_{k-1}'^- = \varnothing$.

We will see that $\Sigma_{t_0} \in \mathcal{H}^\cap$. This is same as saying that Σ_{t_0} is a Heegaard surface of

$$M \setminus \mathrm{Int}(N(K^+ \cup K^-)) \cong \Sigma' \times J,$$

where $N(K^+ \cup K^-)$ is a sufficiently small neighborhood of $K^+ \cup K^-$. Since $\Sigma_{t_0} \cap \Sigma_{k-1}'^- = \Sigma_{t_0} \cap \Sigma_k'^+ = \varnothing$, Σ_{t_0} separates K^+ from K^-. First, we see that Σ_{t_0} is bicompressible in $M \setminus (K^+ \cup K^-)$. By assumption, there exists a loop $\ell \subset \Sigma_{t_0} \cap \Sigma_{k-1}'^+$ bounding a disk $D^- \subset \Sigma_{k-1}'^+$ such that ℓ is nontrivial in Σ_{t_0}. Similarly, there exists a loop $m \subset \Sigma_{t_0} \cap \Sigma_k'^-$ bounding a disk $D^+ \subset \Sigma_k'^-$ such that m is nontrivial in Σ_{t_0}. Since D^- and D^+ are in the opposite side of Σ_{t_0} to each other, Σ_{t_0} is bicompressible in $M \setminus (K^+ \cup K^-)$.

It is known that any genus $g(\Sigma') + 1$ bicompressible surface in $\Sigma' \times J$ separating $\Sigma' \times \{1\}$ from $\Sigma' \times \{-1\}$ must be reducible (see [Johnson 2011]). This means that there exists a 2–sphere $P \subset M \setminus (K^+ \cup K^-)$ intersecting Σ_{t_0} at a single nontrivial loop in Σ_{t_0}. Since M is irreducible, P cuts (M, Σ_{t_0}) into the two Heegaard splittings: one is a genus $g(\Sigma')$ Heegaard splitting of M and the other is a genus 1 Heegaard splitting of S^3. If we denote by S the genus $g(\Sigma')$ surface obtained by cutting Σ_{t_0} along P, then S still separates K^+ from K^-. Thus, S is isotopic to Σ' in the complement of $K^+ \cup K^-$. This shows that Σ_{t_0} is a genus $g(\Sigma') + 1$ Heegaard surface in $M \setminus \mathrm{Int}(N(K^+ \cup K^-))$. Therefore, we conclude that $\Sigma_{t_0} \in \mathcal{H}^\cap$ in this case.

Case 3 $\Sigma_{t_0} \cap \Sigma_{k-1}'^+ = \Sigma_{t_0} \cap \Sigma_{k-1}'^- = \varnothing$.

Let j denote the minimal integer such that for any $j < j' \le k$ and $t \in I_{j'}$, $\Sigma_t \cap \Sigma_{j'}'^+ = \Sigma_t \cap \Sigma_{j'}'^- = \varnothing$. If $j = 0$, then Σ_{t_0} is isotopic to Σ_0 through surfaces disjoint from $K^+ \cup K^-$. This shows that $\Sigma_{t_0} \in \mathcal{H}^\cap$. So we may assume that $j > 0$ in the following. Let $t_1 \in I_j \cap I_{j+1}$.

First, we assume that $\Sigma_{t_1} \cap \Sigma_j'^+ = \varnothing$ and that $\Sigma_{t_1} \cap \Sigma_j'^- \neq \varnothing$. By the assumption of induction, it follows that $\Sigma_{t_1} \in \mathcal{H}^-$. Since Σ_{t_1} and Σ_{t_0} are isotopic in $M \setminus (K^+ \cup K^-)$, we have $\Sigma_{t_0} \in \mathcal{H}^-$ in this case.

Next, we assume that $\Sigma_{t_1} \cap \Sigma_j'^+ \neq \varnothing$ and that $\Sigma_{t_1} \cap \Sigma_j'^- = \varnothing$. Then, there exists a compression disk $D^- \subset \Sigma_j'^+$ for Σ_{t_1}. Since Σ_{t_1} and Σ_{t_0} are isotopic in $M \setminus (K^+ \cup K^-)$, Σ_{t_0} has a compression disk disjoint from K^- as well. On the other hand, as we have seen above, $\Sigma_k'^-$ contains a compression disk

D^+ for Σ_{t_0} lying in the opposite side of Σ_{t_0} to D^-. Thus, Σ_{t_0} is bicompressible in $M \setminus (K^+ \cup K^-)$. Now applying the same argument as in Case 2, we have $\Sigma_{t_0} \in \mathcal{H}^\cap$ and this completes the proof. ☐

The above claim implies that the image of $\tilde{\rho}\colon [0,1] \to \mathcal{H}$ is contained in $\mathcal{H}^\cup$. In particular, $\Theta_{u,1}(\Sigma_u) \in \mathcal{H}^\cup$. Therefore, we conclude that $[\varphi] = 0 \in \pi_2(\mathcal{H}, \mathcal{H}^\cup)$ and this finishes the proof of Lemma 4.2. ☐

5 The isotopy subgroup of a Heegaard splitting of a handlebody

5.1 Proof of Theorem 1.3

We now give a proof of Theorem 1.3. Let V be a genus $g(V) \geq 2$ handlebody, and let (V, Σ) be a genus $g(V)+1$ Heegaard splitting of V. Fix a complete system $E_1, \dots, E_{g(V)}$ of meridian disks for V. Consider a properly embedded, boundary parallel arc I in V that is disjoint from $\bigcup_{i=1}^{g(V)} E_i$. The surface Σ can be viewed as the boundary of a small neighborhood $N(\partial V \cup I)$ of $\partial V \cup I$. In the same spirit in [Johnson and McCullough 2013], we define the space $\mathrm{Unk}(V, I)$ of unknotted arcs to be $\mathrm{Diff}(V)/\mathrm{Diff}(V, I)$. Then, the following holds:

Theorem 5.1 [Scharlemann 2013, Theorem 5.1] *The group* $\mathrm{Isot}(V, \Sigma)$ *is isomorphic to* $\pi_1(\mathrm{Unk}(V, I))$.

Thus, it suffices to show that $\pi_1(\mathrm{Unk}(V, I))$ is finitely presented.

Fix a parallelism disk E for I disjoint from $\bigcup_{i=1}^{g(V)} E_i$. Furthermore, fix a spine K of V such that $K \cap E = \varnothing$ and K intersects each E_i at a single point. We now consider the two subspaces of $\mathrm{Unk}(V, I)$:

$$U_1 := \{I' \in \mathrm{Unk}(V, I) \mid I' \text{ admits a parallelism disk } E' \text{ with } E' \cap K = \varnothing\},$$

$$U_2 := \left\{ I' \in \mathrm{Unk}(V, I) \;\middle|\; I' \cap \bigcup_{i=1}^{g(V)} E_i = \varnothing \right\}.$$

Note that U_1, U_2 and $U_1 \cap U_2$ are all connected.

The group $\pi_1(U_1)$ is identical to the group $\mathfrak{F}_E$ in [Scharlemann 2013], which is called the *freewheeling* subgroup in that paper. This group is an extension of $\pi_1(\partial V)$ by $\mathbb{Z}$, and generated by λ_i, μ_i $(1 \leq i \leq g(V))$ and ρ shown in Figure 4. For each i, λ_i is represented by an isotopy of parallelism disk E along a longitudinal loop that intersects ∂E_i at a single point. Similarly, μ_i is represented by an isotopy of the parallelism disk E along a meridional loop corresponding to ∂E_i. The set $\{\lambda_i, \mu_i \mid 1 \leq i \leq g(V)\}$ corresponds to a generating set of $\pi_1(\partial V)$, and ρ is defined to be the half rotation of the parallelism disk E. Let P denote the planar surface obtained by cutting ∂V along simple closed curves $\partial E_1, \dots, \partial E_{g(V)}$. Then, the group $\pi_1(U_2)$ is isomorphic to the 2–braid group $B_2(P)$ of P. Following [Scharlemann 2013], we define the *anchored* subgroup $\mathfrak{A}_{E_1,\dots,E_{g(V)}}$ of $\pi_1(U_2)$ as follows. This is generated by $2g(V)$ elements α_i and α_i' $(1 \leq i \leq g(V))$ shown in Figure 5. Here each of α_i and α_i' is represented by an isotopy of I that moves the one endpoint p_1 of I along a meridional loop and fixes the other endpoint p_0. Note that we can write $\alpha_i' = \lambda_i^{-1}\alpha_i\lambda_i$ as elements of $\pi_1(\mathrm{Unk}(V, I))$. The group $\pi_1(U_2)$ is generated by $\mathfrak{A}_{E_1,\dots,E_{g(V)}}$ and ρ.

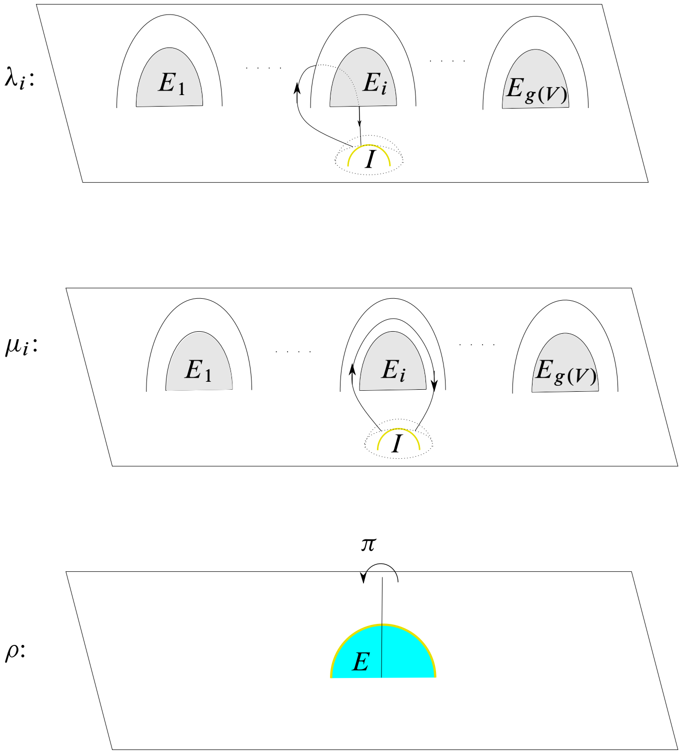


Figure 4: The group $\pi_1(U_1)$ is generated by $2g(V)+1$ elements.

The groups $\pi_1(U_1)$, $\pi_1(U_2)$ and $\pi_1(U_1 \cap U_2)$ are all finitely presented. By van Kampen's theorem, the proof is finished if the following is shown:

Lemma 5.2 *The inclusion* $U_1 \cup U_2 \to \mathrm{Unk}(V, I)$ *is a homotopy equivalence.*

In fact, by the same argument as in Section 4, it is easily seen that $\pi_k(U_1 \cup U_2) = \pi_k(\mathrm{Unk}(V, I)) = 0$ for $k \geq 2$. (And of course, this fact is unnecessary for our present purpose.) So we will see that the natural map $\pi_1(U_1 \cup U_2) \to \pi_1(\mathrm{Unk}(V, I))$ is an isomorphism.

Proof For brevity, set $U := \mathrm{Unk}(V, I)$. In [Scharlemann 2013], it was shown that $\pi_1(U)$ is generated by the two subgroups $\pi_1(U_1)$ $(= \mathfrak{F}_E)$ and $\mathfrak{A}_{E_1,\dots,E_{g(V)}}$ $(\subset \pi_1(U_2))$. It follows from this fact that the map

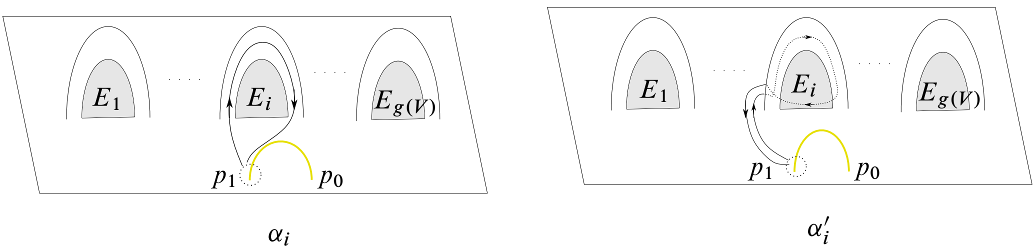


Figure 5: The group $\mathfrak{A}_{E_1,\dots,E_{g(V)}} \subset \pi_1(U_2)$ is generated by $2g(V)$ elements.

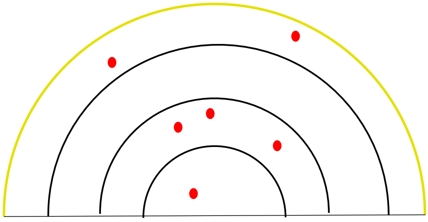


Figure 6: For $u \in \partial D^2$, the parallelism disk E_u intersects $\bigcup_{i=1}^{g(V)} E_i$ only at arcs parallel to I_u and intersects K at finitely many points.

$\pi_1(U_1 \cup U_2) \to \pi_1(U)$ is a surjection. We will see that the map $\pi_1(U_1 \cup U_2) \to \pi_1(U)$ is an injection. In other words, we will see that $\pi_2(U, U_1 \cup U_2) = 0$. Let $\varphi\colon (D^2, \partial D^2) \to (U, U_1 \cup U_2)$. Put $I_u := \varphi(u)$ for $u \in D^2$. In the same spirit of the proof of Lemma 3.1, we can show the following.

Claim 1 *There exists a (smooth) family of disks $\{E_u \mid u \in D^2\}$ in V such that E_u is a parallelism disk for I_u.*

Proof By [Scharlemann 2013], the map $\mathrm{Diff}(V) \to \mathrm{Diff}(V)/\mathrm{Diff}(V, \Sigma)$ is homotopy equivalent to $\mathrm{Diff}(V) \to \mathrm{Diff}(V)/\mathrm{Diff}(V, I)$. The former is a fibration [Johnson and McCullough 2013], and so is the latter. Thus, the map $\varphi\colon D^2 \to U$ lifts to a map $\tilde{\varphi}\colon D^2 \to \mathrm{Diff}(V)$. Now define $E_u := \tilde{\varphi}(u)(E)$. □

Since $\varphi(\partial D^2) \subset U_1 \cup U_2$, the isotopy $\{I_u \mid u \in \partial D^2\}$ represents an element of $\pi_1(U_1 \cup U_2)$. So we can write this isotopy as a product $\omega_1 \omega_2 \cdots \omega_n$ of the ω_k's, where each ω_k is either λ_i, μ_i, ρ, α_i, α_i' or their inverses. Corresponding to this factorization, there is a division of ∂D^2 into the intervals $J_1 = [u_0, u_1], \ldots, J_n = [u_{n-1}, u_n]$ with $u_0 = u_n$.

Claim 2 *After a deformation of $\{E_u \mid u \in D^2\}$ near ∂D^2, the following hold for any $u \in \partial D^2$:*

(i) *E_u intersects $\bigcup_{i=1}^{g(V)} E_i$ at finitely many arcs, and E_u intersects K at finitely many points.*

(ii) *Each arc of $E_u \cap \bigcup_{i=1}^{g(V)} E_i$ is parallel to I_u in E_u.*

(iii) *If a and a' are arcs of $E_u \cap \bigcup_{i=1}^{g(V)} E_i$, then a and a' are nested in the following sense: if Δ and Δ' are bigons in E_u cut by a and a' respectively, then either $\Delta \subset \Delta'$ or $\Delta' \subset \Delta$ holds.*

See Figure 6.

Proof The key is the following simple observation. For each interval J_k, there are the three possibilities:

- $\omega_k = \lambda_i^{\epsilon}$ for some $1 \le i \le g(V)$ and $\epsilon = \pm 1$. Then, during the move ω_k, some intersection arcs of $E_u \cap \bigcup_{i=1}^{g(V)} E_i$ are introduced or removed (possibly both may occur). All such arcs are parallel to I_u in E_u. The intersection pattern of $E_u \cap K$ is not changed by ω_k. See Figure 7, top.
- $\omega_k = \alpha_i^{\epsilon}$ or $\omega_k = \alpha_i'^{\epsilon}$ for some $1 \le i \le g(V)$ and $\epsilon = \pm 1$. Then, during the move ω_k, a single intersection point of $E_u \cap K$ is introduced or removed. The intersection pattern of $E_u \cap \bigcup_{i=1}^{g(V)} E_i$ is not changed by ω_k. See Figure 7, bottom.

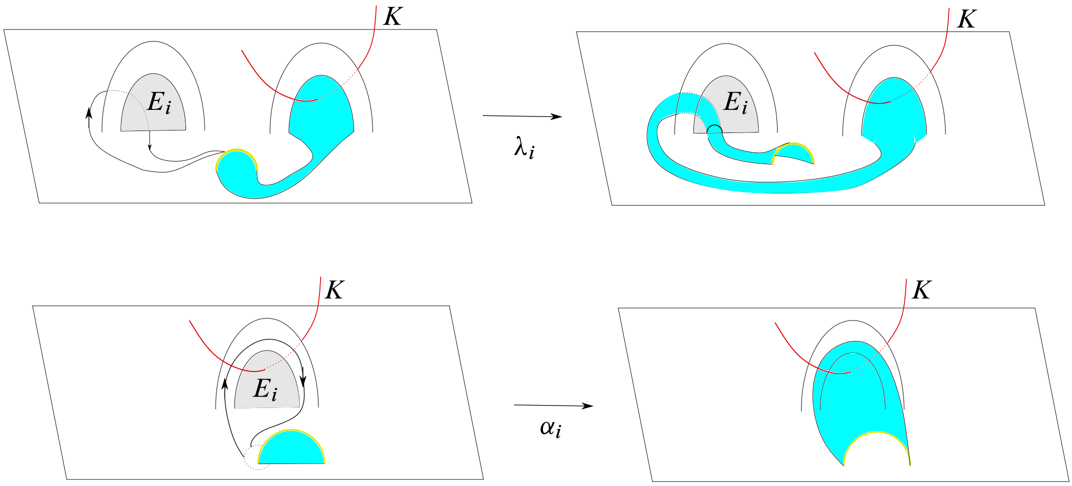


Figure 7: The move λ_i introduces or removes arcs parallel to I_u (top), and the move α_i introduces or removes a single point (bottom).

- $\omega_k = \mu_i^\epsilon$ or $\omega_k = \rho^\epsilon$ for some $1 \le i \le g(V)$ and $\epsilon = \pm 1$. Then, during the move ω_k, the intersection pattern of $E_u \cap \left(\bigcup_{i=1}^{g(V)} E_i \cup K\right)$ does not change.

Recall that $E_{u_0} = E_{u_n} = E$ by definition. In particular, $E_{u_0} \cap \left(\bigcup_{i=1}^{g(V)} E_i \cup K\right) = \varnothing$. By the above observation, it follows that the conditions (i), (ii) and (iii) are satisfied on the interval J_1. By an inductive argument, we can see that these three conditions are satisfied on any interval J_k as well. □

Put $B := \{re^{\sqrt{-1}\theta} \in \mathbb{C} \mid 0 \le r \le 1,\ 0 \le \theta \le \pi\}$. There is a smooth family $\{f_u : E_u \to B \mid u \in D^2\}$ of diffeomorphisms between E_u and B. (More rigorously, this is a consequence of the fact that the space $\mathrm{Diff}(D^2 \operatorname{rel} \partial D^2)$ is contractible [Smale 1959].) Furthermore, by Claim 2, we may choose $\{f_u\}$ such that for any $u \in \partial D^2$, each arc of $E_u \cap \bigcup_{i=1}^{g(V)} E_i$ is mapped to an arc $\{re^{\sqrt{-1}\theta} \in B \mid r = r_0, 0 \le \theta \le \pi\}$ for some $0 < r_0 \le 1$. For $t \in [0,1)$, define $\sigma_t : B \to B$ by $\sigma_t(re^{\sqrt{-1}\theta}) := (1-t)re^{\sqrt{-1}\theta}$. Set $\Theta_{u,t} := f_u^{-1} \circ \sigma_t \circ f_u$ for $u \in D^2$ and $t \in [0,1)$. Then, the isotopy $\Theta_{u,t}$ shrinks I_u along E_u into a small neighborhood of a point in $E_u \cap \partial V$ as $t \to 1$. If t is sufficiently close to 1, then $\Theta_{u,t}(I_u) \in U_1$. Furthermore, by definition, for $u \in \partial D^2$ and $t \in [0,1)$, $\Theta_{u,t}(I_u)$ is disjoint from either K or $\bigcup_{i=1}^{g(V)} E_i$. Let $u \in \partial D^2$ and $t \in [0,1)$. If $\Theta_{u,t}(I_u) \cap K = \varnothing$, then $\Theta_{u,t}(I_u) \in U_1$. On the other hand, if $\Theta_{u,t}(I_u) \cap \bigcup_{i=1}^{g(V)} E_i = \varnothing$, then $\Theta_{u,t}(I_u) \in U_2$. This means that $\Theta_{u,t}(I_u) \in U_1 \cup U_2$ for $u \in \partial D^2$ and $t \in [0,1)$. Therefore, we conclude that $[\varphi] = 0 \in \pi_2(U, U_1 \cup U_2)$. □

5.2 Proof of Corollary 1.4

Proof of Corollary 1.4 By Theorems 1.1 and 1.3, $\mathrm{Isot}(M,\Sigma)$ is finitely presented. It remains to show that $\mathrm{MCG}(M,\Sigma)$ is finitely presented. By definition, there exists an exact sequence

$$1 \to \mathrm{Isot}(M,\Sigma) \to \mathrm{MCG}(M,\Sigma) \to \mathrm{MCG}(M).$$

By Theorem 2.3, M is hyperbolic, and hence $\mathrm{MCG}(M)$ is finite. Therefore, $\mathrm{MCG}(M,\Sigma)$ is finitely presented. □

References

[Boardman 1967] **J M Boardman**, *Singularities of differentiable maps*, Inst. Hautes Études Sci. Publ. Math. 33 (1967) 21–57 MR Zbl

[Cerf 1968] **J Cerf**, *Sur les difféomorphismes de la sphère de dimension trois* $(\Gamma_4 = 0)$, Lecture Notes in Math. 53, Springer (1968) MR Zbl

[Earle and Eells 1969] **C J Earle**, **J Eells**, *A fibre bundle description of Teichmüller theory*, J. Differential Geom. 3 (1969) 19–43 MR Zbl

[Golubitsky and Guillemin 1973] **M Golubitsky**, **V Guillemin**, *Stable mappings and their singularities*, Grad. Texts in Math. 14, Springer (1973) MR Zbl

[Haken 1968] **W Haken**, *Some results on surfaces in* 3*–manifolds*, from "Studies in modern topology" (P J Hilton, editor), Stud. Math. 5, Math. Assoc. America, Concord, CA (1968) 39–98 MR Zbl

[Hatcher 1976] **A Hatcher**, *Homeomorphisms of sufficiently large* P^2*–irreducible* 3*–manifolds*, Topology 15 (1976) 343–347 MR Zbl

[Hatcher 1983] **A E Hatcher**, *A proof of the Smale conjecture,* $\mathrm{Diff}(S^3) \simeq \mathrm{O}(4)$, Ann. of Math. 117 (1983) 553–607 MR Zbl

[Hatcher and Wagoner 1973] **A Hatcher**, **J Wagoner**, *Pseudo-isotopies of compact manifolds*, Astérisque 6, Soc. Math. France, Paris (1973) MR Zbl

[Hempel 2001] **J Hempel**, 3*–Manifolds as viewed from the curve complex*, Topology 40 (2001) 631–657 MR Zbl

[Johnson 2010] **J Johnson**, *Bounding the stable genera of Heegaard splittings from below*, J. Topol. 3 (2010) 668–690 MR Zbl

[Johnson 2011] **J Johnson**, *Mapping class groups of once-stabilized Heegaard splittings*, preprint (2011) arXiv 1108.5302

[Johnson and McCullough 2013] **J Johnson**, **D McCullough**, *The space of Heegaard splittings*, J. Reine Angew. Math. 679 (2013) 155–179 MR Zbl

[Koda and Sakuma 2023] **Y Koda**, **M Sakuma**, *Homotopy motions of surfaces in* 3*–manifolds*, Q. J. Math. 74 (2023) 29–71 MR Zbl

[Rubinstein and Scharlemann 1996] **H Rubinstein**, **M Scharlemann**, *Comparing Heegaard splittings of non-Haken* 3*–manifolds*, Topology 35 (1996) 1005–1026 MR Zbl

[Scharlemann 2013] **M Scharlemann**, *Generating the genus* $g+1$ *Goeritz group of a genus* g *handlebody*, from "Geometry and topology down under" (C D Hodgson, W H Jaco, M G Scharlemann, S Tillmann, editors), Contemp. Math. 597, Amer. Math. Soc., Providence, RI (2013) 347–369 MR Zbl

[Smale 1959] **S Smale**, *Diffeomorphisms of the* 2*–sphere*, Proc. Amer. Math. Soc. 10 (1959) 621–626 MR Zbl

Higashihiroshima, Japan

diguchi00@gmail.com

Received: 4 April 2022 Revised: 8 January 2023

Algebraic & Geometric Topology **24**:7 (2024) 3641–3668
DOI: 10.2140/agt.2024.24.3641
Published: 9 December 2024

On the structure of the top homology group of the Johnson kernel

IGOR A SPIRIDONOV

The Johnson kernel is the subgroup $\mathcal{K}_g$ of the mapping class group $\mathrm{Mod}(\Sigma_g)$ of a genus-g oriented closed surface Σ_g generated by all Dehn twists about separating curves. We study the structure of the top homology group $\mathrm{H}_{2g-3}(\mathcal{K}_g,\mathbb{Z})$. For any collection of $2g-3$ disjoint separating curves on Σ_g, one can construct the corresponding abelian cycle in the group $\mathrm{H}_{2g-3}(\mathcal{K}_g,\mathbb{Z})$; such abelian cycles will be called simple. We describe the structure of a $\mathbb{Z}[\mathrm{Mod}(\Sigma_g)/\mathcal{K}_g]$–module on the subgroup of $\mathrm{H}_{2g-3}(\mathcal{K}_g,\mathbb{Z})$ generated by all simple abelian cycles and find all relations between them.

20F34; 20F36, 20J05, 57M07

1 Introduction

Let Σ_g be a compact oriented genus-g surface. Let $\mathrm{Mod}(\Sigma_g)=\pi_0(\mathrm{Homeo}^+(\Sigma_g))$ be the *mapping class group* of Σ_g, where $\mathrm{Homeo}^+(\Sigma_g)$ is the group of orientation-preserving homeomorphisms of Σ_g. The group $\mathrm{Mod}(\Sigma_g)$ acts on $H=H_1(\Sigma_g,\mathbb{Z})$. This action preserves the algebraic intersection form, so we have the representation $\mathrm{Mod}(\Sigma_g)\to \mathrm{Sp}(2g,\mathbb{Z})$, which is well known to be surjective. The kernel $\mathcal{I}_g$ of this representation is known as the *Torelli group*. This can be written as the short exact sequence

$$1\to\mathcal{I}_g\to\mathrm{Mod}(\Sigma_g)\to\mathrm{Sp}(2g,\mathbb{Z})\to 1. \tag{1}$$

The *Johnson kernel* $\mathcal{K}_g$ is the subgroup of $\mathcal{I}_g$ generated by all Dehn twists about separating curves. Johnson [15] proved that the group $\mathcal{K}_g$ can also be defined as the kernel of the surjective *Johnson homomorphism* $\tau:\mathcal{I}_g\to\bigwedge^3 H/H$, where the inclusion $H\hookrightarrow\bigwedge^3 H$ is given by $x\mapsto x\wedge\Omega$ and $\Omega\in\bigwedge^2 H$ is the inverse tensor of the algebraic intersection form. Therefore we have the short exact sequence

$$1\to\mathcal{K}_g\to\mathcal{I}_g\to\textstyle\bigwedge^3 H/H\to 1. \tag{2}$$

Denote by $\mathcal{G}_g$ the quotient group $\mathrm{Mod}(\Sigma_g)/\mathcal{K}_g$. The exact sequences (1) and (2) imply that $\mathcal{G}_g$ can be presented as the extension

$$1\to\textstyle\bigwedge^3 H/H\to\mathcal{G}_g\to\mathrm{Sp}(2g,\mathbb{Z})\to 1$$

of the symplectic group by the free abelian group $\bigwedge^3 H/H$. The group $H_*(\mathcal{K}_g,\mathbb{Z})$ has the natural structure of a $\mathcal{G}_g$–module.

In the case $g=1$ the representation $\mathrm{Mod}(\Sigma_1)\to\mathrm{Sp}(2,\mathbb{Z})=\mathrm{SL}(2,\mathbb{Z})$ is an isomorphism, so the group $\mathcal{I}_1$ is trivial. Mess [16] proved that the group $\mathcal{I}_2=\mathcal{K}_2$ is free with a countable number of generators. Therefore below we assume that $g\geq 3$ unless explicitly stated otherwise.

A natural problem is to study the homology of the group $\mathcal{K}_g$ for $g \geq 3$. The rational homology group $H_1(\mathcal{K}_g, \mathbb{Q})$ was shown to be finitely generated for $g \geq 4$ by Dimca and Papadima [7]. This group was computed explicitly for $g \geq 6$ by Morita, Sakasai and Suzuki [17] using the description due to Dimca, Hain and Papadima [6]. Recently Ershov and Sue He [8] proved that $\mathcal{K}_g$ is finitely generated in the case $g \geq 12$. This result was extended to any genus $g \geq 4$ by Church, Ershov and Putman [5]. This implies that the group $H_1(\mathcal{K}_g, \mathbb{Z})$ is finitely generated, provided that $g \geq 4$. It is still unknown whether $\mathcal{K}_3$ and $H_1(\mathcal{K}_3, \mathbb{Z})$ are finitely generated.

Bestvina, Bux and Margalit [2] computed the cohomological dimension of the Johnson kernel $\operatorname{cd}(\mathcal{K}_g) = 2g-3$. Gaifullin [11] proved that the top homology group $H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$ contains a free $\mathbb{Z}[\bigwedge^3 H/H]$–module of infinite rank. In particular, $H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$ is not finitely generated.

Recall that for n pairwise commuting elements $h_1, \dots, h_n$ of the group G, one can construct the *abelian cycle* $\mathcal{A}(h_1, \dots, h_n) \in H_n(G, \mathbb{Z})$ defined as follows. Consider the homomorphism $\phi\colon \mathbb{Z}^n \to G$ that maps the generator of the i^{th} factor to h_i. Then $\mathcal{A}(h_1, \dots, h_n) = \phi_*(\mu_n)$, where μ_n is the standard generator of $H_n(\mathbb{Z}^n, \mathbb{Z})$.

By a *curve* we always mean an essential simple closed curve on Σ_g. By an (oriented) *multicurve* we mean a finite union of pairwise disjoint and nonisotopic (oriented) curves on Σ_g. An *ordered multicurve* is a multicurve with a fixed order on its components. Usually we will not distinguish between a curve or a multicurve and its isotopy class. We denote by T_γ the left Dehn twist about a curve γ.

Definition 1.1 An *S–multicurve* is an ordered multicurve consisting of $2g-3$ separating components.

For example, the multicurve $\delta_1 \cup \dots \cup \delta_g \cup \epsilon_2 \cup \dots \cup \epsilon_{g-2}$ in Figure 1 is an S–multicurve. To an S–multicurve $M = \gamma_1 \cup \dots \cup \gamma_{2g-3}$ we assign the abelian cycle $\mathcal{A}(M) = \mathcal{A}(T_{\gamma_1}, \dots, T_{\gamma_{2g-3}}) \in H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$. If an S–multicurve M' is obtained from M by a permutation π of its components, then $\mathcal{A}(M') = (\operatorname{sign} \pi)\mathcal{A}(M)$. An easy computation of the Euler characteristic implies that any S–multicurve separates Σ_g into g one-punctured tori and $g-3$ three-punctured spheres. Throughout the paper we assume that the components of any S–multicurve are ordered so that the curves with numbers $1, \dots, g$ bound one-punctured tori from Σ_g.

Abelian cycles of the form $\mathcal{A}(M) \in H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$ for some S–multicurve M will be called *simple abelian cycles*. Denote by $\mathcal{A}_g \subseteq H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$ the subgroup generated by all simple abelian cycles. The author does not know the answer to the following question, which is an interesting problem itself:

Question 1.2 Is the inclusion $\mathcal{A}_g \subseteq H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$ strict for some $g \geq 3$?

The natural problem is to study the structure of the group $\mathcal{A}_g$. Let us identify Σ_g with the surface shown in Figure 1 and fix the multicurve $\Delta = \delta_1 \cup \dots \cup \delta_g \cup \epsilon_2 \cup \dots \cup \epsilon_{g-2}$ on Σ_g. Denote by $\mathcal{P}_g \subseteq \mathcal{A}_g$ the subgroup generated by all simple abelian cycles $\mathcal{A}(M)$, where $M = \delta_1 \cup \dots \cup \delta_g \cup \epsilon'_2 \cup \dots \cup \epsilon'_{g-2}$ for some separating curves $\epsilon'_2, \dots, \epsilon'_{g-2}$.

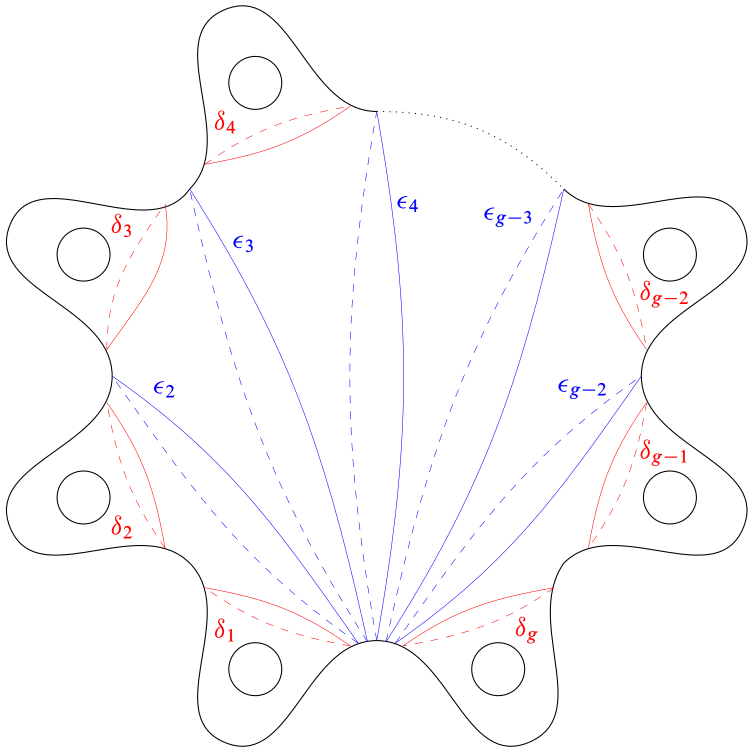


Figure 1: The surface Σ_g.

By an *unordered symplectic splitting* we mean an orthogonal (with respect to the intersection form) decomposition of H into a direct sum of g subgroups of rank 2. Write $N = \delta_1 \cup \dots \cup \delta_g$, and for each i let X_i be the one-punctured torus bounded by δ_i. Set $V_i = H_1(X_i, \mathbb{Z}) \subset H$. Consider the corresponding unordered symplectic splitting $H = \bigoplus_i V_i$. The group $\mathrm{Sp}(2g, \mathbb{Z})$ acts on the set of all unordered symplectic splittings. Denote by $\mathcal{H}_g = \mathrm{SL}(2, \mathbb{Z})^{\times g} \rtimes S_g$ the stabilizer of the unordered splitting $\mathcal{V} = \{V_1, \dots, V_n\}$ in $\mathrm{Sp}(2g, \mathbb{Z})$, where S_g is the symmetric group.

The group $\mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(N)$ preserves the splitting $\mathcal{V}$; therefore the image of the natural homomorphism $\mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(N) \to \mathrm{Sp}(2g, \mathbb{Z})$ coincides with $\mathcal{H}_g$. Consider the corresponding mapping

$$\eta\colon \mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(N) \twoheadrightarrow \mathcal{H}_g.$$

We check in Proposition 2.1 that $\ker(\eta) \subseteq \mathcal{K}_g$, so we have the commutative diagram

(3)
$$\begin{array}{ccccccccc}
1 & \longrightarrow & \mathcal{K}_g & \longrightarrow & \mathrm{Mod}(\Sigma_g) & \longrightarrow & \mathcal{G}_g & \longrightarrow & 1 \\
 & & \uparrow & & \uparrow & & \uparrow & \searrow & \\
 & & & & & & & & \mathrm{Sp}(2g, \mathbb{Z}) \\
 & & & & & & & \nearrow & \\
1 & \longrightarrow & \mathrm{Stab}_{\mathcal{K}_g}(N) & \longrightarrow & \mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(N) & \xrightarrow{\ \eta\ } & \mathcal{H}_g & \longrightarrow & 1
\end{array}$$

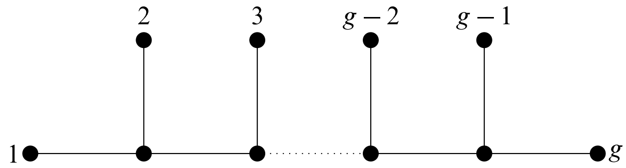


Figure 2: The dual tree $\mathcal{T}_0 = \mathcal{T}(\Delta)$ to the S–multicurve Δ.

Therefore we have the maps $\mathcal{H}_g \to \mathcal{G}_g$. Since the inclusion $\mathcal{H}_g \hookrightarrow \mathrm{Sp}(2g,\mathbb{Z})$ passes through $\mathcal{G}_g$, we have the inclusion $\mathcal{H}_g \hookrightarrow \mathcal{G}_g$. The second row of (3) implies that the group $\mathcal{H}_g = \mathrm{SL}(2,\mathbb{Z})^{\times g} \rtimes S_g$ acts on $\mathcal{P}_g$. The action of $\mathrm{SL}(2,\mathbb{Z})^{\times g}$ is trivial, and therefore $\mathcal{P}_g$ is an S_g–module. The first part of the main result is as follows:

Theorem 1.3 *There is an isomorphism of $\mathcal{G}_g$–modules*

$$\mathcal{A}_g \cong \mathrm{Ind}^{\mathcal{G}_g}_{\mathrm{SL}(2,\mathbb{Z})^{\times g} \rtimes S_g} \mathcal{P}_g.$$

In order to describe the S_g–module $\mathcal{P}_g$, we need to introduce some notation. Denote by $\boldsymbol{T}_g$ the set of trees $\mathcal{T}$ such that

(I) $\mathcal{T}$ has g leaves (vertices of degree 1) marked by $1,\dots,g$, and

(II) degrees of all other vertices of $\mathcal{T}$ equal 3.

We consider such trees up to an isomorphism preserving marking of the leaves. One can prove that $|\boldsymbol{T}_g| = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2g-5)$. For example, $\boldsymbol{T}_3$ consists of a single element.

For each S–multicurve M we consider the *dual tree* $\mathcal{T}(M)$, ie the graph that has a vertex for each connected component of $\Sigma_g \setminus M$ and where two vertices are adjacent if and only if the corresponding connected components are adjacent to each other. Since each component of M is separating, it follows that $\mathcal{T}(M)$ is a tree. The tree $\mathcal{T}(M)$ has g leaves corresponding to one-punctured tori; degrees of all other vertices equal 3. By definition components of M are ordered, so we also have an order on the set of curves that bound one-punctured tori on Σ_g. Each leaf of $\mathcal{T}(M)$ corresponds to a component of M with a number from 1 to g. Therefore the leaves of $\mathcal{T}(M)$ are numbered from 1 to g. Hence $\mathcal{T}(M)$ is an element of $\boldsymbol{T}_g$. For example, the dual tree $\mathcal{T}_0 = \mathcal{T}(\Delta)$ for the multicurve Δ in Figure 1 is shown in Figure 2.

Recall that we have the fixed curves $\delta_1,\dots,\delta_g$ on Σ_g as in Figure 1. For each $\mathcal{T} \in \boldsymbol{T}_g$ we can find a multicurve $\xi_2 \cup \dots \cup \xi_{g-2}$ disjoint from $\delta_1,\dots,\delta_g$ and consisting of separating components such that $\mathcal{T}$ is the dual tree to the multicurve $\Delta_{\mathcal{T}} = \delta_1 \cup \dots \cup \delta_g \cup \xi_2 \cup \dots \cup \xi_{g-2}$. Such a multicurve $\xi_2 \cup \dots \cup \xi_{g-2}$ is not unique, but we will prove that all such multicurves $\delta_1 \cup \dots \cup \delta_g \cup \xi_2 \cup \dots \cup \xi_{g-2}$ belong to the same $\mathcal{K}_g$–orbit; see Proposition 2.4. Therefore the simple abelian cycle $\mathcal{A}_{\mathcal{T}} = \mathcal{A}(\Delta_{\mathcal{T}}) \in H_{2g-3}(\mathcal{K}_g,\mathbb{Z})$ is defined uniquely up to a sign. The sign of $\mathcal{A}_{\mathcal{T}}$ depends on the ordering of the curves $\xi_2,\dots,\xi_{g-2}$.

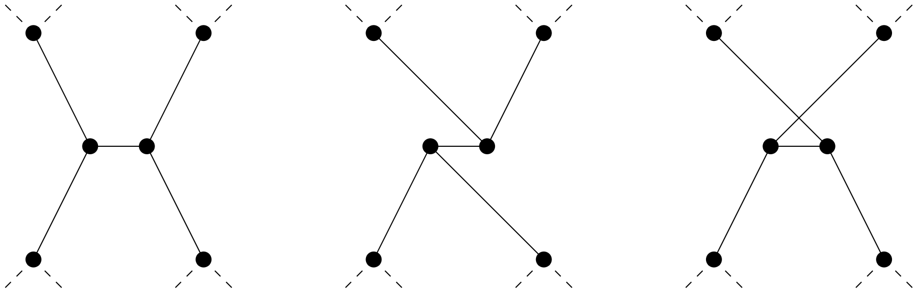


Figure 3: A cyclic triple of trees.

Let $h_1 = 1, h_2, h_3, \ldots \in \mathrm{Sp}(2g, \mathbb{Z})$ be representatives of all left cosets $\mathrm{Sp}(2g, \mathbb{Z})/\mathcal{H}_g$ and let $\hat{h}_1, \hat{h}_2, \hat{h}_3, \ldots \in \mathrm{Mod}(\Sigma_g)$ be mapping classes that go to $h_1, h_2, h_3, \ldots$ under the natural surjective homomorphism $\mathrm{Mod}(\Sigma_g) \twoheadrightarrow \mathrm{Sp}(2g, \mathbb{Z})$. Gaifullin [11, Theorem 1.3] proved that the abelian cycles

$$\hat{h}_s \cdot \mathcal{A}_{\mathcal{T}_0} \quad \text{for } s = 1, 2, 3, \ldots$$

form a basis of a free $\mathbb{Z}\left[\bigwedge^3 H/H\right]$–submodule of $H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$. In particular, these simple abelian cycles are nonzero and generate a free abelian group.

Definition 1.4 A triple of trees $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\} \subseteq \boldsymbol{T}_g$ is called *cyclic* if they differ only as shown in Figure 3 (upper and lower vertices in Figure 3 can be either leaves or not).

Theorem 1.5 *The abelian group $\mathcal{P}_g$ has a presentation where the generators are $\{\mathcal{A}_\mathcal{T} \mid \mathcal{T} \in \boldsymbol{T}_g\}$ and the relations are*

$$\{\mathcal{A}_{\mathcal{T}_1} + \mathcal{A}_{\mathcal{T}_2} + \mathcal{A}_{\mathcal{T}_3} = 0 \mid \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\} \textit{ is a cyclic triple}\}. \tag{4}$$

Remark 1.6 Recall that the signs of the simple abelian cycles $\mathcal{A}_\mathcal{T}$ depend on the order of the components of the corresponding S–multicurve. If $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$ is a cyclic triple, then the corresponding S–multicurves $\Delta_{\mathcal{T}_1}$, $\Delta_{\mathcal{T}_2}$ and $\Delta_{\mathcal{T}_3}$ differ by only one component. In (4) we mean that the components of these three S–multicurves are ordered so that the orderings coincide at $2g-4$ positions.

Remark 1.7 The "hard part" of Theorem 1.5 is the fact that any relation between simple abelian cycles follows from (4). However, the existence of such relations is not hard. For example, one can deduce (4) from the Lantern relation; see Farb and Margalit [9, Proposition 5.1]. Our proof is based on Arnold's relations in the cohomology of the pure braid group.

Also, we find an explicit basis of $\mathcal{P}_g$. For each tree $\mathcal{T} \in \boldsymbol{T}_g$ we say that the leaf with number g is the *root*, so $\mathcal{T}$ is a *rooted tree*. In this case, for each vertex the set of its *descendant leaves* is well defined.

Definition 1.8 Let $\mathcal{T} \in \boldsymbol{T}_g$. A vertex of $\mathcal{T}$ of degree 3 is called *balanced* if the paths from it to the two descendant leaves with the two smallest numbers have no common edges. The tree $\mathcal{T}$ is called *balanced* if all its vertices of degree 3 are balanced. The set of all balanced trees is denoted by $\boldsymbol{T}_g^b \subseteq \boldsymbol{T}_g$.

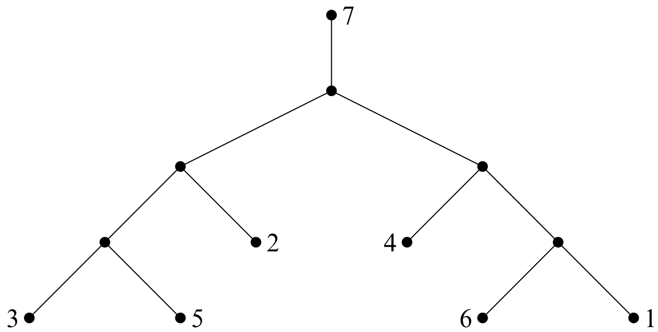


Figure 4: An example of a balanced tree in the case $g = 7$.

An example of a balanced tree for $g = 7$ is shown in Figure 4.

Theorem 1.9 *The simple abelian cycles* $\{\mathcal{A}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}_g^b\}$ *form a basis of* $\mathcal{P}_g$, *and* $\operatorname{rk} \mathcal{P}_g = |\boldsymbol{T}_g^b| = (g-2)!$.

Theorems 1.3, 1.5 and 1.9 provide a complete description of the group $\mathcal{A}_g$.

Acknowledgments The author would like to thank his advisor Alexander A Gaifullin for stating the problem, useful discussions, and constant attention to this work. The author is a winner of the all-Russian mathematical August Möbius contest of graduate and undergraduate student papers and thanks the jury for the high praise of his work.

The author was partially supported by the HSE University Basic Research Program and by the Simons Foundation.

2 Preliminaries and sketch of proof

2.1 Mapping class group of a surface with punctures and boundary components

Let Σ be an oriented surface, possibly with punctures and boundary components. We do not assume that Σ is connected. However, we require $H_*(\Sigma, \mathbb{Q})$ be a finite-dimensional vector space. The mapping class group of Σ is defined as $\operatorname{Mod}(\Sigma) = \pi_0(\operatorname{Homeo}^+(\Sigma, \partial\Sigma))$, where $\operatorname{Homeo}^+(\Sigma, \partial\Sigma)$ is the group of orientation-preserving homeomorphisms of Σ that restrict to the identity on $\partial\Sigma$. By $\operatorname{PMod}(\Sigma) \subseteq \operatorname{Mod}(\Sigma)$ we denote the *pure mapping class group* of Σ, ie the subgroup consisting of those elements fixing each of the punctures and each of the connected components. We have the exact sequence

$$1 \to \operatorname{PMod}(\Sigma_{g,n}^b) \to \operatorname{Mod}(\Sigma_{g,n}^b) \to S_n \to 1, \tag{5}$$

where $\Sigma_{g,n}^b$ denotes the connected genus-g surface with n punctures and b boundary components. For example, the pure mapping class group of the disk with n punctures is precisely the *pure braid group* $\mathrm{PB}_n = \operatorname{PMod}(\Sigma_{0,n}^1)$.

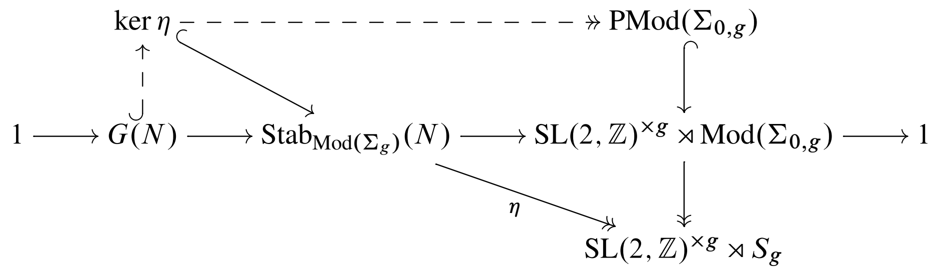
2.2 The Birman–Lubotzky–McCarthy exact sequence

Let M be a multicurve on Σ_g. Then, denoting by $G(M)$ the group generated by Dehn twists about the components of M, we have the Birman–Lubotzky–McCarthy exact sequence (see [3, Lemma 2.1])

$$1 \to G(M) \to \mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(M) \to \mathrm{Mod}(\Sigma_g \setminus M) \to 1. \tag{6}$$

Take $N = \delta_1 \cup \cdots \cup \delta_g$ as in Figure 1 and consider the group $\mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(N)$. We have $\Sigma_g \setminus N = \Sigma_{0,g} \sqcup X_1 \sqcup \cdots \sqcup X_g$, where X_i is the one-punctured torus bounded by δ_i. Since $\mathrm{Mod}(X_i) \cong \mathrm{SL}(2,\mathbb{Z})$, (5) and (6) imply the existence of the following commutative diagram:

(7)



This yields the exact sequence

$$1 \to G(N) \to \ker\eta \to \mathrm{PMod}(\Sigma_{0,g}) \to 1. \tag{8}$$

Proposition 2.1 *The following sequence is exact*:

$$1 \to \mathrm{Stab}_{\mathcal{K}_g}(N) \to \mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(N) \xrightarrow{\eta} \mathrm{SL}(2,\mathbb{Z})^{\times g} \rtimes S_g \to 1. \tag{9}$$

Proof First let us show that $\mathrm{Stab}_{\mathcal{K}_g}(N) \subseteq \ker\eta$. Indeed, any element $\phi \in \mathrm{Stab}_{\mathcal{K}_g}(N)$ stabilizes each component of N, so it also stabilizes each X_i. Since $\mathcal{K}_g$ is contained in the Torelli group, it follows that the restriction of ϕ to $\mathrm{Mod}(X_i) \cong \mathrm{SL}(2,\mathbb{Z}) \subset \mathrm{Sp}(2g,\mathbb{Z})$ is trivial for all i.

Let us prove the opposite inclusion. The groups $G(M)$ and $\mathrm{PMod}(\Sigma_{g,n})$ are generated by Dehn twists about separating curves. The exact sequence (8) implies that the same is true for $\ker\eta$, and therefore $\ker\eta \subseteq \mathcal{K}_g$. Hence $\ker\eta \subseteq \mathrm{Stab}_{\mathcal{K}_g}(N)$. □

Lemma 2.2 *There is an isomorphism*

$$\mathrm{Stab}_{\mathcal{K}_g}(N) \cong \mathbb{Z}^{g-1} \times \mathrm{PB}_{g-1}. \tag{10}$$

Proof We need the following fact:

Fact 2.3 [9, Section 9.3] The center of the group PB_{g-1} is the infinite cyclic group, which is generated by the Dehn twist about the boundary curve. Moreover, we have the split exact sequence

$$1 \to \mathbb{Z} \xrightarrow{j_1} \mathrm{PB}_{g-1} \to \mathrm{PMod}(\Sigma_{0,g}) \to 1,$$

where j_1 is the inclusion of the center of PB_{g-1}.

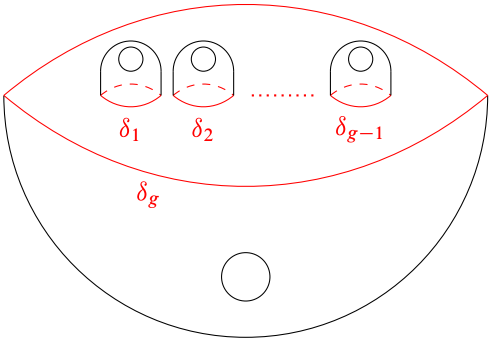


Figure 5

Consider the obvious map

$$j\colon \mathbb{Z}^g \cong \mathbb{Z}^{g-1} \times \mathbb{Z} \hookrightarrow \mathbb{Z}^{g-1} \times \mathrm{PB}_{g-1},$$

where the restriction of j on the first factor is the identity isomorphism and the restriction of j on the second factor is j_1. Fact 2.3 along with the exactness of (8) implies that in order to finish the proof of Lemma 2.2 we need to construct the map $\psi\colon \mathbb{Z}^{g-1} \times \mathrm{PB}_{g-1} \to \mathrm{Stab}_{\mathcal{K}_g}(N)$ such that the following diagram commutes:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathbb{Z}^g & \longrightarrow & \mathbb{Z}^{g-1} \times \mathrm{PB}_{g-1} & \longrightarrow & \mathrm{PMod}(\Sigma_{0,g}) & \longrightarrow & 1 \\ & & \| & & \downarrow \psi & & \| & & \\ 1 & \longrightarrow & G(N) & \longrightarrow & \mathrm{Stab}_{\mathcal{K}_g}(N) & \longrightarrow & \mathrm{PMod}(\Sigma_{0,g}) & \longrightarrow & 1 \end{array}$$

We define ψ as follows. The generator of the i^{th} factor of $\mathbb{Z}^{g-1}$ maps to T_{δ_i}. In order to define the restriction of ψ on the factor PB_{g-1}, let us identify Σ_g with the surface shown in Figure 5. We have the disk bounded by δ_g with $g-1$ handles bounded by $\delta_1, \dots, \delta_{g-1}$. We can replace all these handles by punctures and identify the group PB_{g-1} with the corresponding group $\mathrm{PMod}(\Sigma^1_{0,g-1})$. Then we extend the mapping classes in $\mathrm{PMod}(\Sigma^1_{0,g-1})$ to the handles so that the handles do not rotate.

Since the pure braid group is generated by Dehn twists about separating curves it follows that the image of ψ is contained in $\mathcal{K}_g$. The 5–lemma completes the proof of Lemma 2.2. □

2.3 Simple abelian cycles

Recall that for an S–multicurve $M = \gamma_1 \cup \dots \cup \gamma_{2g-3}$ on Σ_g there is the corresponding simple abelian cycle $\mathcal{A}(M) = \mathcal{A}(T_{\gamma_1}, \dots, T_{\gamma_{2g-3}}) \in H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$. We have already constructed the simple abelian cycles $\mathcal{A}_{\mathcal{T}} = \mathcal{A}(\Delta_{\mathcal{T}})$ for $\mathcal{T} \in \boldsymbol{T}_g$.

Proposition 2.4 *Let $\Delta_{\mathcal{T}} = \delta_1 \cup \dots \cup \delta_g \cup \xi_2 \cup \dots \cup \xi_{g-2}$ and $\Delta'_{\mathcal{T}} = \delta_1 \cup \dots \cup \delta_g \cup \xi'_2 \cup \dots \cup \xi'_{g-2}$ be two S–multicurves with the same dual tree $\mathcal{T}$. Then $\Delta_{\mathcal{T}}$ and $\Delta'_{\mathcal{T}}$ belong to the same $\mathcal{K}_g$–orbit (up to a permutation of the components). In particular, the simple abelian cycles $\{\mathcal{A}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}_g\}$ are well defined.*

Proof Since $\Delta_{\mathcal{T}}$ and $\Delta'_{\mathcal{T}}$ have the same dual tree, there is an element $\phi \in \mathrm{Mod}(\Sigma_g)$ such that after a permutation of $\xi_2, \dots, \xi_{g-2}$ we have $\phi(\xi_i) = \xi'_i$ and $\phi(\delta_j) = \delta_j$ for all i and j. Also, we can assume that $\phi|_{X_j} = \mathrm{id}$ for all $1 \le j \le g$. Then $\phi \in \ker \eta$ (see the exact sequence (9)), so $\phi \in \mathcal{K}_g$. Therefore all such S–multicurves $\Delta_{\mathcal{T}}$ belong to the same $\mathcal{K}_g$–orbit, so the simple abelian cycle $\mathcal{A}_{\mathcal{T}} = \mathcal{A}(\Delta_{\mathcal{T}}) \in H_{2g-3}(\mathcal{K}(\Sigma_g, \mathbb{Z}))$ is defined uniquely up to a sign. □

Proposition 2.5 *The simple abelian cycles* $\{\mathcal{A}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}_g\}$ *generate* $\mathcal{A}_g$ *as a* $\mathbb{Z}[\mathcal{G}_g]$*–module.*

Proof Consider an S–multicurve $M' = \gamma_1 \cup \dots \cup \gamma_{2g-3}$. We can assume that the curves $\gamma_1, \dots, \gamma_g$ bound one-punctured tori on Σ_g. There is an element $\phi \in \mathrm{Mod}(\Sigma_g)$ such that $\phi(\gamma_j) = \delta_j$ for all j. Then $\phi \cdot \mathcal{A}(M') = \pm \mathcal{A}_{\mathcal{T}(M')}$, so $\mathcal{A}(M') = \pm \phi^{-1} \cdot \mathcal{A}_{\mathcal{T}(M')}$. This implies the proposition. □

2.4 Sketches of the proofs of Theorems 1.3, 1.5 and 1.9

By Lemma 2.2 we can consider the simple abelian cycles $\{\mathcal{A}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}_g\}$ as elements of the group $H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(N), \mathbb{Z}) \cong H_{2g-3}(\mathbb{Z}^{g-1} \times \mathrm{PB}_{g-1})$. The following proposition will be proved in Section 3. Its proof is based on Arnold's relations in the cohomology of the pure braid group.

Proposition 2.6 *The abelian group* $H_{2g-3}(\mathrm{PB}_{g-1} \times \mathbb{Z}^{g-1}, \mathbb{Z})$ *is generated by the elements* $\mathcal{A}_{\mathcal{T}}$*, where* $\mathcal{T} \in \boldsymbol{T}_g$*. All relations among this generators follows from the relations*

$$\mathcal{A}_{\mathcal{T}_1} + \mathcal{A}_{\mathcal{T}_2} + \mathcal{A}_{\mathcal{T}_3} = 0$$

for each cyclic triple $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$*.*

The next result will be proved in Sections 4 and 5. The proof is based on the spectral sequence for the action of $\mathcal{K}_g$ on the contractible complex of cycles, introduced by Bestvina, Bux and Margalit in [2], and certain new complexes which will be constructed below.

Proposition 2.7 *Let* $f_1 = 1, f_2, f_3, \dots \in \mathcal{G}_g$ *be representatives of all left cosets* $\mathcal{G}_g/\mathcal{H}_g$ *and let* $\hat{f}_1, \hat{f}_2, \hat{f}_3, \dots \in \mathrm{Mod}(\Sigma_g)$ *be their lifts in* $\mathrm{Mod}(\Sigma_g)$*. Then the inclusions*

$$i_s \colon \mathrm{Stab}_{\mathcal{K}_g}(\hat{f}_s \cdot N) \hookrightarrow \mathcal{K}_g \quad \text{for } s \in \mathbb{N}$$

induce an injective homomorphism

$$\bigoplus_{s \in \mathbb{N}} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{f}_s \cdot N), \mathbb{Z}) \hookrightarrow H_{2g-3}(\mathcal{K}_g, \mathbb{Z}). \tag{11}$$

Proof of Theorem 1.5 By Proposition 2.7, the map

$$i_1 \colon H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(N), \mathbb{Z}) \hookrightarrow H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$$

is injective. Since $\mathcal{P}_g = i_1(H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(N), \mathbb{Z}))$, Proposition 2.6 implies the required result. □

Proof of Theorem 1.3 By Propositions 2.7 and 2.5, (11) induces an isomorphism

$$\bigoplus_{s \in \mathbb{N}} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{f}_s \cdot N), \mathbb{Z}) \cong \mathcal{A}_g,$$

so

$$\mathcal{A}_g = \bigoplus_{s\in\mathbb{N}} \hat{f}_s \cdot \mathcal{P}_g.$$

Therefore, by the definition of an induced module, we have

$$\mathcal{A}_g \cong \mathcal{P}_g \otimes_{\mathcal{H}_g} \mathbb{Z}[\mathcal{G}_g] = \operatorname{Ind}_{\mathcal{H}_g}^{\mathcal{G}_g} \mathcal{P}_g. \qquad \square$$

Theorem 1.9 will be deduced from Proposition 2.6 in Section 3.

3 Proof of Proposition 2.6

3.1 Cohomology of the pure braid group

In order to prove Proposition 2.6 we conveniently consider the pure braid group on $g-1$ strands $q_1,\dots,q_{g-1}$. Let us recall Arnold's results on the structure of the ring $H^*(\mathrm{PB}_{g-1},\mathbb{Z})$. The group PB_{g-1} has a standard set of generators $a_{i,j}$ for $1\le i<j\le g-1$. These elements are the Dehn twists about curves enclosing the i^{th} and j^{th} strands (see Figure 6). Denote by $h_{i,j}\in H_1(\mathrm{PB}_{g-1},\mathbb{Z})$ the corresponding homology classes. We denote by $\{w_{i,j}\}$ the dual basis of $H^1(\mathrm{PB}_{g-1},\mathbb{Z})$. These cohomology classes can be interpreted as the homomorphisms

$$w_{i,j}\colon \mathrm{PB}_{g-1}\to \mathrm{PB}_2\cong\mathbb{Z} \tag{12}$$

given by forgetting all strands besides q_i and q_j. It is convenient to put $w_{j,i}=w_{i,j}$.

Theorem 3.1 [1, Theorem 1] *The ring $H^*(\mathrm{PB}_{g-1},\mathbb{Z})$ is the exterior graded algebra with $\binom{g-1}{2}$ generators $w_{i,j}$ of degree 1, satisfying $\binom{g-1}{3}$ relations*

$$w_{k,l}w_{l,m}+w_{l,m}w_{m,k}+w_{m,k}w_{k,l}=0$$

for all $1\le k<l<m\le n$.

Corollary 3.2 [1, Corollary 3] *The products*

$$w_{k_1,l_1}w_{k_2,l_2}\cdots w_{k_p,l_p} \quad \text{where } k_i<l_i \text{ and } l_1<\cdots<l_p \tag{13}$$

form an additive basis of $H^*(\mathrm{PB}_{g-1},\mathbb{Z})$.

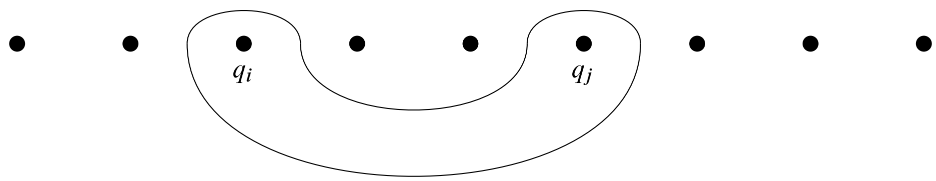


Figure 6: The element $a_{i,j}$ is the Dehn twist about the shown curve.

Corollary 3.2 implies that the products

$$w_{k_1,2}w_{k_2,3}\cdots w_{k_{g-2},g-1} \quad \text{where } k_i \le i \tag{14}$$

form an additive basis of $H^{g-2}(\mathrm{PB}_{g-1},\mathbb{Z})$. We denote the cohomology class (14) by $W_k = W_{k_1,\dots,k_{g-2}}$, where $k=(k_1,\dots,k_{g-2})$. We denote by $\boldsymbol{K}_g$ the set of all sequences $k=(k_1,\dots,k_{g-2})$ satisfying $1\le k_i\le i$.

3.2 Abelian cycles in $H_{g-2}(\mathrm{PB}_{g-1},\mathbb{Z})$

Corollary 3.2 implies that $\mathrm{cd}(\mathrm{PB}_n,\mathbb{Z})\ge n-1$. In fact $\mathrm{cd}(\mathrm{PB}_n)=n-1$. Indeed, let M_n be the ordered configuration space of n points on the disk. This space is aspherical, and $M_n\simeq K(\mathrm{PB}_n,1)$. We have the fiber bundle $M_n\to M_{n-1}$, where the fiber is homotopy equivalent to the wedge of $n-1$ circles. Hence, by induction, we obtain that M_n is homotopy equivalent to an $(n-1)$–dimensional CW–complex. Therefore $\mathrm{cd}(\mathrm{PB}_n,\mathbb{Z})\le n-1$.

So the isomorphism (10) implies

$$H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(N),\mathbb{Z})\cong H_{2g-3}(\mathbb{Z}^{g-1}\times\mathrm{PB}_{g-1},\mathbb{Z})\cong H_{g-2}(\mathrm{PB}_{g-1},\mathbb{Z}). \tag{15}$$

Let us recall the construction of the isomorphism $\mathrm{Stab}_{\mathcal{K}_g}(N)\cong\mathbb{Z}^{g-1}\times\mathrm{PB}_{g-1}$. We consider the surface $\Sigma^1_{0,g-1}$, given by replacing the boundary components corresponding to the curves $\delta_1,\dots,\delta_{g-1}$ on $\Sigma_{0,g}\subset\Sigma_g$ by the punctures $q_1,\dots,q_{g-1}$. Hence we obtain the pure braid group $\mathrm{PB}_{g-1}=\mathrm{PMod}(\Sigma^1_{0,g-1})$. The i^{th} factor in $\mathbb{Z}^{g-1}$ is generated by T_{δ_i}.

Consider a simple abelian cycle

$$\mathcal{A}_{\mathcal{T}}=\mathcal{A}(T_{\delta_1},\dots,T_{\delta_g},T_{\xi_2},\dots,T_{\xi_{g-2}})\in H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(N),\mathbb{Z})$$

for some $\mathcal{T}\in\boldsymbol{T}_g$. Isomorphism (15) sends $\mathcal{A}_{\mathcal{T}}$ to the abelian cycle

$$\mathcal{A}(T_{\delta_g},T_{\xi_2},\dots,T_{\xi_{g-2}})\in H_{g-2}(\mathrm{PB}_{g-1},\mathbb{Z}),$$

Let us set $\xi_1=\delta_g$ and

$$\widehat{\mathcal{A}}_{\mathcal{T}}=\mathcal{A}(T_{\xi_1},T_{\xi_2},\dots,T_{\xi_{g-2}})=\mathcal{A}(T_{\delta_g},T_{\xi_2},\dots,T_{\xi_{g-2}})\in H_{g-2}(\mathrm{PB}_{g-1},\mathbb{Z}).$$

Any simple closed curve on $\Sigma^1_{0,g-1}$ divides it into two parts. We say that a puncture q is *enclosed* by a curve γ on $\Sigma^1_{0,g-1}$ if q is contained in the part which does not contain the boundary component. For $k\in\boldsymbol{K}_g$, define the matrix $X_{k,\mathcal{T}}\in\mathrm{Mat}_{(g-2)\times(g-2)}(\mathbb{Z})$ by

$$(X_{k,\mathcal{T}})_{i,j}=\begin{cases}1 & \text{if the punctures } q_{k_i} \text{ and } q_{i+1} \text{ are enclosed by } \xi_j,\\ 0 & \text{otherwise.}\end{cases} \tag{16}$$

Lemma 3.3 *Let $k\in\boldsymbol{K}_g$ and $\mathcal{T}\in\boldsymbol{T}_g$. Then $\langle W_k,\widehat{\mathcal{A}}_{\mathcal{T}}\rangle=(-1)^{\binom{g-2}{2}}\det(X_{k,\mathcal{T}})$.*

Proof Consider a free abelian group $\mathbb{Z}^{g-2} = \langle c_1, \dots, c_{g-2} \rangle$ and the homomorphism $f : \mathbb{Z}^{g-2} \to \mathrm{PB}_{g-1}$ given by $c_i \mapsto T_{\xi_i}$. Denote by μ_{g-2} the standard generator of the group $H_{g-2}(\mathbb{Z}^{g-2}, \mathbb{Z})$. We have

$$\begin{aligned}\langle W_k, \widehat{\mathcal{A}}_{\mathcal{T}} \rangle &= \langle W_k, f_*(\mu_{g-2}) \rangle = \langle f^* W_k, \mu_{g-2} \rangle = \langle (f^* w_{k_1,2}) \cdots (f^* w_{k_{g-2},g-1}), \mu_{g-2} \rangle \\ &= (-1)^{\binom{g-2}{2}} \det(\langle f^* w_{k_i,i+1}, c_j \rangle)_{i,j=1}^{g-2} = (-1)^{\binom{g-2}{2}} \det(\langle w_{k_i,i+1}, f_* c_j \rangle)_{i,j=1}^{g-2} \\ &= (-1)^{\binom{g-2}{2}} \det(\langle w_{k_i,i+1}, T_{\xi_j} \rangle)_{i,j=1}^{g-2} = (-1)^{\binom{g-2}{2}} \det(X_{k,\mathcal{T}}).\end{aligned}$$

The last equality comes from the following corollary of (12):

$$\langle w_{k,l}, T_{\xi_j} \rangle = \begin{cases} 1 & \text{if the punctures } q_k \text{ and } q_l \text{ are enclosed by } \xi_j, \\ 0 & \text{otherwise.} \end{cases}$$

□

Let us denote by $\{D_k \in H_{g-2}(\mathrm{PB}_{g-1}, \mathbb{Z}) \mid k \in \boldsymbol{K}_g\}$ the dual basis to $\{W_k \mid k \in \boldsymbol{K}_g\}$.

Corollary 3.4 *Let $\mathcal{T} \in \boldsymbol{T}_g$. Then $\widehat{\mathcal{A}}_{\mathcal{T}} = \sum_{k \in \boldsymbol{K}_g} (-1)^{\binom{g-2}{2}} \det(X_{k,\mathcal{T}}) D_k$.*

3.3 Balanced trees

Recall that we consider the elements of $\boldsymbol{T}_g$ as marked trees such that the leaf with number g is the root. Also, we have already defined the subset $\boldsymbol{T}_g^b \subseteq \boldsymbol{T}_g$ of balanced trees. Take any $k \in \boldsymbol{K}_g$. Our goal is to construct a balanced tree $\mathcal{T}_k \in \boldsymbol{T}_g$ such that $\widehat{\mathcal{A}}_{\mathcal{T}_k} = (-1)^{\binom{g-2}{2}} D_k$ and the map $k \mapsto \mathcal{T}_k$ is a bijection between the sets $\boldsymbol{K}_g$ and $\boldsymbol{T}_g^b$. First let us construct the map $k \mapsto \mathcal{T}_k$ (then we will check that $\widehat{\mathcal{A}}_{\mathcal{T}_k} = (-1)^{\binom{g-2}{2}} D_k$; see Theorem 3.6).

Construction 3.5 We construct curves $\xi_1, \dots, \xi_{g-2}$ such that $\widehat{\mathcal{A}}_{\mathcal{T}_k} = \mathcal{A}(T_{\xi_1}, T_{\xi_2}, \dots, T_{\xi_{g-2}})$ by induction on g. The case $g = 3$ is trivial since $|\boldsymbol{T}_3^b| = |\boldsymbol{T}_3| = |\boldsymbol{K}_3| = 1$. Let us prove the induction step from $g-1$ to g. Consider any $k = (k_1, \dots, k_{g-2}) \in \boldsymbol{K}_g$ with $g > 3$. Let ξ_{g-2} be a curve enclosing exactly two points $q_{k_{g-2}}$ and q_{g-1}. Let us remove the curve ξ_{g-2} with its interior and denote the corresponding puncture by $q'_{k_{g-2}}$. Also take $q'_i = q_i$ for $i \le g-2$ and $i \ne k_{g-2}$. We obtain a disk with $g-2$ punctures $q'_1, \dots, q'_{g-2}$ and $k' = (k_1, \dots, k_{g-3}) \in \boldsymbol{K}_{g-1}$. The induction hypothesis implies that there is a balanced tree $\mathcal{T}_{k'} \in \boldsymbol{T}_{g-1}^b$ corresponding to k' given by some curves $\xi_1, \dots, \xi_{g-3}$. Now consider the curves $\xi_1, \dots, \xi_{g-3}, \xi_{g-2}$ and denote the dual tree by $\mathcal{T}_k \in \boldsymbol{T}_g$. It remains to show that $\mathcal{T}_k$ is balanced. Indeed, since the vertex q_{g-1} has the greatest number, all common vertices of $\mathcal{T}_{k'}$ and $\mathcal{T}_k$ are balanced. Also, this property holds for the vertex of $\mathcal{T}_k$ corresponding to the curve ξ_{g-2}, because it has only two descendant leaves. This implies the induction step.

Since for different $k, k' \in \boldsymbol{K}_g$ the corresponding trees $\mathcal{T}_k$ and $T_{k'}$ are also different, it follows that the map $k \mapsto \mathcal{T}_k$ given by Construction 3.5 is injective. Moreover, direct computation shows that $|\boldsymbol{K}_g| = (g-2)!$. Therefore in order to prove that this map is a surjection to $\boldsymbol{T}_g^b$, it suffices to show that $|\boldsymbol{T}_g^b| = (g-2)!$. We use induction on g; the base case $g = 3$ is trivial. Consider a balanced tree $\mathcal{T} \in \boldsymbol{T}_g^b$ with $g \ge 4$. Let

$q_1, \dots, q_{g-1}$ be its leaves (besides the root). Let p be the vertex adjacent to q_{g-1}. Since $\mathcal{T}$ is balanced, another descendant vertex of p is a leaf q_i for some $1 \le i \le g-2$. Let us remove the vertices q_i and q_{g-1} (with the incident edges) and set $q_i = p$; denote the obtained tree by $\mathcal{T}'$. Then $\mathcal{T}' \in \boldsymbol{T}^b_{g-1}$. Since $|\boldsymbol{T}^b_{g-1}| = (g-3)!$ and there are $g-2$ ways to choose i, we have $|\boldsymbol{T}^b_g| = (g-2)!$. This implies the induction step.

Theorem 3.6 *Suppose that $k \in \boldsymbol{K}_g$. Then $\hat{\mathcal{A}}_{\mathcal{T}_k} = (-1)^{\binom{g-2}{2}} D_k$.*

Proof By Corollary 3.4 it suffices to show that for any $k' \in \boldsymbol{K}_g$, we have $\det(X_{k',\mathcal{T}_k}) = 1$ if $k = k'$ and $\det(X_{k',\mathcal{T}_k}) = 0$ otherwise.

Lemma 3.7 *Let $k \in \boldsymbol{K}_g$. Then $\det(X_{k,\mathcal{T}_k}) = 1$.*

Proof Let $\xi_1, \dots, \xi_{g-2}$ be a multicurve with dual tree $\mathcal{T}_k$. By Construction 3.5 the punctures q_{k_i} and q_{i+1} are enclosed by the curve ξ_i for all $1 \le i \le g-2$. Indeed, for $i = g-2$ this follows from the construction of the curve ξ_{g-2}, and for $i < g-2$ this follows by the induction on g. Therefore $(X_{k,\mathcal{T}_k})_{i,i} = 1$ for all i.

Now let us check that $(X_{k,\mathcal{T}_k})_{i,j} = 0$ whenever $i < j$. Indeed, for $j = g-2$ this follows from the construction of the curve ξ_{g-2}, and for $j < g-2$ this follows by the induction on g. Therefore $X_{k,\mathcal{T}_k}$ is lower unitriangular, so $\det(X_{k,\mathcal{T}_k}) = 1$. □

Lemma 3.8 *Let $k, k' \in \boldsymbol{K}_g$ and $k \neq k'$. Then $\det(X_{k',\mathcal{T}_k}) = 0$.*

Proof Define $s = \max\{i \mid k_i \neq k'_i\}$. Let us check that the matrix $X_{k',\mathcal{T}_k}$ has the following form, where the s^{th} column is highlighted:

$$X_{k',\mathcal{T}_k} = \begin{pmatrix} * & * & \cdots & * & 0 & 0 & 0 & \cdots & 0 \\ * & * & \cdots & * & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * & 0 & 0 & 0 & \cdots & 0 \\ * & * & \cdots & * & 0 & 0 & 0 & \cdots & 0 \\ * & * & \cdots & * & * & 1 & 0 & \cdots & 0 \\ * & * & \cdots & * & * & * & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * & * & * & * & \cdots & 1 \end{pmatrix}. \tag{17}$$

This will immediately imply $\det(X_{k',\mathcal{T}_k}) = 0$. We have $k_i = k'_i$ for all $i > s$, so similar arguments as in the proof of Lemma 3.7 show that the last $g-2-s$ columns have the required form.

Let $\mathcal{T}_k$ be given by curves $\xi_1, \dots, \xi_{g-2}$ as in Construction 3.5. Recall the construction of the curve ξ_s. At this step we have the punctures $q'_1, \dots, q'_{s+1}$. Some of them coincide with the q_i, others are identified with the interiors of some curves ξ_i, with $i \ge s+1$. Nevertheless, q_i is enclosed by ξ_s if and only if q'_i is enclosed by ξ_s for all $1 \le i \le s+1$.

By Construction 3.5, ξ_s is a curve enclosing exactly two punctures q'_{s+1} and q'_{k_s} from the set $\{q'_1, \dots, q'_{s+1}\}$. Therefore it does not enclose $q_{k'_s}$ as well as $q'_{k'_s}$, which implies $(X_{k',\mathcal{T}_k})_{s,s} = 0$. Take any j with $1 \le j < s$. The curve ξ_s encloses precisely one puncture among $q'_1, \dots, q'_s$, and so it also encloses precisely one puncture among $q_1, \dots, q_s$. Consequently, the curve ξ_s cannot enclose the punctures q_{j+1} and $q_{k'_j}$ simultaneously since $j+1 \le s$ and $k'_j \le s$. Hence, by formula (16), we have $(X_{k',\mathcal{T}_k})_{j,s} = 0$.

Therefore $(X_{k',\mathcal{T}_k})_{j,s} = 0$ for $1 \le j \le s$. □

Theorem 3.6 immediately follows form Lemmas 3.7 and 3.8. □

Corollary 3.9 *The abelian cycles $\{\hat{\mathcal{A}}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}^b_g\}$ form a basis of the group $H_{g-2}(\mathrm{PB}_{g-1}, \mathbb{Z})$. For any $\mathcal{T} \in \boldsymbol{T}_g$ we have $\hat{\mathcal{A}}_{\mathcal{T}} = \sum_{k \in \boldsymbol{K}_g} (-1)^{\binom{g-2}{2}} \det(X_{k,\mathcal{T}}) \hat{\mathcal{A}}_{\mathcal{T}_k}$.*

Proof The result follows from Corollary 3.4 and Theorem 3.6. □

Proof of Theorem 1.9 By Proposition 2.7 and the first part of Corollary 3.9 there is an isomorphism

$$\mathcal{P}_g \cong H_{2g-3}(\mathbb{Z}^{g-1} \times \mathrm{PB}_{g-1}, \mathbb{Z}) \cong H_{g-2}(\mathrm{PB}_{g-1}, \mathbb{Z}),$$

which maps $\mathcal{A}_{\mathcal{T}}$ to $\hat{\mathcal{A}}_{\mathcal{T}}$ for all $\mathcal{T} \in \boldsymbol{T}_g$. The theorem follows from the second part of Corollary 3.9. □

3.4 Relations

Let $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\} \subseteq \boldsymbol{T}_g$ be a triple of trees. For $l = 1, 2, 3$ denote by $\xi^l_1, \dots, \xi^l_{g-2}$ the corresponding sets of curves given by Construction 3.5. As before, the leaves of $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ (besides the root) are identified with the corresponding punctures and marked by $q_1, \dots, q_{g-1}$. One can check that the trees $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ form a cyclic triple if and only if, after some permutations of the corresponding sets of curves, the following conditions hold:

(a) There exists s with $1 \le s \le g-2$ such that $\xi^1_i = \xi^2_i = \xi^3_i$ for $i \ne s$ and $1 \le i \le g-2$.

(b) There exists $t \ne s$ with $1 \le t \le g-2$ and pairwise disjoint nonempty subsets $B_1, B_2, B_3 \subset \{q_1, \dots, q_{g-1}\}$ such that the set of punctures enclosed by the curve $\xi^1_t = \xi^2_t = \xi^3_t$ coincides with $B_1 \cup B_2 \cup B_3$.

(c) The set of punctures enclosed by ξ^1_s coincides with $B_2 \cup B_3$.

(d) The set of punctures enclosed by ξ^2_s coincides with $B_3 \cup B_1$.

(e) The set of punctures enclosed by ξ^3_s coincides with $B_1 \cup B_2$.

Lemma 3.10 *Let $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\} \subseteq \boldsymbol{T}_g$ be a cyclic triple of trees. Then*

$$\hat{\mathcal{A}}_{\mathcal{T}_1} + \hat{\mathcal{A}}_{\mathcal{T}_2} + \hat{\mathcal{A}}_{\mathcal{T}_3} = 0. \tag{18}$$

Proof It suffices to prove that

$$\langle W_k, \hat{\mathcal{A}}_{\mathcal{T}_1} + \hat{\mathcal{A}}_{\mathcal{T}_2} + \hat{\mathcal{A}}_{\mathcal{T}_3} \rangle = 0 \tag{19}$$

for all $k \in \boldsymbol{K}_g$. The formula (19) is equivalent to

$$\det(X_{k,\mathcal{T}_1}) + \det(X_{k,\mathcal{T}_2}) + \det(X_{k,\mathcal{T}_3}) = 0. \tag{20}$$

We can assume that (a)–(e) hold. The matrices $X_{k,\mathcal{T}_1}$, $X_{k,\mathcal{T}_2}$ and $X_{k,\mathcal{T}_3}$ coincide everywhere besides the s^{th} column. Therefore the left-hand side of (20) equals the determinant of the matrix Y defined as follows. The s^{th} column of Y is the s^{th} column of the matrix $X_{k,\mathcal{T}_1} + X_{k,\mathcal{T}_2} + X_{k,\mathcal{T}_3}$ and all other columns are the corresponding columns of $X_{k,\mathcal{T}_1}$ (or, equivalently, $X_{k,\mathcal{T}_2}$ or $X_{k,\mathcal{T}_3}$). By (c)–(e) we have

$$(X_{k,\mathcal{T}_1})_{i,s} = \begin{cases} 1 & \text{if } i \in B_2 \cup B_3, \\ 0 & \text{otherwise,} \end{cases} \quad (X_{k,\mathcal{T}_2})_{i,s} = \begin{cases} 1 & \text{if } i \in B_3 \cup B_1, \\ 0 & \text{otherwise,} \end{cases} \quad (X_{k,\mathcal{T}_3})_{i,s} = \begin{cases} 1 & \text{if } i \in B_1 \cup B_2, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$Y_{i,s} = (X_{k,\mathcal{T}_1} + X_{k,\mathcal{T}_2} + X_{k,\mathcal{T}_3})_{i,s} = \begin{cases} 2 & \text{if } i \in B_1 \cup B_2 \cup B_3, \\ 0 & \text{otherwise.} \end{cases}$$

By (b) we have

$$Y_{i,t} = (X_{k,\mathcal{T}_1})_{i,t} = \begin{cases} 1 & \text{if } i \in B_1 \cup B_2 \cup B_3, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore the matrix Y has two proportional columns, so $\det(Y) = 0$. This implies (20). □

Lemma 3.11 *All relations between the abelian cycles $\{\hat{\mathcal{A}}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}_g\}$ follow from (18).*

Proof Consider the abelian cycle $\hat{\mathcal{A}}_{\mathcal{T}}$ for some $\mathcal{T} \in \boldsymbol{T}_g$. Corollary 3.9 implies that it suffices to decompose $\hat{\mathcal{A}}_{\mathcal{T}}$ into a linear combination of abelian cycles $\{\hat{\mathcal{A}}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}_g^b\}$ using (18).

Recall that a vertex of $\mathcal{T}$ of degree 3 is called balanced if the paths from it to the descendant leaves with the two smallest numbers have no common edges. If all vertices of $\mathcal{T}$ are balanced there is nothing to prove. Otherwise take any nonbalanced vertex v with the largest height (distance to the root) $h(v)$. Let v_1 and v_2 be its closest descendants and let w be its closest ancestor. Without loss of generality we may assume that the paths from v to the two descendant leaves with the smallest numbers start with the edge (v, v_1). Let u_1 and u_2 be the closest descendants of v_1.

Consider the trees $\mathcal{T}'$ and $\mathcal{T}''$ that differ from $\mathcal{T}$ as shown in Figure 7. The triple $\{\mathcal{T}, \mathcal{T}', \mathcal{T}''\}$ is cyclic, so $\hat{\mathcal{A}}_{\mathcal{T}} = -\hat{\mathcal{A}}_{\mathcal{T}'} - \hat{\mathcal{A}}_{\mathcal{T}''}$. Note that the vertex v_1 is balanced in $\mathcal{T}$, and therefore the vertex v is balanced in $\mathcal{T}'$

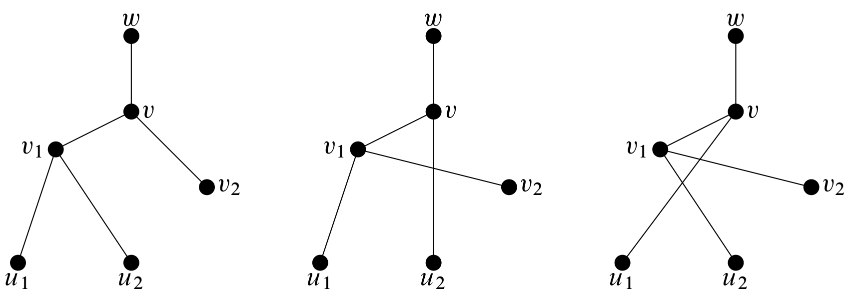


Figure 7: The trees $\mathcal{T}$, $\mathcal{T}'$ and $\mathcal{T}''$.

and $\mathcal{T}''$. Consequently, $\mathcal{T}'$ and $\mathcal{T}''$ have fewer nonbalanced vertices of height $h(v)$ and no nonbalanced vertices of greater height. Repeating this operation, we decompose $\widehat{\mathcal{A}}_\mathcal{T}$ into a linear combination of abelian cycles $\{\widehat{\mathcal{A}}_\mathcal{T} \mid \mathcal{T} \in \boldsymbol{T}_g^b\}$ using (18). □

Proof of Proposition 2.6 By Proposition 2.7 and the first part of Corollary 3.9 there is an isomorphism

$$\mathcal{P}_g \cong H_{2g-3}(\mathbb{Z}^{g-1} \times \mathrm{PB}_{g-1}, \mathbb{Z}) \cong H_{g-2}(\mathrm{PB}_{g-1}, \mathbb{Z}),$$

which maps $\mathcal{A}_\mathcal{T}$ to $\widehat{\mathcal{A}}_\mathcal{T}$ for all $\mathcal{T} \in \boldsymbol{T}_g$. The abelian cycles $\{\widehat{\mathcal{A}}_\mathcal{T} \mid \mathcal{T} \in \boldsymbol{T}_g\}$ generate $H_{g-2}(\mathrm{PB}_{g-1}, \mathbb{Z})$. Therefore the required assertion follows from Lemmas 3.10 and 3.11. □

4 The complex of cycles and the spectral sequence

Consider the commutative diagram

$$\begin{array}{ccccccccc}
1 & \longrightarrow & \bigwedge^3 H/H & \longrightarrow & \mathcal{G}_g & \xrightarrow{\;p\;} & \mathrm{Sp}(2g,\mathbb{Z}) & \longrightarrow & 1 \\
 & & & & \uparrow & \nearrow & & & \\
 & & & & \mathcal{H}_g = \mathrm{SL}(2,\mathbb{Z})^{\times g} \rtimes S_g & & & &
\end{array}$$

Let us choose elements $h_1 = 1, h_2, h_3, \ldots \in \mathcal{G}_g$ such that $1 = p(h_1), p(h_2), p(h_3), \ldots \in \mathrm{Sp}(2g, \mathbb{Z})$ are representatives of all left cosets $\mathrm{Sp}(2g,\mathbb{Z})/\mathcal{H}_g$. Let $\hat{h}_1, \hat{h}_2, \hat{h}_3, \ldots \in \mathrm{Mod}(\Sigma_g)$ be mapping classes that go to $h_1, h_2, h_3, \ldots$ under the natural surjective homomorphism $\mathrm{Mod}(\Sigma_g) \twoheadrightarrow \mathcal{G}_g$.

It is convenient to denote by U_g the abelian group $\bigwedge^3 H/H$ with multiplicative notation. For each $u \in U_g$ let $\hat{u} \in \mathcal{I}_g$ be the mapping class that goes to u under the Johnson homomorphism $\tau\colon \mathcal{I}_g \to U_g$. Let $f_1 = 1, f_2, f_3, \ldots \in \mathcal{G}_g$ be representatives of all left cosets $\mathcal{G}_g/\mathcal{H}_g$. Let $\hat{f}_1, \hat{f}_2, \hat{f}_3, \ldots \in \mathrm{Mod}(\Sigma_g)$ be mapping classes that go to $f_1, f_2, f_3, \ldots$ under the homomorphism $\mathrm{Mod}(\Sigma_g) \twoheadrightarrow \mathcal{G}_g$. For any $s \in \mathbb{N}$ the element f_s can be uniquely decomposed as $f_s = u \cdot h_r$ for some $u \in U_g$ and $r \in \mathbb{N}$. We can choose $\hat{f}_s$ such that $\hat{f}_s = \hat{u} \cdot \hat{h}_r$.

For each $r \in \mathbb{N}$ denote by G_r the subgroup of $H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$ generated by the images of homomorphisms

$$H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot \hat{h}_r \cdot N), \mathbb{Z}) \to H_{2g-3}(\mathcal{K}_g, \mathbb{Z}) \quad \text{for } u \in U_g. \tag{21}$$

In this section we prove the following result:

Lemma 4.1 *The inclusions*

$$G_r \hookrightarrow H_{2g-3}(\mathcal{K}_g, \mathbb{Z}) \quad \textit{for } r \in \mathbb{N}$$

induce an injective homomorphism

$$\bigoplus_{r \in \mathbb{N}} G_r \hookrightarrow H_{2g-3}(\mathcal{K}_g, \mathbb{Z}).$$

In our proof we follow ideas of [11].

4.1 The complex of cycles

Bestvina, Bux and Margalit [2] constructed a contractible CW–complex $\mathcal{B}_g$ called the *complex of cycles* on which the Johnson kernel acts without rotations. "Without rotations" means that if an element $h \in \mathcal{K}_g$ stabilizes a cell σ setwise, then h stabilizes σ pointwise. Let us recall the construction of $\mathcal{B}_g$. More details can be found in [2; 13; 12; 10].

We denote by $\mathcal{C}$ the set of all isotopy classes of oriented nonseparating simple closed curves on Σ_g. Fix any nonzero element $x \in H$. The construction of $\mathcal{B}_g = \mathcal{B}_g(x)$ depends on the choice of the homology class x, however the CW–complexes $\mathcal{B}_g(x)$ are pairwise homeomorphic for different x.

A *basic* 1*–cycle* for a homology class x is a formal linear combination $\gamma = \sum_{i=1}^{n} k_i \gamma_i$, where $\gamma_i \in \mathcal{C}$ and $k_i \in \mathbb{N}$, such that

(I) the homology classes $[\gamma_1], \dots, [\gamma_n]$ are linearly independent,

(II) $\sum_{i=1}^{n} k_i [\gamma_i] = x$, and

(III) the isotopy classes $\gamma_1, \dots, \gamma_g$ contain pairwise disjoint representatives.

The oriented multicurve $\gamma_1 \cup \dots \cup \gamma_g$ is called the *support* of γ.

Denote by $\mathcal{M}(x)$ the set of oriented multicurves $M = \gamma_1 \cup \dots \cup \gamma_s$ such that

(i) no nontrivial linear combination of the homology classes $[\gamma_1], \dots, [\gamma_s]$ with nonnegative coefficients equals zero, and

(ii) for each $1 \le i \le s$ there exists a basic 1–cycle for x whose support is contained in M and contains γ_i.

For each $M \in \mathcal{M}(x)$ let us denote by $P_M \subset \mathbb{R}^{\mathcal{C}}_{\ge 0}$ the convex hull of the basic 1–cycles supported in M. We have that P_M is a convex polytope. By definition, the complex of cycles is the regular CW–complex given by $\mathcal{B}_g(x) = \bigcup_{M \in \mathcal{M}(x)} P_M$. Denote by $\mathcal{M}_0(x) \subseteq \mathcal{M}(x)$ the set of supports of basic 1–cycles for x. Then $\{P_M \mid M \in \mathcal{M}_0(x)\}$ is the set of 0–cells of $\mathcal{B}_g(x)$.

Theorem 4.2 [2, Theorem E] *Let $g \ge 1$ and $0 \ne x \in H_1(\Sigma_g, \mathbb{Z})$. Then $\mathcal{B}_g(x)$ is contractible.*

4.2 The spectral sequence

Suppose that a group G acts cellularly and without rotations on a contractible CW–complex X, let $C_*(X, \mathbb{Z})$ be the cellular chain complex of X and let $\mathcal{R}_*$ be a projective resolution of $\mathbb{Z}$ over $\mathbb{Z}G$. Consider the double complex $B_{p,q} = C_p(X, \mathbb{Z}) \otimes_G \mathcal{R}_q$ with the filtration by columns. The corresponding spectral sequence (see (7.7) in [4, Section VII.7]) has the form

$$E^1_{p,q} \cong \bigoplus_{\sigma \in \mathcal{X}_p} H_q(\mathrm{Stab}_G(\sigma), \mathbb{Z}) \Rightarrow H_{p+q}(G, \mathbb{Z}), \tag{22}$$

where $\mathcal{X}_p$ is a set containing exactly one representative in each G–orbit of p–cells of X. Let us remark that for an arbitrary CW–complex X, the spectral sequence (22) converges to the equivariant homology $H^G_{p+q}(X, \mathbb{Z})$. So for a contractible CW–complex X it converges to $H^G_{p+q}(X, \mathbb{Z}) \cong H_{p+q}(G, \mathbb{Z})$.

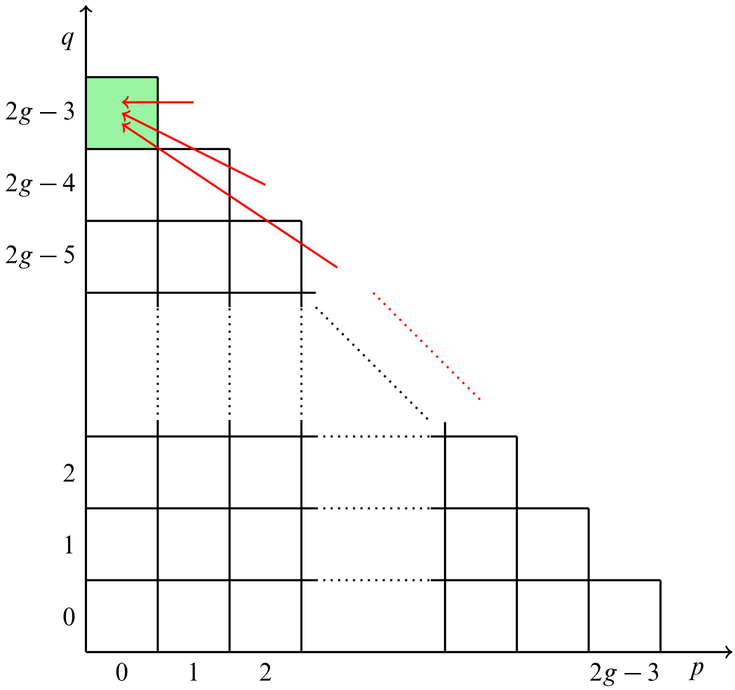


Figure 8

Now let $E^*_{*,*}$ be the spectral sequence (22) for the action of $\mathcal{K}_g$ on $\mathcal{B}_g(x)$ for some $0 \neq x \in H_1(\Sigma_g, \mathbb{Z})$. The fact that $\mathcal{K}_g$ acts on $\mathcal{B}_g$ without rotations follows from a result of Ivanov [14, Theorem 1.2]: if an element $h \in \mathcal{I}_g$ stabilizes a multicurve M then h stabilizes each component of M. Bestvina, Bux and Margalit proved [2, Proposition 6.2] that for each cell $\sigma \in \mathcal{B}_g(x)$ we have

$$\dim(\sigma) + \operatorname{cd}(\operatorname{Stab}_{\mathcal{K}_g}(\sigma)) \leq 2g-3.$$

This immediately implies $E^1_{p,q} = 0$ for $p+q > 2g-3$. Hence all differentials $d^1, d^2, \dots$ to the group $E^1_{0,2g-3}$ are trivial (see Figure 8, where the group $E^1_{0,2g-3}$ is shaded), so $E^1_{0,2g-3} = E^\infty_{0,2g-3}$. Therefore we have the following result:

Proposition 4.3 [11, Proposition 3.2] *Let $\mathfrak{M} \subseteq \mathcal{M}_0(x)$ be a subset consisting of oriented multicurves from pairwise different $\mathcal{K}_g$–orbits. Then the inclusions $\operatorname{Stab}_{\mathcal{K}_g}(M) \subseteq \mathcal{K}_g$, where $M \in \mathfrak{M}$, induce an injective homomorphism*

$$\bigoplus_{M \in \mathfrak{M}} H_{2g-3}(\operatorname{Stab}_{\mathcal{K}_g}(M), \mathbb{Z}) \hookrightarrow H_{2g-3}(\mathcal{K}_g), \mathbb{Z}).$$

Proof of Lemma 4.1 Denote by $X_{r,i} \subset \Sigma_g$ the one-punctured torus bounded by $\hat{h}_r \delta_i$ and $V_{r,i} = H_1(X_{r,i}, \mathbb{Z}) \subset H$. Then for each r we have the symplectic splittings $H = \bigoplus_i V_{r,i}$. Denote this unordered splitting by $\mathcal{V}_r = \{V_{r,1}, \dots, V_{r,g}\}$. Since $\mathcal{H}_g$ is the stabilizer of $\mathcal{V}_1$ in $\operatorname{Sp}(2g, \mathbb{Z})$, it follows that the $\mathcal{V}_r$ are pairwise distinct.

Assume the converse to the statement of Lemma 4.1 and consider a nontrivial linear relation

$$\sum_{r=1}^{k} \lambda_r \theta_r = 0 \quad \text{for } \lambda_r \in \mathbb{Z} \text{ and } \theta_r \in G_r. \tag{23}$$

For any homology class $x \in H_1(\Sigma_g, \mathbb{Z})$ and for any $1 \leq r \leq k$ we have a unique decomposition

$$x = \sum_{i=1}^{g} x_{r,i} \quad \text{for } x_{r,i} \in V_{r,i}.$$

The following result is proved in [11]:

Proposition 4.4 [11, Lemma 4.5] *There is a homology class $x \in H$ such that*

(I) *all homology classes $x_{r,i}$ are nonzero for $1 \leq r \leq k$ and $1 \leq i \leq g$, and*

(II) *for all $1 \leq p \neq q \leq k$ we have $\{x_{p,1}, \dots, x_{p,g}\} \neq \{x_{q,1}, \dots, x_{q,g}\}$ as unordered sets.*

Take any $x \in H$ satisfying the conditions of Proposition 4.4. For any $1 \leq r \leq k$ and $1 \leq i \leq g$ we have $x_{r,i} = n_{r,i} a_{r,i}$ where $a_{r,i} \in H$ is primitive and $n_{r,i} \in \mathbb{N}$. Let us check that for all $1 \leq p \neq q \leq k$ we have $\{a_{p,1}, \dots, a_{p,g}\} \neq \{a_{q,1}, \dots, a_{q,g}\}$ as unordered sets. Indeed, assume that there is a permutation $\pi \in S_g$ with $a_{p,i} = a_{q,\pi(i)}$. Therefore we have

$$\sum_{i=1}^{g} (n_{p,i} - n_{p,\pi(i)}) a_{p,i} = 0. \tag{24}$$

Since $a_{p,1}, \dots, a_{p_g}$ are linearly independent, (24) implies $n_{p,i} = n_{p,\pi(i)}$ for all $1 \leq i \leq g$. Hence $x_{p,i} = x_{p,\pi(i)}$ for all $1 \leq i \leq g$, which contradicts Proposition 4.4(II).

For any $1 \leq r \leq k$ and $1 \leq i \leq g$ let $\alpha_{r,i}$ be a simple curve on $X_{r,i}$ with $[\alpha_{r,i}] = a_{r,i} \in H$. Consider the oriented multicurve $A_r = \alpha_{r,1} \cup \dots \cup \alpha_{r,g}$. By construction $A_r \in \mathcal{M}_0(x)$.

Proposition 2.6 implies that the group $H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot \hat{h}_r \cdot N), \mathbb{Z})$ is generated by the primitive abelian cycles $\{\hat{u} \cdot \hat{h}_r \cdot \mathcal{A}_{\mathcal{T}} \mid \mathcal{T} \in \boldsymbol{T}_g\}$. Therefore for each $u \in U_g$ the homomorphisms (21) can be decomposed as

$$H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot \hat{h}_r \cdot N), \mathbb{Z}) \to H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(A_r), \mathbb{Z}) \to H_{2g-3}(\mathcal{K}_g, \mathbb{Z}). \tag{25}$$

Consequently there exists $\theta'_r \in H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(A_r), \mathbb{Z})$ which maps to θ_r under the second homomorphism in (25).

Proposition 4.3 implies that the inclusions $\mathrm{Stab}_{\mathcal{K}_g}(A_r) \subseteq \mathcal{K}_g$ for $r \in \mathbb{N}$ induce the injective homomorphism

$$\bigoplus_{r \in \mathbb{N}} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(A_r), \mathbb{Z}) \hookrightarrow H_{2g-3}(\mathcal{K}_g, \mathbb{Z}).$$

So (23) implies that $\sum_{r=1}^{k} \lambda_r \theta'_r = 0$ as an element of the direct sum $\bigoplus_{r \in \mathbb{N}} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(A_r), \mathbb{Z})$. Therefore $\lambda_r = 0$ for all r, which gives a contradiction. □

5 Proof of Proposition 2.7

In this section we prove the following lemma, which implies Proposition 2.7. Recall that G_r is the subgroup of $H_{2g-3}(\mathcal{K}_g, \mathbb{Z})$ generated by the images of homomorphisms

$$H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot \hat{h}_r \cdot N), \mathbb{Z}) \to H_{2g-3}(\mathcal{K}_g, \mathbb{Z}) \quad \text{for } u \in U_g.$$

Lemma 5.1 *Let $r \in \mathbb{N}$. Then the inclusions*

$$\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot \hat{h}_r \cdot N) \hookrightarrow \mathcal{K}_g \quad \text{for } u \in U_g$$

induce an injective homomorphism

$$\bigoplus_{u \in U_g} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot \hat{h}_r \cdot N), \mathbb{Z}) \hookrightarrow G_r.$$

Proof of Proposition 2.7 We can prove Proposition 2.7 for an arbitrary choice of $\hat{f}_s$, so we can assume that $\hat{f}_s = \hat{u} \cdot \hat{h}_r$ for some $u \in U_g$ and $r \in \mathbb{N}$. Combining Lemmas 4.1 and 5.1, we obtain

$$\bigoplus_{r \in \mathbb{N}} \bigoplus_{u \in U_g} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot \hat{h}_r \cdot N), \mathbb{Z}) \hookrightarrow \bigoplus_{r \in \mathbb{N}} G_r \hookrightarrow H_{2g-3}(\mathcal{K}_g, \mathbb{Z}). \tag{26}$$

Then the sets $\{s \in \mathbb{N}\}$ and $\{\hat{u} \cdot \hat{h}_r \mid u \in U_g,\ r \in \mathbb{N}\}$ coincide, so (26) implies (11). □

To prove Lemma 5.1 we need to construct a new CW–complex, which will be called the *complex of relative cycles*. The idea is to introduce an analogue of $\mathcal{B}_g$ that makes sense for a sphere (ie the $g = 0$ case) with punctures.

5.1 The complex of relative cycles

Recall that $\Sigma_{0,2g}$ denotes a sphere with $2g$ punctures. In order to construct the complex of relative cycles $\mathcal{B}_{0,2g}$ we need to split the punctures into two disjoint sets: $P = \{p_1, \dots, p_g\}$ and $Q = \{q_1, \dots, q_g\}$.

By an *arc* on $\Sigma_{0,2g}$ we mean an embedded oriented curve with endpoints at punctures. By a *multiarc* we mean a disjoint union of arcs (common endpoints are allowed). We always consider arcs and multiarcs up to isotopy.

Denote by $\mathcal{D}$ the set of isotopy classes of arcs starting at a point in P and finishing at a point in Q. A *relative basic 1–cycle* is a formal sum $\gamma = \gamma_1 + \dots + \gamma_g$ where $\gamma_i \in \mathcal{D}$ such that

(I) $\partial\big(\sum_{i=1}^{g} \gamma_i\big) = \sum_{i=1}^{g} (q_i - p_i)$, and

(II) the isotopy classes $\gamma_1, \dots, \gamma_g$ contain pairwise disjoint representatives.

The multiarc $\gamma_1 \cup \dots \cup \gamma_g$ is called the *support* of γ.

Denote by $\mathcal{L}$ the set of multiarcs $L = \gamma_1 \cup \dots \cup \gamma_n$ (for arbitrary n) such that

(i) for each $1 \le i \le s$ there exists a relative basic 1–cycle, whose support is contained in L and contains γ_i.

For each $L \in \mathcal{L}$ we denote by $P_L \subset \mathbb{R}^{\mathcal{D}}_{\geq 0}$ the convex hull of all relative basic 1–cycles supported in L. We have that P_L is a convex polytope. By definition, the complex of relative cycles is the regular CW–complex given by $\mathcal{B}_{0,2g} = \bigcup_{L \in \mathcal{L}} P_L$. Denote by $\mathcal{L}_0 \subseteq \mathcal{L}$ the set of supports of all relative basic 1–cycles. Then $\{P_L \mid L \in \mathcal{L}_0\}$ is the set of 0–cells of $\mathcal{B}_{0,2g}$.

Remark 5.2 By construction, $\mathcal{B}_{0,2g}$ is the subset of $\mathbb{R}^{\mathcal{D}}_{\geq}$ consisting of the points (formal sums) $\sum_{i=1}^n k_i \gamma_i$ where $\gamma_i \in \mathcal{D}$ and $k_i \in \mathbb{R}_{\geq 0}$ satisfying the following conditions:

(I) $\partial(\sum_{i=1}^n k_i \gamma_i) = \sum_{i=1}^g (q_i - p_i)$.

(II) The isotopy classes $\gamma_1, \dots, \gamma_n$ contain pairwise disjoint representatives.

5.2 Contractability

Theorem 5.3 *Let $g \geq 1$. Then $\mathcal{B}_{0,2g}$ is contractible.*

In our proof we follow ideas of [2, Section 5]. Let us define an auxiliary complex $\widetilde{\mathcal{B}}_{0,2g}$. Denote by $\widetilde{\mathcal{D}}$ the union of $\mathcal{D}$ and the set consisting of the isotopy classes of all oriented simple closed curves on $\Sigma_{0,2g}$ (including contractible curves). Let us define $\widetilde{\mathcal{B}}_{0,2g}$ as the subset of $\mathbb{R}^{\widetilde{\mathcal{D}}}_{\geq 0}$ consisting of all points (formal sums) $\sum_{i=1}^n k_i \gamma_i$ where $\gamma_i \in \widetilde{\mathcal{D}}$ and $k_i \in \mathbb{R}_{\geq 0}$ satisfying the following conditions:

(I) $\partial(\sum_{i=1}^n k_i \gamma_i) = \sum_{i=1}^g (q_i - p_i)$.

(II) The isotopy classes $\gamma_1, \dots, \gamma_n$ contain pairwise disjoint representatives.

Remark 5.2 implies that $\mathcal{B}_{0,2g} \subseteq \widetilde{\mathcal{B}}_{0,2g}$. Denote by Drain: $\widetilde{\mathcal{B}}_{0,2g} \to \mathcal{B}_{0,2g}$ the retraction induced by the natural projection $\mathbb{R}^{\widetilde{\mathcal{D}}}_{\geq 0} \to \mathbb{R}^{\mathcal{D}}_{\geq 0}$.

Let d and d' be two points of $\mathcal{B}_{0,2g} \subseteq \mathbb{R}^{\mathcal{D}}_{\geq 0}$ and $t \in [0,1]$. The point $c = td + (1-t)d' \in \mathbb{R}^{\mathcal{D}}_{\geq 0}$ may not belong to $\mathcal{B}_{0,2g}$, because the arcs can have intersection points. We now explain how to do surgery to convert c into a point $\mathrm{Surger}(c) \in \widetilde{\mathcal{B}}_{0,2g} \subseteq \mathbb{R}^{\widetilde{\mathcal{D}}}_{\geq 0}$ which is canonical up to isotopy.

Let $c = \sum_{i=1}^n k_i c_i$ where the $c_i \in \mathcal{D}$ are in minimal position and $k_i \in \mathbb{R}_{\geq 0}$. We have $\partial\big(\sum_{i=1}^n k_i c_i\big) = \sum_{i=1}^g (q_i - p_i)$. Now it is convenient to replace the punctures $p_1, \dots, p_g, q_1, \dots, q_g$ by closed disks $P_1, \dots, P_g, Q_1, \dots, Q_g$. We thicken each c_i to a rectangle $R_i = [0,1] \times [0, k_i]$ of width k_i with coordinates $x_i \in [0,1]$ and $t_i \in [0, k_i]$ such that the curves $t_i = \mathrm{const}$ for different i are transverse to each other. We assume that the sides of R_i given by $x = 0$ and $x = 1$ are subsets of ∂P_a and ∂Q_b, respectively, where $\partial c_i = q_b - p_a$.

For a path $\alpha\colon [0,1] \to \Sigma_{0,2g}$, define $\mu_i(\alpha) = \int_\alpha dt_i$ and $\mu(\alpha) = \sum_{i=1}^n \mu_i(\alpha)$. Here we assume that $dt_i = 0$ outside R_i. Let us fix an arbitrary point $y_0 \in \Sigma_{0,2g}$. For each point $y \in \Sigma_{0,2g}$ choose a path α_y from y_0 to y and consider the map $\phi\colon \Sigma_{0,2g} \to S^1 = \mathbb{R}/\mathbb{Z}$ given by $\phi(y) = \mu(\alpha_y) \bmod 1$.

Let us check that the map ϕ is well defined. We have that $\phi(x)$ depends only on the homotopy class of α_x. Therefore it suffices to check that $\mu(\partial P_i) \in \mathbb{Z}$ and $\mu(\partial Q_i) \in \mathbb{Z}$ for all i. This follows from the fact that $\partial\big(\sum_{i=1}^n k_i c_i\big) = \sum_{i=1}^g (q_i - p_i)$.

The set of zeros of $d\phi$ is precisely $\Sigma_{0,2g} \setminus \bigcup_{i=1}^{g} R_i$, that is, a finite disjoint union of connected open sets. Therefore the map ϕ has a finite number of critical values separating S^1 into a finite number of intervals $w_1, \dots, w_l$. For any $1 \le j \le l$ take any point $y_j \in w_j$. The preimage $\eta_j = \phi^{-1}(y_j) \subset \Sigma_{0,2g}$ is a smooth 1–dimensional oriented submanifold, where the orientation on η_j is defined so that at each point of η_j the vector $\partial/\partial t_i$ and the positive tangent vector to η_s form a positive basis of the tangent space to the sphere. Moreover, $\eta_1, \dots, \eta_l$ are pairwise disjoint. Define Surger(c) as the formal sum $\sum_{j=1}^{l} |w_j| \eta_j$.

We claim that Surger$(c) \in \mathbb{R}_{\ge 0}^{\widetilde{\mathcal{D}}}$. It suffices to show that each connected component of η_j is either closed or its initial point belongs to ∂P_a for some a and its terminal point belongs to ∂Q_b for some b. This follows from the orientation argument. Indeed, for all i the restrictions $\phi|_{\partial P_i}$ and $\phi|_{\partial Q_i}$ have degrees -1 and 1, respectively. Hence $\phi|_{\partial P_i}$ can only contain initial points of components of η_j, while $\phi|_{\partial Q_i}$ can only contain terminal points of components of η_j. Consequently, no component of Surger(c) connects ∂P_a with ∂P_b or ∂Q_a with ∂Q_b. Moreover, since the restrictions $\phi|_{\partial P_i}$ and $\phi|_{\partial Q_i}$ have degrees -1 and 1, respectively, we obtain $\partial(\text{Surger}(c)) = \sum_{i=1}^{g} (q_i - p_i)$, so Surger$(c) \in \widetilde{\mathcal{B}}_{0,2g}$.

Proof of Theorem 5.3 Take a point $c \in \mathcal{B}_{0,2g}$. Then the map

$$d \mapsto \text{Drain}(\text{Surger}(tc + (1-t)d))$$

is a deformation retraction from $\mathcal{B}_{0,2g}$ to the point c. □

5.3 Stabilizer dimensions

Proposition 5.4 *The group* $\text{PMod}(\Sigma_{0,2g})$ *acts on* $\mathcal{B}_{0,2g}$ *without rotations.*

Proof Assume the converse and consider an element $\phi \in \text{PMod}(\Sigma_{0,2g})$ and a cell corresponding to a multiarc $\gamma = \gamma_1 \cup \dots \cup \gamma_s$ such that $\phi(\gamma_i) = \gamma_{\pi(i)}$ for a nontrivial permutation π. Without loss of generality can assume that there exist arcs γ_1, γ_2 and γ_3 from $p \in P$ to $q \in Q$ satisfying $\gamma_1 \ne \gamma_2$ and $\gamma_2 \ne \gamma_3$ (and possibly $\gamma_1 = \gamma_3$), such that $\phi(\gamma_1) = \gamma_2$ and $\phi(\gamma_2) = \gamma_3$. Denote by $W_1 \subset \Sigma_{0,2g}$ and $W_2 \subset \Sigma_{0,2g}$ the subsurfaces bounded by the loops $\gamma_1\bar{\gamma}_2$ and $\gamma_2\bar{\gamma}_3$, respectively ($\bar{\gamma}_i$ denotes the arc γ_i with opposite direction). We assume that W_1 and W_2 are located on the left sides of $\gamma_1\bar{\gamma}_2$ and $\gamma_2\bar{\gamma}_3$, respectively.

By construction of $\mathcal{B}_{0,2g}$ we see that γ_1 is not isotopic to γ_2, so W_1 contains a nonempty set of punctures $\varnothing \ne Z_1 \subset P \sqcup Q$. Define $\varnothing \ne Z_2 \subset P \sqcup Q$ in a similar way. Since γ_2 separates W_1 from W_2 we have $Z_1 \ne Z_2$. The map f preserves the orientation, therefore $f(W_1) = W_2$ and so $f(Z_1) = Z_2$. However, $f \in \text{PMod}(\Sigma_{0,2g})$ preserves the punctures, so we come to a contradiction. □

Theorem 5.5 *Let* σ *be a cell of* $\mathcal{B}_{0,2g}$. *Then*

$$\dim(\sigma) + \text{cd}(\text{Stab}_{\text{PMod}(\Sigma_{0,2g})}(\sigma)) \le 2g - 3.$$

Proof The cell σ is given by a multiarc $\gamma_1 \cup \dots \cup \gamma_E$. Consider the planar graph Υ on the sphere with the vertices $p_1, \dots, p_g, q_1, \dots, q_g$ and the edges $\gamma_1, \dots, \gamma_E$. It is convenient for us to denote the number of vertices by $V = 2g$. Also let us denote by C the number of connected components of Υ and by F the number of its faces (ie the number of connected components of $\Sigma_{0,2g} \setminus \Upsilon$).

Lemma 5.6 *We have*

$$\dim(\sigma) = \dim(H_1(\Upsilon, \mathbb{R})) = E - V + C. \tag{27}$$

Proof The condition $\partial\big(\sum_{i=1}^{E} k_i\gamma_i\big) = \sum_{i=1}^{g}(q_i - p_i)$ is a nonhomogeneous system of linear equation in $\mathbb{R}^E$. The affine space of its solutions has the same dimension as the space of solutions of the homogeneous system $\partial\big(\sum_{i=1}^{E} k_i\gamma_i\big) = 0$. This space is precisely $H_1(\Upsilon, \mathbb{R})$. The cell σ is given by the intersection of this affine space with $\mathbb{R}^E_{\geq 0}$. Condition (i) in the construction of $\mathcal{B}_{0,2g}$ implies that σ contains a point in the interior of $\mathbb{R}^E_{\geq 0}$. Therefore $\dim(\sigma) = \dim(H_1(\Upsilon, \mathbb{R}))$. The second equality in (27) is trivial. □

Denote by $Y_1, \dots, Y_F$ the connected components of $\Sigma_{0,2g} \setminus \Upsilon$. We have $Y_i \cong \Sigma_{0,k_i}$ for some k_i. Recall that Σ_0^k denotes the sphere with k boundary components.

Proposition 5.7 $\mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\sigma) \cong \mathrm{Mod}(\Sigma_0^{k_1}) \times \dots \times \mathrm{Mod}(\Sigma_0^{k_F})$.

Proof By construction we have $\mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\sigma) = \mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\Upsilon)$. Denote by $\overline{Y}_i$ the closure of $Y_i = \Sigma_{0,k_i}$ in the sphere. Let $\widetilde{Y}_i \cong \Sigma_0^{k_i}$ be the compactification of Y_i given by replacing each puncture by a boundary component. Let $p_i \colon \widetilde{Y}_i \to \overline{Y}_i$ be the natural projection. Then we have the corresponding mapping $\Phi_i \colon \mathrm{Mod}(\widetilde{Y}_i) \to \mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\Upsilon)$. It suffices to prove that the obvious mapping

$$\Phi \colon \mathrm{Mod}(\widetilde{Y}_1) \times \dots \times \mathrm{Mod}(\widetilde{Y}_F) \to \mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\Upsilon)$$

is an isomorphism. We use the Alexander method; see [9, Proposition 2.8]. In the proof we need to distinguish between mapping classes and their representatives; the mapping class of a homeomorphism ψ is denoted by $[\psi]$.

First we prove the surjectivity of Φ. Let $[\psi] \in \mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\Upsilon)$. Then $\psi(\delta)$ is isotopic to δ for each arc δ of Υ. All such arcs are disjoint, so the Alexander method implies that there is a representative $\psi' \in [\psi]$ such that $\psi'(\delta) = \delta$ for each arc δ of Υ. Set $\phi'_i = \psi'|_{\overline{Y}_i}$. Since ϕ'_i is identical on $\partial\overline{Y}_i$, there exist $\phi_i \in \mathrm{Homeo}^+(\widetilde{Y}_i)$ such that $p_i \circ \phi_i = \phi'_i \circ p_i$. Hence $\Phi([\phi_1], \dots, [\phi_F]) = [\psi]$.

Now we prove that Φ is injective. Let $\Phi([\psi_1], \dots, [\psi_F]) = [\mathrm{id}]$. Since for each i the mapping $\psi_i|_{\partial\widetilde{Y}_i}$ is identical, there exists $\psi'_i \in \mathrm{Homeo}^+(\overline{Y}_i)$ such that $p_i \circ \psi_i = \psi'_i \circ p_i$. Consider the mapping $\psi' \in \mathrm{Homeo}^+(\Sigma_{0,2g})$ such that $\psi'|_{\overline{Y}_i} = \psi'_i$ for all i. By assumption ψ' is isotopic to the identity map.

Let $\widetilde{\Upsilon}$ be a planar graph on the sphere obtained from Υ by adding several arcs such that each face of $\widetilde{\Upsilon}$ is a disk. Let us show that there is an isotopy $\Psi_t \colon \Sigma_{0,2g} \to \Sigma_{0,2g}$ with $\Psi_0 = \psi'$ such that Ψ restricts to the identity on Υ and $\Psi_1(\psi'(\delta)) = \delta$ for each arc δ of $\widetilde{\Upsilon}$. It suffices to prove the existence of this isotopy in the case when we add only one arc γ to Υ. Let $\Upsilon' = \Upsilon \cup \{\gamma\}$. We can assume that $\psi'(\gamma)$ is transverse to γ. If $\psi'(\gamma)$ is disjoint from γ then these two arcs bound a disk on $\Sigma_{0,2g}$. This disk is contains no punctures, so it is disjoint from Υ. Hence in this case such an isotopy exists. If $\psi'(\gamma)$ and γ intersect, they form a bigon (see [9, Proposition 1.7]) that is disjoint from Υ for the same reason. So we can decrease the number of intersection points of γ and $\psi'(\gamma)$.

Set $\phi' = \Psi_1$, $\phi'_i = \phi'|_{\overline{Y}_i}$ and $\Psi'_i = \Psi|_{\overline{Y}_i}$. There exist homeomorphisms $\phi_i \in \mathrm{Homeo}(\widetilde{Y}_i)$ and isotopies Ψ_i of $\widetilde{Y}_i$ such that $p_i \circ \phi_i = \phi'_i \circ p_i$ and $p_i \circ \Psi_i = \Psi'_i \circ p_i$. Therefore Ψ_i is an isotopy between ψ_i and ϕ_i. By construction ϕ_i is identical on a collection of arcs that fill $\widetilde{Y}_i$ (*fill* means that each connected component of the complement to this collection is a disk). Hence the Alexander method implies that ϕ_i is isotopic to the identity for each i. Therefore ψ_i is also isotopic to the identity. □

For $k \geq 2$ we have $\mathrm{Mod}(\Sigma^1_{0,k-1}) \cong \mathrm{PB}_{k-1}$. If we replace the punctures on the disk S^1_0 by boundary components, the corresponding mapping class groups will by related to each other via the following exact sequence (see [9, Proposition 3.19]):

$$1 \to \mathbb{Z}^{k-1} \to \mathrm{Mod}(\Sigma^k_0) \to \mathrm{Mod}(\Sigma^1_{0,k-1}) \to 1.$$

Since the tangent bundle to the disk is trivial, this sequence splits. Therefore $\mathrm{Mod}(\Sigma^k_0) \cong \mathbb{Z}^{k-1} \times \mathrm{PB}_{k-1}$. Since $\mathrm{cd}(\mathrm{PB}_{k-1}) = k-2$ we have $\mathrm{cd}(\mathbb{Z}^{k-1} \times \mathrm{PB}_{k-1}) = 2k-3$. When $k = 1$ we have $\mathrm{cd}(\mathrm{Mod}(\Sigma^1_0)) = 0$. Denote by D the number of Y_i that are homeomorphic to the disk. Proposition 5.7 immediately implies the following result:

Corollary 5.8 $$\mathrm{cd}(\mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\sigma)) = \sum_{i=1}^{F}(2k_i - 3) + D.$$

Let us finish the proof of Theorem 5.5. By Lemma 5.6 and Corollary 5.8 we have

$$\dim(\sigma) + \mathrm{cd}(\mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\sigma)) = E - V + C + \sum_{i=1}^{F}(2k_i - 3) + D = E - V + C + D - 3F + 2\sum_{i=1}^{F} k_i.$$

Let $\Theta_1, \dots, \Theta_C$ be the connected components of Υ. Note that

$$\begin{aligned} \sum_{i=1}^{F} k_i &= |\{(Y_i, \Theta_j) \mid Y_i \text{ is adjacent to } \Theta_j\}| = \sum_{j=1}^{C}(\dim(H_1(\Theta_j, \mathbb{R})) + 1) \\ &= \dim(H_1(\Upsilon, \mathbb{R})) + C = E - V + 2C. \end{aligned} \tag{28}$$

Therefore

$$\begin{aligned} E - V + C + D - 3F + 2\sum_{i=1}^{F} k_i &= E - V + C + D - 3F + 2(E - V + 2C) \\ &= 3E - 3V + 5C - 3F + D = 2C + D - 3(V - E + F - C). \end{aligned}$$

By Euler's formula we have

$$V - E + F - C = 1. \tag{29}$$

Therefore

$$\dim(\sigma) + \mathrm{cd}(\mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\sigma)) \leq 2C + D - 3. \tag{30}$$

In order to finish the proof of Theorem 5.5 we need the following result:

Lemma 5.9 *Let a planar graph Υ represent a cell of $\mathcal{B}_{0,2g}$ and $g \geq 2$. Then $2C + D \leq 2g$.*

Proof We proceed by induction on the number of connected components of Υ with only one edge.

Base case: Υ does not have a connected component with only one edge Since $D \leq F$ and $V = 2g$, it suffices to check that

$$2C + F \leq V. \tag{31}$$

Note that Υ is a bipartite graph and does not contain isotopic edges. Since Υ does not have a connected component with only one edge, if Y_i is adjacent to Θ_j for some i and j, then Y_i is adjacent to at least four edges of Θ_j. Then by (28) we have

$$E \geq 2\sum_{i=1}^{F} k_i = 2E - 2V + 4C = 2C + 2F - 2.$$

The last equality follows from (29). Since $C \geq 1$ we have

$$E \geq 2C + 2F - 2 \geq 2F + C - 1.$$

We can rewrite this as

$$2C + F \leq 1 + C - F + E. \tag{32}$$

Equation (29) implies that the right-hand side of (32) equals V. Therefore (31) holds.

Induction step: Υ has a connected component with only one edge In the case $g = 2$ the graph Υ is a disjoint union of two closed intervals, so $C = 2$ and $D = 0$; in this case the required inequality $2C + D \leq 2g$ is obvious. Hence we can assume that $g \geq 3$. Let p_i and q_j form such a component, that is, p_i and q_j are vertices of Υ of degree 1 connected by an edge α. Assume that after removing this component the remaining graph Υ_1 will not contain isotopic edges (and, consequently, will represent some cell of $\mathcal{B}_{0,2g-2}$). Then $C_1 = C - 1$ is the number of connected components of Υ_1. Denote by D_1 the number of faces of Υ_1 homeomorphic to the disk. We have $D_1 \leq D + 1$, since at most one disk can appear. The graph Υ_1 has fewer connected components with only one edge than Υ. Since $g - 1 \geq 2$, by the induction assumption we have

$$2C + D \leq 2C_1 + D_1 + 1 \leq 2g - 2 + 1 < 2g.$$

Now assume that our previous assumption does not hold, that is, after removing the component consisting of one edge, the remaining graph will contain isotopic edges. This means that there exist punctures p_r and q_s and edges β_1 and β_2 between them such that p_i and q_j are the only vertices of Υ located inside of the disks bounded by β_1 and β_2. There exists an arc γ_1 from p_i to q_s and an arc γ_2 from p_r to q_j such that γ_1 and γ_2 are disjoint from Υ and from each other. Consider the graph Υ' obtained from Υ by adding the edges γ_1 and γ_2. Note that Υ' has fewer connected components with exactly one edge than Υ and also represents a cell of $\mathcal{B}_{0,2g}$. Then $C' = C - 1$ is the number of connected components of Υ' and $D' = D + 2$ is the number of faces of Υ' homeomorphic to the disk. Therefore $2C + D \leq 2g$ if and only if $2C' + D' \leq 2g$. The induction assumption concludes the proof. □

Lemma 5.9 and (30) imply that

$$\dim(\sigma) + \mathrm{cd}(\mathrm{Stab}_{\mathrm{PMod}(\Sigma_{0,2g})}(\sigma)) \leq 2g - 3.$$ □

5.4 The spectral sequence

Let $K \subseteq \mathrm{PMod}(\Sigma_{0,2g})$ be a subgroup. Denote by $\widehat{E}^*_{*,*}$ the spectral sequence (22) for the action of K on $\mathcal{B}_{0,2g}$. Since cohomological dimension is monotonic, Theorem 5.5 implies that for any cell σ of $\mathcal{B}_{0,2g}$ we have

$$\dim(\sigma) + \mathrm{cd}(\mathrm{Stab}_K(\sigma)) \leq 2g-3.$$

This immediately implies $\widehat{E}^1_{p,q} = 0$ for $p+q > 2g-3$. Hence all differentials $d^1, d^2, \dots$ to the group $\widehat{E}^1_{0,2g-3}$ are trivial (Figure 8 is also applicable here, where the group $\widehat{E}^1_{0,2g-3}$ is shaded), so $\widehat{E}^1_{0,2g-3} = \widehat{E}^\infty_{0,2g-3}$. Therefore we have the following result:

Proposition 5.10 *Let $\mathfrak{L} \subseteq \mathcal{L}_0$ be a subset consisting of multiarcs from pairwise different K–orbits. Then the inclusions $\mathrm{Stab}_K(L) \subseteq K$, $L \in \mathfrak{L}$ induce the injective homomorphism*

$$\bigoplus_{L \in \mathfrak{L}} H_{2g-3}(\mathrm{Stab}_K(L), \mathbb{Z}) \hookrightarrow H_{2g-3}(K), \mathbb{Z}).$$

Proof of Lemma 5.1 It suffices to prove that the inclusions

$$j_u \colon \mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot N) \hookrightarrow \mathcal{K}_g \quad \text{for } u \in U_g$$

induce the injective homomorphism

$$\bigoplus_{u \in U_g} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u} \cdot N), \mathbb{Z}) \hookrightarrow G_1 \subseteq H_{2g-3}(\mathcal{K}_g, \mathbb{Z}).$$

Assume the converse and consider a nontrivial linear relation

$$\sum_{s=1}^{k} \lambda_s (j_{u_s})_*(\theta_s) = 0 \quad \text{for } \lambda_s \in \mathbb{Z} \text{ and } \theta_s \in H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(\hat{u}_s \cdot N), \mathbb{Z}) \tag{33}$$

for some pairwise different $u_1, \dots, u_s \in U_g$. For each $i = 1, \dots, g$ take an essential simple closed curve $\beta_i = \beta_{1,i}$ on the one-punctured torus X_i. Denote by $b_i = [\beta_{1,i}] \in H_1(\Sigma_g, \mathbb{Z})$ the corresponding homology class. For each $s \in \mathcal{N}$ denote by $\widehat{X}_{s,i} \subset \Sigma_g$ the one-punctured torus bounded by $\hat{u}_s \cdot \delta_i$. Since $\hat{u}_s$ belongs to the Torelli group $\mathcal{I}_g$, we have $H_1(\widehat{X}_{s,i}, \mathbb{Z}) = H_1(\widehat{X}_{t,i}, \mathbb{Z})$ for all $1 \leq s, t \leq k$. Denote by $\beta_{s,i}$ a unique curve on $\widehat{X}_{s,i}$ representing the homology class b_i.

Let $B_s = \beta_{s,1} \cup \dots \cup \beta_{s,g}$. Let $\{B_{d_1}, \dots, B_{d_l}\} \subseteq \{B_1, \dots, B_k\}$ be the maximal subset consisting of the multicurves from pairwise distinct $\mathcal{K}_g$–orbits. Take the homology class $x = \sum_{i=1}^{g} b_i$ and consider the complex of cycles $\mathcal{B}_g(x)$. Proposition 4.3 implies that the inclusions

$$\iota_i \colon \mathrm{Stab}_{\mathcal{K}_g}(B_{d_i}) \hookrightarrow \mathcal{K}_g$$

induce the injective homomorphism

$$\bigoplus_{i=1}^{l} H_{2g-3}(\mathrm{Stab}_{\mathcal{K}_g}(B_{d_i}), \mathbb{Z}) \hookrightarrow H_{2g-3}(\mathcal{K}_g, \mathbb{Z}). \tag{34}$$

Since the curves $\beta_{s,i}$ can be chosen in a unique way, we have the inclusions

$$\hat{j}_{u_s}\colon \mathrm{Stab}_{\mathcal{K}_g}(\hat{u}_s\cdot N)\hookrightarrow \mathrm{Stab}_{\mathcal{K}_g}(B_s).$$

Since $j_{u_i}=\iota_i\circ\hat{j}_{u_i}$, (34) and (33) imply that for each $i=1,\dots,l$ we have

$$\sum_{\{z\mid B_z\in\mathrm{Orb}_{\mathcal{K}_g}(B_{d_i})\}}\lambda_z(j_{u_z})_*(\theta_z)=0. \tag{35}$$

Equality (35) implies that it is sufficient to prove the statement of the lemma in the case where the multicurves $B_1,\dots,B_k$ belong to the same $\mathcal{K}_g$–orbit. Since we can prove Lemma 5.1 for an arbitrary choice of $\widehat{U}$, then by choosing the lifts $\hat{u}$ we can assume that $B_1=\dots=B_k=B$. Let $\zeta_{s,i}$ be a curve on $\widehat{X}_{s,i}$ intersecting β_i once and let $L_s=\zeta_{s,1}\cup\dots\cup\zeta_{s,g}$. Consider the surface $\Sigma_g\setminus B\cong\Sigma_{0,2g}$. Denote by p_i and q_i the punctures on $\Sigma_{0,2g}$ corresponding to the two sides of the curve β_i.

Consider the exacts sequence (6) in the case $M=B$. We have

$$1\to\langle T_{\beta_1},\dots,T_{\beta_g}\rangle\to\mathrm{Stab}_{\mathrm{Mod}(\Sigma_g)}(B)\to\mathrm{Mod}(\Sigma_{0,2g})\to 1.$$

Since the intersection $\langle T_{\beta_1},\dots,T_{\beta_g}\rangle\cap\mathcal{K}_g$ is trivial, we have the inclusion $K=\mathrm{Stab}_{\mathcal{K}_g}(B)\hookrightarrow\mathrm{Mod}(\Sigma_{0,2g})$. The action of $\mathcal{K}_g$ on the homology of Σ_g is trivial, so the image of this inclusion is contained in $\mathrm{PMod}(\Sigma_{0,2g})$. We have $K\hookrightarrow\mathrm{PMod}(\Sigma_{0,2g})$. Denote by $\zeta'_{s,i}$ the arc on $\Sigma_{0,2g}$ from p_i to q_i corresponding to the curve $\zeta_{s,i}$ and let $L'_s=\zeta'_{s,1}\cup\dots\cup\zeta'_{s,g}$. Let us show that $L'_1,\dots,L'_k$ belong to pairwise distinct K–orbits.

Assuming the converse, $f(L'_1)=L'_2$ for some $f\in K$. Then $f(L_1\cup B)=L_2\cup B$. Note that the surface $\Sigma_g\setminus(L_s\cup B)$ has g punctures, and each component of $\hat{u}_s\cdot N$ is homotopic into a neighborhood of its own puncture. Therefore the corresponding components of the multicurves $f(\hat{u}_1\cdot N)$ and $\hat{u}_2\cdot N$ are homotopic into a neighborhood of the same puncture. Consequently, the multicurves $f(\hat{u}_1\cdot N)$ and $\hat{u}_2\cdot N$ are isotopic. Since $\hat{u}_1,\hat{u}_2\in\mathcal{I}_g$, we obtain $\hat{u}_2^{-1}f\hat{u}_1\in\mathrm{Stab}_{\mathcal{I}_g}(N)$. It follows from the exactness of (9) that $\mathrm{Stab}_{\mathcal{I}_g}(N)\subseteq\mathcal{K}_g$. Hence $\hat{u}_2^{-1}f\hat{u}_1\in\mathcal{K}_g$ and we obtain

$$0=\tau(\hat{u}_2^{-1}f\hat{u}_1)=\tau(\hat{u}_1)-\tau(\hat{u}_2),$$

where τ is the Johnson homomorphism. This implies $u_1=u_2$, giving a contradiction.

Therefore $L'_1,\dots,L'_k$ belong to pairwise distinct K–orbits. Proposition 5.10 implies that the inclusions $\mathrm{Stab}_K(L'_s)\subseteq K$, $L'\in\mathfrak{L}$ induce the injective homomorphism

$$\bigoplus_s H_{2g-3}(\mathrm{Stab}_K(L'_s),\mathbb{Z})\hookrightarrow H_{2g-3}(K,\mathbb{Z}).$$

By Proposition 4.3 we also have the inclusion $H_{2g-3}(K,\mathbb{Z})\hookrightarrow H_{2g-3}(\mathcal{K}_g,\mathbb{Z})$ and so $\mathrm{Stab}_K(L'_s)=\mathrm{Stab}_{\mathcal{K}_g}(\hat{u}_s\cdot N)$. Therefore (33) implies $\lambda_s=0$ for all s. This contradiction proves Lemma 5.1. □

References

[1] **V I Arnold**, *The cohomology ring of the colored braid group*, Mat. Zametki 5 (1969) 227–231 Zbl In Russian; translated in Math. Notes Acad. Sci. USSR 5 (1969) 138–140

[2] **M Bestvina**, **K-U Bux**, **D Margalit**, *The dimension of the Torelli group*, J. Amer. Math. Soc. 23 (2010) 61–105 MR Zbl

[3] **J S Birman**, **A Lubotzky**, **J McCarthy**, *Abelian and solvable subgroups of the mapping class groups*, Duke Math. J. 50 (1983) 1107–1120 MR Zbl

[4] **K S Brown**, *Cohomology of groups*, Graduate Texts in Math. 87, Springer (1982) MR Zbl

[5] **T Church**, **M Ershov**, **A Putman**, *On finite generation of the Johnson filtrations*, J. Eur. Math. Soc. 24 (2022) 2875–2914 MR Zbl

[6] **A Dimca**, **R Hain**, **S Papadima**, *The abelianization of the Johnson kernel*, J. Eur. Math. Soc. 16 (2014) 805–822 MR Zbl

[7] **A Dimca**, **S Papadima**, *Arithmetic group symmetry and finiteness properties of Torelli groups*, Ann. of Math. 177 (2013) 395–423 MR Zbl

[8] **M Ershov**, **S He**, *On finiteness properties of the Johnson filtrations*, Duke Math. J. 167 (2018) 1713–1759 MR Zbl

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

[10] **A A Gaifullin**, *On a spectral sequence for the action of the Torelli group of genus* 3 *on the complex of cycles*, Izv. Ross. Akad. Nauk Ser. Mat. 85 (2021) 27–103 MR Zbl In Russian; translated in Izv. Math. 85 (2021) 1060–1127

[11] **A A Gaifullin**, *On the top homology group of the Johnson kernel*, Mosc. Math. J. 22 (2022) 83–102 MR Zbl

[12] **A A Gaifullin**, *On infinitely generated homology of Torelli groups*, Algebra i Analiz 35 (2023) 87–134

[13] **A Hatcher**, **D Margalit**, *Generating the Torelli group*, Enseign. Math. 58 (2012) 165–188 MR Zbl

[14] **N V Ivanov**, *Subgroups of Teichmüller modular groups*, Transl. Math. Monogr. 115, Amer. Math. Soc., Providence, RI (1992) MR Zbl

[15] **D Johnson**, *The structure of the Torelli group, II: A characterization of the group generated by twists on bounding curves*, Topology 24 (1985) 113–126 MR Zbl

[16] **G Mess**, *The Torelli groups for genus* 2 *and* 3 *surfaces*, Topology 31 (1992) 775–790 MR Zbl

[17] **S Morita**, **T Sakasai**, **M Suzuki**, *Torelli group, Johnson kernel, and invariants of homology spheres*, Quantum Topol. 11 (2020) 379–410 MR Zbl

Faculty of Mathematics, National Research University Higher School of Economics
Moscow, Russia

Skolkovo Institute of Science and Technology
Skolkovo, Russia

spiridonovia@ya.ru

Received: 22 April 2022 Revised: 11 February 2023

Algebraic & Geometric Topology **24**:7 (2024) 3669–3691
DOI: 10.2140/agt.2024.24.3669
Published: 9 December 2024

The Heisenberg double of involutory Hopf algebras and invariants of closed 3–manifolds

Serban Matei Mihalache
Sakie Suzuki
Yuji Terashima

We construct an invariant of closed oriented 3–manifolds using a finite-dimensional involutory unimodular and counimodular Hopf algebra H. We use the framework of normal o–graphs introduced by R Benedetti and C Petronio, in which one can represent a branched ideal triangulation via an oriented virtual knot diagram. We assign a copy of the canonical element of the Heisenberg double $\mathcal{H}(H)$ of H to each real crossing, which represents a branched ideal tetrahedron. The invariant takes values in the cyclic quotient $\mathcal{H}(H)/[\mathcal{H}(H), \mathcal{H}(H)]$, which is isomorphic to the base field. In the construction we use only the canonical element and structure constants of H, and not any representations of H. This, together with the finiteness and locality conditions of the moves for normal o–graphs, makes the calculation of our invariant rather simple and easy to understand. When H is the group algebra of a finite group, the invariant counts the number of group homomorphisms from the fundamental group of the 3–manifold to the group.

57K31

1 Introduction

S Baaj and G Skandalis [1] and R M Kashaev [6] found a striking fact: for the Heisenberg double $\mathcal{H}(H)$ of any finite-dimensional Hopf algebra H, there exists a canonical element T in $\mathcal{H}(H)^{\otimes 2}$ satisfying the pentagon equation

$$T_{12}T_{13}T_{23} = T_{23}T_{12}.$$

In [6], Kashaev also showed that the Drinfeld double $\mathcal{D}(H)$ of H can be realized as a subalgebra of $\mathcal{H}(H)^{\otimes 2}$, and observed that the universal R–matrix of $\mathcal{D}(H)$ can be represented as a combination of four copies of T, where the quantum Yang–Baxter equation of the universal R–matrix follows from a sequence of the pentagon equation of T. Using his results, the second author [16] reconstructed the universal quantum $\mathcal{D}(H)$ invariant of framed tangles by assigning a copy of the canonical element T to each branched ideal tetrahedron of the tangle complements, and expected that this construction leads to invariants of pairs of a 3–manifold and geometrical input. We show that this construction defines an invariant of closed oriented 3–manifolds when the Hopf algebra H is involutory unimodular and counimodular.

In the formulation of our invariant, we use a diagrammatic representation of closed oriented 3–manifolds introduced by R Benedetti and C Petronio [3]. Their diagrams, which are called closed normal o–graphs, are oriented virtual knot diagrams satisfying certain conditions. They showed that homeomorphism classes of closed oriented 3–manifolds are identified with equivalence classes of closed normal o–graphs up to certain moves. A crossing of a closed normal o–graph represents a branched ideal tetrahedron in the corresponding 3–manifold, and the orientation of the strand specifies a way to extend these local branching structures to a global one. Our invariant is obtained by assigning a copy of the canonical element T (or its inverse) to each crossing of closed normal o–graphs, and by reading them along the strands. The invariant takes values in the cyclic quotient $\mathcal{H}(H)/[\mathcal{H}(H),\mathcal{H}(H)]$, which is isomorphic to the base field through the character of the Fock space representation.

The proof of the invariance will be performed by checking the invariance under each move for normal o–graphs. There are two important types: the MP–moves and the CP–move. An MP–move represents a Pachner move equipped with a branching structure, which corresponds to a modified pentagon equation in the level of the invariant. Here, the modification is related to the antipode of H, and we need to assume that the antipode is involutive. Up to the MP–moves (and the 0–2 move), a closed normal o–graph represents a closed oriented 3–manifold with a combing. The CP–move is a special move for branched triangulations, which changes the combing. The invariance under the CP–move will be shown using tensor networks, where we need to assume that H is in addition both unimodular and counimodular. Using tensor networks, we will also show a connected sum formula of the invariant.

The main example of finite-dimensional involutory unimodular and counimodular Hopf algebras is the group algebra $\mathbb{C}[G]$ of a finite group G. We show that, in this case, the invariant is same as the number of homomorphisms from the fundamental group of the 3–manifold to G. There are several other examples of Hopf algebras which would be interesting to consider. The restricted enveloping algebras of restricted Lie algebras (see Jacobson [5]) are finite-dimensional involutory Hopf algebras. In [14], S Majid and A Pachol classified Hopf algebras of dimension ≤ 4 over the field of characteristic 2. M Kim [8] also gave some examples of finite-dimensional involutory unimodular and counimodular (commutative and cocommutative) Hopf algebras which are not group algebras.

There are several invariants based on triangulations; see Barrett and Westbury [2] and Turaev and Viro [17]. Ours uses only the structure constants of Hopf algebras and not representation categories, and thus the construction is rather simple. It would be interesting to compare our invariant and the Kuperberg invariant [11; 12], which also uses only the structure constants, and agrees with our invariant on group algebras. We remark that the Kuperberg invariant is constructed based on Heegaard diagrams, and the handle slide moves in Heegaard diagrams are global and infinite. The moves for closed normal o–graphs are local and finite, and each of these finite moves corresponds nicely to an algebraic equation, like the MP–moves and the pentagon equation. Thus the proof of the invariance is rather easy to understand.

The rest of the paper is organized as follows. In Section 2, we discuss Hopf algebras and their Heisenberg doubles. In Section 3, following Benedetti and Petronio [3], we explain how to represent closed oriented 3–manifolds in a combinatorial manner using closed normal o–graphs. In Section 4, we explain the construction of our invariant $Z(M;\mathcal{H}(H)) = Z(\Gamma;\mathcal{H}(H))$ using the Heisenberg double $\mathcal{H}(H)$ and a closed normal o–graph Γ which represents a 3–manifold M. The proof of the invariance will be given in two sections. In Section 5, we prove the invariance under all moves except for the CP–move. In Section 6, we reformulate our invariant using tensor networks, and use them to prove the invariance under the CP–move. In Section 7, we show the connected sum formula, and study the case for the group algebra $\mathbb{C}[G]$ of a finite group G.

Acknowledgments We thank S Baseilhac, R Benedetti, K Hikami, M Ishikawa, R M Kashaev, A Kato, Y Koda and T T Q Lê for valuable discussions. This work is partially supported by JSPS KAKENHI grants JP17K05243 and JP19K14523, and by JST CREST grant JPMJCR14D6.

2 Hopf algebra and Heisenberg double

In this section, we quickly review the definition and some properties of the Heisenberg double of Hopf algebras.

2.1 Hopf algebra

A *Hopf algebra* H over a field $\mathbb{K}$ is a vector space equipped with five linear maps,

$$M\colon H\otimes H\to H,\quad 1\colon \mathbb{K}\to H,\quad \Delta\colon H\to H\otimes H,\quad \epsilon\colon H\to\mathbb{K}\quad\text{and}\quad S\colon H\to H,$$

called multiplication, unit, comultiplication, counit and antipode, respectively, satisfying the standard axioms of Hopf algebras. When the antipode is involutive, ie $S^2=\mathrm{id}_H$, we call H *involutory*. Throughout the paper, H will denote a finite-dimensional Hopf algebra and H^* will denote the *dual Hopf algebra* of H. We will also use the Sweedler notation $\Delta(x)=x_{(1)}\otimes x_{(2)}$ for $x\in H$.

Recall that a *right integral* of a Hopf algebra H is an element $\mu_R\in H^*$ satisfying $\mu_R\cdot f=\mu_R f(1)$ for every $f\in H^*$. A *left integral* is defined similarly. Since H is finite-dimensional, a left (resp. right) integral of the dual Hopf algebra H^* is an element of H and is called a *left* (*resp. right*) *cointegral* of H. It is well known that for a finite-dimensional Hopf algebra, an integral always exists and is unique up to scalar multiplication. We say H is *unimodular* when the left cointegrals are also the right cointegrals, and *counimodular* when the left integrals are also the right integrals. For more details on Hopf algebras and their integrals, see [15; 11].

2.2 Heisenberg double

We use the left action of H on H^* defined by $(a\rightharpoonup f)(x):=f(xa)$, for $a,x\in H$ and $f\in H^*$. The *Heisenberg double*

$$\mathcal{H}(H)=H^*\otimes H$$

of a Hopf algebra H is a $\mathbb{K}$–algebra with unit $\epsilon\otimes 1$ and multiplication given by

$$(f\otimes a)\cdot(g\otimes b)=f\cdot(a_{(1)}\rightharpoonup g)\otimes a_{(2)}b,$$

for $a,b\in H$ and $f,g\in H^*$.

Let $\{e_i\}$ be the basis of H and $\{e^i\}$ its dual basis. Then the canonical element is given by

$$T=\sum_i(\epsilon\otimes e_i)\otimes(e^i\otimes 1)\quad\in\mathcal{H}(H)^{\otimes 2}\tag{2-1}$$

and its inverse by

$$\overline{T}=\sum_i(\epsilon\otimes S(e_i))\otimes(e^i\otimes 1)\quad\in\mathcal{H}(H)^{\otimes 2}.\tag{2-2}$$

In the case of the Drinfeld double $\mathcal{D}(H)$ of H, the canonical element satisfies the quantum Yang–Baxter equation, which in turn produces invariants of links and 3–manifolds [4; 7]. One important feature of the Heisenberg double, which plays a central role in our construction of invariants, is that the canonical element satisfies the pentagon equation.

Proposition 2.1 [1; 6] *The pentagon equation*

$$T_{12}T_{13}T_{23}=T_{23}T_{12}\tag{2-3}$$

holds in $\mathcal{H}(H)^{\otimes 3}$.

Proof Note that

$$\begin{aligned}T_{12}T_{13}T_{23}&=\sum_{i,j,k}(\epsilon\otimes e_i)(\epsilon\otimes e_j)\otimes(e^i\otimes 1)(\epsilon\otimes e_k)\otimes(e^j\otimes 1)(e^k\otimes 1)\\&=\sum_{i,j,k}(\epsilon\otimes e_ie_j)\otimes(e^i\otimes e_k)\otimes(e^je^k\otimes 1)\in\epsilon\otimes(H\otimes H^*)^{\otimes 2}\otimes 1,\\T_{23}T_{12}&=\sum_{i,j}(\epsilon\otimes e_j)\otimes(\epsilon\otimes e_i)(e^j\otimes 1)\otimes(e^i\otimes 1)\\&=\sum_{i,j}(\epsilon\otimes e_j)\otimes(e_{i(1)}\rightharpoonup e^j\otimes e_{i(2)})\otimes(e^i\otimes 1)\in\epsilon\otimes(H\otimes H^*)^{\otimes 2}\otimes 1.\end{aligned}$$

Let us identify $(H\otimes H^*)^{\otimes 2}$ with $\mathrm{End}(H\otimes H)$ through the map

$$\iota\colon x\otimes f\otimes y\otimes g\mapsto(a\otimes b\mapsto f(a)x\otimes g(b)y),$$

for $x,y,a,b\in H$ and $f,g\in H^*$. Then, after identifying $\epsilon\otimes(H\otimes H^*)^{\otimes 2}\otimes 1$ with $(H\otimes H^*)^{\otimes 2}$, we can see that the both elements $T_{12}T_{13}T_{23}$ and $T_{23}T_{12}$ are sent by ι to the same element as follows:

$$\begin{aligned}\iota(T_{12}T_{13}T_{23})(a\otimes b)&=e^i(a)e_ie_j\otimes e^je^k(b)e_k=ae_j\otimes e^j(b_{(1)})e^k(b_{(2)})e_k=ab_{(1)}\otimes b_{(2)},\\\iota(T_{23}T_{12})(a\otimes b)&=e_j(e_{i(1)}\rightharpoonup e^j)(a)\otimes e^i(b)e_{i(2)}=e_je^j(ae_{i(1)})\otimes e^i(b)e_{i(2)}\\&=ae_{i(1)}\otimes e^i(b)e_{i(2)}=ab_{(1)}\otimes b_{(2)}.\end{aligned}$$

□

The Heisenberg double $\mathcal{H}(H)$ has a canonical left module $F(H^*)=H^*$, which we call the *Fock space*, with the action $\phi\colon\mathcal{H}(H)\to\mathrm{End}(H^*)$ given by

$$\phi(f\otimes a)(g)=(f\otimes a)\triangleright g:=f\cdot(a\rightharpoonup g),\tag{2-4}$$

for $(f \otimes a) \in \mathcal{H}(H)$ and $g \in H^*$. Let χ_{Fock} be the character associated to the Fock space. For a $\mathbb{K}$–algebra A, let $[A, A]$ be the subspace spanned by $\{xy - yx \mid x, y \in A\}$ over $\mathbb{K}$. We are interested in the quotient space $\mathcal{H}(H)/[\mathcal{H}(H), \mathcal{H}(H)]$ since this is the space in which our invariant takes values.

Proposition 2.2 *The character of the Fock space*

$$\chi_{\mathrm{Fock}}\colon \mathcal{H}(H)/[\mathcal{H}(H), \mathcal{H}(H)] \to \mathbb{K}$$

is an isomorphism between vector spaces.

Proof In [13, Proposition 6.1] it was shown that for $F \in \mathrm{End}(H^*)$, the element

$$\sum_{i,j} F(e^i)e^j \otimes S^{-1}(e_j)e_i \in \mathcal{H}(H)$$

is in the preimage of F by ϕ, ie for any $g \in H^*$ we have

$$\begin{aligned}\sum_{i,j} \phi(F(e^i)e^j \otimes S^{-1}(e_j)e_i)(g) &= \sum_{i,j} F(e^i)e^j \cdot (S^{-1}(e_j)e_i \rightharpoonup g)\\ &= \sum_{i,j} g_{(3)}(e_i) \cdot F(e^i) \cdot g_{(2)}(S^{-1}(e_j))e^j \cdot g_{(1)}\\ &= F(g_{(3)}) \cdot S^{-1}(g_{(2)}) \cdot g_{(1)} = F(g_{(2)})\epsilon(g_{(1)}) = F(g).\end{aligned}$$

Since $\dim \mathcal{H}(H) = \dim \mathrm{End}(H^*)$, it follows that ϕ is bijective, and hence ϕ is an algebra isomorphism.

Note that $\mathrm{End}(H^*)$ is a matrix algebra, and thus the canonical trace

$$\mathrm{tr}\colon \mathrm{End}(H^*)/[\mathrm{End}(H^*), \mathrm{End}(H^*)] \to \mathbb{K}$$

is also an isomorphism. □

When the antipode S is involutive, χ_{Fock} can be given in terms of integrals. Let $\mu_R \in H^*$ and $e_L \in H$ be a right integral and a left cointegral satisfying $\mu_R(e_L) = 1$.

Proposition 2.3 *For an involutory Hopf algebra H, we have*

$$\chi_{\mathrm{Fock}}(f \otimes a) = f(e_L)\mu_R(a)$$

for $f \otimes a \in \mathcal{H}(H)$.

Proof For $F \in \mathrm{End}(H^*)$, the trace map is given (see [15, Chaper 10]) by

$$\mathrm{tr}(F) = \langle e_L, F(\mu_{R(2)})S(\mu_{R(1)})\rangle.$$

Thus

$$\chi_{\mathrm{Fock}}(f \otimes a) = \langle e_L, f(a \rightharpoonup \mu_{R(2)})S(\mu_{R(1)})\rangle = \langle e_L, \mu_{R(3)}(a) f \cdot \mu_{R(2)} \cdot S(\mu_{R(1)})\rangle = f(e_L)\mu_R(a).$$

The third equality follows from the fact that $x_{(2)}S(x_{(1)}) = S(x_{(1)}S(x_{(2)})) = \epsilon(x)1$ for an involutory Hopf algebra. □

Figure 1: Vertices of type + (left) and − (right).

3 Closed normal o–graph

In order to define an invariant, we first recall the method introduced in [3] by Benedetti and Petronio to represent closed oriented 3–manifolds in a combinatorial manner. This method is based on the theory of branched spines, which are the dual of branched ideal triangulations.

Definition 3.1 [3] A *closed normal o–graph* is an oriented virtual knot diagram, ie a finite connected 4–valent graph Γ immersed in $\mathbb{R}^2$ with the following data and conditions:

- N1 Only two types (+ or −) of 4–valent vertices are considered, which are represented by the over–under notation as in Figure 1.
- N2 Each edge has an orientation such that it matches among two edges which are opposite to each other at a vertex.
- C1 If one removes the vertices and joins the edges which are opposite to each other, the result is a unique oriented circuit.

This diagram satisfies the following additional conditions:

- C2 The trivalent graph obtained from Γ by the rule defined in Figure 2 is connected.
- C3 Consider the disjoint union of oriented circuits obtained from Γ by the rule defined in Figure 3. The number of these circuits is exactly one more than the number of vertices of Γ.

Let $\mathcal{G}$ be the set of closed normal o–graphs and $\mathcal{M}$ the set of oriented closed 3–manifolds up to orientation-preserving homeomorphisms. Given a closed normal o–graph $\Gamma \in \mathcal{G}$, one can canonically construct a 3–manifold $\Phi(\Gamma) \in \mathcal{M}$ as follows. We fix an orientation of $\mathbb{R}^3$ and place a closed normal o–graph on $\mathbb{R}^2 \subset \mathbb{R}^3$. Then we replace each of its vertices with a tetrahedron (with the orientation given by $\mathbb{R}^3$), and

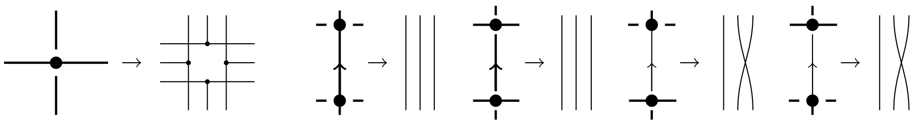


Figure 2

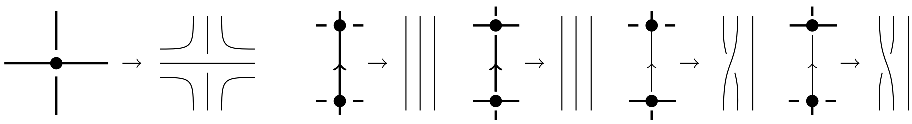


Figure 3

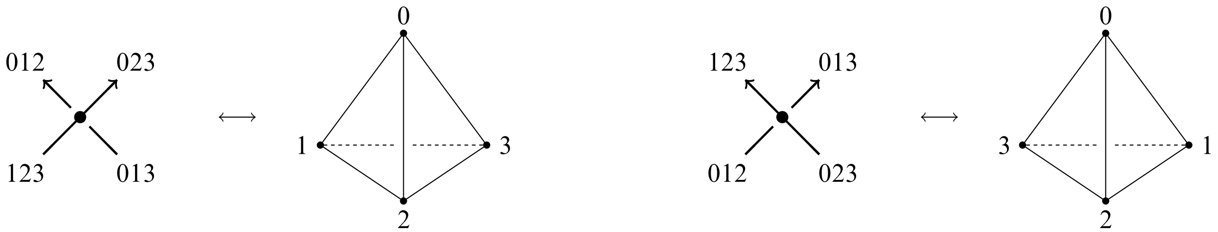


Figure 4

glue the faces of the ideal tetrahedra. This gluing is specified by the order of vertices of ideal tetrahedra defined as in Figure 4: we glue faces by the orientation-reversing map which preserves the order of vertices. Conditions on closed normal o–graphs ensure that the result is an ideally triangulated 3–manifold with S^2 boundary. Then, after we cap the boundary, the result defines an element $\Phi(\Gamma) \in \mathcal{M}$. Here, for the geometrical meaning of the order of the vertices of ideal tetrahedra, see Remark 3.4, where the meaning of the technical conditions C1, C2 and C3 are also explained. We denote the construction map obtained in the above way by $\Phi\colon \mathcal{G} \to \mathcal{M}$.

The map Φ is not one-to-one, and in order to make it so we need the following local moves of the diagrams:

(1) planar isotopy of the diagram and the *Reidemeister-type moves* described in Figure 5, left,

(2) the 0–2 *move* and the *Matveev–Piergallini moves* (*MP–moves*), described in Figure 5, right, and Figure 6, respectively,

(3) the *combinatorial Pontryagin move* (*CP–move*) in Figure 7.

All of the above local moves preserve the axioms of closed normal o–graphs. We say that two closed normal o–graphs are equivalent if one can be obtained from the other by planar isotopy and a finite sequence of moves defined above. Let us denote this equivalence relation by $\sim$. The following was proved in [3]:

Proposition 3.2 *The map*

$$\Phi\colon \mathcal{G}/\!\sim\, \to \mathcal{M} \tag{3-1}$$

is well defined and bijective.

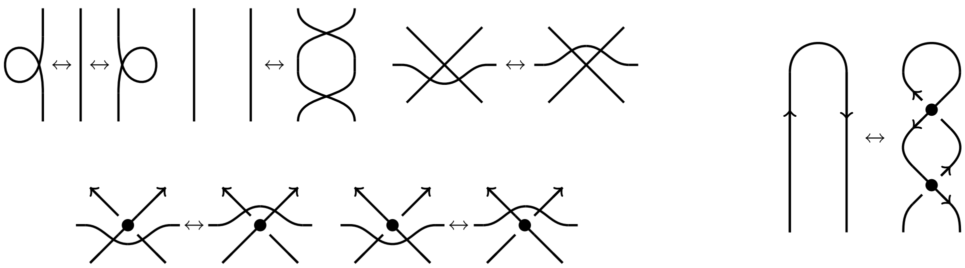


Figure 5: Left: the Reidemeister-type moves. Right: the 0–2 move.

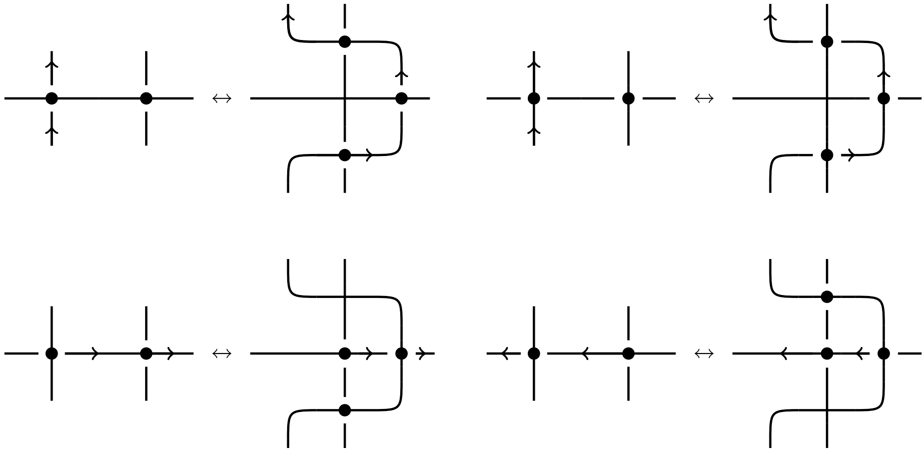


Figure 6: The MP–moves. The orientation of each nonoriented edge in the figure is arbitrary if it matches before and after the move.

Example 3.3 The closed normal o–graph of the lens space $L(p,1)$, for $p \geq 1$, is given by the following graph with p vertices:

Remark 3.4 We briefly remark on the geometrical meaning of the representation of 3–manifolds by closed normal o–graphs; see [3] for more details. The order of vertices of ideal tetrahedra as in Figure 4 specifies a branching structure for the ideal triangulation, which gives a combing, ie a nonvanishing vector field up to homotopy, to the underlying 3–manifold. The technical conditions C1, C2 and C3 ensure that the 3–manifold corresponding to a closed normal o–graph has an S^2 boundary with a nice branching structure, where the associated combing can be extended canonically to the closed 3–manifold after we can cap off the boundary by B^3. In this case, a closed normal o–graph up to the 0–2 move and the MP–moves represents a 3–manifolds with a combing. The CP–move in Figure 7 changes the combing while preserving the underlying 3–manifold; thus one gets the complete representation of $\mathcal{M}$ as in Proposition 3.2. Here the CP–move is an interpretation of the Pontryagin surgery in terms of branched standard spines; see [3, Chaper 6] for details.

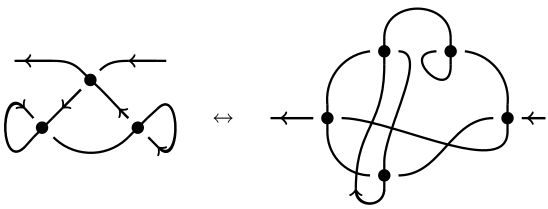


Figure 7: The CP–move.

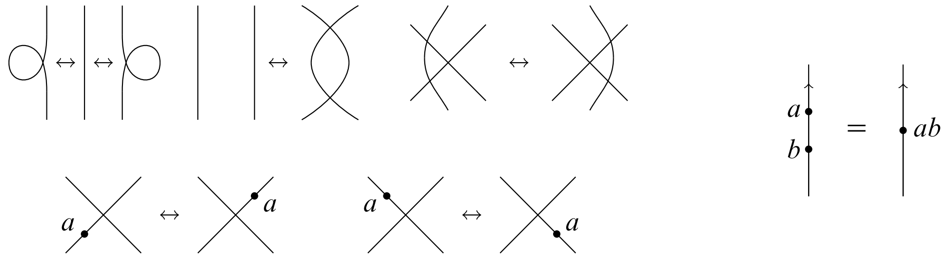


Figure 8

4 Invariant

In this section, we define a scalar $Z(\Gamma; \mathcal{H}(H))$ for a closed normal o–graph Γ. In the following section, we show that this scalar is an invariant of closed oriented 3–manifolds when the Hopf algebra is involutory, unimodular and counimodular.

Let A be a $\mathbb{K}$–algebra. An *A–decorated diagram* is an oriented closed curve immersed in $\mathbb{R}^2$, where the self-intersections are transverse double points, with a finite number of dots, each of which is labeled with an element of A. These dots are called beads. We shall consider the A–decorated diagram up to planar isotopy and moves in Figure 8. We also allow the beads to slide along the curve.

We define the scalar $Z(\Gamma; \mathcal{H}(H))$ as follows. Recall the definition of the canonical element T and its inverse $\overline{T}$ given in (2-1)–(2-2). Using Sweedler notation, we write the canonical element as $T = T_1 \otimes T_2 \in \mathcal{H}(H)^{\otimes 2}$ and its inverse as $\overline{T} = \overline{T}_1 \otimes \overline{T}_2 \in \mathcal{H}(H)^{\otimes 2}$. Given a closed normal o–graph, we replace its vertices with the diagram in Figure 10 to get an $\mathcal{H}(H)$–decorated diagram C_Γ.

Since a closed normal o–graph satisfies axiom C1 in Definition 3.1, we can perform the moves in Figure 8 and slide beads on C_Γ to get a closed circle with a single bead labeled by J_Γ in $\mathcal{H}(H)$. Because one can permute the beads as in Figure 9, J_Γ is well defined in the quotient space $\mathcal{H}(H)/[\mathcal{H}(H), \mathcal{H}(H)]$, which can be identified with $\mathbb{K}$ by Proposition 2.2.

Definition 4.1
$$Z(\Gamma; \mathcal{H}(H)) := \chi_{\mathrm{Fock}}(J_\Gamma).$$

Recall from Section 2.1 that a Hopf algebra H is called unimodular if the left cointegrals are also the right cointegrals, and counimodular if the left integrals are also the right integrals.

Theorem 4.2 *Let H be a finite-dimensional involutory unimodular counimodular Hopf algebra over $\mathbb{K}$, and Γ a closed normal o–graph of a closed oriented 3–manifold M. Then $Z(\Gamma, \mathcal{H}(H))$ is an invariant of M.*

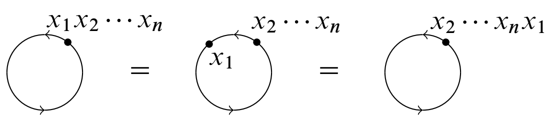
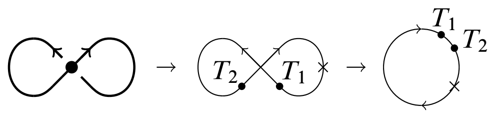


Figure 9

Figure 10: How to associate beads to vertices.

We end this section with some calculations, and prove Theorem 4.2 in Section 5.

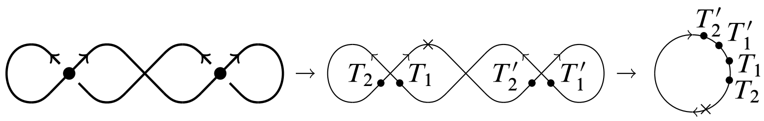
Example 4.3 The invariant of $S^3 = L(1,1)$ is given by



Thus $Z(S^3; \mathcal{H}(H)) = \chi_{\mathrm{Fock}}(T_2 T_1) = \sum_i \chi_{\mathrm{Fock}}(e^i \otimes e_i)$. By Proposition 2.3,

$$Z(S^3; \mathcal{H}(H)) = e^i(e_L)\mu_R(e_i) = \mu_R(e^i(e_L)e_i) = \mu_R(e_L) = 1.$$

Example 4.4 The invariant of $\mathbb{R}P^3 = L(2,1)$ is given by



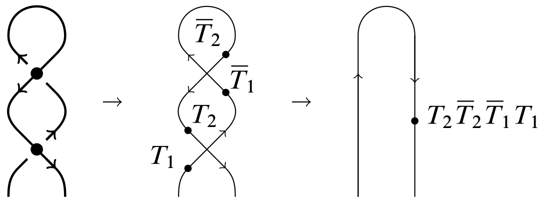
Thus $Z(\mathbb{R}P^3; \mathcal{H}(H)) = \chi_{\mathrm{Fock}}(T_2 T_1 T_1' T_2') = \sum_{i,j} \chi_{\mathrm{Fock}}(e^i \cdot (e_{i(1)} e_{j(1)} \rightharpoonup e^j) \otimes e_{i(2)} e_{j(2)}) = \mathrm{tr}(S)$, where S is the antipode.

5 Main theorem

In this section we prove Theorem 4.2. According to Proposition 3.2, in order to prove that $Z(\Gamma, \mathcal{H}(H))$ is an invariant, we need to show that $Z(\Gamma, \mathcal{H}(H))$ is an invariant under planar isotopy and the local moves (the Reidemeister-type moves, the 0–2 move, the MP–moves and the CP–move) of closed normal o–graphs described in Figures 5–7. In [16], it was essentially proved that the value $Z(\Gamma, \mathcal{H}(H))$ is an invariant under planar isotopy, the Reidemeister-type moves, the 0–2 move and the MP–moves. Here the Reidemeister-type moves are nothing but the *symmetry moves* in [16], the 0–2 move is a special case of the *colored* $(0,2)$ *moves* and the MP–moves are obtained by the *colored Pachner* $(2,3)$ *moves* and the colored $(0,2)$ moves. Even so, since the frameworks are slightly different, we give another proof in this section. The invariance under the CP–move was not observed in [16], and we give the proof in the following section, where we use a framework of tensor networks.

Proof of Theorem 4.2 **Invariance under planar isotopy and the Reidemeister-type moves** It is obvious from the construction that $Z(\Gamma, \mathcal{H}(H))$ is an invariant under planar isotopy and the Reidemeister-type moves described in the Figure 5, left.

Invariance under the 0–2 move Let us calculate the local tensor associated to the right-hand side of Figure 5, right:



So we need to show $T_2\overline{T}_2\overline{T}_1T_1 = \epsilon\otimes 1$. Let $T = \sum_i(\epsilon\otimes e_i)\otimes(e^i\otimes 1)$ and $\overline{T} = \sum_j(\epsilon\otimes S(e_j))\otimes(e^j\otimes 1)$. Then

$$T_2\overline{T}_2\overline{T}_1T_1 = \sum_{i,j}(e^i\otimes 1)(e^j\otimes 1)(\epsilon\otimes S(e_j))(\epsilon\otimes e_i) = \sum_{i,j} e^ie^j\otimes S(e_j)e_i. \tag{5-1}$$

Identifying $H^*\otimes H$ with $\mathrm{End}(H)$, the right side of (5-1) becomes $x\mapsto S(x_{(2)})x_{(1)}$. Since S is assumed to be involutive, $S(x_{(2)})x_{(1)} = \epsilon(x)1$. Thus $T_2\overline{T}_2\overline{T}_1T_1 = \epsilon\otimes 1$.

Invariance under the MP moves There are 16 MP moves. We write the calculations in Figures 11–12. Namely, we need to check the following 16 equalities:

MP 1.1	$T_{23}T_{13} = \overline{T}_{21}T_{13}T_{21}$,	MP 2.1	$\overline{T}_{32}\overline{T}_{31} = T_{12}\overline{T}_{31}\overline{T}_{12}$,
MP 1.2	$\overline{T}_{13}\overline{T}_{23} = \overline{T}_{21}\overline{T}_{13}T_{21}$,	MP 2.2	$T_{31}T_{32} = T_{12}T_{31}\overline{T}_{12}$,
MP 1.3	$\overline{T}_{23}T_{13} = T_2'T_1\overline{T}_2\otimes\overline{T}_1T_1'\otimes T_2$,	MP 2.3	$T_{32}\overline{T}_{31} = \overline{T}_1'\overline{T}_2T_1\otimes T_2\overline{T}_2'\otimes\overline{T}_1$,
MP 1.4	$\overline{T}_{13}T_{23} = T_2\overline{T}_1'\overline{T}_2\otimes\overline{T}_1T_1\otimes\overline{T}_2'$,	MP 2.4	$T_{31}\overline{T}_{32} = \overline{T}_1T_2'T_1\otimes T_2\overline{T}_2\otimes T_1'$,
MP 3.1	$T_{23}\overline{T}_{31} = \overline{T}_{31}T_{23}T_{21}$,	MP 4.1	$T_{31}\overline{T}_{23} = \overline{T}_{21}\overline{T}_{23}T_{31}$,
MP 3.2	$\overline{T}_{23}\overline{T}_{31} = \overline{T}_{31}\overline{T}_{21}\overline{T}_{23}$,	MP 4.2	$T_{31}T_{23} = T_{23}T_{21}T_{31}$,
MP 3.3	$T_{23}T_{31} = \overline{T}_2T_2'\otimes T_1\overline{T}_1\otimes T_1'T_2$,	MP 4.3	$\overline{T}_{31}\overline{T}_{23} = \overline{T}_2T_2\otimes T_1\overline{T}_1'\otimes\overline{T}_2'\overline{T}_1$,
MP 3.4	$\overline{T}_{23}\overline{T}_{31} = T_{21}T_{31}\overline{T}_{23}$,	MP 4.4	$\overline{T}_{31}T_{23} = T_{23}\overline{T}_{31}\overline{T}_{21}$.

We verify that each of these is equivalent to the pentagon equation (2-3). Define $\tau_{\mathcal{H}}\colon \mathcal{H}(H)^{\otimes 2}\to\mathcal{H}(H)^{\otimes 2}$ by $\tau_{\mathcal{H}}(x\otimes y) = y\otimes x$ for $x,y\in\mathcal{H}(H)$. Then, for example, if we multiply MP 1.1 by T_{21} from the left and apply $\tau_{\mathcal{H}}\otimes\mathrm{id}$, the result is exactly the pentagon equation. Similarly, we can reduce MP 1.2, MP 2.1, MP 2.2, MP 3.1, MP 3.2, MP 3.4, MP 4.1, MP 4.2, and MP 4.4 to the pentagon equation.

For the other six equalities, define the map $\mathcal{S}\colon\mathcal{H}(H)\to\mathcal{H}(H)$ by $\mathcal{S}(f\otimes a) = S(f)\otimes S(a)$, where S is the antipode of the Hopf algebra H. Then, for example, we can transform MP 1.3 into MP 1.1 by applying $\mathrm{id}\otimes\mathcal{S}\otimes\mathrm{id}$ to both sides as follows:

$$\begin{aligned}
&(\mathrm{id}\otimes\mathcal{S}\otimes\mathrm{id})(\overline{T}_{23}T_{13}) = (\mathrm{id}\otimes\mathcal{S}\otimes\mathrm{id})(T_2'T_1\overline{T}_2\otimes\overline{T}_1T_1'\otimes T_2)\\
&\iff \sum_{i,j}(\epsilon\otimes e_j)\otimes(\epsilon\otimes S^2(e_i))\otimes(e^ie^j\otimes 1) = \sum_{i,j,k}(e^j(e_{i(1)}\rightharpoonup e^k)\otimes e_{i(2)})\otimes(\epsilon\otimes S(e_j)S^2(e_k))\otimes(e^i\otimes 1)\\
&\iff T_{23}T_{13} = \overline{T}_{21}T_{13}T_{21}.
\end{aligned}$$

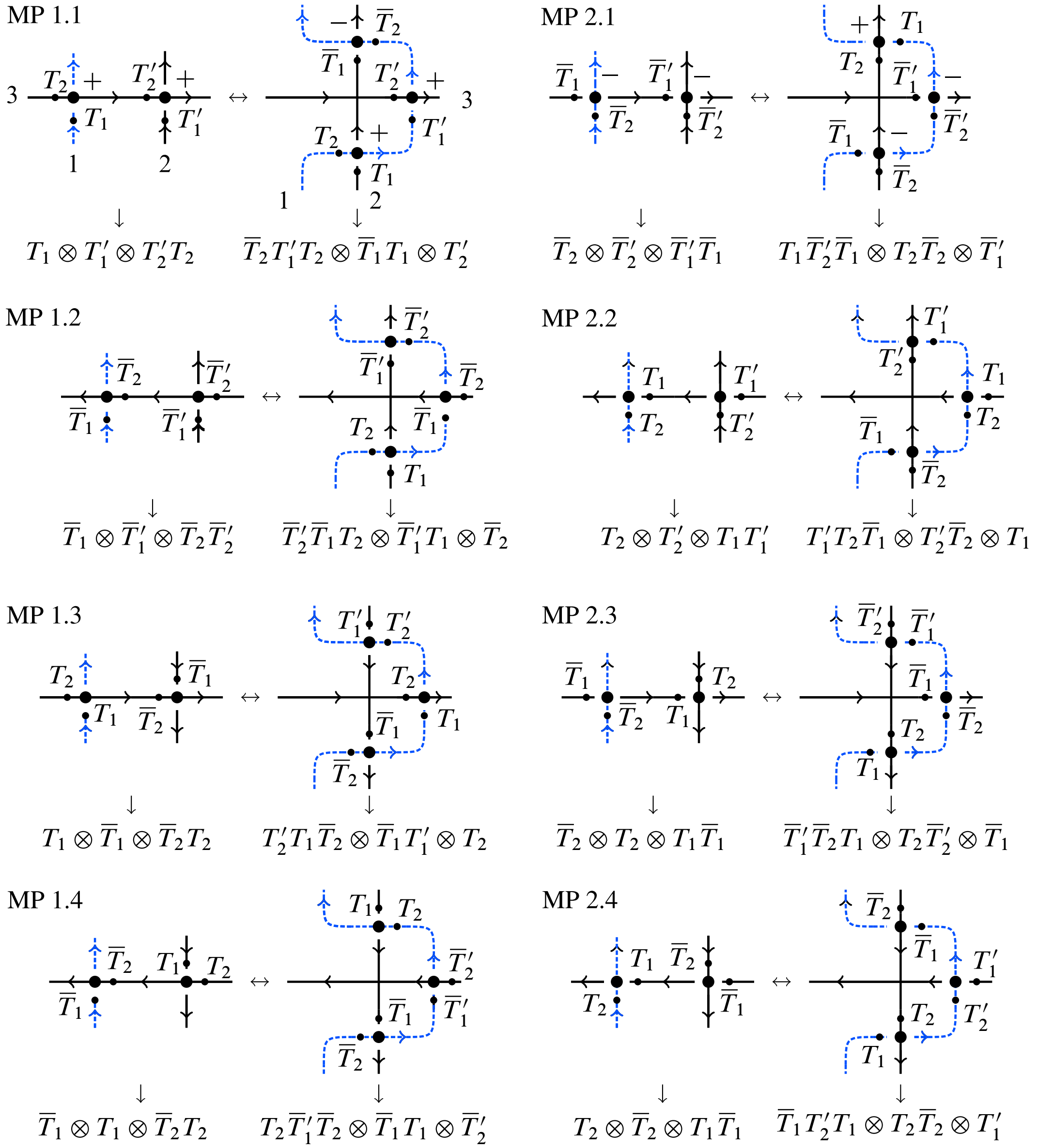


Figure 11

Here we used the involutivity of S for the last equivalence. Thus MP 1.3 is also equivalent to the pentagon equation. We can show that MP 1.4, MP 2.3, MP 2.4, MP 3.3 and MP 4.3 are also equivalent to the pentagon equation in a similar manner.

Since the pentagon equation holds in the Heisenberg double $\mathcal{H}(H)$, we conclude that Z is invariant under the MP–moves.

Invariance under the CP–move This will be proved in Proposition 6.6 using tensor networks. □

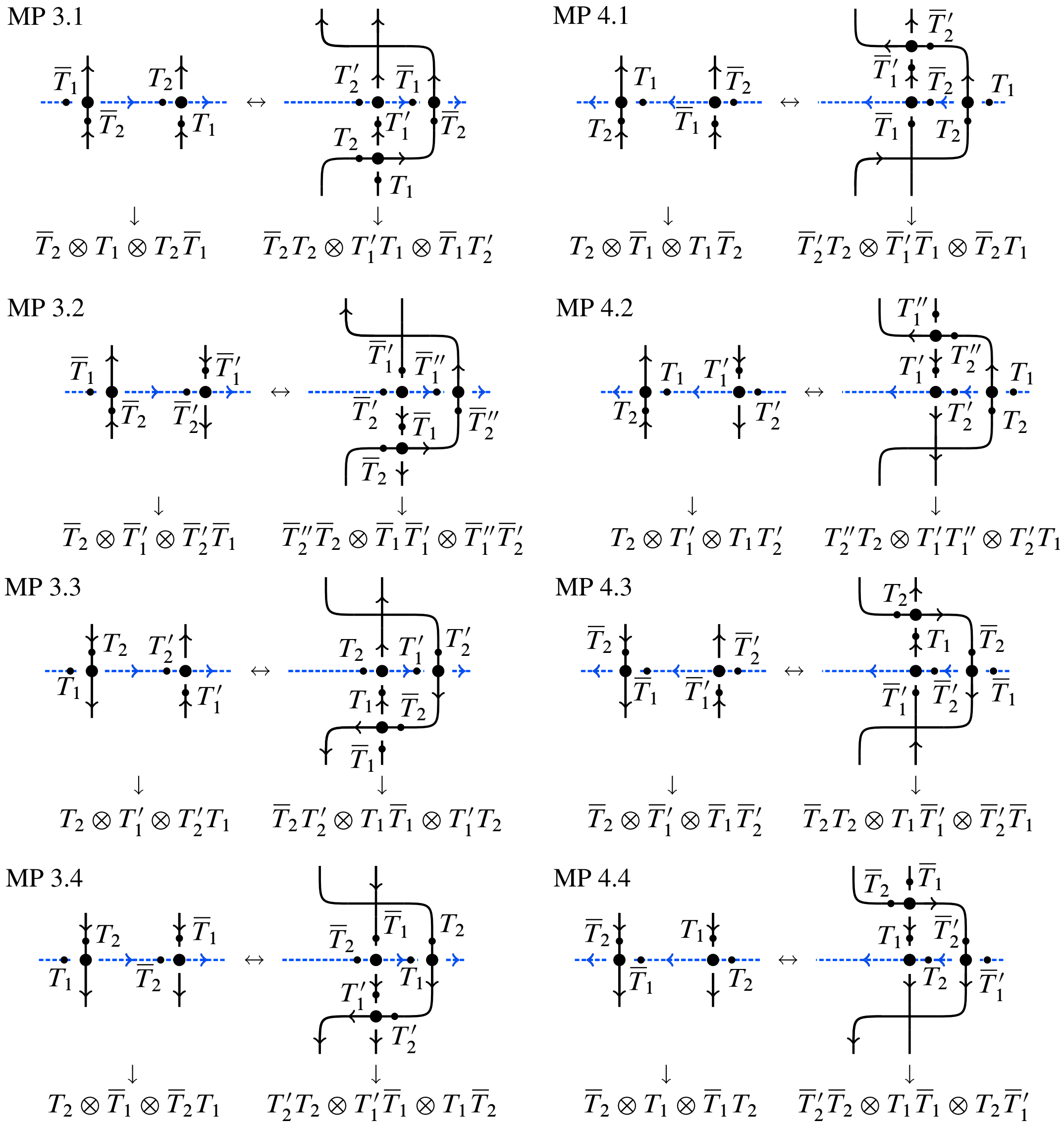


Figure 12

Remark 5.1 The assumption of unimodularity and counimodularity of H will be used only for invariance of the CP–move. Involutivity was used for the 0–2 move and the MP–moves MP 1.3, MP 1.4, MP 2.3, MP 2.4, MP 3.3 and MP 4.3. The other 10 equalities for the MP–moves do not need any restriction on Hopf algebras to hold.

Remark 5.2 In [16, Theorem 5.1], there is an error; even for an involutory Hopf algebra, J is not an invariant under the colored Pachner $(2, 3)$ move in [16, Figure 16], which cannot be obtained by rotating the allowed one. This excluded move corresponds to the MP–moves with some strands reversed, which are not actually the MP–moves.

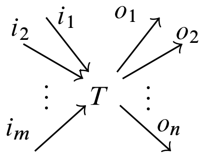


Figure 13

6 Tensor network approach

In this section, we give a quick review of tensor networks, which enable graphical calculus of tensors and linear maps. Then we reformulate the invariant using tensor networks, and prove the invariance under the CP–move.

6.1 Tensor network

A *tensor network* over a vector space V is an oriented graph which represents *a tensor labeled by the set* of open edges, where each incoming (resp. outgoing) edge labels V^* (resp. V). For example, the diagram in Figure 13 presents an (m,n) tensor $T \in (V^*)^{\otimes \mathcal{I}} \otimes V^{\otimes \mathcal{O}} = \mathrm{Hom}(V^{\otimes \mathcal{I}}, V^{\otimes \mathcal{O}})$, where $\mathcal{I} = \{i_1, \dots, i_m\}$ is the set of incoming edges and $\mathcal{O} = \{o_1, \dots, o_n\}$ is the set of outgoing edges.

One important feature of tensor networks, which makes this notion practical, is the contraction of tensors. Given two tensor networks T and S, one gets a new tensor network by connecting an outgoing edge o of T and an incoming edge i of S (see Figure 14), which represents the tensor obtained from $T \otimes S$ by contracting V_o and $(V^*)_i$.

For example, the left diagram in Figure 15 represents the composition $g \circ f$ of two maps $f, g \colon V \to V$ and the right diagram represents the trace $\sum_i f_i^i = \mathrm{tr}(f) \in \mathbb{K}$ of a map $f \colon V \to V$.

Note that a *tensor labeled by a set* $\mathcal{I}$ does not fix the order of the tensorands. More precisely, $V^{\otimes \mathcal{I}}$ is the tensor product constructed from the product $V^{\mathcal{I}} = \{v \colon \mathcal{I} \to V\}$ labeled by $\mathcal{I}$. For example, for $\mathcal{I} = \{a, b, c\}$, the linear space $V^{\otimes \mathcal{I}}$ is isomorphic to $V^{\otimes 3}$, but such an isomorphism is not canonical, ie there is not a canonical order of labeled V to be written. On the other hand, for $\mathcal{I} = \{1, 2, \dots, n\}$, we have the canonical isomorphism $V^{\otimes \mathcal{I}} \to V^{\otimes n}$, $v \mapsto v(1) \otimes \cdots \otimes v(n)$. In order to present calculations of a Hopf algebra $(H, M, 1, \Delta, \epsilon, S)$ using tensor networks over H, we fix an order of the incoming

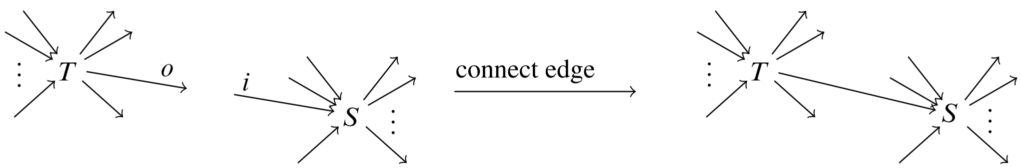


Figure 14

Figure 15

and the outgoing edges of the graphs representing multiplication $M: H^{\otimes 2} \to H$ and comultiplication $\Delta: H \to H^{\otimes 2}$, respectively, as below:

Note that the incoming edges of multiplication are counted in counterclockwise order, and the outgoing edges of comultiplication are counted in clockwise order. Using this notation, for example, the axioms of Hopf algebras are represented in Figure 16, where we use the subscript 1 to make the order clear. This subscript will be omitted sometimes when it is obvious.

6.2 Reformulation of invariant

Assume $H = (H, M, 1, \Delta, \epsilon, S)$ is a finite-dimensional involutory unimodular counimodular Hopf algebra.

An *o–tangle* is an oriented virtual tangle diagram in $[0,1]^2$ such that each boundary point is on the bottom, $[0,1]\times\{0\}$, or on the top, $[0,1]\times\{1\}$. For finite sequences ε and ε' of $\pm$, an $(\varepsilon, \varepsilon')$ o–tangle is an o–tangle having boundary points on the bottom and top compatible to ε and ε', respectively, where compatible means that if an edge is oriented upwards (resp. downwards) then it is connected to $+$ (resp. $-$).

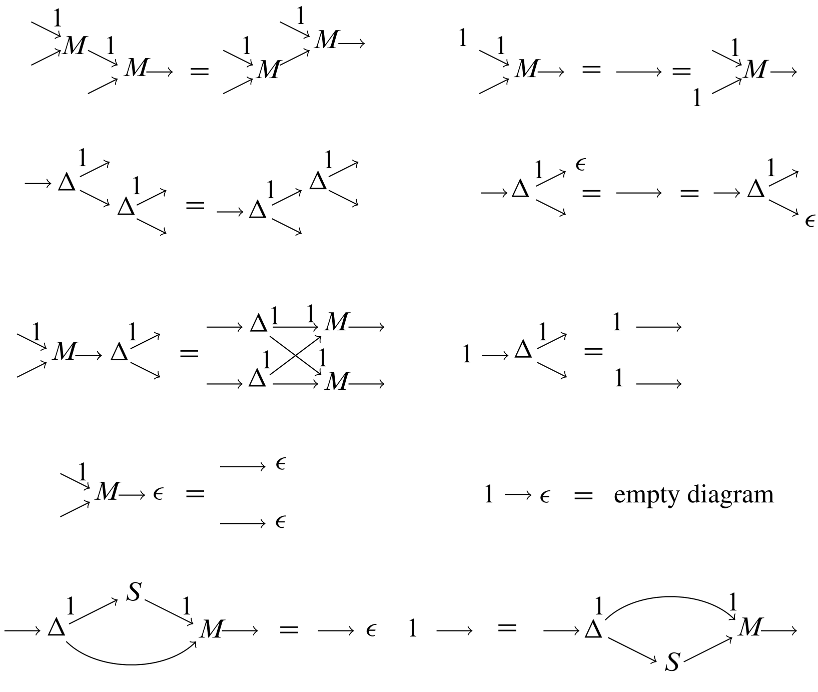
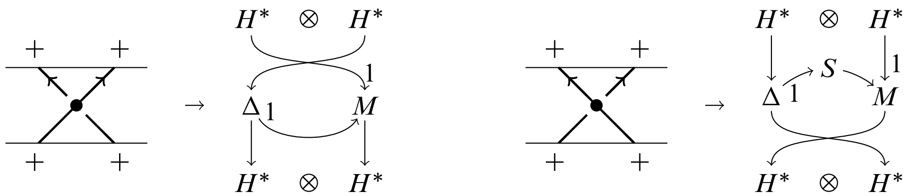


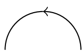
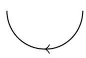
Figure 16

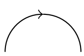
Let $\mathcal{T}_o$ be the *category of o–tangles*, where objects are finite sequences of $\pm$ including the empty sequence $\varnothing$, and morphisms $\mathcal{T}_o(\varepsilon,\varepsilon')$ from ε to ε' are isotopy classes of $(\varepsilon,\varepsilon')$ o–tangles. As usual, $\mathcal{T}_o$ is a strict monoidal category with the unit object $\varnothing$, and the composition $\Gamma'\circ\Gamma$ of $(\varepsilon,\varepsilon')$ o–tangle Γ and $(\varepsilon',\varepsilon'')$ o–tangle Γ' is obtained by connecting the ε'-type boundary points on the top of Γ to these on the bottom of Γ'. We can construct a monoidal functor $Z(\,*\,;H)$ from the category of o–tangles $\mathcal{T}_o$ to the category of finite-dimensional vector spaces $\mathrm{Vect}_{\mathbb{K}}$ as follows.

For the object $+$ (resp. $-$), we set $Z(+;H):=H^*$ (resp. $Z(-;H):=H$). For a sequence $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)$ in $\pm$, let H^{ε} denote $H^{\varepsilon_1}\otimes\cdots\otimes H^{\varepsilon_n}$ with $H^+=H^*$ and $H^-=H$. To a given $(\varepsilon,\varepsilon')$ o–tangle Γ, we associate a tensor network over H, which represents a linear map $Z(\Gamma;H)\in\mathrm{Hom}(H^{\varepsilon},H^{\varepsilon'})$ as follows; we replace each positive (resp. negative) crossing of Γ with the tensor network as in the left (resp. right) picture below, and then connect the boundary points of these tensor networks following the strands of Γ. The boundary edges on the bottom and top of the resulting tensor network are counted from the left, and an input element to $Z(\Gamma;H)\in\mathrm{Hom}(H^{\varepsilon},H^{\varepsilon'})$ is a tensor $T\in H^{\varepsilon}$ labeled by the bottom edges. Then $Z(\Gamma;H)$ sends T to the tensor $Z(\Gamma;H)(T)\in H^{\varepsilon'}$ (labeled by the top edges) which is obtained by concatenating T to the bottom edges of the tensor network. Note that the corresponding strands of the tensor network are oriented in the opposite direction to these of Γ:



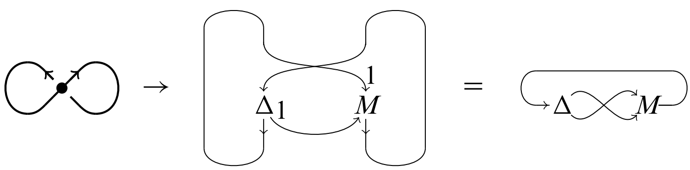
Note that a maximum point plays the role of evaluation map, and a minimum point plays the role of coevaluation map:

 $: H^*\otimes H\to\mathbb{K},\quad f\otimes a\mapsto f(a),$  $:\mathbb{K}\to H\otimes H^*,\quad 1\mapsto\sum e_i\otimes e^i,$

 $: H\otimes H^*\to\mathbb{K},\quad a\otimes f\mapsto f(a),$  $:\mathbb{K}\to H^*\otimes H,\quad 1\mapsto\sum e^i\otimes e_i.$

Since a closed normal o–graph Γ is an $(\varnothing,\varnothing)$ o–tangle, it is sent to an endomorphism $Z(\Gamma;H)$ of $\mathbb{K}$, which is represented by a scalar in $\mathbb{K}$. By abusing notation we also denote the scalar by $Z(\Gamma;H)\in\mathbb{K}$.

Example 6.1 For a closed normal o–graph Γ for S^3, which is an $(\varnothing,\varnothing)$ o–tangle as in the left picture below, the resulting tensor network $Z(\Gamma;H)$ is the right picture below.

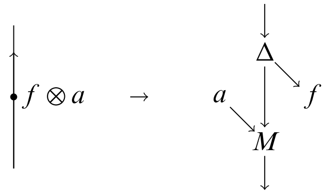


Thus $Z(\Gamma;H)=\mathrm{tr}(M\circ\tau\circ\Delta)\in\mathbb{K}$, where $\tau(x\otimes y)=y\otimes x$.

Let Γ be a closed normal o–graph and $Z(\Gamma;\mathcal{H}(H))$ the invariant defined in Section 4.

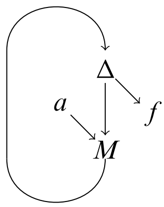
Proposition 6.2 $$Z(\Gamma; \mathcal{H}(H)) = Z(\Gamma; H).$$

Proof First, let us consider an oriented strand with a bead $f \otimes a \in \mathcal{H}(H)$ as in the left picture below. We will think of this strand with a bead as the action of $f \otimes a$ on the Fock space $F(H^*)$, which was given by $\phi(f \otimes a)\colon g \mapsto f(a \rightharpoonup g)$ for $g \in H^*$, where $f(a \rightharpoonup g)\colon x \mapsto f(x_{(1)})g(x_{(2)}a)$ for $x \in H$. Graphically this map $\phi(f \otimes a) \in \mathrm{Hom}(H^*, H^*)$ can be represented by the tensor network in the right picture below.

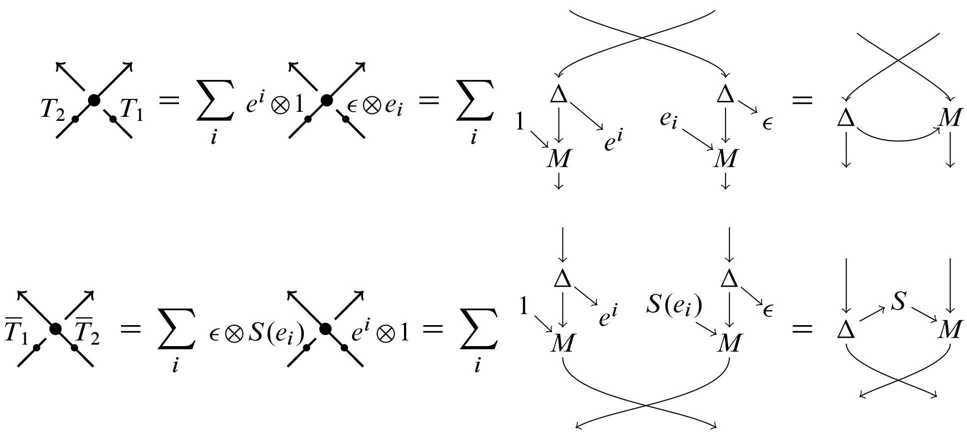


If there are multiple beads, we replace each one with the above tensor network. Since each map is an action, the result is well defined under the move of Figure 8, right.

Then, let us consider the oriented closed curve with a bead $f \otimes a \in \mathcal{H}(H)$. Recall that $\chi_{\mathrm{Fock}}(f \otimes a)$ is the trace of the linear map defined by the action of $f \otimes a$. As remarked in Section 6.1, in terms of tensor networks, taking a trace is just connecting the incoming edge with the outgoing edge. Thus

$\chi_{\mathrm{Fock}}(f \otimes a) =$ 

Finally, let Γ be a closed normal o–graph. Replacing its vertices as in Figure 10 and sliding beads, we get a single bead $J_\Gamma \in \mathcal{H}(H)$, and the invariant $Z(\Gamma, \mathcal{H}(H))$ was defined as $\chi_{\mathrm{Fock}}(J_\Gamma)$. Here, before the sliding process, we replace each bead with a corresponding tensor network as above, and compare the result to $Z(\Gamma; H)$. Since the beads only appear at the vertices of Γ, we just need to look at the associated tensor network for these beads:

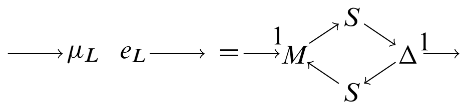


These are the same tensor networks associated with vertices in the definition of $Z(\Gamma; H)$. □

6.3 Invariance under CP–move

We prove the invariance of $Z(\Gamma; H)$ under the CP–move. Refer to Section 2 for the definition of integrals. Let $e_L \in H$ and $\mu_L \in H^*$ be a left cointegral and a left integral satisfying $\mu_L(e_L) = 1$. In terms of tensor networks, we have

Lemma 6.3 [12, Lemma 3.3] *The following equality holds in any Hopf algebra*:



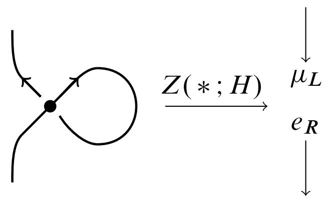
Set $e_R := S(e_L)$. Since S is an antialgebra map, e_R is a (nonzero) right cointegral.

Lemma 6.4 [15, Theorem 10.5.4] *For a finite-dimensional involutory counimodular Hopf algebra*,

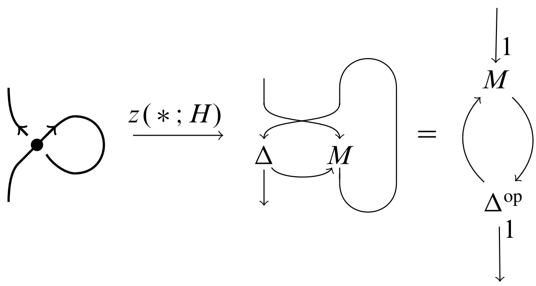
$$\Delta^{\mathrm{op}}(e_R) = \Delta(e_R),$$

where $\Delta^{\mathrm{op}}(x) := x_{(2)} \otimes x_{(1)}$.

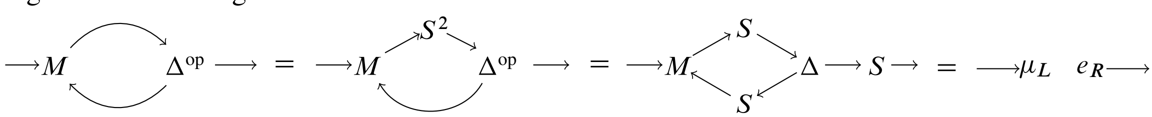
Lemma 6.5 *Let H be an involutory Hopf algebra. Then*



Proof Replacing the vertex of the o–tangle with the corresponding tensor network, we have



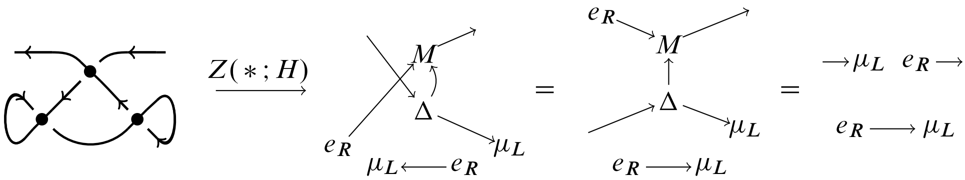
Using Lemma 6.3 we get



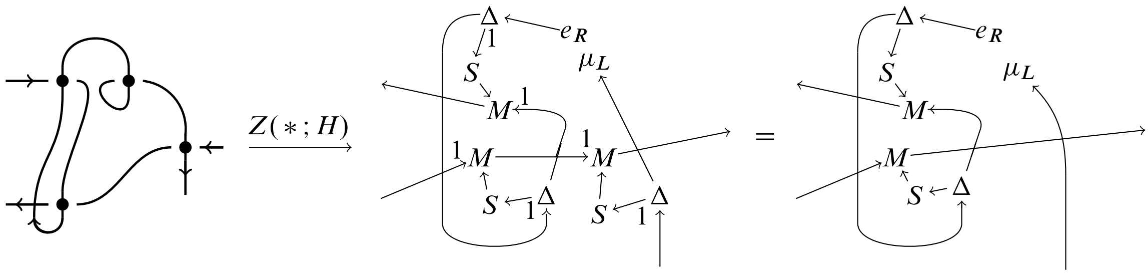
□

Proposition 6.6 *Let H be a finite-dimensional involutory unimodular counimodular Hopf algebra. Then $Z(\Gamma; H)$ is an invariant under the CP–move in Figure 7.*

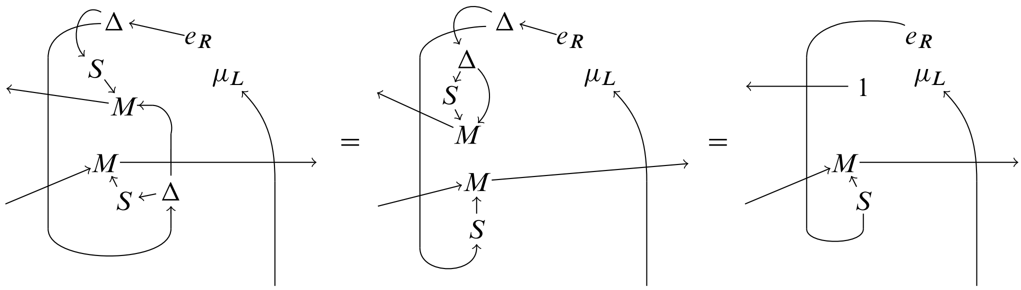
Proof First, we evaluate the left-hand side of the CP–move. Using Lemma 6.5, the twists in the closed normal o–graph can be replaced by integrals, and thus



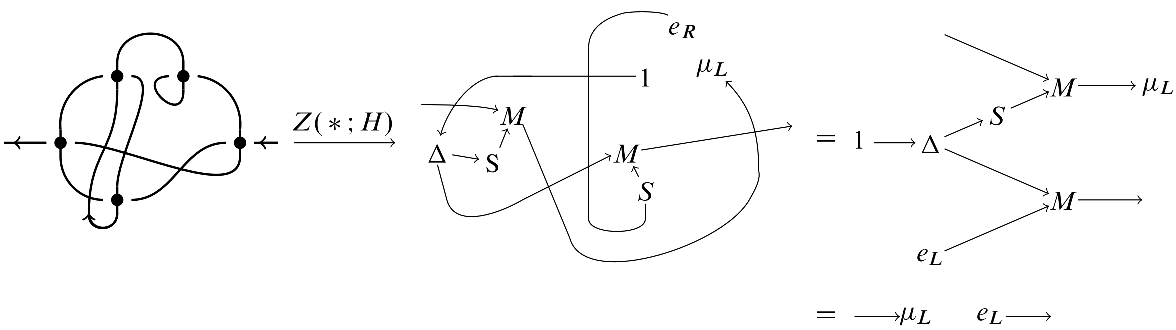
Next, we evaluate a part of the right-hand side of the CP–move:



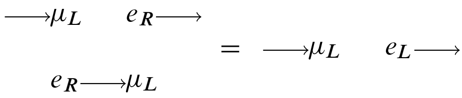
We used the definition of left integral μ_L for the equality above. From Lemma 6.4, e_R is cyclic. Thus this equals



The first equality follows from coassociativity of the comultiplication, and the second follows from the property of the antipode. Thus right-hand side of the CP–move becomes



Finally, we show that the following equality holds for involutory unimodular counimodular Hopf algebra:



Since H is unimodular, $e_R = k e_L$ for some nonzero $k \in \mathbb{K}$. Since $\mu_L(e_L)$ is 1, $\mu_L(e_R) = k$ and the left-hand side of the above tensor network is the same as right-hand side times k^2. Applying S^2 to e_L and using the fact that S is involutive, we see that $k^2 = 1$, and the above equality follows. □

7 Properties

We continue to assume that H is a finite-dimensional involutory unimodular counimodular Hopf algebra over $\mathbb{K}$. For a closed oriented 3–manifold M and a closed normal o–graph Γ representing M, set $Z(M;H) := Z(\Gamma;H)$.

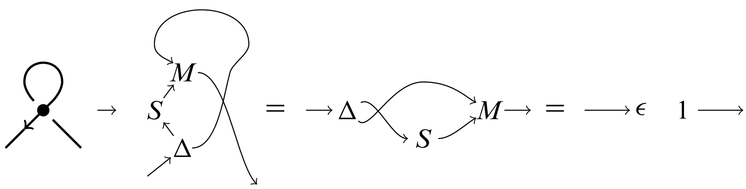
7.1 Connected sum formula

For closed oriented 3–manifolds M and N, let $M \# N$ be their connected sum.

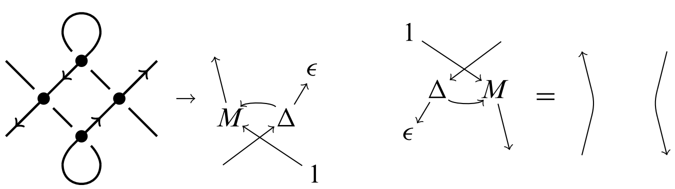
Proposition 7.1 $$Z(M \# N;H) = Z(M;H)Z(N;H).$$

Proof Let Γ_M and Γ_N be closed normal o–graphs representing M and N, respectively. Let $\Gamma_M \# \Gamma_N$ be the connected sum of closed normal o–graphs defined in Figure 17.

In [10], Y Koda showed that for two closed normal o–graphs Γ_M and Γ_N, the 3–manifold represented by $\Gamma_M \# \Gamma_N$ is $M \# N$. We show the assertion by comparing the tensor networks for the closed normal o–graphs in the left and right-hand sides of the Figure 17. Note that



Thus



which implies the assertion. □

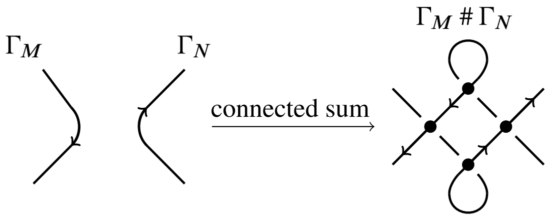


Figure 17

7.2 Group algebra

We show that the invariant $Z(M;\mathbb{C}[G])$ with the group algebra $\mathbb{C}[G]$ of a finite group G counts the number $|\mathrm{Hom}(\pi_1 M, G)|$ of group homomorphisms from the fundamental group $\pi_1 M$ of M to G. The proof essentially follows the line of [7].

The algebra $\mathbb{C}[G]$ has a canonical basis given by $\{g\}_{g\in G}$ and the Hopf algebra structure is given by $\Delta(g) = g \otimes g$, $\epsilon(g) = 1$ and $S(g) = g^{-1}$. Note that $\mathbb{C}[G]$ is involutory, unimodular and counimodular. The dual group algebra $\mathbb{C}(G) := \mathbb{C}[G]^*$ has the dual basis $\{\delta_g\}_{g\in G}$, and the dual Hopf algebra structure is given by

$$\delta_g \cdot \delta_h = \delta_{g,h}\delta_g, \quad 1_{\mathbb{C}(G)} = \sum_{g\in G} \delta_g, \quad \Delta(\delta_g) = \sum_{hk=g} \delta_h \otimes \delta_k, \quad \epsilon(\delta_g) = \delta_g(e), \quad S(\delta_g) = \delta_{g^{-1}},$$

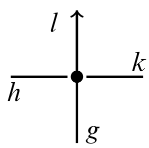
where $\delta_{g,h} \in \{0,1\}$ is 1 if $g = h$ and 0 otherwise, and e is the unit of G.

The left action of $x \in G \subset \mathbb{C}[G]$ on $\delta_g \in \mathbb{C}(G)$ is given by

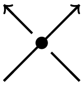
$$x \rightharpoonup \delta_g = \delta_{gx^{-1}}.$$

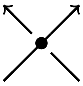
Proposition 7.2 *For a closed oriented* 3*–manifold* M *and a finite group* G, *we have* $Z(M;\mathbb{C}[G]) = |\mathrm{Hom}(\pi_1 M, G)|$.

Proof Let Γ be a closed normal o–graph representing M. In [10; 9], Koda gave an explicit formula for the fundamental group $\pi_1 M$ in terms of the closed normal o–graph Γ. Let E be the set of all edges of Γ. We consider the group generated by E and the relation set R consisting of $g = l$ and $hg = k$ for the edges g, l, h and k around each vertex, as below. Then the resulting group $\langle E \mid R\rangle$ is isomorphic to the fundamental group $\pi_1 M$ of M, and the number $|\mathrm{Hom}(\pi_1 M, G)|$ is equal to the number of edge colorings $c\colon E \to G$ such that $c(R)$ holds in G.



We show that the invariant $Z(M;\mathbb{C}[G])$ indeed counts such edge colorings. As we explained in Section 6.2, each vertex of a closed normal o–graph can be treated as a linear map between $\mathbb{C}(G)^{\otimes 2}$:

 $: \mathbb{C}(G)\otimes\mathbb{C}(G)\to\mathbb{C}(G)\otimes\mathbb{C}(G), \quad \delta_g\otimes\delta_h \mapsto T_1\triangleright\delta_h\otimes T_2\triangleright\delta_g$

 $: \mathbb{C}(G)\otimes\mathbb{C}(G)\to\mathbb{C}(G)\otimes\mathbb{C}(G), \quad \delta_h\otimes\delta_g \mapsto \overline{T}_2\triangleright\delta_g\otimes \overline{T}_1\triangleright\delta_h$

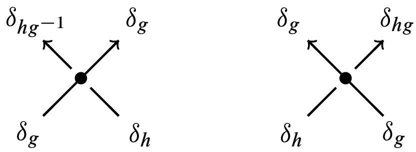
Here

$$T_1\triangleright\delta_h\otimes T_2\triangleright\delta_g=\sum_{x\in G}(\epsilon\otimes x)\triangleright\delta_h\otimes(\delta_x\otimes e)\triangleright\delta_g=\sum_{x\in G}x\rightharpoonup\delta_h\otimes\delta_x\cdot\delta_g=\sum_{x\in G}\delta_{hx^{-1}}\otimes\delta_{x,g}\delta_g=\delta_{hg^{-1}}\otimes\delta_g$$

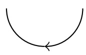
and

$$\overline{T}_2\triangleright\delta_g\otimes\overline{T}_1\triangleright\delta_h=\sum_{x\in G}(\delta_x\otimes e)\triangleright\delta_g\otimes(\epsilon\otimes x^{-1})\triangleright\delta_h=\sum_{x\in G}\delta_x\cdot\delta_g\otimes x^{-1}\rightharpoonup\delta_h$$
$$=\sum_{x\in G}\delta_{x,g}\delta_g\otimes\delta_{hx}=\delta_g\otimes\delta_{hg}.$$

We draw these maps as follows:



Note that the subscripts of δ give nothing but an edge coloring (around the vertices) as desired. Furthermore, to connect those vertices by strands means that we insert maximum and minimum points among them. Recall from Section 6.2 that maximum and minimum points correspond to evaluations and coevaluations, respectively. In the present case they are the maps shown below, which means after all we sum up all edge colorings:

 $:\mathbb{C}(G)\otimes\mathbb{C}[G]\to\mathbb{C}, \quad \delta_g\otimes h\mapsto\delta_{g,h},$  $:\mathbb{C}\to\mathbb{C}[G]\otimes\mathbb{C}(G), \quad 1\mapsto\sum g\otimes\delta_g,$

 $:\mathbb{C}[G]\otimes\mathbb{C}(G)\to\mathbb{C}, \quad h\otimes\delta_g\mapsto\delta_{g,h},$  $:\mathbb{C}\to\mathbb{C}(G)\otimes\mathbb{C}[G], \quad 1\mapsto\sum \delta_g\otimes g.$ □

References

[1] **S Baaj**, **G Skandalis**, *Unitaires multiplicatifs et dualité pour les produits croisés de C^*–algèbres*, Ann. Sci. École Norm. Sup. 26 (1993) 425–488 MR Zbl

[2] **J W Barrett**, **B W Westbury**, *Invariants of piecewise-linear 3–manifolds*, Trans. Amer. Math. Soc. 348 (1996) 3997–4022 MR Zbl

[3] **R Benedetti**, **C Petronio**, *Branched standard spines of 3–manifolds*, Lecture Notes in Math. 1653, Springer (1997) MR Zbl

[4] **M Hennings**, *Invariants of links and 3–manifolds obtained from Hopf algebras*, J. Lond. Math. Soc. 54 (1996) 594–624 MR Zbl

[5] **N Jacobson**, *Lie algebras*, Interscience Tracts Pure Appl. Math. 10, Interscience, New York (1962) MR Zbl

[6] **R M Kashaev**, *The Heisenberg double and the pentagon relation*, Algebra i Analiz 8 (1996) 63–74 MR Zbl

[7] **L H Kauffman**, **D E Radford**, *Invariants of 3–manifolds derived from finite-dimensional Hopf algebras*, J. Knot Theory Ramifications 4 (1995) 131–162 MR Zbl

[8] **M Kim**, *Homology theory valued in the category of bicommutative Hopf algebras*, preprint (2020) arXiv 2005.04652

[9] **Y Koda**, *Spines, Heegaard splittings and the Reidemeister–Turaev torsion*, Tokyo J. Math. 30 (2007) 417–439 MR Zbl

[10] **Y Koda**, *O–graphic study of closed 3–manifolds*, conference paper, Hakone Seminar (2010) Available at `http://hakone-seminar.com/Koda2010.pdf`

[11] **G Kuperberg**, *Involutory Hopf algebras and 3–manifold invariants*, Int. J. Math. 2 (1991) 41–66 MR Zbl

[12] **G Kuperberg**, *Noninvolutory Hopf algebras and 3–manifold invariants*, Duke Math. J. 84 (1996) 83–129 MR Zbl

[13] **J-H Lu**, *Hopf algebroids and quantum groupoids*, Int. J. Math. 7 (1996) 47–70 MR Zbl

[14] **S Majid**, **A Pachoł**, *Digital quantum groups*, J. Math. Phys. 61 (2020) art. id. 103510 MR Zbl

[15] **D E Radford**, *Hopf algebras*, Series on Knots and Everything 49, World Sci., Hackensack, NJ (2012) MR Zbl

[16] **S Suzuki**, *The universal quantum invariant and colored ideal triangulations*, Algebr. Geom. Topol. 18 (2018) 3363–3402 MR Zbl

[17] **V G Turaev**, **O Y Viro**, *State sum invariants of 3–manifolds and quantum 6j–symbols*, Topology 31 (1992) 865–902 MR Zbl

Department of Mathematics, Tohoku University
Sendai, Japan

Department of Mathematical and Computing Science, Tokyo Institute of Technology
Tokyo, Japan

Department of Mathematics, Tohoku University
Sendai, Japan

`matei.mihalache.q3@dc.tohoku.ac.jp`, `sakie@c.titech.ac.jp`, `yujiterashima@tohoku.ac.jp`

Received: 4 May 2022 Revised: 4 December 2022

Algebraic & Geometric Topology **24**:7 (2024) 3693–3717
DOI: 10.2140/agt.2024.24.3693
Published: 9 December 2024

A closed ball compactification of a maximal component via cores of trees

Giuseppe Martone
Charles Ouyang
Andrea Tamburelli

We show that, in the character variety of surface group representations into the Lie group $\mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$, the compactification of the maximal component introduced by the second author is a closed ball upon which the mapping class group acts. We study the dynamics of this action. Finally, we describe the boundary points geometrically as $(\overline{A_1 \times A_1}, 2)$–valued mixed structures.

53C43, 57K20

1 Introduction

A recurring theme in higher Teichmüller theory is to relate surface group representations into higher-rank Lie groups with geometric objects. Taking its cue from classical Teichmüller theory, one is often interested in studying the degeneration of these associated geometric objects when the representation leaves all compact sets in the character variety. The celebrated Thurston compactification of Teichmüller space regards Fuchsian representations as marked hyperbolic metrics, where degenerating families of hyperbolic metrics subconverge to projectivized measured laminations. One key aspect of this compactification is that it is a closed ball upon which the mapping class group acts. In years following, there have been numerous different perspectives of the Thurston compactification, using a variety of methods, topological, geometric, analytic and algebraic (see [Bonahon 1988; Bestvina 1988; Paulin 1988; Wolf 1989; Morgan and Shalen 1984; Brumfiel 1988]).

When the Lie group $\mathrm{PSL}(2,\mathbb{R})$ is replaced with a higher-rank one, the relevant geometric object is not always immediately clear. In rank 2 however, combined work of Schoen [1993], Labourie [2017], Loftin [2001], Collier [2016], Alessandrini and Collier [2019], and Collier, Tholozan and Toulisse [Collier et al. 2019] provides a geometric interpretation to representations in the various distinguished components of the relevant character variety. These components are usually maximal components or Hitchin components, which maximize a topological quantity, the Toledo invariant, or contain a deformation of the classical Teichmüller space. Parreau [2012] compactifies them by attaching at infinity surface group actions on a Euclidean building.

This paper will primarily be concerned with the rank-2 semisimple split Lie group $G = \mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$. The product structure of G makes our study more amenable towards techniques from classical Teichmüller theory. For S a closed, orientable, smooth surface of genus $g > 1$, work of Goldman [1988] shows the connected components of the character variety $\chi(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))$ are determined by the Euler number. In particular, the distinguished component with maximal Euler number of $2g-2$ is the Teichmüller space $\mathrm{Teich}(S)$. If we denote the character variety for $G = \mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$ by $\chi(\pi_1(S), G)$, then the connected components are merely products of the connected components of $\chi(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))$. The maximal component $\mathrm{Max}(S, G)$ of $\chi(\pi_1(S), G)$ is the collection of conjugacy classes of pairs of representations, each of which is a Fuchsian representation. Hence $\mathrm{Max}(S) := \mathrm{Max}(S, \mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R}))$ is the product of two copies of Teichmüller space.

Elements in the component $\mathrm{Max}(S)$ have a number of related geometric interpretations. Schoen [1993] has shown these representations correspond to equivariant minimal Lagrangians in $\mathbb{H}^2 \times \mathbb{H}^2$. At the same time, the group $G = \mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$ is the isometry group of AdS^3, and Mess [2007] has shown the holonomy representations of GHMC-AdS^3 manifolds are precisely the ones in $\mathrm{Max}(S)$. Krasnov and Schlenker [2007] have shown to each GHMC-AdS^3 manifold there is a unique equivariant space-like maximal surface, whose image under the Gauss map is the aforementioned minimal Lagrangian.

In seeking a compactification of $\mathrm{Max}(S)$ via degeneration of geometric objects, the second author in his thesis [Ouyang 2023] showed the natural limits to the minimal Lagrangians were given by cores of $\mathbb{R}$–trees dual to measured laminations. These are topologically and group-theoretically defined distinguished subcomplexes of the product of two trees, where some parts are two-dimensional and the remaining parts are one-dimensional. Denote by $\mathrm{Core}(\mathcal{T}, \mathcal{T})$, the space of cores in the product of trees dual to measured laminations. Observe that there is a natural $\mathbb{R}^+$–action on $\mathrm{Core}(\mathcal{T}, \mathcal{T})$ and denote by $\mathbb{P}\,\mathrm{Core}(\mathcal{T}, \mathcal{T})$ the resulting projectivization. We equip $\mathrm{Max}(S)$ and $\mathbb{P}\,\mathrm{Core}(\mathcal{T}, \mathcal{T})$ with the equivariant Gromov–Hausdorff topology. One natural question one might ask is what exactly is the topology of the resulting compactification. Our first main result is the following.

Theorem A *The disjoint union*

$$\mathfrak{B} = \mathrm{Max}(S) \sqcup \mathbb{P}\,\mathrm{Core}(\mathcal{T}, \mathcal{T})$$

is homeomorphic to a closed ball of dimension $12g-12$.

More precisely, we will show that the interior of $\mathfrak{B}$ can be identified with $\mathrm{Teich}(S) \times \mathrm{Teich}(S)$ and its boundary with $\mathbb{P}(\mathcal{MF}(S) \times \mathcal{MF}(S))$. A point in $\mathfrak{B}$ will thus be represented by a pair (x_1, x_2), where x_1 and x_2 are either both marked hyperbolic structures or both measured foliations up to simultaneous projective equivalence.

The key new contribution of Theorem A is the description of the topology of a compactification of a higher Teichmüller space. Even in the case of Teichmüller space, Thurston's original proof requires the construction of charts in order to show that the compactified space has the structure of a manifold with boundary and then uses the Schönflies theorem (see [Fathi et al. 2012, pages 162–164]). We overcome these difficulties in proving Theorem A by considering a more analytic approach inspired by the compactification of Teichmüller space using harmonic maps in [Wolf 1989]: we naturally identify $\mathrm{Max}(S)$ with a unit ball in a vector space of pairs of holomorphic quadratic differentials and $\mathbb{P}\,\mathrm{Core}(\mathcal{T}, \mathcal{T})$ with its boundary. To the best of our knowledge, this is the first example of a higher Teichmüller component of a closed surface that is compactified to a closed ball.

It is not too difficult to see from the construction of this compactification that the action of the mapping class group extends continuously to the boundary. Following Thurston, we study the action of the mapping class group $\mathrm{MCG}(S)$ on our compactification $\mathfrak{B}$.

Proposition 1.1 *Suppose $\phi \in \mathrm{MCG}(S)$ and $\phi(x) = x$ for some $x = (x_1, x_2) \in \mathfrak{B}$, where $\mathfrak{B}$ is as defined in Theorem A.*

1. *If ϕ is periodic, then x_1 and x_2 are any two points fixed by ϕ in the Thurston compactification of Teichmüller space such that $(x_1, x_2) \in \mathfrak{B}$.*
2. *If ϕ is pseudo-Anosov, then $(x_1, x_2) \in \partial\mathfrak{B}$ and $x_1 = 0$, or $x_2 = 0$ or $x_1 = x_2$.*

The action of the mapping class group appears to be more interesting if we consider its action on a natural quotient of $\mathfrak{B}$. In fact, given a maximal representation ρ, there is a unique equivariant minimal Lagrangian $\widetilde{\Sigma}_\rho$ in $\mathbb{H}^2 \times \mathbb{H}^2$. The induced metric on $\widetilde{\Sigma}_\rho$ descends to a negatively curved Riemannian metric on S. We denote by $\mathrm{Ind}(S)$ the space of such metrics. It turns out that $\mathrm{Ind}(S) = \mathrm{Max}(S)/S^1$, since there is an S^1–family of maximal representations with intrinsically isometric equivariant minimal Lagrangians. (However, these minimal Lagrangians are not extrinsically isometric in $\mathbb{H}^2 \times \mathbb{H}^2$: their second fundamental form, which is completely determined by a holomorphic quadratic differential on S, differs under rotation; see [Ouyang 2023, Proposition 4.3]). Similarly, the distance on the core of the product of two trees dual to a pair of measured laminations can be recovered from a mixed structure, that is, a hybrid geometric object on S that is in part a measured lamination and in part a finite-area flat metric induced by a meromorphic quadratic differential on subsurfaces glued along annuli. The space of projectivized mixed structures can then be identified with the boundary of $\mathrm{Ind}(S)$ in the length spectrum topology [Ouyang 2023]. The mapping class group acts on $\overline{\mathrm{Ind}(S)}$ and we prove the following:

Theorem B *Assume* $\phi \in \mathrm{MCG}(S)$ *fixes* $\mu \in \partial\overline{\mathrm{Ind}(S)}$.

(1) *If* μ *is* ***purely flat***, *ie* μ *is a mixed structure without laminar pieces*, *then* ϕ *is periodic.*

(2) *If* μ *is* ***properly mixed***, *ie* μ *is a mixed structure with at least one flat subsurface and one laminar part*, *then* ϕ *is not pseudo-Anosov.*

Note that the remaining case of μ a *purely laminar* mixed structure, in other words a genuine measured lamination on S, is handled by the Nielson–Thurston classification theorem. Theorem 5.12 will give a more detailed description of item (2) in Theorem B when ϕ is reducible. In particular, we will show that the subdivision of S induced by μ is a refinement of the one induced by ϕ if μ has no trivial parts.

The absence of a product structure for the other simple split Lie groups of rank 2 makes the study of the topology of any compactification considerably more difficult. Furthermore, for $\mathrm{PSL}(2,\mathbb{R})\times\mathrm{PSL}(2,\mathbb{R})$, quadratic differentials are intimately related to pairs of measured laminations, and for higher-order differentials, which appear for the other rank-2 cases, there are no obvious analogous topological objects. However, it is possible to describe our compactification without explicit references to $\mathbb{R}$–trees, and we conjecture this perspective can be extended to the other rank-2 Lie groups. In particular, given any Lie algebra $\mathfrak{g}$ with Cartan subalgebra $\mathfrak{a}$ and positive Weyl chamber $\mathfrak{a}^+$, we define $\overline{\mathfrak{a}^+}$–valued measured laminations and $(\overline{\mathfrak{a}^+}, k)$–mixed structures obtained by gluing these vector-valued laminations together with $1/k$–translation surfaces of finite area along annuli (see Section 6 for details). We will consider this notion for the Lie algebra $\mathfrak{sl}(2,\mathbb{R})\times\mathfrak{sl}(2,\mathbb{R})$: in this case its Cartan subalgebra is of type $A_1\times A_1$ and we denote by $\overline{A_1^+\times A_1^+}$ the closure of a fixed positive Weyl chamber. Concretely, in this case the Cartan subalgebra can be chosen to be the space of pairs of 2×2 traceless diagonal matrices, so it is homeomorphic to $\mathbb{R}^2$ and $\overline{A_1^+\times A_1^+}$ is homeomorphic to a quadrant. We can rephrase our main result as follows:

Theorem C *The boundary of* $\mathrm{Max}(S)$ *can be identified with the space of* $(\overline{A_1^+\times A_1^+}, 2)$*–mixed structures on S*, *which is thus topologically a sphere of dimension* $12g-13$.

Moreover, we prove in Lemma 6.8 that $(\overline{A_1^+\times A_1^+}, 2)$–mixed structures are dual to the subcomplexes of a Euclidean building introduced and studied in [Parreau 2022]. Theorem C has the advantage of being easily adaptable to other higher Teichmüller components (see Conjecture 6.7 for the precise statements in rank 2).

Historical remarks

In analogy with the classical case, compactifications of higher Teichmüller spaces are fruitfully studied using different techniques and perspectives. Parreau [2012] compactifies the character variety of surface group representations into noncompact semisimple connected real Lie groups with finite center using Euclidean buildings. For Hitchin and maximal connected components, one can obtain additional information on the boundary points by using the (Θ-)positivity properties of the representations as in [Alessandrini 2008; Burger and Pozzetti 2017; Fock and Goncharov 2006; Le 2016; Martone 2019a; 2019b;

Parreau 2022]. For rank-2 Lie groups, the second and third authors used analytic methods to study degenerations of geometric objects associated to these representations in [Ouyang 2023; Ouyang and Tamburelli 2021; 2023]. In a series of papers, Burger, Iozzi, Parreau and Pozzetti [Burger et al. 2017; 2021a; 2024] use geodesic currents and real algebrogeometric methods to study the Weyl chamber length spectrum compactification of general character varieties introduced in [Parreau 2012]. Their results apply in particular to Hitchin and maximal components, which are fundamental examples of higher Teichmüller spaces, and establish several structural properties of the boundary points. While we refer to their announcement [Burger et al. 2021b] for an account of their general framework and results, here we describe in greater detail their independent work [Burger et al. 2021c] on the compactification of n–copies of $\mathrm{Teich}(S)$. Burger, Iozzi, Parreau and Pozzetti identify the boundary of the Weyl chamber length spectrum compactification of $\mathrm{Teich}(S)^n$ with the projectivization of $\mathcal{MF}(S)^n$, which is a sphere of dimension $n(6g-6)-1$. In addition, they show that $\mathrm{MCG}(S)$ acts properly discontinuously on the space of *positive joint systole* n–tuples of measured foliations [Burger et al. 2021c, Theorem 1.1]. This result provides a new geometric description of the domain of discontinuity introduced in [Burger et al. 2021a] for the $\mathrm{MCG}(S)$ action on the boundary of the Weyl chamber length spectrum compactification in the case of the Lie group $\mathrm{PSL}(2,\mathbb{R})^n$. Finally, when $n=2$, they describe the boundary points as vector-valued mixed structures (in their language, $\mathbb{R}^2$–mixed structures) and associate to these objects a dual tree-graded $\mathbb{R}^2$–space in the sense of [Druţu and Sapir 2005] (see Theorems 1.2 and 1.3 in [Burger et al. 2021c]). Their results lead to an (a priori different) compactification of $\mathrm{Max}(S)$.

2 Background

2.1 Foliations, laminations and $\mathbb{R}$–trees

We recall some classical facts about measured foliations and laminations. This material can be found in [Fathi et al. 2012]. Let S be a closed, orientable, smooth surface of genus $g>1$. A *measured foliation* is a singular foliation (with k–pronged singularities) equipped with a measure on transverse arcs, invariant under transverse homotopy.

If S is given a hyperbolic metric σ, then a *measured lamination* is a closed set of disjoint simple geodesics on (S,σ) together with a transverse measure. There is a natural homeomorphism between the space $\mathcal{MF}(S)$ of measured foliations on S and the space $\mathcal{ML}(S)$ of measured laminations on (S,σ), so that the role of σ is an auxiliary one. Thurston showed $\mathcal{MF}(S)$ is topologically trivial, being a ball of dimension $6g-6$. The space $\mathbb{P}\mathcal{MF}(S)$ is the boundary of Teichmüller space under the Thurston compactification.

If S is given a complex structure J, then to any holomorphic quadratic differential $q=q(z)\,dz^2$, one may consider the foliation obtained by integrating the line field $q(v,v)>0$. When further given the transverse measure defined by $\int_\alpha |\mathrm{Im}(\sqrt{q})|$, the resulting measured foliation is called the *horizontal foliation* of q. Likewise integrating the line field $q(v,v)<0$ and taking the measure $\int_\alpha |\mathrm{Re}(\sqrt{q})|$ gives the *vertical foliation* of q. The theorem of [Hubbard and Masur 1979] states that for a fixed Riemann surface (S,J)

and any measured foliation $\mathcal{F}$ on S, there is a unique holomorphic quadratic differential q, whose horizontal foliation is Whitehead equivalent (ie it differs at most by isotopies or expanding or collapsing pronged singularities along straight arcs) to $\mathcal{F}$. Any measured foliation $\mathcal{F}$ on S lifts to a measured foliation $\widetilde{\mathcal{F}}$ on the universal cover $\widetilde{S}$. Taking the leaf space of $\widetilde{\mathcal{F}}$ together with a distance induced by the pushforward of the transverse measure gives an $\mathbb{R}$–tree. When an $\mathbb{R}$–tree is constructed from a measured foliation in this way, the $\mathbb{R}$–tree comes equipped with a $\pi_1(S)$–action from $\widetilde{\mathcal{F}}$. This action is *small*, that is, the stabilizer of an arc never contains a free group of rank 2, and *minimal*, that is, the action does not preserve any proper subtree. A result of [Skora 1996] says that any $\mathbb{R}$–tree with a $\pi_1(S)$–action which is both small and minimal is constructed from a measured foliation on S. Such $\mathbb{R}$–trees are said to be dual to a measured foliation, and for our purposes, all $\mathbb{R}$–trees we consider will be dual to a measured foliation.

2.2 Half-translation surfaces, flat metrics and mixed structures

A Riemann surface equipped with a holomorphic quadratic differential q is called a half-translation surface. This terminology comes from the fact these can be realized by gluing polygons in $\mathbb{C}$ via translations or rotations of angle π.

A half-translation surface is naturally endowed with a singular flat metric $|q|$, where the singularities are at the zeros of q. Duchin, Leininger and Rafi [Duchin et al. 2010] have studied the degeneration of unit-area quadratic differential metrics, and have shown the limits are precisely projectivized (quadratic) mixed structures. A *mixed structure* is a collection of integrable meromorphic quadratic differential metrics on subsurfaces and measured laminations on other subsurfaces, glued along flat annuli to recover the surface S. Trivial examples of mixed structures include singular flat metrics on S and measured laminations on S. We say that a mixed structure is *properly mixed* if it has a flat piece but it is not a singular flat metric. Mixed structures, when the meromorphic differential is cubic or quartic, appear in the compactification of Hitchin components for $\mathrm{SL}(3,\mathbb{R})$ and $\mathrm{Sp}(4,\mathbb{R})$ (see [Ouyang and Tamburelli 2021; 2023]).

A measured lamination λ on S is said to *fill* if the complement $S \setminus \lambda$ is a disjoint union of topological disks. A pair $\mathcal{F}_1, \mathcal{F}_2$ of measured foliations on S is said to *fill* or is *transverse* if, for any third foliation $\mathcal{G}$, one has $i(\mathcal{F}_1,\mathcal{G}) + i(\mathcal{F}_2,\mathcal{G}) > 0$. Here $i(\cdot\,,\cdot)$ denotes the Bonahon intersection pairing, which generalizes the topological intersection number between curves. We remark that the intersection number for the corresponding measured laminations is the same; therefore we can define filling for a pair of measured laminations analogously. Notice that given a holomorphic quadratic differential q, the vertical and horizontal foliations of q fill. Conversely, the result of [Gardiner and Masur 1991] says that, given any pair of filling measured foliations, there exists a unique Riemann surface structure and a unique holomorphic quadratic differential which realizes the original pair as its vertical and horizontal foliation (up to Whitehead equivalence). In particular, a pair of filling measured foliations will determine a unique half-translation surface structure and consequently a unique singular flat quadratic differential metric.

2.3 Minimal Lagrangians in $\mathbb{H}^2 \times \mathbb{H}^2$

A *minimal Lagrangian* $\widetilde{\Sigma}$ in $\mathbb{H}^2 \times \mathbb{H}^2$ is a minimal surface which is Lagrangian with respect to the symplectic form $\omega \oplus -\omega$, where ω is the standard Kähler form on $\mathbb{H}^2$. Any $\rho \in \mathrm{Max}(S)$ acts on $\mathbb{H}^2 \times \mathbb{H}^2$, and Schoen [1993] has shown to each such ρ there is a unique ρ–equivariant minimal Lagrangian $\widetilde{\Sigma}_\rho$ in $\mathbb{H}^2 \times \mathbb{H}^2$, thereby providing a geometric interpretation to representations in $\mathrm{Max}(S)$. The second author [Ouyang 2023] has studied the degeneration of these minimal Lagrangians and has shown that one may interpret the space $\mathbb{P}\,\mathrm{Core}(\mathcal{T}, \mathcal{T})$ as the boundary of the maximal component $\mathrm{Max}(S)$.

2.4 Induced metrics and projectivized mixed structures

The induced metric on the unique ρ–equivariant minimal Lagrangian descends to a metric on S. It is not too difficult (see [Ouyang 2023, Proposition 4.2]) to see this metric is in fact negatively curved. Hence, by the result of [Otal 1990], its marked length spectrum determines the metric. The marked length spectrum is the data of both the curve class and the length of its geodesic representative in the given homotopy class. Let $\mathrm{Ind}(S)$ denote the space of induced metrics coming from the ρ–equivariant minimal Lagrangians. Then in fact one may embed $\mathrm{Ind}(S)$ into the space of projectivized marked length spectra. Its closure is then determined to be precisely the space $\mathrm{Ind}(S)$ together with the projectivized mixed structures [Duchin et al. 2010, Theorem 5; Ouyang 2023, Theorem 5.5].

3 Core of a product of trees

In this section we recall the notion of core of a product of trees and describe its geometry in the case of trees dual to measured laminations. The core of a product of two $\mathbb{R}$–trees can actually be defined for any pair of $\mathbb{R}$–trees each admitting a $\pi_1(S)$–action. It is not necessary that the $\mathbb{R}$–trees be dual to measured foliations. However, we will specifically mention when particular properties of cores are germane only to $\mathbb{R}$–trees dual to measured foliations. The main reference for the material covered here is [Guirardel 2005].

Given an $\mathbb{R}$–tree T, a *direction* δ based at a point $p \in T$ is a connected component of $T \setminus \{p\}$. For a product $T_1 \times T_2$ of $\mathbb{R}$–trees, a *quadrant* Q based at $(p_1, p_2) \in T_1 \times T_2$ is a product $\delta_1 \times \delta_2$ of directions. If the $\mathbb{R}$–trees T_1, T_2 are equipped with a $\pi_1(S)$–action by isometries, then we say a quadrant Q is *heavy* if there exists a sequence $\{\gamma_n\} \subset \pi_1(S)$ for which, for $i = 1, 2$,

(i) $\gamma_n \cdot p_i \in \delta_i$, and

(ii) $d_i(\gamma_n \cdot p_i, p_i) \to \infty$ as $n \to \infty$.

Otherwise the quadrant is said to be *light*. Following [Guirardel 2005], the *core* $\mathcal{C}(T_1, T_2)$ of $T_1 \times T_2$ is

$$T_1 \times T_2 \setminus \bigsqcup_{Q \text{ light}} Q.$$

When T_1 and T_2 are dual to measured laminations, the core $\mathcal{C}(T_1, T_2)$ is always nonempty since the $\pi_1(S)$–actions are irreducible [Guirardel 2005, Proposition 3.1].

However, even when T_1 and T_2 are dual to measured foliations, one pathology may still occur: $\mathcal{C}(T_1, T_2)$ may be disconnected. This happens, for instance, when $T_1 = T_2$ and T_1 is dual to a multicurve. However, in such cases, Guirardel introduced a canonical way of extending the core to a connected subset of $T_1 \times T_2$ with convex fibers. (Here, a subset $E \subset T_1 \times T_2$ has convex fibers if for every $x \in T_i$ the set $E \cap p_i^{-1}(x)$ is convex, where $p_i : T_1 \times T_2 \to T_i$ denotes the canonical projection.) With abuse of terminology, we will still refer to this canonical extension as the core of $T_1 \times T_2$. The following result completely characterizes when this extension needs to be considered.

Definition 3.1 Given two real trees T and T' endowed with an action of $\pi_1(S)$, we say that T is a *refinement* of T' if there is an equivariant map $f : T \to T'$ such that for all $x, y, z \in T$ if z lies in the geodesic $[x, y]$ connecting x and y, then $f(z)$ belongs to $[f(x), f(y)]$.

Proposition 3.2 [Guirardel 2005, Proposition 4.14] *Let T_1 and T_2 be trees dual to measured laminations. Then the core $\mathcal{C}(T_1, T_2)$ is disconnected if and only if T_1 and T_2 are refinements of a common nontrivial simplicial tree T.*

For example the assumptions of Proposition 3.2 are satisfied if T_1 and T_2 are dual to measured laminations λ_1 and λ_2 with common isolated leaves.

When T_1 and T_2 are both dual to measured laminations λ_1 and λ_2, we can actually realize the core $\mathcal{C}(T_1, T_2)$ more concretely. Before describing this construction, we need the following result, which can be seen as a special case of the decomposition theorem for general geodesic currents in [Burger et al. 2017] (see also [Burger et al. 2021a]) about how two measured laminations interact on subsurfaces. Here, when we refer to measured laminations on open surfaces S', usually arising as subsurfaces of S, we will always assume them to be compactly supported in S'.

Lemma 3.3 *Let λ_1 and λ_2 be measured laminations on S. Then there is a system of nontrivial, pairwise nonhomotopic, disjoint, simple closed curves $\gamma_1, \dots, \gamma_n$ such that on each connected component S' of $S \setminus \bigcup_j \gamma_j$ either*

(i) *$\lambda_1 + \lambda_2$ is a (possibly zero) measured lamination on S', or*

(ii) *λ_1 and λ_2 are transverse and fill S'; ie for all measured laminations ν on S' we have*

$$i(\lambda_1, \nu) + i(\lambda_2, \nu) \neq 0.$$

Proof Consider a maximal collection of nontrivial, pairwise nonhomotopic, disjoint, simple closed curves γ_j such that

$$i(\lambda_1, \gamma_j) + i(\lambda_2, \gamma_j) = 0.$$

We claim that this collection of curves satisfies the requirement of the lemma. Indeed, let S' be a connected component of $S \setminus \bigcup_j \gamma_j$. We need to show that if the pair (λ_1, λ_2) does not fill the subsurface S', then

$\lambda_1+\lambda_2$ is a lamination on S', or, equivalently, λ_1 and λ_2 are nowhere transverse on S'. The claim is clearly true if the support of either λ_1 or λ_2 does not intersect S', so we can assume that both have support on S'. Because the pair (λ_1,λ_2) does not fill S' by assumption, there is a measured lamination ν on S' such that $i(\lambda_1,\nu)+i(\lambda_2,\nu)=0$. On the other hand, by hypothesis, $i(\lambda_1,\gamma)+i(\lambda_2,\gamma)\neq 0$ for all nonperipheral simple closed curves γ on S'. Therefore, the measured lamination ν does not contain isolated closed leaves. Let us first consider the case in which ν fills the subsurface S', in the sense that the complement of ν (in S') only consists of disks and annuli. We note that then necessarily the support of ν must contain the support of λ_1 and λ_2 because otherwise λ_1 and λ_2 would intersect ν transversely somewhere. But this implies that λ_1 and λ_2 are nowhere transverse, being both contained in the support of a measured lamination. We now reduce the general case to this, by showing that ν must fill S'. Assume the opposite, and let $S''\subset S'$ be a subsurface filled by ν. Note that at least one between λ_1 and λ_2 intersects the boundaries of S'' transversely. Without loss of generality we assume it is λ_1. Since ν fills S'', the support of λ_1 intersects ν transversely, but this contradicts the fact that $i(\nu,\lambda_1)=0$. □

The last ingredient we need is an explicit realization of a tree T_λ dual to a measured lamination λ. The construction goes as follows (see [Morgan and Otal 1993] for more details). Fix an auxiliary hyperbolic metric on S and identify $\widetilde{S}$ with $\mathbb{H}^2$. Let $\tilde\lambda$ be the lift of λ under the covering map $\pi\colon\mathbb{H}^2\to S$. We define the metric space $\mathrm{pre}(T_\lambda)$, where points of $\mathrm{pre}(T_\lambda)$ are the connected components of $\mathbb{H}^2\setminus\tilde\lambda$ and the distance is computed as follows: if $x,y\in\mathrm{pre}(T_\lambda)$ correspond to connected components C_x, C_y of $\mathbb{H}^2\setminus\tilde\lambda$ then

$$d_\lambda(x,y)=\inf\Big\{\int_\gamma d\tilde\lambda \mid \gamma\colon[0,1]\to\mathbb{H}^2,\ \gamma(0)\in C_x,\ \gamma(1)\in C_y\Big\}.$$

The tree T_λ is then the unique $\mathbb{R}$–tree that contains $\mathrm{pre}(T_\lambda)$ such that any point of T_λ lies in a segment with vertices in $\mathrm{pre}(T_\lambda)$. Note that we have a natural projection map $p_\lambda\colon\mathbb{H}^2\setminus\tilde\lambda\to T_\lambda$. If λ has no isolated leaves, this map extends continuously to a map, still denoted by p_λ, defined on the entire $\mathbb{H}^2$. Otherwise, the continuous extension is obtained by first replacing each isolated leaf ℓ in $\tilde\lambda$ with a strip $\ell\times[-\epsilon,\epsilon]$ endowed with a uniform measure with total mass equal to $\tilde\lambda_{|\ell}$.

There is also another way of realizing the tree dual to a measured lamination using the language of measured foliations. Let $\mathcal{F}$ denote the measured foliation corresponding to the measured lamination λ under the homeomorphism between $\mathcal{MF}(S)$ and $\mathcal{ML}(S)$. Let $\widetilde{\mathcal{F}}$ be its lift to $\mathbb{H}^2$. Then the tree T_λ can be defined as the quotient $\mathbb{H}^2/{\sim}$, where $\sim$ denotes the equivalence relation

$$x\sim y \iff d_{\mathcal{F}}(x,y)=0$$

and

$$d_{\mathcal{F}}(x,y)=\inf\{i(\widetilde{\mathcal{F}},\gamma)\mid \gamma\colon[0,1]\to\mathbb{H}^2,\ \gamma(0)=x,\ \gamma(1)=y\}.$$

More concretely, T_λ identifies with the leaf space of $\widetilde{\mathcal{F}}$ with distance given by integrating the measure of $\widetilde{\mathcal{F}}$ along arcs transverse to the leaves. We denote by π_λ the natural projection $\pi_\lambda\colon\mathbb{H}^2\to T_\lambda$.

We are now ready to describe the core of a product of two trees T_1 and T_2 dual to measured laminations λ_1 and λ_2 on S. Lemma 3.3 furnishes a decomposition of S into subsurfaces that we lift to a decomposition of $\mathbb{H}^2$. The regions of this decomposition come in two flavors according to whether they project to subsurfaces where $\lambda_1 + \lambda_2$ is a lamination or to subsurfaces where the pair (λ_1, λ_2) fills. Following the statement of Lemma 3.3, we call these regions of type i and type ii, respectively. On the regions $\Omega \subset \mathbb{H}^2$ of type i, the union of the lifts $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ can be regarded as the lift of the measured lamination $\lambda_0 = \lambda_1 + \lambda_2$. We denote by T_0 the tree dual to λ_0. Note that, for each region $\Omega \subset \mathbb{H}^2$ of type i, we have a map $p_0 := p_{\lambda_0}$, as defined before, and two natural collapsing maps $c_j : T_0 \to T_j$ for $j = 1, 2$. On the regions of type ii, we replace the measured laminations $\tilde{\lambda}_i$ with the corresponding measured foliations $\tilde{\mathcal{F}}_i$ and consider the projections $\pi_i := \pi_{\lambda_i}$ as described previously. Following [Guirardel 2005, Example 4; 2005, Proposition 6.1], the core $\mathcal{C}(T_1, T_2)$ is the image of the map $F : \widetilde{S} \to T_1 \times T_2$ defined as follows:

$$F(x) = \begin{cases} (\pi_1 \times \pi_2)(x) & \text{if } x \text{ belongs to a region of type } ii, \\ (c_1 \times c_2)(p_0(x)) & \text{if } x \text{ belongs to a region of type } i. \end{cases} \tag{1}$$

Note that F is well-defined and continuous on the boundary $\tilde{\gamma}$ between two different regions of $\widetilde{S}$ because $\tilde{\gamma}$ is the lift of a curve γ_j given by Lemma 3.3 which, by definition, has vanishing intersection number with λ_0, $\mathcal{F}_1$, and $\mathcal{F}_2$; hence $(\pi_1 \times \pi_2)(\tilde{\gamma})$ and $(c_1 \times c_2)(p_0(\tilde{\gamma}))$ is a single point.

It follows from this explicit description of $\mathcal{C}(T_1, T_2)$ that the core is, in general, a 2–dimensional subcomplex of $T_1 \times T_2$ that is invariant under the diagonal action of $\pi_1(S)$. Moreover, the 2–dimensional pieces of $\mathcal{C}(T_1, T_2)$ are exactly the images of regions of type ii and are foliated by two families of transverse foliations. Their quotients under the group action are the union of the subsurfaces of S in which λ_1 and λ_2 fill, endowed with the foliations $\mathcal{F}_1$ and $\mathcal{F}_2$ [Guirardel 2005, Example 4]. In particular, the 2–dimensional pieces of $\mathcal{C}(T_1, T_2)$ are the universal covers of half-translation surfaces. On the other hand, the images under F of regions Ω of type i are 1–dimensional subcomplexes of $T_1 \times T_2$. Each such Ω can be seen as the universal cover of a subsurface S' of S where the restriction of $\lambda_1 + \lambda_2$ is a measured lamination. Let $T_1' \subset T_1$ and $T_2' \subset T_2$ be the corresponding subtrees. It turns out [Guirardel 2005, Section 6] that $F(\Omega)$ is an $\mathbb{R}$–tree that is a common refinement of T_1' and T_2' if endowed with the distance

$$d_0(x, y) = d_1(x_1, y_1) + d_2(x_2, y_2), \quad x = (x_1, x_2),\ y = (y_1, y_2) \in T_1' \times T_2',$$

where d_j denotes the distance on T_j.

Lemma 3.4 *The $\mathbb{R}$–tree $(F(\Omega), d_0)$ is isometric to the tree dual to the measured lamination $\lambda_0 = \lambda_1 + \lambda_2$ restricted to S'.*

Proof The tree $F(\Omega)$ inherits from $T_1' \times T_2'$ an isometric action of $\pi_1(S')$. We can define a length function

$$\ell : \pi_1(S') \to \mathbb{R}^+, \quad \gamma \mapsto \lim_{n \to +\infty} \frac{1}{n} d_0(x, \gamma^n \cdot x),$$

where x is any point in $F(\Omega)$ (the definition is independent of the choice of x). The limit in the formula above is well-defined and coincides, indeed, with the minimal translation distance of $\gamma \in \pi_1(S')$ [Guirardel and Levitt 2017, Section A.3]. Since the action of $\pi_1(S')$ on $F(\Omega)$ is minimal and irreducible, by [Guirardel and Levitt 2017, Theorem A.5], the isometry class of $(F(\Omega), d_0)$ is completely determined by its length function. However, it is clear from the definition of ℓ and d_0 that $\ell = \ell_0 := \ell_1 + \ell_2$, where ℓ_j denotes the analogously defined length functions on T_1' and T_2'. On the other hand, ℓ_0 is exactly the length function of the tree dual to the measured lamination λ_0, and the claim follows. □

The ambient space $T_1 \times T_2$ has, however, another natural distance defined by

$$d(x,y) = \sqrt{d_1(x_1,y_1)^2 + d_2(x_2,y_2)^2}, \quad x = (x_1,x_2),\ y = (y_1,y_2) \in T_1 \times T_2.$$

This induces a path metric $d_{\mathcal{C}}$ on the core $\mathcal{C}$ of $T_1 \times T_2$, where the $d_{\mathcal{C}}$–distance between two points in the core is the infimum of the length of all paths connecting the points and entirely contained in the core, where the length is computed using the distance d. Guirardel [2005, Proposition 4.9] showed that the core is a CAT(0) space if endowed with this path distance $d_{\mathcal{C}}$. In particular, since $F(\Omega)$ does not contain topological circles by Lemma 3.4, we can conclude that $F(\Omega)$ endowed with the restriction of $d_{\mathcal{C}}$ is still an $\mathbb{R}$–tree.

We will denote by $\mathrm{Core}(\mathcal{T},\mathcal{T})$ the space of cores of the product of two trees dual to measured laminations on S endowed with this path distance.

Proposition 3.5 $\mathrm{Core}(\mathcal{T},\mathcal{T})$ *is homeomorphic to* $\mathcal{ML}(S) \times \mathcal{ML}(S)$.

Proof Since the core of a product of trees is uniquely determined by the two factors, the result follows immediately from the homeomorphism between the space of trees dual to measured laminations and $\mathcal{ML}(S)$. □

We note that there is a natural $\mathbb{R}^+$–action on $\mathrm{Core}(\mathcal{T},\mathcal{T})$ given by rescaling the induced metric on the core, which, under the homeomorphism above, corresponds to the diagonal action of $\mathbb{R}^+$ by scalar multiplication on the measures. We denote by $\mathbb{P}\,\mathrm{Core}(\mathcal{T},\mathcal{T})$ the quotient $\mathrm{Core}(\mathcal{T},\mathcal{T})/\mathbb{R}^+$. It follows that $\mathbb{P}\,\mathrm{Core}(\mathcal{T},\mathcal{T})$ is homeomorphic to $\mathbb{P}(\mathcal{ML}(S) \times \mathcal{ML}(S))$. In particular, it is topologically a sphere of dimension $12g-13$.

4 Thurston's compactification

Recall that we denote by $\mathrm{Max}(S)$ the space of conjugacy classes of representations $\rho = (\rho_1,\rho_2)$ of the fundamental group of a closed connected oriented surface S of negative Euler characteristic into the Lie group $\mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$ such that $e(\rho_1) + e(\rho_2) = 4g-4$. Here, e denotes the Euler number of the representation. It follows from [Goldman 1988] that ρ_1 and ρ_2 are both Fuchsian representations. Therefore, as $\mathrm{Max}(S)$ may be thought of as the product of two copies of Teichmüller space, it is homeomorphic to an open cell of dimension $12g-12$.

The main goal of this section is to prove Theorem A from the Introduction, which we restate below for the convenience of the reader.

Theorem 4.1 *The disjoint union*

$$\mathfrak{B} = \mathrm{Max}(S) \sqcup \mathbb{P}\,\mathrm{Core}(\mathcal{T},\mathcal{T})$$

is homeomorphic to a closed ball of dimension $12g-12$.

We begin by recalling the topology placed on $\mathfrak{B}$. The maximal component $\mathrm{Max}(S)$ is naturally homeomorphic to the product of two copies of Teichmüller space. This in turn, by the result of Schoen, is homeomorphic to the space of equivariant minimal Lagrangians in $\mathbb{H}^2\times\mathbb{H}^2$. Under the Gromov–Hausdorff topology, diverging sequences of minimal Lagrangians subconverge to the (projective) core of a product of two trees [Ouyang 2023, Theorem 8.1]. The two trees are dual to a pair of measured laminations, and the topology on $\mathfrak{B}$ is compatible with the Thurston compactification on $\mathrm{Teich}(S)\times\mathrm{Teich}(S)$ in the following way: if $(\rho_{1,n},\rho_{2,n})\to[\lambda_1,\lambda_2]$, then the associated minimal Lagrangians converge to the core of $T_1\times T_2$, where T_i is dual to λ_i.

Fix a complex structure J on S and denote by X the Riemann surface (S,J). Then for any hyperbolic metric $h\in\mathrm{Teich}(S)$ there is a unique harmonic map $w_h\colon X\to(S,h)$ in the homotopy class of the identity [Eells and Sampson 1964; Hartman 1967]. Harmonicity of w_h ensures that the Hopf differential $q_h=(w_h^*h)^{(2,0)}$ is a holomorphic quadratic differential on X. The vector space $\mathrm{QD}(X)$ of holomorphic quadratic differentials on X has a natural norm given by the L^2–norm with respect to the uniformizing hyperbolic metric σ of X. With an abuse of notation, we will still denote by X the hyperbolic surface (S,σ). The map which assigns to a point in Teichmüller space its corresponding Hopf differential is a homeomorphism [Wolf 1989].

Proof of Theorem 4.1 By Theorem 6.13 of [Ouyang 2023], the space $\mathrm{Max}(S)\sqcup\mathbb{P}\,\mathrm{Core}(\mathcal{T},\mathcal{T})$ is naturally homeomorphic to $\mathrm{Teich}(S)\times\mathrm{Teich}(S)\sqcup\mathbb{P}(\mathcal{MF}(S)\times\mathcal{MF}(S))$, so it suffices to prove the latter is homeomorphic to a closed ball of dimension $12g-12$.

As $\mathbb{P}(\mathcal{MF}(S)\times\mathcal{MF}(S))$ is homeomorphic to a sphere of dimension $12g-13$, the remainder of the proof consists of describing how to attach this topological space to the open cell $\mathrm{Teich}(S)\times\mathrm{Teich}(S)$ to obtain a closed ball.

We start by fixing a complex structure J on S. Let $X=(S,J)$ be the resulting Riemann surface. By the Wolf parametrization [1989]

$$\mathrm{Teich}(S)\times\mathrm{Teich}(S)\cong\mathrm{QD}(X)\oplus\mathrm{QD}(X)$$

via the map $\Phi(\rho_1,\rho_2)=(q_{\rho_1},q_{\rho_2})$. We equip $\mathrm{QD}(X)\oplus\mathrm{QD}(X)$ with the norm

$$\|q\|=\max(\|q_1\|,\|q_2\|),$$

and consider

$$\mathrm{BPQD}(X)=\{q=(q_1,q_2):\|q\|<1\},$$

which is, topologically, a ball of dimension $12g-12$. We will need the following lemma.

Lemma 4.2 *The map*

$$\beta\colon\mathrm{QD}(X)\oplus\mathrm{QD}(X)\to\mathrm{BPQD}(X),\quad q=(q_1,q_2)\mapsto\frac{4q}{1+4\|q\|},$$

is continuous, injective, and proper. Hence β is a homeomorphism.

Proof Suppose $\beta(q_1, q_2) = \beta(\phi_1, \phi_2)$. It follows then that $q_1 = k\phi_1$ and $q_2 = k\phi_2$ for some $k \in \mathbb{R}$. Writing out $\beta(q_1, q_2) = \beta(q_1/k, q_2/k)$, basic algebra shows $k = 1$. Continuity and properness follow by inspection. □

We will now describe the attaching map. Consider the map

$$\psi\colon \mathrm{Teich}(S) \times \mathrm{Teich}(S) \sqcup \mathbb{P}(\mathcal{MF}(S) \times \mathcal{MF}(S)) \to \overline{\mathrm{BPQD}(X)}$$

defined by

$$\psi(x) = \begin{cases} \beta(\Phi(x)) & \text{if } x \in \mathrm{Teich}(S) \times \mathrm{Teich}(S), \\ \lim_{n\to+\infty} \beta(\Phi(x_n)) & \text{if } x \in \mathbb{P}(\mathcal{MF}(S) \times \mathcal{MF}(S)) \text{ and } x_n \to x. \end{cases}$$

We show first that the map ψ is well-defined. Suppose $x_n = (X_{1,n}, X_{2,n}) \to x$ and $x'_n = (X'_{1,n}, X'_{2,n}) \to x$, where $x = [\lambda_1, \lambda_2] \in \mathbb{P}(\mathcal{MF}(S) \times \mathcal{MF}(S))$. That is to say, there exist sequences of real numbers c_n, d_n for which the rescaled hyperbolic surfaces $\widetilde{X}_{i,n}/c_n$ and $\widetilde{X}'_{i,n}/d_n$ converge to $\mathbb{R}$–trees T_i, T'_i dual to laminations λ_i and λ'_i such that $[(\lambda_1, \lambda_2)] = [(\lambda'_1, \lambda'_2)]$. By [Wolf 1989], the sequences c_n and d_n can be taken to be $\|\Phi(x_n)\|$ and $\|\Phi(x'_n)\|$. Note that, a priori, $\lambda'_i = k\lambda_i$ for some $k > 0$. With such rescaling, the harmonic maps $h_{i,n}\colon \widetilde{X} \to \widetilde{X}_{i,n}/c_n$ converge to the harmonic map $h_i\colon \widetilde{X} \to T_i$ given by projection onto the leaf space of the measured foliation $\widetilde{\mathcal{F}}_i$ corresponding to $\tilde{\lambda}_i$ [Wolf 1995, Corollary 5.2]. Moreover the sequence of Hopf differentials $q_{i,n}$ of $h_{i,n}$ converges to the Hopf differential q_i of h_i (here take the quotient so that q_i is a holomorphic quadratic differential on X and not $\widetilde{X}$). Finally, the differential q_i is the unique holomorphic quadratic differential on X whose horizontal foliation is Whitehead equivalent to $\mathcal{F}_i$. Likewise the sequence of harmonic maps $h'_{i,n}\colon \widetilde{X} \to \widetilde{X}'_{i,n}/d_n$ converges to the harmonic map $h'\colon \widetilde{X} \to T'_i$, whose Hopf differential q'_i is the limit of the Hopf differentials $q'_{i,n}$ of $h'_{i,n}$ and has horizontal foliation $\mathcal{F}'_i$ corresponding to the lamination λ'_i. Notice, in addition, that $(q_{1,n}, q_{2,n}) = \Phi(x_n)/\|\Phi(x_n)\|$ and similarly $(q'_{1,n}, q'_{2,n}) = \Phi(x'_n)/\|\Phi(x'_n)\|$. It follows that the limits of $\beta(\Phi(x_n))$ and $\beta(\Phi(x'_n))$ as $n \to +\infty$ exist and coincide with (q_1, q_2) and (q'_1, q'_2). As the distance functions d_i and d'_i on T_i and T'_i satisfy $d_i = k \cdot d'_i$, by homogeneity of the Hopf differential, one has $q_i = k \cdot q'_i$. Since the pairs (q_1, q_2) and (q'_1, q'_2) both have unit norm, we conclude that $k = 1$ and the limits of $\beta(\Phi(x_n))$ and $\beta(\Phi(x'_n))$ as $n \to +\infty$ are equal.

Continuity follows almost immediately: the map $\beta \circ \Phi$ is continuous on the interior and extends continuously to the boundary by a diagonal argument. Indeed, we can approximate a sequence along the boundary by sequences in the interior.

Bijectivity of ψ on the interior also follows by [Wolf 1989] and Lemma 4.2. On the boundary, given $q = (q_1, q_2)$ with $\|q\| = 1$, if $X_{i,t}$ is the hyperbolic surface corresponding to the rays tq_i in Wolf's parameterization of $\mathrm{Teich}(S)$, we have that $\beta(\Phi(X_{1,t}, X_{2,t})) \to q$ as $t \to \infty$; thus ψ is surjective on the boundary. Since the limit of $\beta(\Phi(x_n))$ along diverging sequences in $x_n \in \mathrm{Teich}(S) \times \mathrm{Teich}(S)$ only depends on the projective class of the limit of x_n and not on the particular sequence, we deduce that ψ is injective on the boundary, because every point in $\mathbb{P}(\mathcal{ML}(S) \times \mathcal{ML}(S))$ can be obtained as a limit along a ray defined above and the limit of $\beta \circ \Phi$ along distinct rays is different.

It remains to prove ψ^{-1} is continuous. We can actually write the inverse explicitly:

$$\psi^{-1}(q_1,q_2)=\begin{cases}\Phi^{-1}(\beta^{-1}(q_1,q_2)) & \text{if } \|(q_1,q_2)\|<1,\\ [\lambda_1,\lambda_2] & \text{if } \|(q_1,q_2)\|=1,\end{cases}$$

where λ_i is the measured lamination corresponding to the horizontal foliation of q_i. Continuity of ψ^{-1} on $\mathrm{BPQD}(X)$ is then a consequence of Lemma 4.2 and Wolf's parameterization. Continuity on the boundary follows from the Hubbard–Masur theorem [1979]. In general, if $q_n=(q_{1,n},q_{2,n})\in\mathrm{BPQD}(X)$ converges to $(q_1,q_2)\in\partial\overline{\mathrm{BPQD}(X)}$, then there is a sequence of scaling factors c_n such that the pair of hyperbolic surfaces $x_n=\psi^{-1}(q_{1,n},q_{2,n})$ rescaled by c_n converges to real trees T_1,T_2 dual to measured laminations λ_1,λ_2. We need to show that $\psi^{-1}(q_1,q_2)$ is equal to $[\lambda_1,\lambda_2]$. Assume not; then we would have, by injectivity and continuity of ψ,

$$(q_1,q_2)=\psi(\psi^{-1}(q_1,q_2))\neq\psi([\lambda_1,\lambda_2])=\lim_{n\to+\infty}\psi(x_n)=\lim_{n\to+\infty}(q_{1,n},q_{2,n}),$$

which contradicts the assumption on $(q_{1,n},q_{2,n})$.

Finally, we remark the compactification in [Ouyang 2023] is independent of the choice of a base point, so that the role of the base point (S,J) is merely an auxiliary one. This completes the proof of the theorem. □

5 Fixed point for the mapping class group action

In this section, we study the action of the mapping class group $\mathrm{MCG}(S)$ on the compactification $\mathfrak{B}=\overline{\mathrm{Max}(S)}$ constructed in Theorem 4.1. We wish to study the fixed points of this action. We will need the following observations.

Lemma 5.1 *The action of the mapping class group on* $\mathrm{Max}(S)$ *extends continuously to the closure* $\mathfrak{B}=\overline{\mathrm{Max}(S)}$.

Corollary 5.2 *For every* $\phi\in\mathrm{MCG}(S)$, *there exists* $x\in\mathfrak{B}$ *such that* $\phi(x)=x$.

The first main goal of this section is to analyze these fixed points via the celebrated Nielsen–Thurston classification, which we recall for future reference.

Theorem 5.3 (Nielsen–Thurston classification; see [Farb and Margalit 2012, Chapter 13]) *Any diffeomorphism* ϕ *on* S *is isotopic to a map* ϕ' *satisfying one of the following mutually exclusive conditions*:

(1) ***Periodic*** ϕ' *is of finite order.*

(2) ***Reducible*** ϕ' *is not periodic, and there is a nonempty set* $\{c_1,\dots,c_r\}$ *of isotopy classes of essential pairwise disjoint simple closed curves in* S *such that* $\{\phi'(c_i)\}_{i=1}^r=\{c_i\}_{i=1}^r$.

(3) ***Pseudo-Anosov/pA*** *There exist* $\lambda>1$ *and two transverse measured foliations* $\mathcal{F}$ *and* $\mathcal{F}'$ *such that*

$$\phi'(\mathcal{F})=\lambda\mathcal{F}\quad\text{and}\quad\phi'(\mathcal{F}')=\frac{1}{\lambda}\mathcal{F}'.$$

Remark 5.4 Note that our definition of reducible mapping class is nonstandard as we assume that if ϕ is reducible, then it is not periodic. We do so to improve our exposition. The set $\{c_1, \dots, c_r\}$ in item (2) is a *reduction system* of ϕ. The *canonical reduction system* $\{\bar{c}_1, \dots, \bar{c}_k\}$ of ϕ reducible is the intersection of all the maximal (with respect to inclusion) reduction systems. Equivalently, each $\bar{c}_j$ is part of a reduction system and if $i(\bar{c}_j, c) \neq 0$ and $n \neq 0$, then $\phi^n(c) \neq c$.

Remark 5.5 The Nielsen–Thurston classification theorem also applies to surfaces S' with boundary [Fathi et al. 2012, Theorem 11.6]. In this case a diffeomorphism of S' is considered up to isotopies that do not necessarily fix pointwise the boundary components. We can thus still talk about pseudo-Anosov diffeomorphisms of S', which are exactly the mapping classes that are neither reducible nor periodic and preserve two transverse measured foliations on S'.

We are ready to characterize the fixed points of a mapping class acting on $\mathfrak{B}$ and establish Proposition 1.1 from the Introduction.

Proposition 5.6 *Suppose $\phi \in \mathrm{MCG}(S)$ and $\phi(x) = x$ for some $x = (x_1, x_2) \in \mathfrak{B}$.*

1. *If ϕ is periodic, then x_1 and x_2 are any two points fixed by ϕ in the Thurston compactification of Teichmüller space such that $(x_1, x_2) \in \mathfrak{B}$.*
2. *If ϕ is pA, then $(x_1, x_2) \in \partial\mathfrak{B}$ and $x_1 = 0$, or $x_2 = 0$ or $x_1 = x_2$.*

Proof (1) If ϕ fixes x projectively, there exists $\alpha > 0$ such that $\phi(x_1, x_2) = (\alpha x_1, \alpha x_2)$. Since ϕ is periodic, we can check that $\alpha = 1$.

(2) Since ϕ fixes the projective class of (x_1, x_2), there exists $\alpha > 0$ such that $\phi(x_1, x_2) = (\alpha x_1, \alpha x_2)$. On the other hand, since ϕ is pseudo-Anosov, there exist two measured laminations y_1 and y_2 and $\lambda > 1$ such that $\phi(y_1) = \lambda y_1$ and $\phi(y_2) = (1/\lambda) y_2$. Since ϕ does not fix any other projective class of measured laminations [Fathi et al. 2012, Corollary 12.4], it follows that $x_i = 0$, y_1 or y_2 for $i = 1, 2$. We claim that $x \neq (y_1, y_2)$ (and, symmetrically, $x \neq (y_2, y_1)$). Otherwise, because $i(y_1, y_2) \neq 0$,

$$\lambda \cdot i(y_1, y_2) = i(\phi(y_1), y_2) = \alpha \cdot i(y_1, y_2) = i(y_1, \phi(y_2)) = \frac{1}{\lambda} \cdot i(y_1, y_2),$$

which is a contradiction. □

There is a natural continuous projection map $\pi \colon \mathfrak{B} \to \overline{\mathrm{Ind}(S)}$ defined as follows. For $x \in \mathrm{Max}(S)$, consider the corresponding equivariant minimal Lagrangian $\widetilde{\Sigma}_x$. Then, $\pi(x)$ is the induced metric on $\widetilde{\Sigma}_x$. Otherwise, if $x \in \partial\mathfrak{B}$, consider the core of the tree corresponding to $x \in \mathbb{P}(\mathcal{ML}(S) \times \mathcal{ML}(S))$: its length spectrum coincides with that of a mixed structure μ on S. Set $\pi(x) = \mu$. This projection π is continuous then by [Ouyang 2023, Theorem 6.13]. We consider the corresponding action of $\mathrm{MCG}(S)$ on $\overline{\mathrm{Ind}(S)}$ given by push-forward.

Lemma 5.7 *The actions of $\mathrm{MCG}(S)$ on $\mathfrak{B}$ and $\overline{\mathrm{Ind}(S)}$ commute. In other words, for every $\phi \in \mathrm{MCG}(S)$*

$$\pi \circ \phi = \phi \circ \pi.$$

Proof If $x = (x_1, x_2)$ is in the interior of $\mathfrak{B}$, then $\pi(\phi(x)) = \phi(\pi(x))$ because $\phi(\pi(x))$ has the same length spectrum as the induced metric on the minimal Lagrangian associated to $\phi(x_1)$ and $\phi(x_2)$. Suppose $x \in \partial\mathfrak{B}$ and consider a sequence $(x_n)_{n\in\mathbb{N}} \subset \mathrm{Max}(S)$ such that $x_n \to x$. Since, $\pi(\phi(x_n)) = \phi(\pi(x_n))$ for all $n \in \mathbb{N}$, the result follows by continuity of ϕ and π. □

We are now ready to establish the main theorems of this section. In particular, Theorem 5.8 below is Theorem B from the Introduction.

Theorem 5.8 *Assume $\phi \in \mathrm{MCG}(S)$ fixes $\mu \in \partial\overline{\mathrm{Ind}(S)}$.*

(1) *If μ is **purely flat**, then ϕ is periodic.*

(2) *If μ is **properly mixed**, then ϕ is not pA.*

Proof For item (1), if ϕ fixes projectively a geodesic current coming from a flat metric, then ϕ rescales the flat metric by some positive constant. Therefore, it is an automorphism of the underlying conformal structure, and hence is of finite order by the Hurwitz automorphism theorem.

We establish item (2). Suppose μ is properly mixed, ie μ is not flat but it has at least one flat piece. We can decompose S as

$$(\{S_\alpha\}_{\alpha\in\mathcal{A}}, \{d_\beta\}_{\beta\in\mathcal{B}}, \{\mu_\alpha\}_{\alpha\in\mathcal{A}}),$$

where μ_α is a flat structure or a (possibly zero) laminar structure on S_α and d_β is a maximal collection of closed geodesics so that

$$i(d_\beta, d_{\beta'}) = 0 \quad \text{and} \quad i(d_\beta, \mu) = 0$$

for all $\beta, \beta' \in \mathcal{B}$ and for every c that intersects some d_β transversely, $i(c, \mu) > 0$. Note that there exists a unique set $\{d_\beta\}_{\beta\in\mathcal{B}}$ with these properties (see [Burger et al. 2017, Theorem 1.1]).

Claim 5.9 *The map ϕ fixes the set $\{d_\beta\}_{\beta\in\mathcal{B}}$.*

Proof Observe that

$$i(\phi(d_\beta), \phi(d_{\beta'})) = i(d_\beta, d_{\beta'}) = 0 \quad \text{and} \quad i(\mu, \phi(d_\beta)) = i(\phi^{-1}(\mu), d_\beta) = 0.$$

If c is a curve that intersects $\phi(d_\beta)$ transversely, then

$$i(\phi^{-1}(c), d_\beta) = i(c, \phi(d_\beta)) > 0 \quad \text{and} \quad i(c, \mu) = i(\phi^{-1}(c), \phi^{-1}(\mu)) > 0.$$

Thus, by uniqueness, $\{\phi(d_\beta)\}_{\beta\in\mathcal{B}} = \{d_\beta\}_{\beta\in\mathcal{B}}$. □

Item (2) now follows immediately from the claim above as ϕ must fix the set of closed curves $\{d_\beta\}_{\beta\in\mathcal{B}}$, but pseudo-Anosov diffeomorphisms do not preserve any closed curve. □

Remark 5.10 For an explicit example of μ purely flat and ϕ periodic such that $\phi(\mu) = \mu$, consider a singular flat metric on a surface of genus 2 obtained by doubling a singular flat metric on a torus with boundary.

Remark 5.11 Theorem 5.8(1) holds more generally, and with the same proof, in the case in which S has punctures and μ gives a conformal class of metrics with a finite group of conformal automorphisms. In particular, conformal structures on a surface (with or without punctures) with negative Euler characteristic will have finite conformal automorphism group; see [Oikawa 1956]. From this, we deduce that if $\phi \in \mathrm{MCG}(S)$ fixes a purely flat structure μ on a surface S (possibly with punctures), then ϕ is necessarily periodic.

Theorem 5.12 *Suppose* $\phi \in \mathrm{MCG}(S)$ *is reducible and fixes* $\mu \in \partial\overline{\mathrm{Ind}(S)}$, *which is properly mixed. Let* $S = (S_\alpha, \{d_\beta\}_{\beta\in\mathcal{B}}, \mu_\alpha)$ *be the subdivision of* S *induced by* μ.

(1) *If, for some* $N > 0$, *we have* $\psi_\alpha = (\phi^N)_{|S_\alpha} : S_\alpha \to S_\alpha$ *is pA, then* $\mu_\alpha = 0$.

(2) *If* $\mu_\alpha \neq 0$ *for all* $\alpha \in \mathcal{A}$, *then the canonical reduction system of* ϕ *is contained in* $\{d_\beta\}_{\beta\in\mathcal{B}}$.

Proof By the hypotheses, we can decompose S as $(\{S_\alpha\}_{\alpha\in\mathcal{A}}, \{d_\beta\}_{\beta\in\mathcal{B}}, \{\mu_\alpha\}_{\alpha\in\mathcal{A}})$. By Claim 5.9, there exists $N > 0$ such that ϕ^N fixes d_β for all $\beta \in \mathcal{B}$ and $\phi^N(S_\alpha) = S_\alpha$. Set $\psi_\alpha = (\phi^N)_{|S_\alpha} : S_\alpha \to S_\alpha$.

In order to prove item (1), we need to consider three cases.

(a) If $\mu_\alpha = 0$, then ψ_α can be any element in $\mathrm{MCG}(S_\alpha)$.

(b) If (S_α, μ_α) is purely flat (there exists at least one α for which this happens), then ψ_α can only be periodic by Theorem 5.8 and Remark 5.11, as incompressibility of the subsurfaces rules out the case of the once-punctured sphere and annuli.

(c) Suppose (S_α, μ_α) is purely laminar and nonzero. Since μ has a flat piece μ_β, we know that ψ_β is periodic and hence it fixes μ_β (not just projectively). We deduce that $\phi^N(\mu) = \mu$; otherwise we could find $z \neq 1$ such that $\psi_\alpha(\mu_\alpha) = z\mu_\alpha$, but then ϕ^N would not fix μ projectively. We can now conclude that ψ_α cannot be pA. This is because if c is a curve such that $i(\mu_\alpha, c) > 0$, then

$$i(\mu_\alpha, c) = i(\psi_\alpha^{-1}(\mu_\alpha), c) = i(\mu_\alpha, \psi_\alpha(c)),$$

but $i(\mu_\alpha, \psi_\alpha(c)) \neq i(\mu_\alpha, c)$ because ψ_α would change the length of curves transverse to μ_α.

This completes the proof of item (1).

For item (2), we wish to prove that the canonical reduction system $\{\bar{c}_1, \dots, \bar{c}_k\}$ of ϕ is a subset of $\{d_\beta\}_{\beta\in\mathcal{B}}$ under the additional assumption that $\mu_\alpha \neq 0$ for all $\alpha \in \mathcal{A}$. First, observe that by Claim 5.9 $\{d_\beta\}_{\beta\in\mathcal{B}}$ is contained in a maximal reduction system for ϕ. In particular $i(\bar{c}_j, d_\beta) = 0$ for all j and β. Moreover, since μ is properly mixed, there exists β such that μ_β is flat; hence ψ fixes μ, not just its projective class, as observed before.

Assume, by contradiction, $\bar{c}_j \notin \{d_\beta\}_{\beta\in\mathcal{B}}$. Suppose $\bar{c}_j$ is contained in a purely flat piece (S_α, μ_α). Then, by Theorem 5.8 and Remark 5.11, ψ_α is necessarily periodic. But this contradicts the property that if $i(\bar{c}_j, c) \neq 0$ and $n \neq 0$, then $\phi^n(c) \neq c$ since there exists m such that ψ_α^m is the identity. Therefore $\bar{c}_j$ is contained in a purely laminar piece μ_α.

By the definition of $\{d_\beta\}_{\beta\in\mathcal{B}}$, ψ_α fixes a measured lamination $\mathcal{F}$ which is filling in S_α. Hence, by the Nielsen–Thurston classification theorem, ψ_α is necessarily pseudo-Anosov or periodic. If ψ_α is pA, then this contradicts item (1). Assume that ψ_α is periodic, so that there exists $m > 0$ such that ψ_α^m is the identity. Then, we achieve again a contradiction because there would exist c such that $i(\bar{c}_j, c) \neq 0$ but $\phi^m(c) = c$. Therefore, $\bar{c}_j$ cannot be contained in a purely laminar part either. By the definition of $\{d_\beta\}_{\beta\in\mathcal{B}}$, this forces the curve $\bar{c}_j$ to be one of the d_β's. □

6 $\overline{\mathfrak{a}^+}$–valued measured laminations and mixed structures

In this final section we introduce Weyl-chamber-valued measured laminations and use them to refine the notion of mixed structures on a closed surface defined in [Duchin et al. 2010], and generalized to higher-order differentials in [Ouyang and Tamburelli 2021; 2023]. We show that the core of the product of two trees dual to measured laminations is dual to such a mixed structure, thus giving a new interpretation of the boundary objects in our compactification of $\mathrm{Max}(S)$.

Let $\mathfrak{g}$ be a real semisimple Lie algebra. The choice of a maximal compact subalgebra $\mathfrak{k}$ induces an orthogonal decomposition of $\mathfrak{g}$ for the Killing form:

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{m}.$$

A Cartan subalgebra $\mathfrak{a} \subset \mathfrak{g}$ is a maximal abelian subspace of $\mathfrak{m}$. This induces a decomposition of $\mathfrak{g}$ in $\mathrm{ad}(\mathfrak{a})$–eigenspaces

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \bigoplus_{\alpha\in\Sigma} \mathfrak{g}_\alpha.$$

Elements of $\Sigma \subset \mathfrak{a}^* = \mathrm{Hom}(\mathfrak{a}, \mathbb{R})$ are called restricted roots of $\mathfrak{a}$ in $\mathfrak{g}$. Here we can extract a subset Δ of *simple* roots with the property that any $\alpha \in \Sigma$ can be expressed as a linear combination of simple roots with coefficients all of the same sign. This distinguishes, thus, a subset of positive roots that we denote by $\Sigma^+ \subset \Sigma$. The closed positive Weyl chamber of $\mathfrak{a}$ associated to Σ^+ is then the cone

$$\overline{\mathfrak{a}^+} = \{X \in \mathfrak{a} \mid \alpha(X) \geq 0 \text{ for all } \alpha \in \Sigma^+\}.$$

We also denote by W the Weyl group of $\mathfrak{g}$, ie $W = N(\mathfrak{a})/\mathfrak{a}$, and by r the opposition involution. Moreover, recall that $\mathfrak{a}$ has a partial order: if $x, y \in \mathfrak{a}$, then $x \leq y$ if $x - y \in \overline{\mathfrak{a}^+}$. The following definition is due to [Parreau 2012, Section 2.2.3]:

Definition 6.1 A function $d_{\mathfrak{a}^+} : Y \times Y \to \overline{\mathfrak{a}^+}$ on a topological space Y is an $\overline{\mathfrak{a}^+}$–valued distance if

(i) $d_{\mathfrak{a}^+}(x, y) = 0$ if and only if $x = y$,

(ii) $d_{\mathfrak{a}^+}(x, y) = r(d_{\mathfrak{a}^+}(y, x))$ for all $x, y \in Y$,

(iii) $d_{\mathfrak{a}^+}(x, y) \leq d_{\mathfrak{a}^+}(x, z) + d_{\mathfrak{a}^+}(y, z)$ for all $x, y, z \in Y$.

We introduce the notion of Weyl-chamber-valued measured lamination.

Definition 6.2 An $\overline{\mathfrak{a}^+}$–valued measured lamination on a (not necessarily closed) surface S is a geodesic lamination λ on S that supports a measure μ on transverse arcs that takes values in $\overline{\mathfrak{a}^+}$ and satisfies the following properties:

(a) $\mu(\gamma) \neq 0$ if γ intersects λ transversely.

(b) If γ and γ' are homotopic arcs transverse to λ and there is a homotopy between them that preserves transversality at every time, then $\mu(\gamma) = \mu(\gamma')$.

(c) μ is additive on concatenation of paths, ie $\mu(\gamma\gamma') = \mu(\gamma) + \mu(\gamma')$ for all γ and γ' transverse to λ such that concatenation is defined.

Remark 6.3 If $\mathfrak{g} = \mathfrak{sl}(2,\mathbb{R})$, then we can identify the closed positive Weyl chamber with $\mathbb{R}_{\geq 0}$. Thus, in this case, Definition 6.2 recovers the standard notion of measured laminations. Similarly, if $\mathfrak{g} = \mathfrak{sl}(2,\mathbb{R}) \oplus \mathfrak{sl}(2,\mathbb{R})$, then $\overline{\mathfrak{a}^+}$–valued laminations can be identified with ordered pairs (λ_1, λ_2) such that λ_1, λ_2 and $\lambda_1 + \lambda_2$ are measured laminations (ie λ_1 and λ_2 are nowhere transverse).

We can also extend the classical notion of trees dual to a measured lamination to this context.

Definition 6.4 Let (T, d) be an $\mathbb{R}$–tree acted upon by the fundamental group of S. We say that the action of $\pi_1(S)$ is dual to an $\overline{\mathfrak{a}^+}$–valued measured lamination μ if there is an equivariant map $p\colon \widetilde{S} \to T$ and an $\overline{\mathfrak{a}^+}$–valued distance $d_{\mathfrak{a}^+}\colon T \times T \to \overline{\mathfrak{a}^+}$ such that:

(a) For all $x, y \in \widetilde{S}$, we have $d_{\mathfrak{a}^+}(p(x), p(y)) = \mu(\gamma)$ for some (hence any) arc $\gamma\colon [0,1] \to \widetilde{S}$ transverse to the support of μ with $\gamma(0) = x$ and $\gamma(1) = y$.

(b) Given a geodesic path $\gamma\colon [0,1] \to T$, we have $d(\gamma(0), \gamma(1)) \geq \|d_{\mathfrak{a}^+}(\gamma(0), \gamma(1))\|$. Here $\|\cdot\|$ denotes the standard Euclidean norm of a vector in $\overline{\mathfrak{a}^+}$.

We now combine $\overline{\mathfrak{a}^+}$–valued measured laminations with the classical notion of $1/k$–translation surfaces in order to define a hybrid structure on S.

Definition 6.5 Let $\overline{\mathfrak{a}^+}$ be a closed Weyl chamber and $k \geq 1$ an integer. An $(\overline{\mathfrak{a}^+}, k)$–mixed structure on a closed surface S is the datum of

(a) a collection of nonhomotopically trivial, pairwise nonhomotopic, disjoint simple closed curves $\gamma_1, \dots, \gamma_n$ on S;

(b) for each connected component S' of $S \setminus \bigcup_j \gamma_j$ either
   - an $\overline{\mathfrak{a}^+}$–valued measured lamination λ, where we allow each γ_j to be in the support; or
   - a meromorphic k–differential of finite area that endows S' with a $1/k$–translation surface structure.

These $(\overline{\mathfrak{a}^+}, k)$–mixed structures can be interpreted as dual to the $(\mathfrak{a}, W)$–complexes studied by Anne Parreau [2022] in the context of $\mathfrak{g} = \mathfrak{sl}(3,\mathbb{R})$. Let us recall briefly how these complexes are defined and explain in which sense these notions can be considered dual to each other.

Following [Parreau 2022], an $(\mathfrak{a}, W)$–complex K is the union of (possibly degenerate) polygons in $\mathfrak{a}$ glued together along boundary segments via elements of $W_{\mathrm{aff}} = W \rtimes \mathbb{R}$. More precisely, there is a family of affine simplices $P_\mu \subset \mathfrak{a}$ and injective maps $\phi_\mu \colon P_\mu \to K$ such that if $K_\mu = \phi_\mu(P_\mu)$ and $K_{\mu'} = \phi_{\mu'}(P_{\mu'})$ have nonempty intersection then there is $w_{\mu,\mu'} \in W_{\mathrm{aff}}$ such that $\phi_\mu(x) = \phi_{\mu'}(x')$ if and only if $x' = w_{\mu,\mu'}(x)$ and $P_\mu \cap w^{-1}_{\mu,\mu'}(P_{\mu'})$ is a face in P_μ. We only consider connected and simply connected $(\mathfrak{a}, W)$–complexes acted upon by $\pi_1(S)$. Note that, since the gluing maps between simplices are Euclidean isometries, the Euclidean distance on $\mathfrak{a}$ induces a distance on K. We will only work with $(\mathfrak{a}, W)$–complexes whose induced distance is CAT(0). Similarly, K is also endowed with an $\overline{\mathfrak{a}^+}$–valued distance inherited from $\mathfrak{a}$.

Examples of $(\mathfrak{a}, W)$–complexes are subcomplexes of an Euclidean building modeled on W_{aff}. We will see that cores of products of two trees dual to measured laminations are indeed $(\mathfrak{a}, W)$–complexes, where $\mathfrak{a}$ is the Cartan subalgebra of $\mathfrak{sl}(2,\mathbb{R}) \oplus \mathfrak{sl}(2,\mathbb{R})$ and $W = \{\pm \mathrm{Id}\}$.

Definition 6.6 We say that an $(\mathfrak{a}, W)$–complex K acted upon by $\pi_1(S)$ is dual to an $(\overline{\mathfrak{a}^+}, k)$–mixed structure μ on S if we can decompose K into a 1–dimensional part K_1 and a 2–dimensional part K_2 such that

- K_1 is the union of $\mathbb{R}$–trees dual to the laminar part of μ,
- K_2 is endowed with a $1/k$–translation surface structure isomorphic to the universal cover of the flat parts of μ.

Note that the 2–dimensional part of an $(\mathfrak{a}, W)$–complex can be endowed with a $1/k$–translation surface structure only if W contains the subgroup generated by rotations of angle $2\pi/k$.

We believe that these mixed structures naturally appear in a harmonic map compactification of the Hitchin and maximal components of the character variety for real Lie groups G of rank 2. In this context, Labourie [2017], Collier [2016] and Collier, Tholozan and Toulisse [Collier et al. 2019] proved that given a Hitchin or maximal representation $\rho \colon \pi_1(S) \to G$ there exists a unique ρ–equivariant minimal surface $\widetilde{\Sigma}_\rho$ in G/K, where K is a maximal compact subgroup of G. One could then find a compactification of these components by studying the limiting behavior of $\widetilde{\Sigma}_{\rho_n}$ when ρ_n leaves all compact sets in the character variety. Up to subsequences, and after rescaling the metric on G/K appropriately, $\widetilde{\Sigma}_{\rho_n}$ should converge to a subcomplex $\widetilde{\Sigma}_\infty \subset B$, where B is a nondiscrete Euclidean building modeled on the affine Weyl group of G. We conjecture that Σ_∞ is dual to a mixed structure as in Definition 6.6, where $\overline{\mathfrak{a}^+}$ is a Cartan subalgebra of the Lie algebra of G and k depends on the particular group. More precisely, we conjecture the following:

Conjecture 6.7 (a) *Let G be a real split semisimple Lie group of rank 2. Then the boundary of* $\mathrm{Hit}(S, G)$ *can be identified with the space of projective classes of $(\overline{\mathfrak{a}^+}, k)$–mixed structures where:*

- *If $G = \mathrm{SL}(3,\mathbb{R})$, then $\mathfrak{a} = A_2$ and $k = 3$.*
- *If $G = \mathrm{Sp}(4,\mathbb{R})$, then $\mathfrak{a} = B_2$ and $k = 4$.*
- *If $G = G_2^{\mathbb{R}}$, then $\mathfrak{a} = G_2$ and $k = 6$.*

(b) *Let G be a real semisimple Lie group of Hermitian type and rank 2. Then the boundary of* $\mathrm{Max}(S,G)$ *can be identified with the space of projective classes of* $(\overline{\mathfrak{a}^+},k)$*–mixed structures where*:

- *If* $G=\mathrm{SL}(2,\mathbb{R})\times\mathrm{SL}(2,\mathbb{R})$, *then* $\mathfrak{a}=A_1\times A_1$ *and* $k=2$.
- *If* $G=\mathrm{SO}(2,n)$ *with* $n\geq 3$, *then* $\mathfrak{a}=B_2$ *and* $k=4$.

In support of this conjecture, we show that the core of the product of two trees dual to measured laminations is dual to an $(\overline{A_1^+\times A_1^+},2)$–mixed structure and that we can identify $\mathrm{Core}(\mathcal{T},\mathcal{T})$ with the space of such structures, thus proving the conjecture for $G=\mathrm{SL}(2,\mathbb{R})\times\mathrm{SL}(2,\mathbb{R})$. Moreover, in [Loftin et al. 2022], Loftin, Wolf, and the third author give further evidence towards Conjecture 6.7 by describing the geometry of the harmonic maps to buildings arising from some diverging sequences of $\mathrm{SL}(3,\mathbb{R})$–Hitchin representations. It would be interesting to introduce a higher-rank version of our vector-valued mixed structures, at least for the case of $\mathrm{SL}(d,\mathbb{R})$–Hitchin components, and relate it to the subspaces of the Euclidean building studied in [Le 2016; Martone 2019a].

Lemma 6.8 *Let T_1 and T_2 be real trees dual to measured laminations λ_1 and λ_2 and let C be the core of $T_1\times T_2$. Then C is an $(A_1\times A_1,\{\pm\mathrm{Id}\})$–complex dual to an $(\overline{A_1^+\times A_1^+},2)$–mixed structure on S.*

Proof We already saw in Section 3 that C is the union of a 1–dimensional subcomplex C_1 and a 2–dimensional subcomplex C_2 of $T_1\times T_2$. Moreover, we showed that each connected component of C_2 is the universal cover of a half-translation surface structure on a subsurface S' of S, on which the laminations λ_1 and λ_2 fill. Thus, it only remains to show that each connected component C_1' of C_1 is a tree dual to an $\overline{A_1^+\times A_1^+}$–valued measured lamination.

Recall from Section 3 that C_1' is the image under the map F defined in (1) of a domain $\Omega\subset\mathbb{H}^2$ that can be identified with the universal cover of a subsurface S' of S on which λ_1 and λ_2 are nowhere transverse. Moreover, we observe that C_1' has a natural distance d induced by the ambient space

$$d((x_0,y_0),(x_1,y_1))=\sqrt{d_1(x_0,y_0)^2+d_2(x_1,y_1)^2}$$

and a natural $\overline{A_1^+\times A_1^+}$–valued distance $\vec{d}$ defined by

$$\vec{d}((x_0,y_0),(x_1,y_1))=(d_1(x_0,y_0),d_2(x_1,y_1)).$$

We claim that (C_1',d) is an $\mathbb{R}$–tree dual to the $\overline{A_1^+\times A_1^+}$–valued measured lamination $\vec{\lambda}=(\lambda_1,\lambda_2)$ (see Remark 6.3). By Lemma 3.4, C_1' can be identified with the $\mathbb{R}$–tree dual to the measured lamination $\lambda_0=\lambda_1+\lambda_2$ if endowed with the distance d_0 introduced in Section 3. In particular, there is a continuous $\pi_1(S')$–equivariant map $p:=p_{\lambda_0}\colon\Omega\to C_1'$. It follows immediately from the definitions and the fact that T_1 and T_2 are dual to the laminations λ_1 and λ_2 that for all $x,y\in\Omega'$ we have

$$\vec{d}(p(x),p(y))=\vec{\lambda}(\gamma)$$

for all $\gamma\colon[0,1]\to\Omega$ transverse to the support of λ_0 with $\gamma(0)=x$ and $\gamma(1)=y$.

Property (b) in Definition 6.4 also holds. Indeed, a geodesic path $\gamma = (\gamma_1, \gamma_2)\colon [0,1] \to C_1' \subset T_1 \times T_2$, seen in the quadrant $\gamma_1 \times \gamma_2$, consists of a concatenation of horizontal, vertical or diagonal paths in which the projections onto the two factors are always nondecreasing. Hence,

$$d(\gamma(0), \gamma(1)) \geq \|\vec{d}(\gamma(0), \gamma(1))\|,$$

and the proof is complete. □

Theorem 6.9 *The space of $(\overline{A_1^+ \times A_1^+}, 2)$–mixed structures on S is homeomorphic to* $\mathrm{Core}(\mathcal{T}, \mathcal{T})$.

Proof Let Y denote the set of $(\overline{A_1^+ \times A_1^+}, 2)$–mixed structures on S. We still need to define a topology on Y. We will construct a bijection

$$\varphi\colon Y \to \mathcal{ML}(S) \times \mathcal{ML}(S)$$

with the property that for all $y \in Y$ the core of the product of trees corresponding to $\varphi(y)$ is dual to the $(\overline{A_1^+ \times A_1^+}, 2)$–mixed structure y. We then give Y the topology that makes φ a homeomorphism, thus proving the result.

Given $y \in Y$, let $\gamma_1, \dots, \gamma_n$ be the simple closed curves subdividing S into its laminar and flat parts, as in Definition 6.5. Let S_i for $i = 1, \dots, m$ denote the connected components of $S \setminus \bigcup_j \gamma_j$. If S_i is endowed with a half-translation surface structure induced by a meromorphic quadratic differential q_i of finite area, then the horizontal and vertical foliations of q_i determine a pair of measured laminations $(\lambda_1^i, \lambda_2^i)$. Here we are implicitly using the well-known homeomorphism between the space of measured foliations arising this way and the space of measured laminations; see for instance [Levitt 1983; Lindenstrauss and Mirzakhani 2008]. On the other hand, by Remark 6.3, if S_i carries an $\overline{\mathfrak{a}^+}$–valued measured lamination, then this is equivalent to a pair of measured laminations $(\lambda_1^i, \lambda_2^i)$ possibly containing some boundary curves γ_j in their support. We can then associate to $y \in Y$ the pair of measured laminations $(\lambda_1, \lambda_2) \in \mathcal{ML}(S) \times \mathcal{ML}(S)$ defined as $\lambda_j = \sum_i^m \lambda_j^i$ for $j = 1, 2$. Since the horizontal and vertical measured foliations uniquely determine a meromorphic quadratic differential of finite area [Gardiner and Masur 1991], using Remark 6.3 and Lemma 3.3, it is clear that φ is a bijection.

Moreover, comparing the definition of the map φ with Lemma 6.8, it is easy to verify that the core of the product of trees dual to the pair $\varphi(y)$ is dual to the $(\overline{A_1^+ \times A_1^+}, 2)$–mixed structure y we started with. □

Acknowledgements

Part of this work was carried out when Martone and Ouyang were visiting Rice University during Summer 2021. We thank the Mathematics Department for their hospitality. We thank Francis Bonahon for helpful comments on an earlier version of this manuscript, and Beatrice Pozzetti for helpful comments on this manuscript and for pointing out a mistake in a previous version of the statement of Theorem B and Lemma 6.8. Finally, we thank the anonymous referee for many helpful comments. Ouyang and Tamburelli acknowledge support from the National Science Foundation under grants NSF-DMS:2202832 and NSF-DMS:2005501, respectively.

References

[Alessandrini 2008] **D Alessandrini**, *Tropicalization of group representations*, Algebr. Geom. Topol. 8 (2008) 279–307 MR Zbl

[Alessandrini and Collier 2019] **D Alessandrini**, **B Collier**, *The geometry of maximal components of the* $\mathrm{PSp}(4,\mathbb{R})$ *character variety*, Geom. Topol. 23 (2019) 1251–1337 MR Zbl

[Bestvina 1988] **M Bestvina**, *Degenerations of the hyperbolic space*, Duke Math. J. 56 (1988) 143–161 MR Zbl

[Bonahon 1988] **F Bonahon**, *The geometry of Teichmüller space via geodesic currents*, Invent. Math. 92 (1988) 139–162 MR Zbl

[Brumfiel 1988] **G W Brumfiel**, *The real spectrum compactification of Teichmüller space*, from "Geometry of group representations", Contemp. Math. 74, Amer. Math. Soc., Providence, RI (1988) 51–75 MR Zbl

[Burger and Pozzetti 2017] **M Burger**, **M B Pozzetti**, *Maximal representations, non-Archimedean Siegel spaces, and buildings*, Geom. Topol. 21 (2017) 3539–3599 MR Zbl

[Burger et al. 2017] **M Burger**, **A Iozzi**, **A Parreau**, **M B Pozzetti**, *A structure theorem for geodesic currents and length spectrum compactifications*, preprint (2017) arXiv 1710.07060

[Burger et al. 2021a] **M Burger**, **A Iozzi**, **A Parreau**, **M B Pozzetti**, *Currents, systoles, and compactifications of character varieties*, Proc. Lond. Math. Soc. 123 (2021) 565–596 MR Zbl

[Burger et al. 2021b] **M Burger**, **A Iozzi**, **A Parreau**, **M B Pozzetti**, *The real spectrum compactification of character varieties: characterizations and applications*, C. R. Math. Acad. Sci. Paris 359 (2021) 439–463 MR Zbl

[Burger et al. 2021c] **M Burger**, **A Iozzi**, **A Parreau**, **M B Pozzetti**, *Weyl chamber length compactification of the* $\mathrm{PSL}(2,\mathbb{R})\times\mathrm{PSL}(2,\mathbb{R})$ *maximal character variety*, preprint (2021) arXiv 2112.13624

[Burger et al. 2024] **M Burger**, **A Iozzi**, **A Parreau**, **M B Pozzetti**, *Positive crossratios, barycenters, trees and applications to maximal representations*, Groups Geom. Dyn. 18 (2024) 799–847 MR Zbl

[Collier 2016] **B Collier**, *Maximal* $\mathrm{Sp}(4,\mathbb{R})$ *surface group representations, minimal immersions and cyclic surfaces*, Geom. Dedicata 180 (2016) 241–285 MR Zbl

[Collier et al. 2019] **B Collier**, **N Tholozan**, **J Toulisse**, *The geometry of maximal representations of surface groups into* $\mathrm{SO}_0(2,n)$, Duke Math. J. 168 (2019) 2873–2949 MR Zbl

[Druţu and Sapir 2005] **C Druţu**, **M Sapir**, *Tree-graded spaces and asymptotic cones of groups*, Topology 44 (2005) 959–1058 MR Zbl

[Duchin et al. 2010] **M Duchin**, **C J Leininger**, **K Rafi**, *Length spectra and degeneration of flat metrics*, Invent. Math. 182 (2010) 231–277 MR Zbl

[Eells and Sampson 1964] **J Eells, Jr**, **J H Sampson**, *Harmonic mappings of Riemannian manifolds*, Amer. J. Math. 86 (1964) 109–160 MR Zbl

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

[Fathi et al. 2012] **A Fathi**, **F Laudenbach**, **V Poénaru**, *Thurston's work on surfaces*, Math. Notes 48, Princeton Univ. Press (2012) MR Zbl

[Fock and Goncharov 2006] **V Fock**, **A Goncharov**, *Moduli spaces of local systems and higher Teichmüller theory*, Publ. Math. Inst. Hautes Études Sci. 103 (2006) 1–211 MR Zbl

[Gardiner and Masur 1991] **F P Gardiner**, **H Masur**, *Extremal length geometry of Teichmüller space*, Complex Variables Theory Appl. 16 (1991) 209–237 MR Zbl

[Goldman 1988] **W M Goldman**, *Topological components of spaces of representations*, Invent. Math. 93 (1988) 557–607 MR Zbl

[Guirardel 2005] **V Guirardel**, *Cœur et nombre d'intersection pour les actions de groupes sur les arbres*, Ann. Sci. École Norm. Sup. 38 (2005) 847–888 MR Zbl

[Guirardel and Levitt 2017] **V Guirardel**, **G Levitt**, *JSJ decompositions of groups*, Astérisque 395, Soc. Math. France, Paris (2017) MR Zbl

[Hartman 1967] **P Hartman**, *On homotopic harmonic maps*, Canadian J. Math. 19 (1967) 673–687 MR Zbl

[Hubbard and Masur 1979] **J Hubbard**, **H Masur**, *Quadratic differentials and foliations*, Acta Math. 142 (1979) 221–274 MR Zbl

[Krasnov and Schlenker 2007] **K Krasnov**, **J-M Schlenker**, *Minimal surfaces and particles in 3–manifolds*, Geom. Dedicata 126 (2007) 187–254 MR Zbl

[Labourie 2017] **F Labourie**, *Cyclic surfaces and Hitchin components in rank 2*, Ann. of Math. 185 (2017) 1–58 MR Zbl

[Le 2016] **I Le**, *Higher laminations and affine buildings*, Geom. Topol. 20 (2016) 1673–1735 MR Zbl

[Levitt 1983] **G Levitt**, *Foliations and laminations on hyperbolic surfaces*, Topology 22 (1983) 119–135 MR Zbl

[Lindenstrauss and Mirzakhani 2008] **E Lindenstrauss**, **M Mirzakhani**, *Ergodic theory of the space of measured laminations*, Int. Math. Res. Not. 2008 (2008) art. id. rnm126 MR Zbl

[Loftin 2001] **J C Loftin**, *Affine spheres and convex $\mathbb{RP}^n$–manifolds*, Amer. J. Math. 123 (2001) 255–274 MR Zbl

[Loftin et al. 2022] **J Loftin**, **A Tamburelli**, **M Wolf**, *Limits of cubic differentials and buildings*, preprint (2022) arXiv 2208.07532

[Martone 2019a] **G Martone**, *Positive configurations of flags in a building and limits of positive representations*, Math. Z. 293 (2019) 1337–1368 MR Zbl

[Martone 2019b] **G Martone**, *Sequences of Hitchin representations of tree-type*, preprint (2019) arXiv 1910.13585

[Mess 2007] **G Mess**, *Lorentz spacetimes of constant curvature*, Geom. Dedicata 126 (2007) 3–45 MR Zbl

[Morgan and Otal 1993] **J W Morgan**, **J-P Otal**, *Relative growth rates of closed geodesics on a surface under varying hyperbolic structures*, Comment. Math. Helv. 68 (1993) 171–208 MR Zbl

[Morgan and Shalen 1984] **J W Morgan**, **P B Shalen**, *Valuations, trees, and degenerations of hyperbolic structures, I*, Ann. of Math. 120 (1984) 401–476 MR Zbl

[Oikawa 1956] **K Oikawa**, *Notes on conformal mappings of a Riemann surface onto itself*, Kōdai Math. Sem. Rep. 8 (1956) 23–30 MR Zbl

[Otal 1990] **J-P Otal**, *Le spectre marqué des longueurs des surfaces à courbure négative*, Ann. of Math. 131 (1990) 151–162 MR Zbl

[Ouyang 2023] **C Ouyang**, *High-energy harmonic maps and degeneration of minimal surfaces*, Geom. Topol. 27 (2023) 1691–1746 MR Zbl

[Ouyang and Tamburelli 2021] **C Ouyang**, **A Tamburelli**, *Limits of Blaschke metrics*, Duke Math. J. 170 (2021) 1683–1722 MR Zbl

[Ouyang and Tamburelli 2023] **C Ouyang**, **A Tamburelli**, *Length spectrum compactification of the* $\mathrm{SO}_0(2,3)$*–Hitchin component*, Adv. Math. 420 (2023) art. id. 108997 MR Zbl

[Parreau 2012] **A Parreau**, *Compactification d'espaces de représentations de groupes de type fini*, Math. Z. 272 (2012) 51–86 MR Zbl

[Parreau 2022] **A Parreau**, *Invariant weakly convex cocompact subspaces for surface groups in* A_2*–buildings*, Trans. Amer. Math. Soc. 375 (2022) 2293–2339 MR Zbl

[Paulin 1988] **F Paulin**, *Topologie de Gromov équivariante, structures hyperboliques et arbres réels*, Invent. Math. 94 (1988) 53–80 MR Zbl

[Schoen 1993] **R M Schoen**, *The role of harmonic mappings in rigidity and deformation problems*, from "Complex geometry", Lecture Notes in Pure and Appl. Math. 143, Dekker, New York (1993) 179–200 MR Zbl

[Skora 1996] **R K Skora**, *Splittings of surfaces*, J. Amer. Math. Soc. 9 (1996) 605–616 MR Zbl

[Wolf 1989] **M Wolf**, *The Teichmüller theory of harmonic maps*, J. Differential Geom. 29 (1989) 449–479 MR Zbl

[Wolf 1995] **M Wolf**, *Harmonic maps from surfaces to* $\mathbb{R}$*–trees*, Math. Z. 218 (1995) 577–593 MR Zbl

Department of Mathematics and Statistics, Sam Houston State University
Huntsville, TX, United States

Department of Mathematics, Washington University
St Louis, MO, United States

Department of Mathematics, University of Pisa
Pisa, Italy

gxm120@shsu.edu, ouyang@math.wustl.edu, andrea_tamburelli@libero.it

Received: 29 May 2022 Revised: 18 August 2023

Algebraic & Geometric Topology **24**:7 (2024) 3719–3758
DOI: 10.2140/agt.2024.24.3719
Published: 9 December 2024

An algorithmic discrete gradient field and the cohomology algebra of configuration spaces of two points on complete graphs

Emilio J González
Jesús González

We introduce and study an algorithm that constructs a discrete gradient field on any simplicial complex. With a computational complexity similar to that of existing methods, our algorithmic gradient field is always maximal and in a number of cases even optimal. We make a thorough analysis of the resulting gradient field in the case of Munkres discrete model for $\mathrm{Conf}(K_m, 2)$, the configuration space of ordered pairs of noncolliding particles moving on the complete graph K_m on m vertices. This allows us to describe in full the cohomology algebra $H^*(\mathrm{Conf}(K_m, 2); R)$ for any commutative unital ring R. As an application we prove that, although $\mathrm{Conf}(K_m, 2)$ is outside the "stable" regime, all its topological complexities are maximal when $m \geq 4$.

55R80, 57Q70; 57M15

1 Introduction

Since the development of discrete Morse theory (DMT) by R Forman [15], the concept of a discrete gradient field (DGF) has played an important role in a wide range of areas of mathematics and the sciences alike. The idea arose as a combinatorial analogue of the concept of a smooth gradient field in differential topology, and has proven to be just as important as its smooth predecessor. In particular, DGFs have become one of the main tools in the relatively recent growth of computational topology techniques. For instance, Forman's DMT has been successfully used to deal with noise-reduction problems by Bauer, Lange and Wardetzky [6], as well as in topological data analysis by Harker, Mischaikow, Mrozek and Nanda [22], and within topological visualization and mesh compression applications by Lewiner, Lopes and Tavares [26]. DMT has also seen important applications in the purely theoretical realm, for instance, in the establishment of minimal cellular structures with the homotopy type of the complement of hyperplane arrangements and, more generally, of different sorts of configuration spaces; see Farley [10], Mori and Salvetti [28], Salvetti and Settepanella [32] and Severs and White [33]. DGFs have also been used in the determination of explicit homology bases for complexes of two-connected graphs, objects that play a relevant role in Vassiliev's study of knots in the standard 3–sphere; see Shareshian [34] and Vassiliev [35; 36; 37].

We review the basics on Forman's DMT in Section 2.2. For the purposes of this introduction, the nonspecialized reader should keep in mind that a DGF encodes an organized recipe to stretch the structure of a CW complex X, without changing its homotopy type, with the aim of simplifying the original cell structure. During the stretching process, a typical (regular) cell α gets squeezed by pushing one of its faces β towards the interior of α. The DGF consists of all such pairs (α, β), the "Morse pairings". Cells that are not of the squeezable type nor the pushable type are called *critical* and carry much of the homotopy information of X. Although the roots of the idea go back to Whitehead's simple homotopy theory in the 1930s, DGF technology currently stands as an important alternative to homotopy-minded methods in algebraic topology, especially when the heart of the topological phenomenon under consideration has a combinatorial origin.

In a typical application, the goal is to construct a DGF that renders an efficient and tractable simplification of the cell structure of a given complex. Actually, in each of the applications noted above, the efficiency goal is attained by constructing a suitable though ad-hoc DGF. In contrast, our first main contribution is the description, in Section 3, of an algorithm that constructs, for *any* finite ordered abstract simplicial complex $(K, \preceq)$, a DGF W that reaches reasonable (even notable, in a number of cases) DGF-efficiency goals:

Theorem 1.1 *The discrete gradient field W on $(K, \preceq)$ constructed in Section 3 is maximal. Indeed, all faces and all cofaces of a W–critical face are involved in a Morse pairing.*

In particular, W is a steepness pairing in the sense of Lampret [25, Lemma 2.2]. More importantly, it turns out that in many cases W is either optimal (perhaps after a convenient selection of the vertex ordering $\preceq$), or close to being so. Here optimality refers to the fact that in every dimension $k \geq 0$, the resulting Morse complex, which is homotopy equivalent to $|K|$, has exactly as many k–cells as the k^{th} Betti number of the geometric realization $|K|$ of K. Indeed, the algorithm constructing W can be thought of as a generalization of the inclusion–exclusion (IE) process with respect to a chosen vertex. For instance, the IE process gives an optimal gradient field collapsing a full simplex to the chosen vertex, and our algorithm remains optimal for many other complexes. In fact, for a general ordered simplicial complex $(K, \preceq)$, the vertex ordering $\preceq$ plays a heuristic role that guides the IE process.

The flexibility and generality of our method should lead to many more applications of the sort discussed in the first paragraph of this introduction, both in the theoretical and applied realms. So, in addition to illustrating the efficiency/optimality feature of our algorithmic DGF in a number of standard examples, as our second main contribution we obtain in Section 4 a full description of the cohomology ring of configuration spaces of ordered pairs of points in complete graphs. This is attained through a thorough study of the corresponding algorithmic DGF. Our results in this direction are described in the next paragraphs, after placing our work in context.

Configuration spaces

$$\mathrm{Conf}(X, n) = \{(x_1, \ldots, x_n) \in X^n : x_i \neq x_j \text{ for } i \neq j\}$$

are important ubiquitous objects in mathematics and its applications. They are reasonably well understood when $X = M$, a manifold of dimension at least two. For $X = \Gamma$ a graph, $\mathrm{Conf}(\Gamma, n)$ has attracted much attention in recent years due to its role in geometric group theory, and also because graph configuration spaces provide natural models for the problem of planning collision-free motion of multiple agents performing on a system of tracks; see Farber [8], Ghrist [17] and Ghrist and Koditschek [19]. Yet, the current understanding of the topology of $\mathrm{Conf}(\Gamma, n)$ appears to be far more limited than that of the higher-dimensional case $\mathrm{Conf}(M, n)$. This is due in part to the lack of Fadell–Neuwirth fibrations relating graph configuration spaces for different values of n. Informally, unlike the higher-dimensional counterpart $\mathrm{Conf}(M, n)$, one-dimensional motion planning actually requires global knowledge of the ambient graph. Thus, while additive information about the homology of graph configuration spaces is already available in the literature (see for instance Abrams [1], An, Drummond-Cole and Knudsen [3], Chettih and Lütgehetmann [7], Farber and Hanbury [9], Ghrist [18], Ko and Park [24], Maciążek and Sawicki [27] and Ramos [31]), explicit cup product descriptions seem to be scarcer. Notable exceptions are the work of Farley and Sabalka [11; 13; 14] (see also González and Hoekstra-Mendoza [21]) and Barnett and Farber [4]. The former relates the cohomology algebra of (unordered) configurations on trees to exterior face rings, while the latter describes in full the rational cohomology algebra of ordered pairs of points on planar graphs. We close the gap by focusing on a family of graphs which is diametrically different to that considered by Barnett and Farber. Indeed, we give a full description of the cohomology algebra, with any ring coefficients, of the configuration space of ordered pairs of points on a complete graph K_m with m vertices. The complete description is slightly technical and, for the purposes of this introduction, it is more useful to offer the following detailed navigational chart for Section 4, where the cohomology ring $H^*(\mathrm{Conf}(|K_m|, 2))$ is fully determined.

We start by reviewing a standard combinatorial homotopy model for $\mathrm{Conf}(|K_m|, 2)$ in the introductory Section 2.1. The corresponding algorithmic DGF is described in Proposition 4.1, while the resulting Morse (co)differential is described in Proposition 4.3. Bases of Morse cocycles are described in Definition 4.4 and Proposition 4.6 (for dimension 1), and in Definition 4.8 (for dimension 2). Corresponding cohomological bases are derived in Corollaries 4.7 (for dimension 1) and 4.11 (for dimension 2). The Morse-theoretic cup product is fully determined at the cocycle level by (43) and Propositions 4.13 and 4.14. The cohomological cup product can then be read off from (9) using the full power of Corollary 4.11, which gives explicit formulae that allow us to recover the basis expression of the cohomology class represented by any given Morse 2–cocycle. This renders a complete and fully computer-implementable description of the ring $H^*(\mathrm{Conf}(|K_m|, 2))$.

For the reader's benefit we spell out in Example 4.15 the above navigational chart in the case of the complete graph on five vertices. Our description of the cohomology algebra of $\mathrm{Conf}(|K_5|, 2)$ reflects the well-known fact that this space is homotopy equivalent to a closed orientable surface of genus six. More interestingly, Corollary 4.16 is a simple though partial description of the cup product structure in the cohomology of any $\mathrm{Conf}(|K_m|, 2)$. In such terms, it is clear that certain cup product aspects coming from the homotopy manifold structure of $\mathrm{Conf}(|K_5|, 2)$ are kept for $\mathrm{Conf}(|K_m|, 2)$ when $m > 5$.

We close with an application to motion planning in topological robotics. Namely, after reviewing in Section 5 the basics of Farber and Rudyak's sequential topological complexity TC_s, we use Corollary 4.16

to compute Farber and Rudyak's homotopy invariant in the case of two ordered point-type robots moving without collisions on a track system in the shape of a complete graph:

Theorem 1.2 *For $m \geq 4$ and $s \geq 2$,*

$$\mathrm{TC}_s(\mathrm{Conf}(|K_m|,2)) = s\,\mathrm{hdim}(\mathrm{Conf}(|K_m|,2)) = \begin{cases} s & \text{if } m=4, \\ 2s & \text{otherwise.} \end{cases}$$

Here hdim stands for homotopy dimension. The relevance of Theorem 1.2 is fully discussed in the paragraph of Section 5 containing (46)–(48).

2 Preliminaries

2.1 The Munkres model for 2–particle configuration spaces

Let D be a full subcomplex of a given abstract simplicial complex X, ie assume that every simplex of X whose vertices lie in D is itself a simplex of D. Consider the (necessarily full) subcomplex C of X consisting of the simplices σ of X whose geometric realization $|\sigma|$ is disjoint from $|D|$. The vertices of X are partitioned into those of D and those of C and, as observed in [30, Lemma 70.1], the linear homotopy

$$H\colon (|X|-|D|)\times[0,1]\to |X|-|D|, \quad H(x,s) = (1-s)x + s\sum_{i=1}^{r}\frac{t_i}{\sum_{k=1}^{r} t_k}c_i$$

exhibits $|C|$ as a strong deformation retract of $|X|-|D|$. Here $x=\sum_{i=1}^{r} t_i c_i + \sum_{j=1}^{\rho}\tau_j d_j$ is the barycentric expression of $x\in|X|-|D|$ having $t_i>0<\tau_j$ for all i and j, with $c_1,\dots,c_r$ vertices of C for $r\geq 1$ and $d_1,\dots,d_\rho$ vertices of D for $\rho\geq 0$.

Let K be a finite abstract *ordered* simplicial complex, meaning the vertex set V of K comes equipped with a partial ordering $\preceq$ which is linear upon restriction to any face. We will be interested in Munkres model C above when $X=K\times K$ is the ordered product, with D corresponding to the subcomplex whose geometric realization is the diagonal $\Delta_{|K|}$ in $|K\times K| = |K|\times|K|$. The vertex set of $K\times K$ is $V\times V$, with elements denoted by columns, while a k–simplex of $K\times K$ is a matrix array

$$\begin{bmatrix} v_{0,1} & v_{1,1} & \cdots & v_{k,1} \\ v_{0,2} & v_{1,2} & \cdots & v_{k,2}\end{bmatrix} \tag{1}$$

of elements in V satisfying:

- For $i=1,2$, $v_{0,i}\preceq v_{1,i}\preceq\cdots\preceq v_{k,i}$ with $\{v_{0,i},v_{1,i},\dots,v_{k,i}\}$ an l–face of K (possibly with $l\leq k$).
- For $j=0,1,\dots,k-1$, at least one of the inequalities $v_{j,1}\preceq v_{j+1,1}$ or $v_{j,2}\preceq v_{j+1,2}$ is strict.

Such a matrix-type simplex belongs to D provided its two rows are repeated: $v_{j,1}=v_{j,2}$ for $j=0,1,\dots,k$. In particular, D is a full subcomplex of $K\times K$, and we get a homotopy equivalence

$$|C|\simeq \mathrm{Conf}(|K|,2). \tag{2}$$

Note that a simplex (1) belongs to C precisely when $v_{j,1}\neq v_{j,2}$ for $j=0,1,\dots,k$. In particular, the vertex set of C is $V\times V\setminus\Delta_V$ (with elements denoted by column matrices).

2.2 Discrete Morse theory

We review the notation and facts we need from Forman's discrete Morse theory. See [15; 16] for details.

As in the previous subsection, let K be a finite abstract ordered simplicial complex with ordered vertex set $(V, \preceq)$. Let $(\mathcal{F}, \subseteq)$ be the face poset of K, that is, $\mathcal{F}$ is the set of faces of K partially ordered by inclusion. For a face $\alpha \in \mathcal{F}$, we write $\alpha^{(p)}$ to indicate that α is p–dimensional, and use the notation $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_p]$, where

$$\alpha_0 \prec \alpha_1 \prec \cdots \prec \alpha_p \tag{3}$$

is the ordered list of vertices of α. We choose the orientation on α determined by (3). For faces $\alpha^{(p)} \subset \beta^{(p+1)}$, consider the incidence number $\iota_{\alpha,\beta}$ of α and β, (the coefficient ± 1 of α in the expression of $\partial(\beta)$). Here ∂ stands for the standard boundary operator in the oriented simplicial chain complex $C_*(K)$,

$$\partial([v_0, v_1, \dots, v_i]) = \sum_{0 \le j \le i} (-1)^j \partial_{v_j}([v_0, v_1, \dots, v_i]) = \sum_{0 \le j \le i} (-1)^j [v_0, \dots, \hat{v}_j, \dots, v_i],$$

where $\partial_{v_j}([v_0, v_1, \dots, v_i]) = [v_0, \dots, \hat{v}_j, \dots, v_i]$ is the face obtained by removing v_j from $[v_0, v_1, \dots, v_i]$.

We think of the Hasse diagram $H_{\mathcal{F}}$ of $\mathcal{F}$ as a directed graph: the vertex set of $H_{\mathcal{F}}$ is $\mathcal{F}$ and the directed edges are the ordered pairs $(\alpha^{(p+1)}, \beta^{(p)})$ with $\beta \subset \alpha$. Such a directed edge will be denoted by $\alpha^{(p+1)} \searrow \beta^{(p)}$.

Definition 2.1 A *partial matching* W on $H_{\mathcal{F}}$ is a directed subgraph of $H_{\mathcal{F}}$ whose vertices have degree one. The *W–modified Hasse diagram* $H_{\mathcal{F},W}$ is the directed graph obtained from $H_{\mathcal{F}}$ by reversing the orientation of all edges of W.

Note that the vertex set of W may be a proper subset of $\mathcal{F}$. In such a case, faces in $\mathcal{F}$ that are not vertices of W are called *W–critical.* On the other hand, a reversed edge is denoted by $\beta^{(p)} \nearrow \alpha^{(p+1)}$, in which case α is said to be *W–collapsible* and β is said to be *W–redundant.* The words "critical", "collapsible" and "redundant" will also be used when the partial matching W is implicit from the context.

Definition 2.2 Let W be a partial matching on $H_{\mathcal{F}}$. A *W–path* is an alternating chain of up-going and down-going directed edges of $H_{\mathcal{F},W}$ of either of the two forms

$$\alpha_0 \nearrow \beta_1 \searrow \alpha_1 \nearrow \cdots \nearrow \beta_k \searrow \alpha_k \quad \text{or} \quad \gamma_0 \searrow \delta_1 \nearrow \gamma_1 \searrow \cdots \searrow \delta_k \nearrow \gamma_k. \tag{4}$$

A W–path as the one on the left (resp. right) side of (4) is called an *upper* (resp. *lower*) W–path, and the W–path is called *elementary* (resp. *constant*) when $k = 1$ (resp. when $k = 0$). A *mixed* W–path $\tilde{\lambda}$ from a face $\beta^{(p+1)}$ to a face $\alpha^{(p)}$ is the concatenation of a directed edge $\beta \searrow \gamma$ in $H_{\mathcal{F},W}$ and an upper W–path λ from γ to α.

As above, we use the term "path" as a synonym of "W–path" when the partial matching is implicit from the context. The sets of upper and lower paths that start on a p–cell α and end on a p–cell β are denoted by $\overline{\Gamma}(\alpha, \beta)$ and $\underline{\Gamma}(\alpha, \beta)$, respectively. Note that concatenation of upper/lower paths yields product maps

$$\overline{\Gamma}(\alpha, \beta) \times \overline{\Gamma}(\beta, \gamma) \to \overline{\Gamma}(\alpha, \gamma) \quad \text{and} \quad \underline{\Gamma}(\alpha, \beta) \times \underline{\Gamma}(\beta, \gamma) \to \underline{\Gamma}(\alpha, \gamma). \tag{5}$$

For instance, any nonconstant upper/lower path is a product of corresponding elementary paths.

Definition 2.3 The *multiplicity* of a constant path γ is $\mu(\gamma) := 1$, and of elementary upper/lower paths is

$$\mu(\alpha_0 \nearrow \beta_1 \searrow \alpha_1) := -\iota_{\alpha_0,\beta_1}\iota_{\alpha_1,\beta_1} \quad \text{and} \quad \mu(\gamma_0 \searrow \delta_1 \nearrow \gamma_1) := -\iota_{\delta_1,\gamma_0}\iota_{\delta_1,\gamma_1}.$$

The multiplicity of nonelementary nonconstant paths is defined to be a multiplicative function with respect to the product maps (5). Likewise,

$$\mu(\tilde{\lambda}) := \iota_{\gamma,\beta}\mu(\lambda)$$

defines the multiplicity of the mixed path $\tilde{\lambda}$ given by the concatenation of the edge $\beta \searrow \gamma$ and the upper path $\lambda \in \overline{\Gamma}(\gamma, \alpha)$.

Our central tools are discrete gradient fields:

Definition 2.4 A nonconstant path as in (4) is called a cycle if $\alpha_0 = \alpha_k$ in the upper case, or $\gamma_0 = \gamma_k$ in the lower case. Note that the cycle condition can only hold with $k > 1$. A partial matching W is said to be a *gradient field* on K if no nonconstant path is a cycle. In such a case, paths are referred as *gradient paths.*

Note that W is a gradient field if and only if $H_{\mathcal{F},W}$ has no cycles (as a directed graph).

We close this preliminary section by recalling (Definition 2.5 and Proposition 2.7) the way in which the structure of critical faces and gradient paths between them can be used to assemble a (co)chain complex that recovers the (co)homology of K.

Definition 2.5 Let R be a commutative unital ring.[1] As a graded additive R–module, the *Morse chain complex* $(\mu_*(K), \partial)$ is degreewise R–free, with basis in dimension $p \geq 0$ given by the oriented critical faces $\alpha^{(p)}$ of K, and with *Morse boundary map* $\partial\colon \mu_*(K) \to \mu_{*-1}(K)$ given at a critical face $\alpha^{(p)}$ by

$$\partial(\alpha^{(p)}) = \sum_{\beta^{(p-1)}} \Big(\sum_{\tilde{\lambda}} \mu(\tilde{\lambda})\Big)\beta, \tag{6}$$

where the outer summation runs over all critical faces $\beta^{(p-1)}$, and the inner summation runs over all mixed gradient paths $\tilde{\lambda}$ from α to β. The *Morse cochain complex* $(\mu^*(K), \delta)$ is the R–dual[2] of $(\mu_*(K), \partial)$.

Thus $\mu^p(K)$ is R–free with basis given by the duals of the oriented critical faces $\alpha^{(p)}$ of K. The value of the Morse coboundary map $\delta\colon \mu^*(K) \to \mu^{*+1}(K)$ at a (dualized) critical face $\alpha^{(p)}$ is

$$\delta(\alpha^{(p)}) = \sum_{\beta^{(p+1)}} \Big(\sum_{\tilde{\lambda}} \mu(\tilde{\lambda})\Big)\beta, \tag{7}$$

where the outer summation runs over all (dualized) critical faces $\beta^{(p+1)}$, and the inner summation runs over all mixed gradient paths $\tilde{\lambda}$ from β to α.

[1] We restrict to ring coefficients as we will ultimately be interested in cup products.

[2] For the sake of brevity, we will consistently omit writing asterisks for dualized objects; the context clarifies the intended meaning.

Remark 2.6 For critical faces $\gamma_1^{(p)}$ and $\gamma_2^{(p+1)}$, the multiplicity-counted number of mixed gradient paths λ from γ_2 to γ_1, ie the sum $[\gamma_1;\gamma_2] := \sum_\lambda \mu(\lambda)$, is called the *Morse-theoretic incidence number* of γ_1 and γ_2. In these terms, (6) and (7) take the more familiar forms

$$\partial(\alpha^{(p)}) = \sum_{\beta^{(p-1)}} [\beta;\alpha]\beta \quad \text{and} \quad \delta(\alpha^{(p)}) = \sum_{\beta^{(p+1)}} [\alpha,\beta]\beta.$$

Gradient paths yield a homotopy equivalence between the Morse cochain complex $\mu^*(K)$ and the usual simplicial cochain complex $C^*(K)$. For our purposes we need:

Proposition 2.7 *The formulae*

$$\begin{aligned}\overline{\Phi}(\alpha^{(p)}) &= \sum_{\beta^{(p)}}\Bigg(\sum_{\lambda\in\overline{\Gamma}(\beta,\alpha)} \mu(\lambda)\Bigg)\beta \quad \textit{for } \alpha \textit{ critical and } \beta \textit{ arbitrary,}\\ \underline{\Phi}(\beta^{(p)}) &= \sum_{\alpha^{(p)}}\Bigg(\sum_{\lambda\in\underline{\Gamma}(\alpha,\beta)} \mu(\lambda)\Bigg)\alpha \quad \textit{for } \beta \textit{ arbitrary and } \alpha \textit{ critical,}\end{aligned} \tag{8}$$

determine cochain maps $\overline{\Phi}\colon \mu^*(K) \to C^*(K)$ *and* $\underline{\Phi}\colon C^*(K) \to \mu^*(K)$ *inducing cohomology isomorphisms* $\overline{\Phi}^*$ *and* $\underline{\Phi}^*$ *with* $(\underline{\Phi}^*)^{-1} = \overline{\Phi}^*$.

In particular, cup products can be evaluated directly at the level of the Morse cochain complex $\mu^*(K)$. Indeed, for Morse cocycles $x, y \in \mu^*(K)$ representing respective cohomology classes $x', y' \in H^*(\mu^*(K))$, the Morse-theoretic cohomology cup product $x' \cdot y'$ is represented by the Morse cocycle

$$x \overset{\mu}{\smile} y := \underline{\Phi}(\overline{\Phi}(x) \smile \overline{\Phi}(y)) \in \mu^*(K), \tag{9}$$

where $\smile$ stands for the simplicial cup product.

3 Algorithmic gradient fields

Let K be a finite abstract ordered simplicial complex of dimension d with ordered vertex set $(V,\preceq)$. Recall that the partial order $\preceq$ is required to restrict to a linear order on simplices of K. In this section, we describe and study an algorithm $\mathcal{A}$ that constructs a discrete gradient field W (which depends on $\preceq$) on K.

By the order-extension principle, we may assume $\preceq$ is linear from the outset. Let $\mathcal{F}^i$ denote the set of i–dimensional faces of K. Recall that a face $\alpha^{(i)} \in \mathcal{F}^i$ is identified with the ordered tuple $[\alpha_0, \alpha_1, \ldots, \alpha_i]$, with $\alpha_0 \prec \alpha_1 \prec \cdots \prec \alpha_i$, of its vertices. In such a setting, we say that α_r appears in position r of α. The ordered tuple notation allows us to lexicographically extend $\preceq$ to a linear order (also denoted by $\preceq$) on the set $\mathcal{F}$ of faces of K. We write $\prec$ for the strict version of $\preceq$.

For a vertex $v \in V$, a face $\alpha \in \mathcal{F}^i$ and an integer $r \geq 0$, let

$$\iota_r(v,\alpha) = \begin{cases} \alpha \cup \{v\} & \text{if } \alpha\cup\{v\} \in \mathcal{F}^{i+1} \text{ with } v \text{ appearing in position } r \text{ of } \alpha\cup\{v\},\\ \varnothing & \text{otherwise.}\end{cases}$$

3.1 Acyclicity

At the start of the algorithm we set $W := \varnothing$ and initialize auxiliary variables $F^i := \mathcal{F}^i$ for $0 \le i \le d$ which, at any moment of the algorithm, keep track of i–dimensional faces not taking part in a pairing in W. Throughout the algorithm $\mathcal{A}$, pairings $(\alpha, \beta) \in \mathcal{F}^i \times \mathcal{F}^{i+1}$ are added to W by means of a family of processes $\mathcal{P}^i$ running for $i = d-1, d-2, \dots, 1, 0$ (in that order), where $\mathcal{P}^i$ is executed provided (at the relevant moment) both F^i and F^{i+1} are not empty (so there is a chance to add new pairings to W). Process $\mathcal{P}^i$ consists of three levels of nested subprocesses:

(i) At the most external level, $\mathcal{P}^i$ consists of a family of processes $\mathcal{P}^{i,r}$ for $i+1 \ge r \ge 0$, executed in descending order with respect to r.

(ii) In turn, each $\mathcal{P}^{i,r}$ consists of a family of subprocesses $\mathcal{P}^{i,r,v}$ for $v \in V$, executed from the $\preceq$–largest vertex to the smallest.

(iii) At the innermost level, each process $\mathcal{P}^{i,r,v}$ consists of a family of instructions $\mathcal{P}^{i,r,v,\alpha}$ for $\alpha \in \mathcal{F}^i$, executed following the $\preceq$–lexicographic order.

Instruction $\mathcal{P}^{i,r,v,\alpha}$ checks whether, at the moment of its execution, $(\alpha, \iota_r(v,\alpha))$ is in $F^i \times F^{i+1}$, that is, whether $(\alpha, \iota_r(v,\alpha))$ is "available" as a new pairing. If so, the pairing $\alpha \nearrow \iota_r(v,\alpha)$ is added to W, while α and $\iota_r(v,\alpha)$ are removed from F^i and F^{i+1}, respectively. Two immediate consequences stand from the above construction. Namely, at the end of the algorithm, the resulting family of pairs W is a partial matching in $\mathcal{F}$, and all faces and cofaces of an unpaired cell are involved in a W–paring. The former fact is part of the far more important Proposition 3.1 which, together with the latter, yields Theorem 1.1.

Proposition 3.1 *W is a gradient field.*

In preparation for the proof of Proposition 3.1, we need:

Definition 3.2 Let $W_{i,r,v}$ denote the collection of pairings $\alpha \nearrow \beta$ in W constructed during $\mathcal{P}^{i,r,v}$. Consider also the collection $P_{i,r,v}$ of pairs $(\alpha, \beta) \in \mathcal{F}^i \times \mathcal{F}^{i+1}$ such that $\beta \setminus \alpha = \{v\}$ with v appearing in position r of β. Thus $W_{i,r,v} = P_{i,r,v} \cap W$.

We start by proving that, at the moment that $\mathcal{A}$ constructs a pairing $\alpha \nearrow \beta$, α is in fact the smallest (with respect to $\preceq$) of the facets of β that remain unpaired.

Lemma 3.3 *Let $\alpha = [\alpha_0, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_i] \nearrow \beta = [\alpha_0, \dots, \alpha_r, \beta_0, \alpha_{r+1}, \dots, \alpha_i]$ be a pairing in $W_{i,r+1,\beta_0}$ and let γ be a face of β with $\gamma = [\alpha_0, \dots, \alpha_r, \beta_0, \alpha_{r+1}, \dots, \widehat{\alpha}_j, \dots, \alpha_i]$ for $r+1 \le j \le i$, ie $\gamma \prec \alpha$. Then there is an integer $l \in \{j+1, j+2, \dots, i+1\}$ and a vertex δ_0 with $\alpha_j \prec \delta_0$ such that*

$$\gamma \nearrow \delta := [\alpha_0, \dots, \alpha_r, \beta_0, \alpha_{r+1}, \dots, \alpha_{j-1}, \widehat{\alpha}_j, \dots, \delta_0, \dots]$$

lies in W_{i,l,δ_0}. In particular, the pairing $\gamma \nearrow \delta$ is constructed by $\mathcal{A}$ before the pairing $\alpha \nearrow \beta$.

Proof Previous to the instruction $\mathcal{P}^{i,r+1,\beta_0,\alpha}$ that constructs $\alpha \nearrow \beta$, the algorithm $\mathcal{A}$ executes the instruction $\mathcal{P}^{i,j+1,\alpha_j,\gamma}$ that evaluates the potential pair $(\gamma,\beta) \in P_{i,j+1,\alpha_j}$. This is not an element of W, as β remains available until a later stage in $\mathcal{A}$. So γ must be paired by an instruction $\mathcal{P}^{i,l,\delta_0,\gamma}$ previous to $\mathcal{P}^{i,j+1,\alpha_j,\gamma}$, which forces the conclusion. □

Proof of Proposition 3.1 Assume for a contradiction that there is a W–cycle

(10) $$\alpha^0 \nearrow \beta^0 \searrow \alpha^1 \nearrow \beta^1 \searrow \alpha^2 \nearrow \cdots \nearrow \beta^n \searrow \alpha^{n+1} = \alpha^0$$

(the condition $n \geq 1$ is forced by the definition of a gradient path). Without loss of generality, we can assume that $\alpha^0 \nearrow \beta^0$ is constructed by $\mathcal{A}$ before any other pairing $\alpha^j \nearrow \beta^j$ with $1 \leq j \leq n$. So Lemma 3.3 forces the start of the cycle to have the form

$$\begin{aligned}\alpha^0 &= [\alpha^0_0, \dots, \alpha^0_{j_0}, \alpha^0_{j_0+1}, \dots, \alpha^0_k],\\ \beta^0 &= [\alpha^0_0, \dots, \alpha^0_{j_0}, \beta^0_0, \alpha^0_{j_0+1}, \dots, \alpha^0_k],\\ \alpha^1 &= [\alpha^0_0, \dots, \widehat{\alpha}^0_l, \dots, \alpha^0_{j_0}, \beta^0_0, \alpha^0_{j_0+1}, \dots, \alpha^0_k].\end{aligned}$$

Assume inductively $\alpha^j = [\dots, \beta^0_0, \alpha^0_{j_0+1}, \dots, \alpha^0_k]$ with β^0_0 appearing in position j_0 (so $\alpha^j \neq \alpha^0$). The choosing of $\alpha^0 \nearrow \beta^0$ implies that β^j is obtained from α^j by inserting a vertex v on the left of β^0_0 ($v < \beta^0_0$). A new application of Lemma 3.3 (together with the choosing of $\alpha^0 \nearrow \beta^0$) then shows that α^{j+1} must be obtained from β^j by removing a vertex other than $\beta^0_0, \alpha^0_{j_0+1}, \dots, \alpha^0_k$. Thus $\alpha^{j+1} = [\dots, \beta^0_0, \alpha^0_{j_0+1}, \dots, \alpha^0_k]$, which is again different from α_0. Iterating, we get a situation incompatible with the equality in (10). □

We have noted that, when K is a full simplex, $\mathcal{A}$ constructs the standard (and optimal) gradient field determined by inclusion–exclusion of a fixed vertex (the largest one in the selected order $\preceq$). As illustrated in Examples 3.4, optimality is reached in other standard situations. Example 3.6 and Corollary 4.2 deal with slightly less standard instances, while [20] deals with novel situations in which our gradient field is optimal.

Examples 3.4 Figure 1, left, gives a triangulation of the projective plane $\mathbb{R}P^2$. The gradient field shown by the heavy arrows is determined by $\mathcal{A}$ using the indicated ordering of vertices. The only critical faces are [6] (in dimension 0), [2, 5] (in dimension 1) and [1, 3, 4] (in dimension 2), so optimality of the field

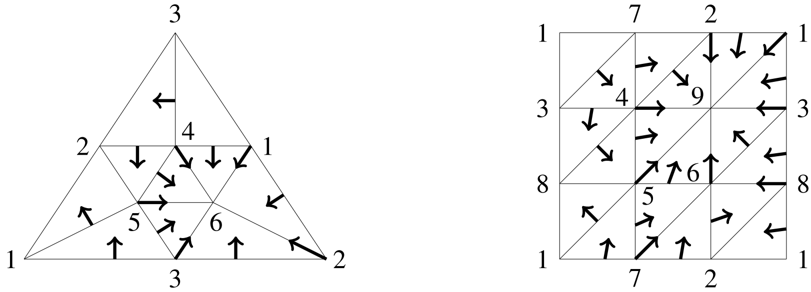


Figure 1: Algorithmic gradient fields for the projective plane (left) and the 2–torus (right).

follows from the known mod-2 homology of $\mathbb{R}P^2$. Although the gradient field depends on the ordering of vertices, we have verified with the help of a computer that, in this case, all possible 720 gradient fields (coming from the corresponding 6! possible orderings of vertices) are optimal. A corresponding optimal gradient field on the 2–torus (and the vertex order rendering it) is shown in Figure 1, right. This time the critical faces are $[9]$ (in dimension 0), $[2,8]$ and $[5,8]$ (in dimension 1) and $[1,3,7]$ (in dimension 2). The torus case is interesting in that there are vertex orderings that yield nonoptimal gradient fields. In general, a plausible strategy for choosing a convenient ordering of vertices consists of assuring the largest possible number of vertices with high $\preceq$–tag so that no two such vertices lie on a common face. For instance, in our torus example, no pair of vertices taken from 7, 8 and 9 lie on a single face.

We address the option $\alpha \prec \gamma$ ruled out by the hypotheses in Lemma 3.3:

Lemma 3.5 *Let $\alpha = [\alpha_0, \dots, \alpha_i, \alpha_{i+1}, \dots, \alpha_k] \nearrow \beta = [\alpha_0, \dots, \alpha_i, \beta_0, \alpha_{i+1}, \dots, \alpha_k]$ lie in $W_{k,i+1,\beta_0}$ and let γ be a face of β with $\alpha \prec \gamma$, ie $\gamma = [\alpha_0, \dots, \widehat{\alpha}_j, \dots, \alpha_i, \beta_0, \alpha_{i+1}, \dots, \alpha_k]$ for $0 \le j \le i$. Assume $\gamma \nearrow \delta$ is a pairing constructed* ***after*** *the pairing $\alpha \nearrow \beta$. Then δ is obtained from γ by inserting a vertex δ_0 which is $\prec$–smaller than β_0, ie $\delta = (\dots, \delta_0, \dots, \beta_0, \alpha_{i+1}, \dots, \alpha_k)$.*

Proof The assertion follows from the definition of the algorithm $\mathcal{A}$, noticing that α_{i+1} appears in position $i+1$ in γ. □

3.2 Gradient fields via a faster algorithm

The proof of Proposition 3.1 makes critical use of "timing" in the construction of W–pairs within the algorithm $\mathcal{A}$. We will modify this characteristic to get a more efficient and faster version of $\mathcal{A}$. While the timing of the W–pairs construction will be altered, we shall show that the new algorithm constructs the same gradient field.

The algorithm $\bar{\mathcal{A}}$ in this subsection, initialized with auxiliary variables $\overline{W}$ and $\bar{F}^i$ analogous to those for its counterpart $\mathcal{A}$, consists of a family of processes $\overline{\mathcal{P}}^i$ running for $i = d-1, d-2, \dots, 1, 0$ (in that order). Each $\overline{\mathcal{P}}^i$ is executed under the same conditions (with respect to $\bar{F}^i$ and $\bar{F}^{i+1}$) as its analogue $\mathcal{P}^i$, but consists only of two (rather than three) levels of nested subprocess. Namely, at the most external level, $\overline{\mathcal{P}}^i$ consists of a family of processes $\overline{\mathcal{P}}^{i,v}$ for $v \in V$, executed from the $\preceq$–largest vertex to the smallest. In turn, each process $\overline{\mathcal{P}}^{i,v}$ consists of a family of instructions $\overline{\mathcal{P}}^{i,v,\alpha}$ for $\alpha \in \mathcal{F}^i$, executed following the $\preceq$–lexicographic order. Instruction $\overline{\mathcal{P}}^{i,v,\alpha}$ checks whether, at that moment, $(\alpha, \{v\} \cup \alpha) \in \bar{F}^i \times \bar{F}^{i+1}$ (availability). If so, the pairing $\alpha \nearrow \{v\} \cup \alpha$ is added to $\overline{W}$, while α and $\{v\} \cup \alpha$ are removed from $\bar{F}^i$ and $\bar{F}^{i+1}$, respectively. Thus the difference with the algorithm $\mathcal{A}$ is that, in order to construct a pairing $\alpha \nearrow \{v\} \cup \alpha$ in $\overline{W}$, we do not care about the position of v in $\{v\} \cup \alpha$. As we will explain next, such a situation means that algorithm $\bar{\mathcal{A}}$ constructs some gradient pairings $\alpha \nearrow \beta$ earlier than they would be constructed by $\mathcal{A}$, thus avoiding the need to perform subsequent testing instructions related to α or β.

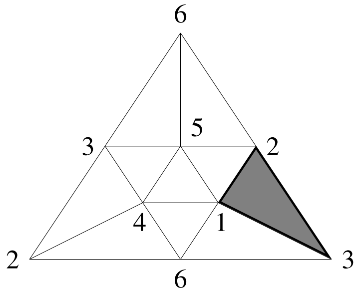


Figure 2: Vertex order in the projective plane with the facet [1, 2, 3] removed.

Example 3.6 Consider the triangulation of the punctured projective plane shown in Figure 2. In the algorithm $\mathcal{A}$, the pairing $[2,3] \nearrow [2,3,4]$, which is constructed during the process $\mathcal{P}^{1,2,4}$, comes before the pairing $[1,5] \nearrow [1,4,5]$, which is constructed during the process $\mathcal{P}^{1,1,4}$. Instead, these two pairings arise in the opposite order in the algorithm $\bar{\mathcal{A}}$, and they both are constructed during the process $\mathcal{Q}^{1,4}$. As the reader can easily check, the (common) resulting gradient field has only two critical faces, namely [6] and [4, 5], and is thus optimal (for the punctured projective plane has the homotopy type of the circle S^1).

The goal of this subsection is to prove Theorem 3.7, which states that $W = \overline{W}$ at the end of both algorithms. The proof is best organized by setting $\overline{W}_{i,r,v} := P_{i,r,v} \cap \overline{W}$ (cf Definition 3.2), as well as

$$W_{k,v} = \bigsqcup_r W_{k,r,v} \quad \text{and} \quad \overline{W}_{k,v} = \bigsqcup_r \overline{W}_{k,r,v}.$$

Theorem 3.7 *The pairings constructed by $\mathcal{A}$ and $\bar{\mathcal{A}}$ agree: $W_{k,r,v} = \overline{W}_{k,r,v}$ for all relevant indices k, r and v. In particular, $\overline{W}$ is acyclic.*

The proof of Theorem 3.7 uses the following elementary observations for vertices v and w with $v \preceq w$:

$$\begin{aligned}(11)\qquad &(\alpha,\beta) \in P_{k,r,v} \text{ and } (\alpha,\gamma) \in P_{k,s,w} \implies r \le s, \text{ with equality provided } v = w,\\ &(\alpha,\beta) \in P_{k,r,v} \text{ and } (\gamma,\beta) \in P_{k,s,w} \implies r \le s, \text{ with equality provided } v = w.\end{aligned}$$

Remark 3.8 In the proof of Theorem 3.7, it will be convenient to keep in mind the following closer view of the central part of algorithms $\mathcal{A}$ and $\bar{\mathcal{A}}$. In the case of $\mathcal{A}$, an efficient way to execute a process $\mathcal{P}^{k,r,v}$ is by assembling the set $N_{k,r,v}$ of $(k+1)$–dimensional faces γ having v in position r and such that both γ and $\partial_v(\gamma)$ are available (neither γ nor $\partial_v(\gamma)$ have been paired previous to the start of $\mathcal{P}^{k,r,v}$). With such a preparation, $\mathcal{P}^{k,r,v}$ simply adds[3] to W all pairs $(\partial_v(\gamma),\gamma)$ with $\gamma \in N_{k,r,v}$ (construction of new pairings), and removes all faces γ and $\partial_v(\gamma)$, for $\gamma \in N_{k,r,v}$, from the corresponding lists of unpaired faces (update of available faces). Likewise, an efficient way to execute process $\overline{\mathcal{P}}^{k,v}$ in $\bar{\mathcal{A}}$ is by assembling the set $\overline{N}_{k,v}$ of $(k+1)$–dimensional faces γ containing v as a vertex (in any position) and such that both γ and $\partial_v(\gamma)$ are available at the start of $\overline{\mathcal{P}}^{k,v}$. With such a preparation, Lemma 3.9 shows that $\overline{\mathcal{P}}^{k,v}$ simply adds to $\overline{W}$ (in lexicographic order) all pairings $(\partial_v(\gamma),\gamma)$ with $\gamma \in \overline{N}_{k,v}$ (construction of new

[3]The adding of pairs is done following the $\preceq$–lexicographic order (cf. Lemma 3.9), though this much is immaterial at this point.

pairings), and removes all faces γ and $\partial_v(\gamma)$, for $\gamma \in \overline{N}_{k,v}$, from the corresponding lists of unpaired faces (update of available faces). In particular, $N_{k,r,v}$ (resp. $\overline{N}_{k,v}$) is the set of collapsible faces for the block of pairings constructed by $\mathcal{P}^{k,r,v}$ (resp. $\overline{\mathcal{P}}^{k,v}$), while the faces $\partial_v(\beta)$ for $\beta \in N_{k,r,v}$ (resp. $\beta \in \overline{N}_{k,v}$) are the corresponding redundant faces.

Lemma 3.9 *Let α and β be $(k+1)$–dimensional faces each containing v as a vertex (in any position). In terms of the $\preceq$–lexicographic order, the condition $\alpha \prec \beta$ holds if and only if $\partial_v(\alpha) \prec \partial_v(\beta)$.*

Proof We provide a detailed proof for completeness. The lexicographic order is linear (we have assumed so at the vertex level), so it suffices to show that $\partial_v(\alpha) \prec \partial_v(\beta)$ provided $\alpha \prec \beta$. Say v appears in positions i and j in α and β, respectively. The result is obvious if $i = j$ (this is why we did not need the lemma in our closer look at $\mathcal{A}$), or if the lexicographic decision for the inequality $\alpha \prec \beta$ is taken at a position smaller than $m := \min\{i, j\}$. Thus we can assume $i \neq j$ with α and β being identical up to and including position $m-1$. The inequality $\alpha \prec \beta$ then forces $i > j = m$. Thus $\partial_v(\alpha)$ and $\partial_v(\beta)$ are identical up to position $j-1$, while in position j,

- $\partial_v(\alpha)$ has the vertex α_j, which is smaller than $v = \alpha_i$, and
- $\partial_v(\beta)$ has the vertex β_{j+1}, which is larger than $v = \beta_j$.

Consequently $\partial_v(\alpha) \prec \partial_v(\beta)$. □

Proof of Theorem 3.7 Recall that d denotes the dimension of the simplicial complex under consideration. Fix $i \in \{0, 1, \dots, d-1\}$ and assume

(12) $$\text{the equality } W_{k,r,v} = \overline{W}_{k,r,v} \text{ is valid whenever } k > i,$$

for all relevant values of r and v. The inductive goal is to prove

(13) $$W_{i,r,v} = \overline{W}_{i,r,v} \quad \text{for all } v \in V \text{ and all } r \in \{0, 1, \dots, i+1\}.$$

(The induction is vacuously grounded by the fact that $W_{d,r,v} = \varnothing = \overline{W}_{d,r,v}$ at the start of both algorithms.) We start by arguing the case $r = i+1$ in (13), which in turn will be done by induction on the reverse ordering of vertices (starting from the largest vertex $v_{\max}$) and through a comparison of the corresponding actions of $\mathcal{A}$ and $\overline{\mathcal{A}}$ during *simultaneous* execution of these algorithms. In detail:

Case I ($r = i+1$ and $v = v_{\max}$ in (13)) Pairings in $W_{i,i+1,v_{\max}}$ are constructed during the execution of process $\mathcal{P}^{i,i+1,v_{\max}}$, while those in $\overline{W}_{i,i+1,v_{\max}}$ are constructed during the execution of process $\overline{\mathcal{P}}^{i,v_{\max}}$. In principle, the latter process would also construct pairings outside $\overline{W}_{i,i+1,v_{\max}}$. However, such a possibility is prevented by the fact that $v_{\max}$ can only appear in the last position of any face. Taking into account the inductive assumption (12), this means that processes $\mathcal{P}^{i,i+1,v_{\max}}$ in $\mathcal{A}$ and $\overline{\mathcal{P}}^{i,v_{\max}}$ in $\overline{\mathcal{A}}$ construct the same new pairings, and consequently perform the same updating of sets of available faces (this justifies the abuse of notation $\mathcal{P}^{i,i+1,v_{\max}} = \overline{\mathcal{P}}^{i,v_{\max}}$). Furthermore, after these processes conclude, no further pairings can be constructed by insertion of $v_{\max}$ (either in $\mathcal{A}$ or in $\overline{\mathcal{A}}$). Thus in fact

(14) $$W_{i,v_{\max}} = W_{i,i+1,v_{\max}} = \overline{W}_{i,i+1,v_{\max}} = \overline{W}_{i,v_{\max}},$$

which in particular grounds the inductive (on the vertices) argument for the case $r = i+1$ in (13). As explained in the paragraph preceding Lemma 3.9, the redundant entries in (14) are the i–dimensional faces α such that $\alpha \cup \{v_{\max}\}$ is an $(i+1)$–dimensional face (so $v_{\max} \notin \alpha$) available at the start of process $\mathcal{P}^{i,i+1,v_{\max}} = \overline{\mathcal{P}}^{i,v_{\max}}$ (all i–dimensional faces are available at this point), whereas the collapsible entries in (14) are the $(i+1)$–dimensional faces available at the start of $\mathcal{P}^{i,i+1,v_{\max}} = \overline{\mathcal{P}}^{i,v_{\max}}$ that contain $v_{\max}$ as a vertex.

The above situation changes slightly in later stages of the algorithms and, in order to better appreciate subtleties, it is highly illustrative to spend a little time analyzing in detail a few of the pairings constructed right after (14).

Case II ($r = i+1$ and $v = v_{\max -1}$ in (13)) Let $v_1, v_2, \dots, v_{\max -1}, v_{\max}$ be the elements of the vertex set V listed increasingly according to $\preceq$. Pairings in $W_{i,i+1,v_{\max -1}}$ (resp. $\overline{W}_{i,i+1,v_{\max -1}}$) are constructed during the execution of the process $\mathcal{P}^{i,i+1,v_{\max -1}}$ (resp. $\overline{\mathcal{P}}^{i,v_{\max -1}}$). In both processes, the construction is done by considering the insertion of $v_{\max -1}$ among available faces (these are common to both algorithms up to this point), either in position $i+1$ in the case of $\mathcal{A}$, or in any position in the case of $\overline{\mathcal{A}}$. As in Case I,

$$\overline{\mathcal{P}}^{i,v_{\max -1}} \text{ might construct pairings outside } \overline{W}_{i,i+1,v_{\max -1}} \tag{15}$$

and

$$\text{any such a pairing would have to lie in } \overline{W}_{i,i,v_{\max -1}}, \tag{16}$$

as $v_{\max -1}$ cannot appear in a position smaller than i in an $(i+1)$–dimensional face. In terms of the notation introduced in the paragraph previous to Lemma 3.9, the possibility in (15) translates into a strict inclusion $N_{i,i+1,v_{\max -1}} \subset \overline{N}_{i,v_{\max -1}}$. However, an element in $\overline{N}_{i,v_{\max -1}} \setminus N_{i,i+1,v_{\max -1}}$ is forced to be an $(i+1)$–dimensional face which, in addition to being available at the start of $\mathcal{P}^{i,i+1,v_{\max -1}}$ and $\overline{\mathcal{P}}^{i,v_{\max -1}}$, has $v_{\max}$ appearing in the last position (for, as indicated in (16), $v_{\max -1}$ appears in the next-to-last position). Such a situation conflicts with the description of collapsible faces noted at the end of Case I, ruling out the possibility in (15). Thus, as above, $\mathcal{P}^{i,i+1,v_{\max -1}} = \overline{\mathcal{P}}^{i,v_{\max -1}}$ and

$$W_{i,v_{\max -1}} = W_{i,i+1,v_{\max -1}} = \overline{W}_{i,i+1,v_{\max -1}} = \overline{W}_{i,v_{\max -1}}. \tag{17}$$

While Cases I and II are essentially identical, the construction of subsequent pairings has a twist whose solution is better appreciated by taking a quick glance at the next block of pairings (those constructed by $\mathcal{P}^{i,i+1,v_{\max -2}}$ in the case of $\mathcal{A}$ and by $\overline{\mathcal{P}}^{i,v_{\max -2}}$ in the case of $\overline{\mathcal{A}}$). Namely, this time the inclusion

$$N_{i,i+1,v_{\max -2}} \subseteq \overline{N}_{i,v_{\max -2}} \tag{18}$$

may actually fail to be an equality, as illustrated in Example 3.6. As a result, the particularly strong forms of assertions (14) and (17) no longer hold true for subsequent blocks of pairings. In any case, what we do recover from (18) — and the discussion previous to Lemma 3.9 — is the fact that $W_{i,i+1,v_{\max -2}} = \overline{W}_{i,i+1,v_{\max -2}}$. We next inductively extend this conclusion to other vertices, and then explain how early pairings constructed in $\overline{\mathcal{A}}$ are eventually recovered in $\mathcal{A}$.

Case III (inductive step settling (13) for $r = i + 1$) Fix a vertex $v \in V$ and assume

(19) $$\overline{W}_{i,i+1,w} = W_{i,i+1,w}$$

whenever $v \prec w$, allowing the possibility that process $\overline{\mathcal{P}}^{i,w}$ in $\bar{\mathcal{A}}$ constructs more pairings than those constructed by the corresponding process $\mathcal{P}^{i,i+1,w}$ in $\mathcal{A}$. In such a setting, faces available at the start of $\overline{\mathcal{P}}^{i,v}$ are necessarily available at the start of $\mathcal{P}^{i,i+1,v}$, so

(20) $$\overline{W}_{i,i+1,v} \subseteq W_{i,i+1,v}.$$

Assume for a contradiction that this inclusion is strict, and pick a pairing

(21) $$(\alpha, \beta) \text{ in } W_{i,i+1,v} \text{ and not in } \overline{W}_{i,i+1,v}.$$

This means that α or β (or both) are not available at the start of $\overline{\mathcal{P}}^{i,v}$; in view of (12), this can only happen provided either

(i) $(\alpha, \beta') \in \overline{W}_{i,r,w} \subseteq P_{i,r,w}$ for some face β', some vertex w and some position r, or

(ii) $(\alpha', \beta) \in \overline{W}_{i,r,w} \subseteq P_{i,r,w}$ for some face α', some vertex w and some position r,

where in either case $v \prec w$ and $r \le i+1$. But $(\alpha, \beta) \in W_{i,i+1,v} \subseteq P_{i,i+1,v}$, so (11) yields in fact $i+1 = r$. Thus α or β is part of a pairing in $\overline{W}_{i,r,w} = \overline{W}_{i,i+1,w} = W_{i,i+1,w}$, where the latter equality comes from (19) but contradicts (21). Thus (20) is an equality. Note that the above argument does not rule out the possibility that $\overline{\mathcal{P}}^{i,v}$ constructs more pairings (by inserting v at a position smaller than $i+1$) than are constructed by $\mathcal{P}^{i,i+1,v}$.

The conclusion of the proof of Theorem 3.7 — the proof of (13) for $r \le i$ — proceeds by (inverse) induction on r, with the above discussion for $r = i+1$ grounding the induction. The new inductive argument requires an entirely different viewpoint coming from the following fact: in $\mathcal{A}$, after $\mathcal{P}^{i,i+1,v_1}$ is over, process $\mathcal{P}^i$ continues with many more subprocesses, the first of which is $\mathcal{P}^{i,i,v_{\max}}$. Yet in $\bar{\mathcal{A}}$, process $\overline{\mathcal{P}}^i$ finishes as soon $\overline{\mathcal{P}}^{i,v_1}$ is over, ie when the final inductive stage in Case III concludes. Therefore, our proof strategy from this point on requires pausing $\bar{\mathcal{A}}$ in order to analyze the rest of the actions in $\mathcal{P}^i$. In particular, we explain next how $\mathcal{P}^i$ catches up with all the "early" pairings $\bigcup_v (\overline{W}_{i,v} \setminus W_{i,i+1,v})$ constructed by $\overline{\mathcal{P}}^i$.

Case IV (double inductive step settling (13) for any r) Fix $r \in \{0, 1, \dots, i\}$ and assume inductively that, as $\mathcal{P}^i$ progresses, $\mathcal{P}^{i,\rho}$ yields $\overline{W}_{i,\rho,w} = W_{i,\rho,w}$ for any vertex w and any position of insertion $\rho > r$. (The induction is grounded by Case III above.) The goal is to prove

(22) $$\overline{W}_{i,r,w} = W_{i,r,w} \quad \text{for all vertices } w.$$

Since $r \le i$, we get $\overline{W}_{i,r,v_{\max}} = \varnothing = W_{i,r,v_{\max}}$. We can therefore assume in a second inductive level that, for some vertex v with $v \prec v_{\max}$, (22) holds true for all vertices w with $v \prec w$. The updated goal is to prove $\overline{W}_{i,r,v} = W_{i,r,v}$.

Inclusion $\overline{W}_{i,r,v} \subseteq W_{i,r,v}$ Suppose for a contradiction that

$$(\alpha,\beta) \in \overline{W}_{i,r,v} \tag{23}$$

is an "early" pairing (constructed during the execution of $\overline{\mathcal{P}}^{i,v}$) that cannot be constructed during the execution of $\mathcal{P}^{i,r,v}$. Then α or β (or both) must be involved as a pairing of some $W_{i,s,w}$ with $s \geq r$, and in addition with $v \prec w$ if in fact $s = r$. The double inductive equality $W_{i,s,w} = \overline{W}_{i,s,w}$, the dynamics of $\bar{\mathcal{A}}$ and (23) then force $v = w$, and consequently $s > r$. But this inequality contradicts (11) since $\overline{W}_{i,r,v} \subseteq P_{i,r,v}$ and $\overline{W}_{i,s,w} \subseteq P_{i,s,w}$.

Inclusion $W_{i,r,v} \subseteq \overline{W}_{i,r,v}$ Suppose for a contradiction that

$$(\alpha,\beta) \in W_{i,r,v} \tag{24}$$

is not one of the "early" pairings constructed during the execution of $\overline{\mathcal{P}}^{i,v}$. Then α or β (or both) must be involved in a pairing of some $\overline{W}_{i,s,w}$ with $v \prec w$. As in the previous paragraph, (11) then yields $r \leq s$. In turn, the double inductive hypothesis gives $\overline{W}_{i,s,w} = W_{i,s,w}$, which thus contains a pairing involving α or β, in contradiction to (24). □

3.3 Computational complexity and performance

Designing efficient algorithms and implementing fast software for the homological processing of large data sets is a lively technological challenge. With this in mind, we now study the computational complexity of our algorithm, and compare it with a closely related technique used within the realm of current applications.

Harker et al. [22] describe and study an efficient way of computing homology of complexes and their maps. At the core of their method there is an algorithm $\mathcal{H}$ for constructing discrete gradient fields on a well-suited class of complexes (à la Tucker). The idea is based on a Morse theory extension of the coreduction method introduced in [29]. Namely, cells α and β form a coreduction pair of a complex K provided α is a codimension-1 free face of β in K. Initializing K to be the whole initial complex, the algorithm $\mathcal{H}$ constructs Morse pairings $\alpha \nearrow \beta$ whenever α and β form a coreduction pair in K. Each time such a coreduction pair is found, its entries are removed from K before looking for the next pair of coreduction cells. If at any moment no coreduction pairs exist in K, faces of K with no boundary in K are declared to be critical (and removed from K) until creating new coreduction pairs. The algorithm repeats until K is empty, which completes the basic (iterative) building process $\mathcal{H}_0$ of $\mathcal{H}$.

Although both $\bar{\mathcal{A}}$ and $\mathcal{H}$ are based on a heuristic search of Morse pairs, the corresponding gradient fields bear no resemblance to each other. Indeed, the coreduction heuristic in $\mathcal{H}$ is replaced in $\bar{\mathcal{A}}$ by an inclusion–exclusion strategy guided by the chosen vertex ordering $\preceq$. More precisely, we highlight the main conceptual differences and apparent similarities between $\bar{\mathcal{A}}$ and $\mathcal{H}$. For starters, note that there is no reason to expect (and indeed it is usually not the case) that $\mathcal{H}_0$ yields a reasonably efficient gradient field on the original complex. Nonetheless the Tucker viewpoint of complexes allows Harker et al. to iterate $\mathcal{H}_0$ and, in doing so, the eventually stabilized gradient field turns out to be reasonably efficient — optimal in some cases. On the other hand, as illustrated in Examples 3.4 and formalized by Theorem 1.1, the corresponding efficiency property is reached by $\bar{\mathcal{A}}$ as a result of the $\preceq$–guided search of Morse pairs. But the most important issue to stress when comparing $\bar{\mathcal{A}}$ and $\mathcal{H}$ is given in terms of computational complexity. Aside from

(a) the cost of iterating $\mathcal{H}_0$ until reaching a stable gradient field, and

(b) the cost of processing the resulting Morse complex at the conclusion of each application of $\mathcal{H}_0$ in order to gather the Tucker information needed for the next application of $\mathcal{H}_0$,

the computational cost of applying $\mathcal{H}_0$ (say for the first time) is linear on the starting *complex mass*

$$m_K = \text{card}\{(\alpha^{p+1}, \beta^p) : p \geq 0 \text{ and } \alpha, \beta \text{ are faces of } K \text{ with } \beta \subset \alpha\}. \tag{25}$$

See [22, Proposition 5.1]. We prove (Proposition 3.10) that, if we think of the basic $\bar{\mathcal{A}}$–instruction $\overline{\mathcal{P}}^{i,v,\alpha}$ in Section 3.2 as being performed in $O(1)$ time,[4] then $\bar{\mathcal{A}}$ also executes in $O(m_K)$ time. Consequently, for practical implementations, a profitable strategy reducing computational costs coming from (a) and (b) above can be based on a combination of algorithms $\bar{\mathcal{A}}$ and $\mathcal{H}$. In fact, the maximality condition of the gradient field W resulting from an initial application of $\bar{\mathcal{A}}$ can potentially be bypassed (and possibly turned into an optimality condition) by means of a subsequent (and then much quicker) application of $\mathcal{H}$ on the Morse–Tucker complex resulting from W.

Proposition 3.10 *For a finite abstract ordered simplicial complex K with complex mass (25), algorithm $\bar{\mathcal{A}}$ executes in $O(m_K)$ time.*

Proof An efficient implementation of $\bar{\mathcal{A}}$ requires initializing a couple of functions, $f(\sigma)$ and $g(i,v)$. The former function is binary and answers, at any moment of the algorithm, the question of whether a given face σ of the original complex belongs to the set of "available" faces $\overline{F}^{\dim(\sigma)}$ (here and below we reuse the notation set forth in Section 3.2). The latter function reports, at any moment of the algorithm, the list of available faces in a given dimension i that contain a given vertex v. The cost of initializing f (with values True) is linear on the number of faces of the original complex, and therefore can be safely neglected for the purposes of this proof. On the other hand, for a given dimension i, we start by setting $g(i,v) = \varnothing$ for all vertices v. Then, for each face $\sigma \in \mathcal{F}^i$ and for each vertex $v \in \sigma$, we append σ to $g(i,v)$. This last task takes $O\big(\sum_{\sigma \in \mathcal{F}^i} m_\sigma\big)$ time, where m_σ stands for the *boundary mass* of σ, ie the cardinality of the set of facets of σ. Thus initializing g (and f) takes $O(m_K)$ time. With this preparation, $\bar{\mathcal{A}}$ can then be executed in $O(m_K)$ time following the indications in Remark 3.8. Namely, to execute process $\overline{\mathcal{P}}^{i,v}$, select the faces $\sigma \in g(i+1,v)$ with $f(\partial_v(\sigma)) = \text{True}$, so that

$$\partial_v(\sigma) \nearrow \sigma \tag{26}$$

is a new Morse pair — in which case the corresponding values of f and g have to be updated. Naturally, some faces $\sigma \in g(i+1,v)$ will not lead to the "v–type" Morse pairing (26), and will have to be accounted for by later processes $\overline{\mathcal{P}}^{i,w}$ (with $w < v$). But, just as in the initialization of g, a fixed such σ will have to be processed at most m_σ times. Therefore the actual algorithm $\bar{\mathcal{A}}$ is executed in $O(m_K)$ time too. □

[4] Such an assumption is easily achieved through an efficient implementation of $\bar{\mathcal{A}}$; see the proof of Proposition 3.10.

Note that the estimation $O(m_K)$ for the complexity in the final part of the previous proof is rather coarse. Indeed, none of the i–dimensional redundant faces paired during the execution of $\overline{\mathcal{P}}^i$ will have to be processed during the execution of $\overline{\mathcal{P}}^{i-1}$. It is in this sense that the maximality of our algorithm (Theorem 1.1) leads, paradoxically, to a computational complexity that, in practical situations, is lower than what is estimated here.

3.4 Collapsibility conditions

This section is devoted to theoretical aspects of our gradient field. Precisely, we describe a set of "local" conditions that allow us to identify gradient pairings without having to actually run any of the two versions of our algorithm. Our local conditions determine in full the gradient field in a number of instances.[5] The main result in this section (Theorem 3.19) is presented through a series of preliminary complexity-increasing results in order to isolate the role of each of the condition ingredients.

Definition 3.11 A vertex α_i of a face $\alpha = [\alpha_0, \dots, \alpha_k] \in \mathcal{F}^k$ is said to be maximal in α if $\partial_{\alpha_i}(\alpha) \cup \{v\} \notin \mathcal{F}^k$ for all vertices v with $\alpha_i \prec v$. When α_i is nonmaximal in α, we write $\alpha(i) := \partial_{\alpha_i}(\alpha) \cup \{\alpha^i\}$, where

$$\alpha^i := \max\{v \in V : \alpha_i \prec v \text{ and } \partial_{\alpha_i}(\alpha) \cup \{v\} \in \mathcal{F}^k\}.$$

Note that α^i is maximal in $\alpha(i)$, and that α^i is not a vertex of α. Iterating the construction, for a given face $\alpha = [\alpha_0, \dots, \alpha_k] \in \mathcal{F}^k$ and a sequence of integers $0 \le i_1 < i_2 < \cdots < i_p \le k$, we say that the face $[\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_p}]$ is nonmaximal in α provided:

- α_{i_1} is nonmaximal in α, so we can form the face $\alpha(i_1)$.
- α_{i_2} is nonmaximal in $\alpha(i_1)$, so we can form the face $\alpha(i_1, i_2) := \alpha(i_1)(i_2)$.

  $\vdots$
- α_{i_p} is nonmaximal in $\alpha(i_1, \dots, i_{p-1})$, so we can form the face $\alpha(i_1, \dots, i_p) := \alpha(i_1, \dots, i_{p-1})(i_p)$.

When $p = 0$ (so there is no constructing process), $\alpha(i_1, i_2, \dots, i_p)$ is interpreted as α.

Lemma 3.12 *No vertex of a redundant k–face $\alpha \in \mathcal{F}^k$ is maximal in α.*

Proof Assume a pairing $\alpha = [\alpha_0, \dots, \alpha_k] \nearrow \beta = [\alpha_0, \dots, \alpha_{r-1}, \beta_0, \alpha_r, \dots, \alpha_k]$ and consider a vertex α_i of α. If $i < r$, the k–face $\partial_{\alpha_i}(\beta) = [\alpha_0, \dots, \widehat{\alpha}_i, \dots, \alpha_{r-1}, \beta_0, \alpha_r, \dots, \alpha_k]$ shows that the vertex α_i is nonmaximal in α. If $i \ge r$, Lemma 3.3 gives a pairing

$$\gamma := [\alpha_0, \dots, \alpha_{r-1}, \beta_0, \alpha_r, \dots, \widehat{\alpha}_i, \dots, \alpha_k] \nearrow [\alpha_0, \dots, \alpha_{r-1}, \beta_0, \alpha_r, \dots, \widehat{\alpha}_i, \dots, \delta_0, \dots] =: \delta$$

by insertion of a vertex δ_0 with $\alpha_i \prec \delta_0$, so that the k–face $\partial_{\beta_0}(\delta)$ shows that vertex α_i is nonmaximal in α. □

While maximal vertices in a face α can be thought of as giving obstructions for redundancy of α, maximality of the largest vertex in α is actually equivalent to collapsibility of α in a specific way:

[5] This holds, for instance, in the case of the projective plane and the torus in Examples 3.4, as well as in our application to spaces of ordered pairs of points on complete graphs; see Section 4.1.

Corollary 3.13 *The following conditions are equivalent for a k–face $\alpha = [\alpha_0, \dots, \alpha_k] \in \mathcal{F}^k$:*

(i) *α_k is maximal in α.*

(ii) *$\partial_{\alpha_k}(\alpha) \nearrow \alpha$.*

Proof Assuming (i), both α and $\partial_{\alpha_k}(\alpha)$ are available at the start of process $\mathcal{P}^{k-1,k,\alpha_k}$; the former face in view of Lemma 3.12, and the latter face by the maximality hypothesis. The W–pairing in (ii) is therefore constructed by the process $\mathcal{P}^{k-1,k,\alpha_k}$. On the other hand, if (i) fails, there is a vertex v of K which is maximal with respect to the conditions $\alpha_k \prec v$ and $\partial_{\alpha_k}(\alpha) \cup \{v\} \in \mathcal{F}^k$. As v is maximal in $\partial_{\alpha_k}(\alpha) \cup \{v\} = [\alpha_0, \dots, \alpha_{k-1}, v]$, the argument in the previous paragraph gives $[\alpha_0, \dots, \alpha_{k-1}] \nearrow [\alpha_0, \dots, \alpha_{k-1}, v]$, thus ruling out the W–pairing in (ii). □

Under additional restrictions (spelled out in (27)), maximality of other vertices also forces collapsibility in a specific way. We start with the case of the next-to-last vertex, where the additional restrictions are simple, yet the if and only if situation in Corollary 3.13 is lost; see Remark 3.15.

Proposition 3.14 *Let $\alpha = [\alpha_0, \dots, \alpha_k] \in \mathcal{F}^k$. If α_{k-1} is maximal in α but α_k is not, then $\partial_{\alpha_{k-1}}(\alpha) \nearrow \alpha$.*

Proof By Lemma 3.12, α is available at the start of process $\mathcal{P}^{k-1}$ and, in fact, at the start of process $\mathcal{P}^{k-1,k-1,\alpha_{k-1}}$, in view of Corollary 3.13 and the hypothesis on α_k. The asserted pairing follows since $\partial_{\alpha_{k-1}}(\alpha) = [\alpha_0, \dots, \widehat{\alpha}_{k-1}, \alpha_k]$ is also available at the start of process $\mathcal{P}^{k-1,k-1,\alpha_{k-1}}$. Indeed, a potential pairing $\partial_{\alpha_{k-1}}(\alpha) \nearrow \partial_{\alpha_{k-1}}(\alpha) \cup \{v\}$ constructed at a stage before $\mathcal{P}^{k-1,k-1,\alpha_{k-1}}$ would have $\alpha_{k-1} \prec v$, contradicting the maximality of α_{k-1} in α. □

Remark 3.15 Consider the gradient field on the projective plane in Examples 3.4. Neither 5 nor 2 are maximal in $[1,2,5]$ (due to the faces $[1,2,6]$ and $[1,3,5]$), yet the pairing $[1,5] \nearrow [1,2,5]$ holds.

More generally,

Proposition 3.16 *For a face $\alpha = [\alpha_0, \dots, \alpha_k] \in \mathcal{F}^k$ and an integer $r \in \{0, 1, \dots, k\}$ with α_r maximal in α, the pairing $\partial_{\alpha_r}(\alpha) \nearrow \alpha$ holds provided*

(27) *for any sequence $r+1 \le t_1 < \cdots < t_p \le k$, the face $[\alpha_{t_1}, \dots, \alpha_{t_p}]$ is nonmaximal in α.*

Proof We argue by decreasing induction on $r = k, k-1, \dots, 0$. The grounding cases $r = k$ and $r = k-1$ are covered by Corollary 3.13 and Proposition 3.14, respectively. For the inductive step, the maximality of α_r in α assures both that α is available at the start of $\mathcal{P}^{k-1}$ (Lemma 3.12), and that $\partial_{\alpha_r}(\alpha)$ is available at the start of $\mathcal{P}^{k-1,r,\alpha_r}$. It thus suffices to note that (27) implies that α is also available at the start of $\mathcal{P}^{k-1,r,\alpha_r}$. But a potential pairing $[\alpha_0, \dots, \widehat{\alpha}_{t_1}, \dots, \alpha_k] \nearrow [\alpha_0, \dots, \alpha_k]$ previous in $\mathcal{A}$ to the intended pairing $\partial_{\alpha_r}(\alpha) \nearrow \alpha$, ie with $t_1 \in \{r+1, \dots, k\}$, is inductively ruled out by the (yet previous in $\mathcal{A}$) pairing

$$[\alpha_0, \dots, \widehat{\alpha}_{t_1}, \dots, \alpha_k] \nearrow \alpha(t_1) = [\alpha_0, \dots, \widehat{\alpha}_{t_1}, \{\alpha^{t_1}, \alpha_{t_1+1}, \dots, \alpha_k\}],$$

where the use of curly braces is meant to indicate that α^{t_1} may occupy any position among the ordered vertices $\alpha_{t_1+1}, \dots, \alpha_k$. □

Not all conditions in (27) would be needed in concrete instances of Proposition 3.16. For instance, this will be (recursively) the case if, in the previous proof, some α^{t_1} turns out to be larger than some of the vertices $\alpha_{t_1+1}, \dots, \alpha_k$.

Example 3.17 The pairing $\partial_{\alpha_{k-2}}(\alpha) \nearrow \alpha = [\alpha_0, \dots, \alpha_k]$ holds provided

(i) α_{k-2} is maximal in α,

(ii) α_{k-1} is nonmaximal in α, and

(iii) α_k is nonmaximal in α as well as in $\alpha(k-1)$.

Note that (ii) is used in order to state (iii).

Theorem 3.19, a far-reaching extension of Proposition 3.16, provides sufficient conditions that allow us to identify "exceptional" pairings such as the one noted in Remark 3.15.

Definition 3.18 A vertex α_r of a face $\alpha = [\alpha_0, \dots, \alpha_k] \in \mathcal{F}^k$ is said to be collapsing in α provided

(i) the face α is not redundant,

(ii) condition (27) holds, and

(iii) for every v with $\alpha_r \prec v$ and $\partial_{\alpha_r}(\alpha) \cup \{v\} \in \mathcal{F}^k$, there is a vertex α_j of α with $v \prec \alpha_j$ such that α_j is collapsing in $\partial_{\alpha_r}(\alpha) \cup \{v\}$.

Definition 3.18(i) and (iii) hold when α_r is maximal in α. Note the recursive nature of Definition 3.18.

Theorem 3.19 *If α_r is collapsing in α, then $\partial_{\alpha_r}(\alpha) \nearrow \alpha$.*

Proof The proof is parallel to that of Proposition 3.16. This time the induction is grounded by Corollary 3.13 and the observation that, when $r = k$, Definition 3.18(iii) implies that α_k is maximal in α. The rest of the argument in the proof of Proposition 3.16 applies with two minor adjustments. First, Lemma 3.12 is not needed — neither can it be applied — in view of condition (i). Second, the fact that $\partial_{\alpha_r}(\alpha)$ is available at the start of $\mathcal{P}^{k-1,r,\alpha_r}$ comes directly from (iii) and induction. □

4 Application to configuration spaces

We use the gradient field in the previous section to describe the cohomology ring of the configuration space of ordered pairs of points on a complete graph.

4.1 Gradient fields on the Munkres homotopy simplicial model

Let K_m be the 1–dimensional skeleton of the full $(m-1)$–dimensional simplex on vertices $V_m = \{1, 2, \dots, m\}$. Thus $|K_m|$ is the complete graph on the m vertices. The homotopy type of $\mathrm{Conf}(|K_m|, 2)$ is well understood for $m \leq 3$, so we assume $m \geq 4$ from now on. We think of K_m as an ordered simplicial complex with the natural order on V_m, and study $\mathrm{Conf}(|K_m|, 2)$ through its simplicial homotopy model

C_m in (2). The condition $m \geq 4$ implies that C_m is a pure 2–dimensional complex, ie all of its maximal faces have dimension 2. Furthermore, 2–dimensional faces of C_m have one of the forms

$$\begin{bmatrix} a & a & d \\ b & c & c \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a' & c' & c' \\ b' & b' & d' \end{bmatrix} \tag{28}$$

where

$$d > a \notin \{b, c\}, \quad b < c \neq d, \quad d' > b' \notin \{a', c'\} \quad \text{and} \quad a' < c' \neq d'. \tag{29}$$

Note that the matrix-type notation in (28) is compatible with the notation $\alpha = [\alpha_0, \dots, \alpha_k]$ in previous sections; each α_i now stands for a column-type vertex $\substack{a \\ b}$ (with $a \neq b$). In what follows, the conditions in (29) on the integers $a, b, c, d, a', b', c', d' \in V_m$ will generally be implicit and omitted when writing a 2–simplex or one of its faces. For instance, the forced relations $a \neq b < d \neq a$ are omitted in item (a) of:

Proposition 4.1 *Let W_m be the gradient field on C_m constructed in Section 3 with respect to the lexicographic order on the vertices $\substack{a \\ b} = (a, b) \in V_m \times V_m \setminus \Delta_{V_m}$ of C_m. The full list of W_m–pairings is:*

(a) $\left[\begin{smallmatrix} a & a \\ b & d \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & a & m \\ b & d & d \end{smallmatrix}\right]$ *for* $a < m > d$.

(b) $\left[\begin{smallmatrix} a & a \\ b & m \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & a & m-1 \\ b & m & m \end{smallmatrix}\right]$ *for* $a < m-1$.

(c) $\left[\begin{smallmatrix} a & c \\ b & b \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & c & c \\ b & b & m \end{smallmatrix}\right]$ *for* $b < m > c$.

(d) $\left[\begin{smallmatrix} a & m \\ b & b \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & m & m \\ b & b & m-1 \end{smallmatrix}\right]$ *for* $b < m-1$.

(e) $\left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & c & c \\ b & b & d \end{smallmatrix}\right]$ *for* $a < c, b < d, b \neq c$ *and either* $c < m > d$ *or* $c = m > d+1$.

(f) $\left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ *for* $a < c, b < d, a \neq d$ *and either* $b = c < m > d$ *or* $c+1 < m = d$.

(g) $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & m \\ b & m-1 \end{smallmatrix}\right]$ *for either* $b < m-1$ *or* $a < m-1 = b$.

(h) $\left[\begin{smallmatrix} a \\ m \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} a & m-1 \\ m & m \end{smallmatrix}\right]$ *for* $a < m-1$.

(i) $\left[\begin{smallmatrix} m-1 \\ m \end{smallmatrix}\right] \nearrow \left[\begin{smallmatrix} m-1 & m-1 \\ m-2 & m \end{smallmatrix}\right]$.

In particular:

(j) *In dimension* 0, *the critical face is the vertex* $\left[\begin{smallmatrix} m \\ m-1 \end{smallmatrix}\right]$.

(k) *In dimension* 1, *the critical faces are the simplices*

(k.1) $\left[\begin{smallmatrix} a & m-1 \\ b & m \end{smallmatrix}\right]$ *with either* $a = m-1 > b+1$ *or* $a < m-1 \geq b$,

(k.2) $\left[\begin{smallmatrix} m & m \\ b & d \end{smallmatrix}\right]$ *with* $d < m-1$,

(k.3) $\left[\begin{smallmatrix} a & c \\ m & m \end{smallmatrix}\right]$ *with* $c < m-1$.

(l) *In dimension* 2, *the critical faces are the simplices* $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ *with* $b \neq c < m > d$.

Note that the condition $a \neq d$ in (f) is forced to hold in the stronger form $a < d$.

Proof All pairings, except for the one in (f) when $b \neq c$ (so that $c+1 < m = d$), are given by Corollary 3.13 and Proposition 3.14. The exceptional case requires the stronger Theorem 3.19. On the other hand, direct inspection shows that the faces listed as critical are precisely those not taking part in the list of W_m–pairings. The proof is then complete by observing that the criticality of any d–dimensional face α in (j)–(l) is forced by the fact that all possible $(d-1)$–faces and all possible $(d+1)$–cofaces of α are involved in one of the pairings (a)–(i). For instance, a face $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ in (l) is not collapsible since the three potential pairings

$$\begin{bmatrix} a & a \\ b & d \end{bmatrix} \nearrow \begin{bmatrix} a & a & c \\ b & d & d \end{bmatrix}, \quad \begin{bmatrix} a & c \\ b & d \end{bmatrix} \nearrow \begin{bmatrix} a & a & c \\ b & d & d \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a & c \\ d & d \end{bmatrix} \nearrow \begin{bmatrix} a & a & c \\ b & d & d \end{bmatrix}$$

are ruled out by (a), (e) and (c), respectively. □

The next-to-last sentence in the proof above reflects the maximality of W_m in Theorem 1.1. On the other hand, a straightforward counting shows that the number c_d of critical faces in dimension $d \in \{0,1,2\}$ is given by

$$c_0 = 1, \quad c_1 = 2(m-2)^2 - 1 \quad \text{and} \quad c_2 = \tfrac{1}{4}(m-1)(m-2)(m-3)(m-4). \tag{30}$$

In particular, the Euler characteristic of $\mathrm{Conf}(|K_m|,2)$ is given by

$$\tfrac{1}{4}m(m^3 - 10m^2 + 27m - 18),$$

which yields an explicit expression for the conclusion of [4, Corollary 1.2] in the case of complete graphs. Note in particular that the gradient field W_4 is optimal:

Corollary 4.2 *There is a homotopy equivalence* $\mathrm{Conf}(|K_4|,2) \simeq \bigvee_7 S^1$.

Corollary 4.2 should be compared to the fact that the configuration space of *unordered* pairs of points in $|K_4|$ has the homotopy type of $\bigvee_4 S^1$; see [12, Example 4.5].

4.2 The Morse cochain complex

The Morse coboundary map $\delta\colon \mu^i(C_m) \to \mu^{i+1}(C_m)$ is forced to vanish for $i = 0$ since $c_0 = 1$. It is more interesting to describe the situation for $i = 1$:

Proposition 4.3 *The coboundary $\delta\colon \mu^1(C_m) \to \mu^2(C_m)$ vanishes on the duals of the critical faces of types $(k.2)$ and $(k.3)$ in Proposition 4.1. For the duals of the critical faces of type $(k.1)$ we have*

$$\delta\left(\begin{bmatrix} a & m-1 \\ b & m \end{bmatrix}\right) = \sum \begin{bmatrix} a & a & x \\ y & b & b \end{bmatrix} - \sum \begin{bmatrix} a & a & x \\ b & y & y \end{bmatrix} + \sum \begin{bmatrix} x & x & a \\ b & y & y \end{bmatrix} - \sum \begin{bmatrix} x & x & a \\ y & b & b \end{bmatrix}, \tag{31}$$

where all four summands on the right hand-side of (31) run over all integers x and y that render critical 2–faces. Explicitly, $a < x < m$ in the first and second summations, $x < a$ in the third and fourth summations, $b < y < m$ in the second and third summations, $y < b$ in the first and fourth summations, and $b \neq x \neq y \neq a$ in all four summations.

Note that the first two summations in (31) are empty when $a = m-1$ (so $b < m-2$).

Proof The complete trees of mixed paths $\beta \searrow\nearrow \cdots \nearrow\searrow \alpha$ from critical 2–dimensional faces β to either critical or collapsible 1–dimensional faces α are given in Figures 3–11, where we indicate a positive (resp. negative) face with a bold (resp. regular) arrow. Types of pairings involved are indicated using the item

$$\left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} a & c \\ d & m \end{smallmatrix}\right] \overset{\text{(f)}}{\nearrow} \left[\begin{smallmatrix} a & a & c \\ d & m & m \end{smallmatrix}\right] \left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} a & a \\ d & m \end{smallmatrix}\right] \overset{\text{(b)}}{\nearrow} \left[\begin{smallmatrix} a & a & m-1 \\ d & m & m \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} a & m-1 \\ d & m \end{smallmatrix}\right] & (+) \\ \searrow \left[\begin{smallmatrix} a & m-1 \\ m & m \end{smallmatrix}\right] & \text{(h)} \end{array}\right. \\
\searrow \left[\begin{smallmatrix} a & c \\ m & m \end{smallmatrix}\right] \quad (+)
\end{array}\right. \\
\searrow \left[\begin{smallmatrix} c & c \\ d & m \end{smallmatrix}\right] \overset{\text{(b)}}{\nearrow} \left[\begin{smallmatrix} c & c & m-1 \\ d & m & m \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} c & m-1 \\ d & m \end{smallmatrix}\right] & (-) \\ \searrow \left[\begin{smallmatrix} c & m-1 \\ m & m \end{smallmatrix}\right] & \text{(h)} \end{array}\right.
\end{array}\right.$$

Figure 3: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & c \\ d & d \end{smallmatrix}\right] \overset{\text{(c)}}{\nearrow} \left[\begin{smallmatrix} a & c & c \\ d & d & m \end{smallmatrix}\right]$ for $b \neq c \leq m-2 \geq d$.

names (a)–(i) in Proposition 4.1. At the end of each branch we indicate either the type of paring that shows α is collapsible or, if α is critical, the multiplicity with which the path must be accounted for in (6) and (7).

The first assertion follows by observing that, in Figures 3–11, there are two mixed paths departing from a fixed critical 2–dimensional face and arriving to a fix critical 1–dimensional face of the form (k.2) or (k.3). These two mixed paths have opposite multiplicities, so they cancel each other out in (7). For instance, each mixed path from $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} a & c \\ m & m \end{smallmatrix}\right]$ in Figure 3 cancels out the corresponding path in Figure 4.

To get at (31), start by noticing from Figures 3–11 that there are only four types of mixed paths departing from a given critical 2–dimensional face $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ that arrive to some critical 1–dimensional faces of type (k.1). Namely:

- There is a mixed path $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \searrow\nearrow \cdots \nearrow\searrow \left[\begin{smallmatrix} a & m-1 \\ d & m \end{smallmatrix}\right]$ with multiplicity $+1$; see Figures 3, 6 and 9,
- There is a mixed path $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \searrow\nearrow \cdots \nearrow\searrow \left[\begin{smallmatrix} a & m-1 \\ b & m \end{smallmatrix}\right]$ with multiplicity -1; see Figures 4, 7 and 10,
- There is a mixed path $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \searrow\nearrow \cdots \nearrow\searrow \left[\begin{smallmatrix} c & m-1 \\ b & m \end{smallmatrix}\right]$ with multiplicity $+1$; see Figures 4, 7 and 10.
- There is a mixed path $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \searrow\nearrow \cdots \nearrow\searrow \left[\begin{smallmatrix} c & m-1 \\ d & m \end{smallmatrix}\right]$ with multiplicity -1 provided $(c,d) \neq (m-1, m-2)$; see Figures 3, 6 and 9.

$$\left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} a & c \\ b & b \end{smallmatrix}\right] \overset{\text{(c)}}{\nearrow} \left[\begin{smallmatrix} a & c & c \\ b & b & m \end{smallmatrix}\right] \left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} a & c \\ b & m \end{smallmatrix}\right] \overset{\text{(f)}}{\nearrow} \left[\begin{smallmatrix} a & a & c \\ b & m & m \end{smallmatrix}\right] \left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} a & a \\ b & m \end{smallmatrix}\right] \overset{\text{(b)}}{\nearrow} \left[\begin{smallmatrix} a & a & m-1 \\ b & m & m \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} a & m-1 \\ b & m \end{smallmatrix}\right] & (-) \\ \searrow \left[\begin{smallmatrix} a & m-1 \\ m & m \end{smallmatrix}\right] & \text{(h)} \end{array}\right. \\
\searrow \left[\begin{smallmatrix} a & c \\ m & m \end{smallmatrix}\right] \quad (-)
\end{array}\right. \\
\searrow \left[\begin{smallmatrix} c & c \\ b & m \end{smallmatrix}\right] \overset{\text{(b)}}{\nearrow} \left[\begin{smallmatrix} c & c & m-1 \\ b & m & m \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} c & m-1 \\ b & m \end{smallmatrix}\right] & (+) \\ \searrow \left[\begin{smallmatrix} c & m-1 \\ m & m \end{smallmatrix}\right] & \text{(h)} \end{array}\right.
\end{array}\right. \\
\searrow \left[\begin{smallmatrix} c & c \\ b & d \end{smallmatrix}\right] \overset{\text{(a)}}{\nearrow} \left[\begin{smallmatrix} c & c & m \\ b & d & d \end{smallmatrix}\right] \left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} c & m \\ b & d \end{smallmatrix}\right] \overset{\text{(e)}}{\nearrow} \left[\begin{smallmatrix} c & m & m \\ b & b & d \end{smallmatrix}\right] \left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} c & m \\ b & b \end{smallmatrix}\right] \overset{\text{(d)}}{\nearrow} \left[\begin{smallmatrix} c & m & m \\ b & b & m-1 \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} c & m \\ b & m-1 \end{smallmatrix}\right] & \text{(g)} \\ \searrow \left[\begin{smallmatrix} m & m \\ b & m-1 \end{smallmatrix}\right] & \text{(g)} \end{array}\right. \\
\searrow \left[\begin{smallmatrix} m & m \\ b & d \end{smallmatrix}\right] \quad (-)
\end{array}\right. \\
\searrow \left[\begin{smallmatrix} c & m \\ d & d \end{smallmatrix}\right] \overset{\text{(d)}}{\nearrow} \left[\begin{smallmatrix} c & m & m \\ d & d & m-1 \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} c & m \\ d & m-1 \end{smallmatrix}\right] & \text{(g)} \\ \searrow \left[\begin{smallmatrix} m & m \\ d & m-1 \end{smallmatrix}\right] & \text{(g)} \end{array}\right.
\end{array}\right.
\end{array}\right.$$

Figure 4: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right] \overset{\text{(e)}}{\nearrow} \left[\begin{smallmatrix} a & c & c \\ b & b & d \end{smallmatrix}\right]$ for $b \neq c \leq m-2 \geq d$.

$$\left\{\begin{array}{l}
\searrow \left[\begin{smallmatrix} a & m \\ b & d \end{smallmatrix}\right] \overset{(e)}{\nearrow} \left[\begin{smallmatrix} a & m & m \\ b & b & d \end{smallmatrix}\right] \left\{\begin{array}{l} \searrow \left[\begin{smallmatrix} a & m \\ b & b \end{smallmatrix}\right] \overset{(d)}{\nearrow} \left[\begin{smallmatrix} a & m & m \\ b & b & m-1 \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} a & m \\ b & m-1 \end{smallmatrix}\right] & \text{(g)} \\ \searrow \left[\begin{smallmatrix} m & m \\ b & m-1 \end{smallmatrix}\right] & \text{(g)} \end{array}\right. \\ \searrow \left[\begin{smallmatrix} m & m \\ b & d \end{smallmatrix}\right] \quad (+) \end{array}\right. \\
\searrow \left[\begin{smallmatrix} a & m \\ d & d \end{smallmatrix}\right] \overset{(d)}{\nearrow} \left[\begin{smallmatrix} a & m & m \\ d & d & m-1 \end{smallmatrix}\right] \left\{\begin{array}{ll} \searrow \left[\begin{smallmatrix} a & m \\ d & m-1 \end{smallmatrix}\right] & \text{(g)} \\ \searrow \left[\begin{smallmatrix} m & m \\ d & m-1 \end{smallmatrix}\right] & \text{(g)} \end{array}\right.
\end{array}\right.$$

Figure 5: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & a \\ b & d \end{smallmatrix}\right] \overset{(a)}{\nearrow} \left[\begin{smallmatrix} a & a & m \\ b & d & d \end{smallmatrix}\right]$ for $b \neq c \leq m-2 \geq d$.

Therefore the value of the boundary map $\partial\colon \mu_2(C_m) \to \mu_1(C_m)$ at a critical face $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ for $b \neq c < m > d$ with $(c,d) \neq (m-1, m-2)$ is

$$\partial\left(\begin{bmatrix} a & a & c \\ b & d & d \end{bmatrix}\right) = \begin{bmatrix} a & m-1 \\ d & m \end{bmatrix} - \begin{bmatrix} a & m-1 \\ b & m \end{bmatrix} + \begin{bmatrix} c & m-1 \\ b & m \end{bmatrix} - \begin{bmatrix} c & m-1 \\ d & m \end{bmatrix}, \tag{32}$$

whereas, for $(c,d) = (m-1, m-2)$,

$$\partial\left(\begin{bmatrix} a & a & m-1 \\ b & m-2 & m-2 \end{bmatrix}\right) = \begin{bmatrix} a & m-1 \\ m-2 & m \end{bmatrix} - \begin{bmatrix} a & m-1 \\ b & m \end{bmatrix} + \begin{bmatrix} m-1 & m-1 \\ b & m \end{bmatrix}. \tag{33}$$

(Note that (32) is valid when $(c,d) = (m-1, m-2)$ provided the fourth noncritical term

$$\begin{bmatrix} m-1 & m-1 \\ m-2 & m \end{bmatrix}$$

is omitted.) That (31) follows by dualizing (32) and (33) is a straightforward exercise left to the reader. □

4.3 Cohomology bases

By Corollary 4.2, we can assume $m \geq 5$ throughout the rest of the paper. We start by identifying (in Corollary 4.7) an explicit basis for $H^1(\mathrm{Conf}(|K_m|, 2))$, ie for the kernel of the Morse coboundary $\delta\colon \mu^1(C_m) \to \mu^2(C_m)$. By Proposition 4.3, it is enough to focus on the submodule $\mu_0^1(C_m)$ of $\mu^1(C_m)$ generated by the duals of the basis elements of type (k.1). Thus $\mu_0^1(C_m)$ is free on elements $\left\{\begin{smallmatrix} a \\ b \end{smallmatrix}\right\}$ satisfying $a < m > b \neq a$ and $(a,b) \neq (m-1, m-2)$, where $\left\{\begin{smallmatrix} a \\ b \end{smallmatrix}\right\}$ stands for the dual of $\left[\begin{smallmatrix} a & m-1 \\ b & m \end{smallmatrix}\right]$.

Definition 4.4 Consider the elements $\left\langle\begin{smallmatrix} a \\ b \end{smallmatrix}\right\rangle \in \mu_0^1(C_m)$ defined for $a < m > b \neq a$ and $(a,b) \neq (m-1, m-2)$ according to the following cases (see Figure 12):

(R_1) For $1 \leq a \leq 2$, or for $1 \leq a \leq m-3$ with $b = m-1$, or for $(a,b) = (3,1)$,

$$\left\langle\begin{matrix} a \\ b \end{matrix}\right\rangle := \sum_{a \neq j \leq b} \left\{\begin{matrix} a \\ j \end{matrix}\right\}.$$

$$\left\{\begin{array}{lll}
\searrow \left[\begin{smallmatrix} a & m-1 \\ d & m \end{smallmatrix}\right] & (+) & \\
\searrow \left[\begin{smallmatrix} m-1 & m-1 \\ d & m \end{smallmatrix}\right] & (-) & \text{if } d < m-2 \\
 & \text{(i)} & \text{if } d = m-2
\end{array}\right.$$

Figure 6: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & m-1 \\ b & d & d \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & m-1 \\ d & d \end{smallmatrix}\right] \overset{(c)}{\nearrow} \left[\begin{smallmatrix} a & m-1 & m-1 \\ d & d & m \end{smallmatrix}\right]$ for $b \neq m-1 > d$.

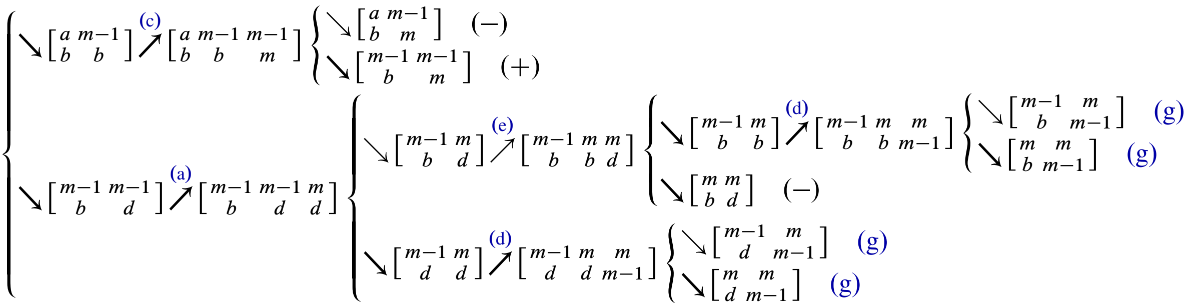


Figure 7: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & m-1 \\ b & d & d \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & m-1 \\ b & d \end{smallmatrix}\right] \overset{(e)}{\nearrow} \left[\begin{smallmatrix} a & m-1 & m-1 \\ b & b & d \end{smallmatrix}\right]$ for $b \neq m-1 > d$.

(R_2) For $4 \le a \le m-1$ with $b = 1$, or for $a = m-1$ with $1 \le b \le m-4$,

$$\left\langle {a \atop b} \right\rangle := \sum_{b \neq i \le a} \left\{ {i \atop b} \right\}.$$

(R_3) $\left\langle {3 \atop 2} \right\rangle := \left\{ {3 \atop 2} \right\} + \left\{ {3 \atop 1} \right\} + \left\{ {2 \atop 3} \right\} + \left\{ {2 \atop 1} \right\} + \left\{ {1 \atop 3} \right\} + \left\{ {1 \atop 2} \right\}$.

(R_4) For $4 \le a \le m-2$,

$$\left\langle {a \atop 2} \right\rangle := \sum_{2 \neq i \le a} \left\{ {i \atop 2} \right\} + \sum_{i \le a-1} \left\{ {i \atop a} \right\}.$$

(R_5) For $a, b \in \{3, 4, \dots, m-2\}$,

$$\left\langle {a \atop b} \right\rangle := \sum_{j \neq i \le a \neq j \le b} \left\{ {i \atop j} \right\} + \sum_{i \le a-1} \left\{ {i \atop a} \right\}.$$

(R_6) $\left\langle {m-1 \atop m-3} \right\rangle := \left\{ {m-1 \atop m-3} \right\} - \sum_{i,j} \left\{ {i \atop j} \right\}$, where the sum runs over i and j with $m-3 \neq j \neq i \le m-2 \ge j$.

(R_7) $\left\langle {m-2 \atop m-1} \right\rangle := \left\{ {m-2 \atop m-1} \right\} - \sum_{i,j} \left\{ {i \atop j} \right\}$, where the sum runs over i and j with $m-2 \ge j \neq i \le m-3$.

Direct inspection of the defining formulae yields:

Lemma 4.5 *The following relations hold under the indicated conditions*:

(i) $\left\langle {a \atop b} \right\rangle - \left\langle {a \atop b-1} \right\rangle = \sum_{b \neq i \le a} \left\{ {i \atop b} \right\}$, *provided* $3 \le a \le m-2$ *and* $4 \le b \le m-2$ *with* $a \neq b \neq a+1$.

(ii) $\left\langle {a \atop 3} \right\rangle - \left\langle {a \atop 2} \right\rangle - \left\langle {a \atop 1} \right\rangle = \sum_{3 \neq i \le a} \left\{ {i \atop 3} \right\}$, *provided* $4 \le a \le m-2$.

(iii) $\left\langle {b-1 \atop b} \right\rangle - \left\langle {b-1 \atop b-2} \right\rangle = \sum_{i < b} \left\{ {i \atop b} \right\}$, *provided* $5 \le b \le m-2$.

(iv) $\left\langle {3 \atop 4} \right\rangle - \left\langle {3 \atop 2} \right\rangle = \sum_{i \le 3} \left\{ {i \atop 4} \right\}$, *provided* $m \ge 6$.

Proposition 4.6 *The elements* $\left\langle {a \atop b} \right\rangle$ *for* $a < m > b \neq a$ *and* $(a, b) \neq (m-1, m-2)$ *yield a basis of* $\mu_0^1(C_m)$.

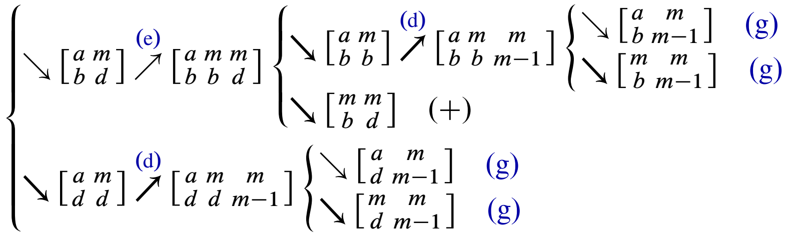


Figure 8: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & m-1 \\ b & d & d \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & a \\ b & d \end{smallmatrix}\right] \overset{\text{(a)}}{\nearrow} \left[\begin{smallmatrix} a & a & m \\ b & d & d \end{smallmatrix}\right]$ for $b \neq m-1 > d$.

Proof In view of the one-to-one correspondence $\left\{\begin{smallmatrix} a \\ b \end{smallmatrix}\right\} \leftrightarrow \left\langle\begin{smallmatrix} a \\ b \end{smallmatrix}\right\rangle$, it suffices to check that

(34) each $\left\{\begin{smallmatrix} a \\ b \end{smallmatrix}\right\}$ is a $\mathbb{Z}$–linear combination of the elements $\left\langle\begin{smallmatrix} a' \\ b' \end{smallmatrix}\right\rangle$.

In most cases (34) follows by a simple recursive argument based on the observation that, in all cases,

$$\left\{\begin{matrix} a \\ b \end{matrix}\right\} = \left\langle\begin{matrix} a \\ b \end{matrix}\right\rangle + \sum_{(a',b') \neq (a,b)} \pm 1 \left\{\begin{matrix} a' \\ b' \end{matrix}\right\}. \tag{35}$$

Namely, the recursive argument applies for (R_1) when $a \leq 2$ or $b = 1$, for (R_2) when $b = 1$, and for (R_3). The recursive argument also applies in the cases (R_6) and (R_7), as well as in the remaining instances of (R_1) and (R_2) provided

(36) (34) holds true when a and b fall in cases (R_4) and (R_5).

Since (R_4) and (R_5) are empty for $m = 5$, we only need to verify (36) assuming $m \geq 6$.

Direct computation gives $\left\{\begin{smallmatrix} 4 \\ 2 \end{smallmatrix}\right\} = \left\langle\begin{smallmatrix} 4 \\ 2 \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right\rangle + \left\langle\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right\rangle + \left\langle\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right\rangle + \left\langle\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right\rangle$, while Lemma 4.5(iii) and (iv) yield

$$\left\{\begin{matrix} a \\ 2 \end{matrix}\right\} = \left(\left\langle\begin{matrix} a \\ 2 \end{matrix}\right\rangle - \left\langle\begin{matrix} a-1 \\ 2 \end{matrix}\right\rangle\right) + \left(\left\langle\begin{matrix} a-2 \\ a-1 \end{matrix}\right\rangle - \left\langle\begin{matrix} a-2 \\ a-3 \end{matrix}\right\rangle\right) - \left(\left\langle\begin{matrix} a-1 \\ a \end{matrix}\right\rangle - \left\langle\begin{matrix} a-1 \\ a-2 \end{matrix}\right\rangle\right)$$

for $5 \leq a \leq m-2$, which establishes (36) in the case of (R_4). The validity of (36) in the case $a = 3$ of (R_5) is established in a similar way. Note first that the idea in the recursive argument based on (35) works to give (36) for $(a, b) = (3, 4)$; then use Lemma 4.5(i) to get

$$\left\{\begin{matrix} 3 \\ b \end{matrix}\right\} = \left(\left\langle\begin{matrix} 3 \\ b \end{matrix}\right\rangle - \left\langle\begin{matrix} 3 \\ b-1 \end{matrix}\right\rangle\right) - \left\langle\begin{matrix} 2 \\ b \end{matrix}\right\rangle + \left\langle\begin{matrix} 2 \\ b-1 \end{matrix}\right\rangle - \left\langle\begin{matrix} 1 \\ b \end{matrix}\right\rangle + \left\langle\begin{matrix} 1 \\ b-1 \end{matrix}\right\rangle$$

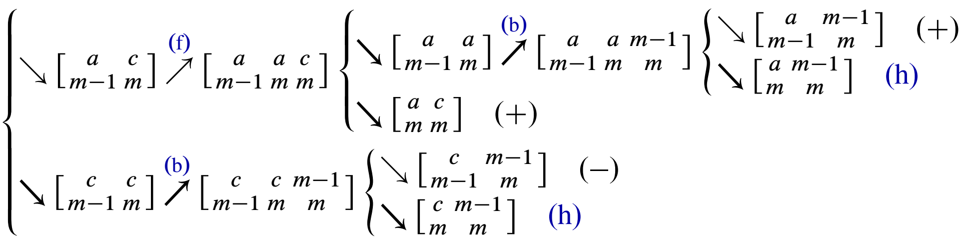


Figure 9: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & c \\ b & m-1 & m-1 \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & c \\ m-1 & m-1 \end{smallmatrix}\right] \overset{\text{(c)}}{\nearrow} \left[\begin{smallmatrix} a & c & c \\ m-1 & m-1 & m \end{smallmatrix}\right]$ for $b \neq c < m-1$.

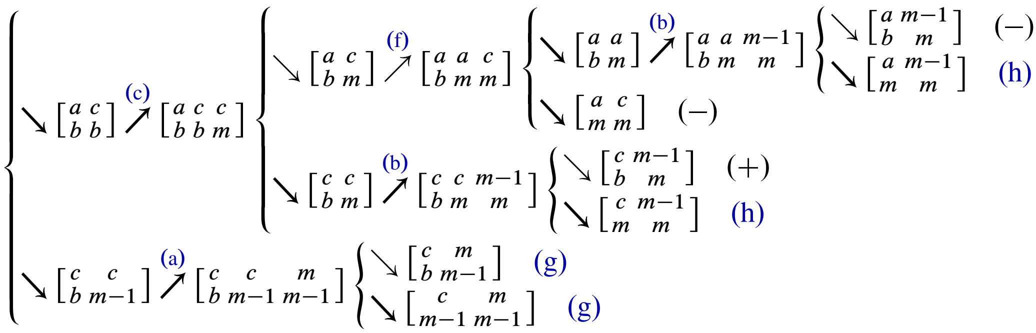


Figure 10: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & c \\ b & m-1 & m-1 \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & c \\ b & m-1 \end{smallmatrix}\right] \overset{(e)}{\nearrow} \left[\begin{smallmatrix} a & c & c \\ b & b & m-1 \end{smallmatrix}\right]$ for $b \neq c < m-1$.

for $5 \leq b \leq m-2$. The validity of (36) in the case $b = 3$ of (R_5) is established using the idea in the recursive argument at the beginning of the proof, except that (35) is replaced by the expression

$$\begin{Bmatrix} a \\ 3 \end{Bmatrix} = \left\langle \begin{matrix} a \\ 3 \end{matrix} \right\rangle - \left\langle \begin{matrix} a \\ 2 \end{matrix} \right\rangle - \left\langle \begin{matrix} a \\ 1 \end{matrix} \right\rangle - \sum_{3 \neq i < a} \begin{Bmatrix} i \\ 3 \end{Bmatrix}$$

coming from Lemma 4.5(ii). Lastly, the validity of (36) in the remaining case $a, b \in \{4, \dots, m-2\}$ of (R_5) is established by the formulae

- $\left\{\begin{smallmatrix} a \\ b \end{smallmatrix}\right\} = \left(\left\langle\begin{smallmatrix} a \\ b \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} a \\ b-1 \end{smallmatrix}\right\rangle\right) - \left(\left\langle\begin{smallmatrix} a-1 \\ b \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} a-1 \\ b-1 \end{smallmatrix}\right\rangle\right)$, when $a+1 \neq b \neq a-1$,
- $\left\{\begin{smallmatrix} a \\ b \end{smallmatrix}\right\} = \left(\left\langle\begin{smallmatrix} a \\ a-1 \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} a \\ a-2 \end{smallmatrix}\right\rangle\right) - \left(\left\langle\begin{smallmatrix} a-2 \\ a-1 \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} a-2 \\ a-3 \end{smallmatrix}\right\rangle\right)$, when $b = a-1$ (so $a \geq 5$),
- $\left\{\begin{smallmatrix} a \\ b \end{smallmatrix}\right\} = \left(\left\langle\begin{smallmatrix} a \\ a+1 \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} a \\ a-1 \end{smallmatrix}\right\rangle\right) - \left(\left\langle\begin{smallmatrix} a-1 \\ a+1 \end{smallmatrix}\right\rangle - \left\langle\begin{smallmatrix} a-1 \\ a \end{smallmatrix}\right\rangle\right)$, when $b = a+1$ (so $a \leq m-3$),

which use Lemma 4.5(i), (iii) and (iv). □

Recall that the Morse coboundary $\delta\colon \mu^0(C_m) \to \mu^1(C_m)$ is trivial, so that the 1–dimensional cohomology of $\mathrm{Conf}(|K_m|, 2)$ is given by the kernel of $\delta\colon \mu^1(C_m) \to \mu^2(C_m)$.

Corollary 4.7 *A basis for* $H^1(\mathrm{Conf}(|K_m|, 2))$ *is given by*

(i) *the duals of the critical* 1*–dimensional faces of type* (k.2) *and* (k.3) *in* Proposition 4.1, *and*

(ii) *the* (*already dualized*) *elements* $\left\langle\begin{smallmatrix} a \\ b \end{smallmatrix}\right\rangle$ *satisfying* $a = m-1$, $b = m-1$ or $(a, b) = (m-2, m-3)$.

Proof Recall $m \geq 5$. A straightforward counting shows that the number of elements in (i) and (ii) is $(m-1)(m-2)$, which is also the first Betti number of $\mathrm{Conf}(|K_m|, 2)$; see [9, Corollary 23]. Since the

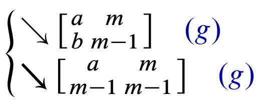


Figure 11: Gradient paths evolving from $\left[\begin{smallmatrix} a & a & c \\ b & m-1 & m-1 \end{smallmatrix}\right] \searrow \left[\begin{smallmatrix} a & a \\ b & m-1 \end{smallmatrix}\right] \overset{(a)}{\nearrow} \left[\begin{smallmatrix} a & a & m \\ b & m-1 & m-1 \end{smallmatrix}\right]$ for $b \neq c < m-1$.

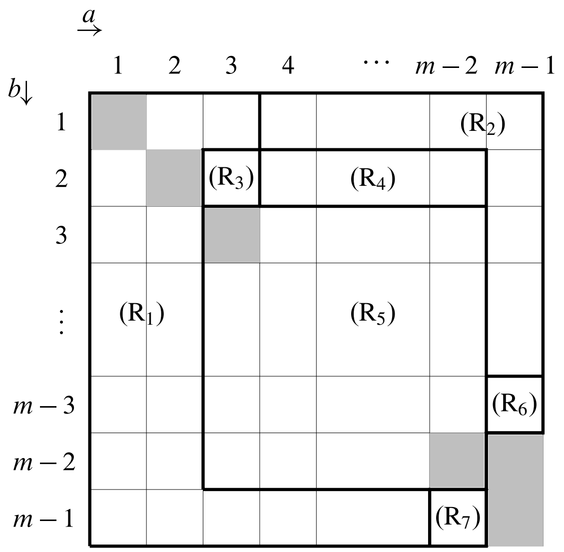


Figure 12: Defining regions for the basis elements $\left\langle{a \atop b}\right\rangle$.

homology of $\mathrm{Conf}(|K_m|,2)$ is torsion-free [9, Proposition 2], Proposition 4.6 implies that the proof will be complete once it is checked that $\delta\colon \mu^1(C_m)\to\mu^2(C_m)$ vanishes on each of the elements in (i) and (ii). Indeed, this actually implies that δ is injective on the submodule generated by the basis elements $\left\langle{a \atop b}\right\rangle$ not included in (ii).

The vanishing of δ on the elements in (i) comes directly from Proposition 4.3, whereas the vanishing of δ on the elements in (ii) is verified by direct calculation using the expression of δ in Proposition 4.3. The arithmetic manipulations needed are illustrated next in a representative case, namely, that of $\left\langle{m-1 \atop m-3}\right\rangle$.

Use Proposition 4.3 and the defining formula (R_6) to get

$$\delta\left\langle{m-1 \atop m-3}\right\rangle = \sum_{x,y}\begin{bmatrix} x & x & m-1 \\ m-3 & y & y\end{bmatrix} - \sum_{x,y}\begin{bmatrix} x & x & m-1 \\ y & m-3 & m-3\end{bmatrix} - \sum_{m-3\neq j\neq i\leq m-2\geq j}\left(\sum_{x,y}\begin{bmatrix} i & i & x \\ y & j & j\end{bmatrix} - \sum_{x,y}\begin{bmatrix} i & i & x \\ j & y & y\end{bmatrix} + \sum_{x,y}\begin{bmatrix} x & x & i \\ j & y & y\end{bmatrix} - \sum_{x,y}\begin{bmatrix} x & x & i \\ y & j & j\end{bmatrix}\right).$$

The summands with $y=m-1$ in the second inner summation cancel out the corresponding ones in the third inner summation. (The corresponding fact for $y<m-1$, dealt with below, is more subtle since $i\leq m-2$ in the third inner summation, while $x\leq m-1$ in the second inner summation.) Noticing in addition that $y=m-2$ is forced in the first outer summation, we then get

$$\delta\left\langle{m-1 \atop m-3}\right\rangle = \sum_{x}\begin{bmatrix} x & x & m-1 \\ m-3 & m-2 & m-2\end{bmatrix} - \sum_{x,y}\begin{bmatrix} x & x & m-1 \\ y & m-3 & m-3\end{bmatrix} - \sum_{m-3\neq j\neq i\leq m-2\geq j}\left(\sum_{x,y}\begin{bmatrix} i & i & x \\ y & j & j\end{bmatrix} - \sum_{\substack{x \\ y\leq m-2}}\begin{bmatrix} i & i & x \\ j & y & y\end{bmatrix} + \sum_{\substack{x \\ y\leq m-2}}\begin{bmatrix} x & x & i \\ j & y & y\end{bmatrix} - \sum_{x,y}\begin{bmatrix} x & x & i \\ y & j & j\end{bmatrix}\right).$$

In the last expression, the summands with $x \le m-2$ in the second inner summation cancel out the third inner summation, so

$$\delta\begin{pmatrix} m-1 \\ m-3 \end{pmatrix} = \sum_{x} \begin{bmatrix} x & x & m-1 \\ m-3 & m-2 & m-2 \end{bmatrix} - \sum_{x,y} \begin{bmatrix} x & x & m-1 \\ y & m-3 & m-3 \end{bmatrix} - \sum_{m-3\ne j\ne i\le m-2\ge j} \Biggl(\sum_{x,y} \begin{bmatrix} i & i & x \\ y & j & j \end{bmatrix} - \sum_{y\le m-2} \begin{bmatrix} i & i & m-1 \\ j & y & y \end{bmatrix} - \sum_{x,y} \begin{bmatrix} x & x & i \\ y & j & j \end{bmatrix} \Biggr).$$

Likewise, in the last expression, summands with $x \le m-2$ in the first inner summation cancel out the third inner summation, so

$$\delta\begin{pmatrix} m-1 \\ m-3 \end{pmatrix} = \sum_{x} \begin{bmatrix} x & x & m-1 \\ m-3 & m-2 & m-2 \end{bmatrix} - \sum_{x,y} \begin{bmatrix} x & x & m-1 \\ y & m-3 & m-3 \end{bmatrix} - \sum_{m-3\ne j\ne i\le m-2\ge j} \Biggl(\sum_{y} \begin{bmatrix} i & i & m-1 \\ y & j & j \end{bmatrix} - \sum_{y\le m-2} \begin{bmatrix} i & i & m-1 \\ j & y & y \end{bmatrix} \Biggr).$$

Lastly, merge the first (resp. second) outer summation and the second (resp. first) inner summation in the last expression to get

$$\delta\begin{pmatrix} m-1 \\ m-3 \end{pmatrix} = \sum_{j\ne i\le m-2\ge y} \begin{bmatrix} i & i & m-1 \\ j & y & y \end{bmatrix} - \sum_{j\ne i\le m-2\ge j} \begin{bmatrix} i & i & m-1 \\ y & j & j \end{bmatrix} = 0,$$

as asserted. □

We next identify (in Corollary 4.11) an explicit basis for $H^2(\mathrm{Conf}(|K_m|,2))$, ie for the cokernel of the Morse coboundary $\delta\colon \mu^1(C_m) \to \mu^2(C_m)$. In what follows, the conditions $1 \le a < c < m$, $1 \le b < d < m$ and $c \ne d \ne a \ne b \ne c < m > d$ for critical faces $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ identified in Proposition 4.1(l) will be implicit (and generally omitted).

Definition 4.8 Let $\mathcal{C}$ be the collection of the critical faces $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ of one of the four types

(37) $\left[\begin{smallmatrix} 1 & 1 & c \\ 2 & d & d \end{smallmatrix}\right]$ with $c, d \in \{3, 4, \dots, m-1\}$,

(38) $\left[\begin{smallmatrix} 1 & 1 & 2 \\ 3 & d & d \end{smallmatrix}\right]$ with $d \in \{4, 5, \dots, m-1\}$,

(39) $\left[\begin{smallmatrix} 2 & 2 & c \\ 1 & 3 & 3 \end{smallmatrix}\right]$ with $c \in \{4, 5, \dots, m-1\}$,

(40) $\left[\begin{smallmatrix} 2 & 2 & 3 \\ 1 & 4 & 4 \end{smallmatrix}\right]$,

and let $\mathcal{B}$ stand for the collection of all other critical faces $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$.

The following change of basis is used to show that the duals of critical faces in $\mathcal{B}$ form a basis of $H^2(\mathrm{Conf}(|K_m|,2))$:

Definition 4.9 For each $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right] \in \mathcal{B}$, consider the element $\left\langle\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right\rangle \in \mu_2(C_m)$ defined through:

- (i) Case $a = 1$ and $c \ge 3$ with $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ not fitting in (37):
  - (a) $\left\langle\begin{smallmatrix} 1 & 1 & c \\ b & d & d \end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 1 & 1 & c \\ b & d & d \end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c \\ 2 & d & d \end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c \\ 2 & b & b \end{smallmatrix}\right]$, for $b \ge 3$.

(ii) Case $a=1$ and $c=2$ with $\left[\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right]$ not fitting in (38):

(b) $\left\langle\begin{smallmatrix} 1 & 1 & 2\\ b & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 1 & 1 & 2\\ b & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & b & b\end{smallmatrix}\right]$, for $b \geq 4$.

(iii) Case $a=2$, $b=1$ and $c \geq 4$ with $\left[\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right]$ not fitting in (39):

(c) $\left\langle\begin{smallmatrix} 2 & 2 & c\\ 1 & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 3 & 3\end{smallmatrix}\right]$, for $d \geq 4$.

(iv) Case $a=2$, $b=1$ and $c=3$ with $\left[\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right]$ not fitting in (40):

(d) $\left\langle\begin{smallmatrix} 2 & 2 & 3\\ 1 & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 2 & 2 & 3\\ 1 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & 3\\ 1 & 4 & 4\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & 4 & 4\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & 4 & 4\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & d & d\end{smallmatrix}\right]$, for $d \geq 5$.

(v) Case $a=2$ and $b \geq 3$:

(e) $\left\langle\begin{smallmatrix} 2 & 2 & c\\ 3 & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 2 & 2 & c\\ 3 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 3 & 3\end{smallmatrix}\right]$,

(f) $\left\langle\begin{smallmatrix} 2 & 2 & c\\ b & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 2 & 2 & c\\ b & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & b & b\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & b & b\end{smallmatrix}\right]$, for $b \geq 4$.

(vi) Case $a=3$ and $b=1$:

(g) $\left\langle\begin{smallmatrix} 3 & 3 & c\\ 1 & 2 & 2\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 3 & 3 & c\\ 1 & 2 & 2\end{smallmatrix}\right] + \left[\begin{smallmatrix} 2 & 2 & 3\\ 1 & 4 & 4\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & 4 & 4\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & 4 & 4\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 3 & 3\end{smallmatrix}\right]$,

(h) $\left\langle\begin{smallmatrix} 3 & 3 & c\\ 1 & 4 & 4\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 3 & 3 & c\\ 1 & 4 & 4\end{smallmatrix}\right] + \left[\begin{smallmatrix} 2 & 2 & 3\\ 1 & 4 & 4\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & 4 & 4\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 4 & 4\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 3 & 3\end{smallmatrix}\right]$,

(i) $\left\langle\begin{smallmatrix} 3 & 3 & c\\ 1 & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 3 & 3 & c\\ 1 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 2 & 2 & 3\\ 1 & 4 & 4\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 2\\ 3 & 4 & 4\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 3 & 3\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & 4 & 4\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & d & d\end{smallmatrix}\right]$, for $d \geq 5$.

(vii) Case $a=3$ and $b \geq 2$:

(j) $\left\langle\begin{smallmatrix} 3 & 3 & c\\ 2 & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 3 & 3 & c\\ 2 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & d & d\end{smallmatrix}\right]$,

(k) $\left\langle\begin{smallmatrix} 3 & 3 & c\\ b & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} 3 & 3 & c\\ b & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & b & b\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & b & b\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & 3\\ 2 & d & d\end{smallmatrix}\right]$, for $b \geq 4$.

(viii) Case $a \geq 4$:

(l) $\left\langle\begin{smallmatrix} a & a & c\\ 1 & 2 & 2\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} a & a & c\\ 1 & 2 & 2\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & a\\ 2 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 3 & 3\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 2 & 2 & a\\ 1 & 3 & 3\end{smallmatrix}\right]$,

(m) $\left\langle\begin{smallmatrix} a & a & c\\ 1 & 3 & 3\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} a & a & c\\ 1 & 3 & 3\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 2 & 2 & a\\ 1 & 3 & 3\end{smallmatrix}\right]$,

(n) $\left\langle\begin{smallmatrix} a & a & c\\ 1 & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} a & a & c\\ 1 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 2 & c\\ 1 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 2 & 2 & a\\ 1 & 3 & 3\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & 3 & 3\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & a\\ 2 & 3 & 3\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & a\\ 2 & d & d\end{smallmatrix}\right]$, for $d \geq 4$,

(o) $\left\langle\begin{smallmatrix} a & a & c\\ 2 & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} a & a & c\\ 2 & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & a\\ 2 & d & d\end{smallmatrix}\right]$,

(p) $\left\langle\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right\rangle := \left[\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & d & d\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & c\\ 2 & b & b\end{smallmatrix}\right] - \left[\begin{smallmatrix} 1 & 1 & a\\ 2 & b & b\end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 1 & a\\ 2 & d & d\end{smallmatrix}\right]$, for $b \geq 3$.

Direct inspection shows that

(41) each $\left\langle\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right\rangle - \left[\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right]$ is a linear combination of basis elements in $\mathcal{C}$.

Therefore $\mathcal{B}' \cup \mathcal{C}$ is a new basis of $\mu_2(C_m)$, where $\mathcal{B}'$ stands for the collection of elements $\left\langle\begin{smallmatrix} a & a & c\\ b & d & d\end{smallmatrix}\right\rangle$. Further, routine verifications using (32) and (33) show that $\mathcal{B}'$ lies in the kernel of $\partial\colon \mu_2(C_m) \to \mu_1(C_m)$. In fact:

Lemma 4.10 *$\mathcal{B}'$ is a basis of the kernel of the Morse boundary map $\partial\colon \mu_2(C_m) \to \mu_1(C_m)$.*

Proof The argument is parallel to that in the proof of Corollary 4.7. Namely, by direct counting, the cardinality of $\mathcal{C}$ is $|\mathcal{C}| = m^2 - 5m + 5$. In view of (30), this leads to

$$|\mathcal{B}'| = |\mathcal{B}| = \tfrac{1}{4} m(m-2)(m-3)(m-5) + 1,$$

which is the second Betti number of $\mathrm{Conf}(|K_m|,2)$; see [9, Corollary 23]. Hence $\partial\colon \mu_2(C_m)\to\mu_1(C_m)$ is forced to be injective on the submodule spanned by $\mathcal{C}$ and, in particular, $\mathcal{B}'$ spans (and is thus a basis of) $\ker(\partial\colon \mu_2(C_m)\to\mu_1(C_m))$. $\square$

The two sequences

$$0\to H_2(\mathrm{Conf}(|K_m|,2))\overset{\iota}{\hookrightarrow}\mu_2(C_m)\overset{\partial}{\to}\mu_1(C_m),\qquad 0\leftarrow H^2(\mathrm{Conf}(|K_m|,2))\overset{\iota^*}{\twoheadleftarrow}\mu^2(C_m)\overset{\delta}{\leftarrow}\mu^1(C_m), \tag{42}$$

are exact; the first one by definition, and the second one, which is the dual of the first one, because $H_1(\mathrm{Conf}(|K_m|,2))$ is torsion-free. Thus the cohomology class represented by an element $e\in\mu^2(C_m)$ is given by the ι^*–image of e. Such an interpretation of cohomology classes is used in the proof of:

Corollary 4.11 *A basis of $H^2(\mathrm{Conf}(|K_m|,2))$ is given by the classes represented by the duals of the critical faces in $\mathcal{B}$. Furthermore, the expression (as a linear combination of basis elements) of the cohomology class represented by the dual of a critical face in $\mathcal{C}$ is obtained from the following equations, which are congruences modulo the image of $\delta\colon\mu^1(C_m)\to\mu^2(C_m)$:*

(E$_1$) $\sum_x\left[\begin{smallmatrix}x&x&c\\2&3&3\end{smallmatrix}\right]-\sum_x\left[\begin{smallmatrix}c&c&x\\2&3&3\end{smallmatrix}\right]\equiv\sum_{\substack{y\neq3\\ z\in\{1,3\}}}\Big(\sum_x\left[\begin{smallmatrix}x&x&c\\z&y&y\end{smallmatrix}\right]-\sum_x\left[\begin{smallmatrix}c&c&x\\z&y&y\end{smallmatrix}\right]\Big)$ *for* $c>3$.

(E$_2$) $\sum_x\left[\begin{smallmatrix}x&x&3\\2&4&4\end{smallmatrix}\right]-\sum_x\left[\begin{smallmatrix}3&3&x\\2&4&4\end{smallmatrix}\right]\equiv\sum_{\substack{y\neq4\\ z\in\{1,4\}}}\Big(\sum_x\left[\begin{smallmatrix}x&x&3\\z&y&y\end{smallmatrix}\right]-\sum_x\left[\begin{smallmatrix}3&3&x\\z&y&y\end{smallmatrix}\right]\Big)$.

(E$_3$) $\sum_{x,y}\left[\begin{smallmatrix}x&x&c\\y&d&d\end{smallmatrix}\right]-\sum_{x,y}\left[\begin{smallmatrix}c&c&x\\y&d&d\end{smallmatrix}\right]\equiv\sum_{x,y}\left[\begin{smallmatrix}x&x&c\\d&y&y\end{smallmatrix}\right]-\sum_{x,y}\left[\begin{smallmatrix}c&c&x\\d&y&y\end{smallmatrix}\right]$ *for* $3\le c\neq d\ge4$ *with* $(c,d)\neq(3,4)$.

(E$_4$) $\sum_{x,y}\left[\begin{smallmatrix}x&x&2\\y&4&4\end{smallmatrix}\right]-\sum_{(x,y)\neq(3,1)}\left[\begin{smallmatrix}2&2&x\\y&4&4\end{smallmatrix}\right]\equiv\Big(\sum_{x,y}\left[\begin{smallmatrix}x&x&2\\4&y&y\end{smallmatrix}\right]-\sum_{x,y}\left[\begin{smallmatrix}2&2&x\\4&y&y\end{smallmatrix}\right]\Big)-\Big(\sum_{\substack{y>4\\ x}}\left[\begin{smallmatrix}x&x&3\\1&y&y\end{smallmatrix}\right]-\sum_{x,y}\left[\begin{smallmatrix}3&3&x\\1&y&y\end{smallmatrix}\right]\Big)$.

(E$_5$) $\sum_{x,y}\left[\begin{smallmatrix}x&x&2\\y&d&d\end{smallmatrix}\right]-\sum_{x,y}\left[\begin{smallmatrix}2&2&x\\y&d&d\end{smallmatrix}\right]\equiv\sum_{x,y}\left[\begin{smallmatrix}x&x&2\\d&y&y\end{smallmatrix}\right]-\sum_{x,y}\left[\begin{smallmatrix}2&2&x\\d&y&y\end{smallmatrix}\right]$ *for* $d>4$.

(E$_6$) $\sum_{x,y}\left[\begin{smallmatrix}x&x&c\\1&y&y\end{smallmatrix}\right]\equiv\sum_{x,y}\left[\begin{smallmatrix}c&c&x\\1&y&y\end{smallmatrix}\right]$ *for* $c>2$.

Note that (E$_5$) says that the congruence in (E$_3$) also holds for $c=2$ provided $d>4$. Likewise, (E$_1$) and (E$_2$) can be stated simultaneously as

$$\sum_x\begin{bmatrix}x&x&c\\2&q&q\end{bmatrix}-\sum_x\begin{bmatrix}c&c&x\\2&q&q\end{bmatrix}\equiv\sum_{\substack{y\neq q\\ z\in\{1,q\}}}\left(\sum_x\begin{bmatrix}x&x&c\\z&y&y\end{bmatrix}-\sum_x\begin{bmatrix}c&c&x\\z&y&y\end{bmatrix}\right),$$

with $c>3=q$ or $c=3=q-1$. We have chosen the structure stated in (E$_1$)–(E$_6$) for proof-organization purposes; see Figure 13.

Proof The first assertion follows from (41) and (42). For the second assertion, start by noting that the listed congruences are obtained by dualizing the 16 formulae (in Definition 4.9) that describe the inclusion ι. Indeed, the validity of the congruences is obtained by a straightforward verification (left as an exercise for the reader) of the fact that both sides of each congruence evaluate the same at each basis element $\left\langle\begin{smallmatrix}a&a&c\\b&d&d\end{smallmatrix}\right\rangle$. Furthermore, direct inspection shows that, in each equation (E$_\mathrm{i}$), there is a single

i	1	2	3	4	5	6	6
s_i	$\left[\begin{smallmatrix}1&1&c\\2&3&3\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}1&1&3\\2&4&4\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}1&1&c\\2&d&d\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}1&1&2\\3&4&4\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}1&1&2\\3&d&d\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}2&2&c\\1&3&3\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}2&2&3\\1&4&4\end{smallmatrix}\right]$
type	(37)	(37)	(37)	(38)	(38)	(39)	(40)
restrictions	$c>3$		$3\le c\ne d\ge 4$ $(c,d)\ne(3,4)$		$d>4$	$c>3$	

Figure 13: Elements coming from $\mathcal{C}$ in the congruences ($\mathrm{E_i}$) of Corollary 4.11.

summand s_i (spelled out in Figure 13) that fails to come from $\mathcal{B}$. Therefore ($\mathrm{E_i}$) can be thought of as expressing the cohomology class represented by s_i as a $\mathbb{Z}$–linear combination of basis elements. The second assertion of the corollary then follows by observing, from Figure 13, that each element in $\mathcal{C}$ arises as one, and only one, of the special summands s_i. □

4.4 Cohomology ring

In previous sections we have described explicit cocycles in $\mu^*(C_m)$ representing basis elements in cohomology. We now make use of (9) to assess the corresponding cup products at both the critical cochain and homology levels. Since cup products in $C^*(C_m)$ are elementary (see Remark 4.12), the bulk of the work amounts to giving a (suitable) description of the cochain maps $\overline{\Phi}\colon \mu^*(C_m)\to C^*(C_m)$ and $\underline{\Phi}\colon C^*(C_m)\to\mu^*(C_m)$.

Remark 4.12 Recall that basis elements in $C^1(C_m)$ are given by the dualized 1–dimensional faces $\left[\begin{smallmatrix}a&c\\b&d\end{smallmatrix}\right]$. (As in earlier parts of the paper, upper stars for dualized elements are omitted, and arithmetic restrictions among the numbers assembling critical faces are usually not written down.) From the usual formula for cup products in the simplicial setting, we see that the only nontrivial products in $C^*(C_m)$ have the form

(43) $$\begin{bmatrix}a&a\\b&d\end{bmatrix}\smile\begin{bmatrix}a&c\\d&d\end{bmatrix}=\begin{bmatrix}a&a&c\\b&d&d\end{bmatrix}\quad\text{or}\quad\begin{bmatrix}a&c\\b&b\end{bmatrix}\smile\begin{bmatrix}c&c\\b&d\end{bmatrix}=\begin{bmatrix}a&c&c\\b&b&d\end{bmatrix}.$$

(So every 2–face is uniquely a product of two 1–faces.) In particular, for the purposes of applying (9), all basis elements $\left[\begin{smallmatrix}a&c\\b&d\end{smallmatrix}\right]$ with $a<c$ and $b<d$ can be ignored in the expression for $\overline{\Phi}$.

Proposition 4.13 *The values of the cochain map* $\overline{\Phi}\colon\mu^1(C_m)\to C^1(C_m)$ *on the basis elements* (k.1)–(k.3) *of Proposition 4.1 satisfy the following family of congruences taken modulo basis elements* $\left[\begin{smallmatrix}a&c\\b&d\end{smallmatrix}\right]$ *with* $a<c$ *and* $b<d$*:*

(i) $\overline{\Phi}\left(\left[\begin{smallmatrix}a&m-1\\b&m\end{smallmatrix}\right]\right)\equiv\left[\begin{smallmatrix}a&a\\b&m\end{smallmatrix}\right]+\sum\left[\begin{smallmatrix}a&x\\b&b\end{smallmatrix}\right]-\sum\left[\begin{smallmatrix}y&a\\b&b\end{smallmatrix}\right]$, *where the first summation runs over* $x\in\{1,\dots,m-1\}$ *with* $a<x\ne b$*, and the second summation runs over* $y\in\{1,\dots,m-1\}$ *with* $b\ne y<a$*.*

(ii) $\overline{\Phi}\left(\left[\begin{smallmatrix}m&m\\b&d\end{smallmatrix}\right]\right)\equiv\sum\left[\begin{smallmatrix}x&x\\b&d\end{smallmatrix}\right]$, *where the summation runs over* $x\in\{1,\dots,m\}$ *with* $b\ne x\ne d$*.*

(iii) $\overline{\Phi}\left(\left[\begin{smallmatrix}a&c\\m&m\end{smallmatrix}\right]\right)\equiv\sum\left[\begin{smallmatrix}a&c\\y&y\end{smallmatrix}\right]$, *where the summation runs over* $y\in\{1,\dots,m\}$ *with* $a\ne y\ne c$*.*

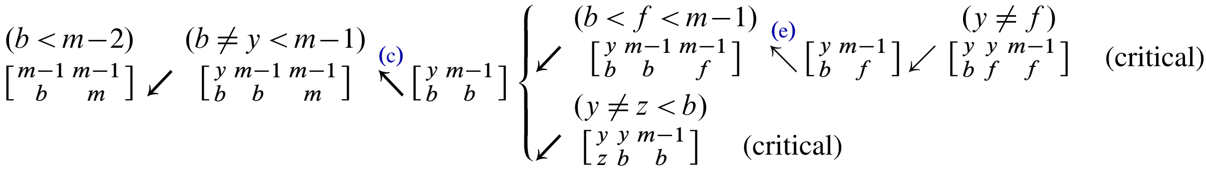


Figure 14: Gradient paths landing on a critical cell $\left[\begin{smallmatrix} m-1 & m-1 \\ b & m \end{smallmatrix}\right]$ of type (k.1) with $b<m-2$.

Proof The congruences follow from (8) and from direct inspection of Figures 14–17, where we spell out the complete trees of gradient paths landing on critical 1–dimensional faces. Here we follow the notational conventions used in Figures 3–11, except that we now keep track of relevant numerical restrictions and, at the start of each path, we indicate the reason that prevents the path from pulling back one further step. □

Proposition 4.14 *For critical* 1*–faces x and y, the product $\overline{\Phi}(x)\smile\overline{\Phi}(y)$ appearing in* (9) *is a linear combination $\sum \pm z$ of dualized* 2*–faces z, each of which has one of the following forms*:

(i) $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ *for $a<c<m>d>b$ and $b\neq a\neq d\neq c$, with trivial $\underline{\Phi}$–image unless $b\neq c$, in which case*

$$\underline{\Phi}\left(\begin{bmatrix} a & a & c \\ b & d & d \end{bmatrix}\right)=\begin{bmatrix} a & a & c \\ b & d & d \end{bmatrix},$$

(ii) $\left[\begin{smallmatrix} a & c & c \\ b & b & d \end{smallmatrix}\right]$ *for $a<c\leq m-1>d>b$ and $a\neq b\neq c\neq d$, with trivial $\underline{\Phi}$–image unless $a\neq d$, in which case*

$$\underline{\Phi}\left(\begin{bmatrix} a & c & c \\ b & b & d \end{bmatrix}\right)=-\begin{bmatrix} a & a & c \\ b & d & d \end{bmatrix},$$

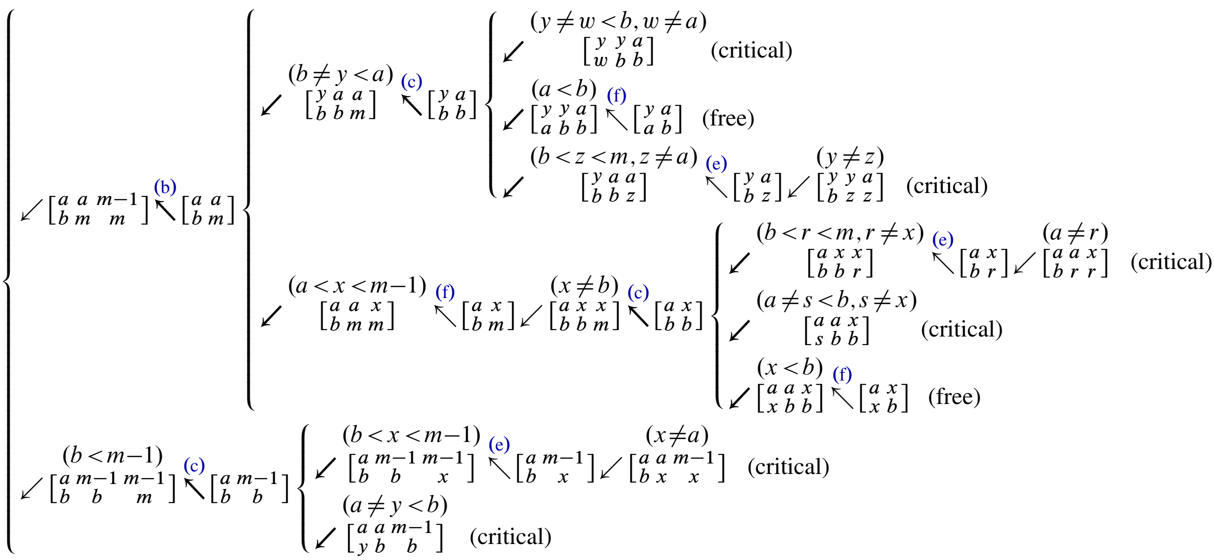


Figure 15: Gradient paths landing on a critical cell $\left[\begin{smallmatrix} a & m-1 \\ b & m \end{smallmatrix}\right]$ of type (k.1) with $a<m-1\geq b$.

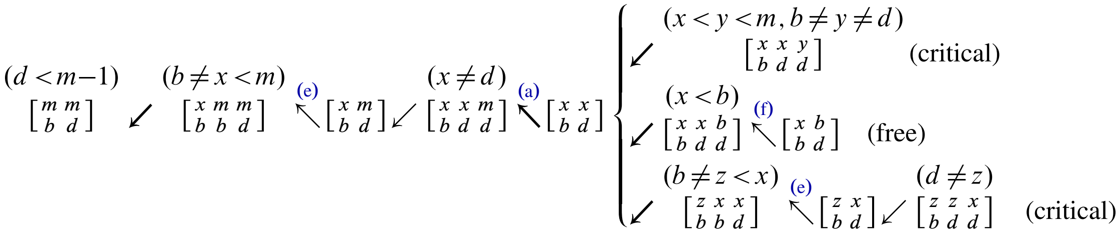


Figure 16: Gradient paths landing on a critical cell $\left[\begin{smallmatrix} m & m \\ b & d \end{smallmatrix}\right]$ of type (k.2).

(iii) $\left[\begin{smallmatrix} a & a & c \\ b & m & m \end{smallmatrix}\right]$ *for* $a < c < m-1 \geq b$ *and* $b \neq a \neq m \neq c$, *with trivial* $\underline{\Phi}$*–image unless* $b \neq c$, *in which case*

$$\underline{\Phi}\left(\begin{bmatrix} a & a & c \\ b & m & m \end{bmatrix}\right) = \sum_{\substack{y<b \\ a \neq y \neq c}} \begin{bmatrix} a & a & c \\ y & b & b \end{bmatrix} - \sum_{\substack{b<x<m \\ a \neq x \neq c}} \begin{bmatrix} a & a & c \\ b & x & x \end{bmatrix},$$

(iv) $\left[\begin{smallmatrix} a & c & c \\ b & b & m \end{smallmatrix}\right]$ *for* $a < c \leq m-1 \geq b$, $a \neq b \neq c$ *and either* $c < m-1$ *or* $c = m-1 > b+1$, *with* $\underline{\Phi}$*–image*

$$\underline{\Phi}\left(\begin{bmatrix} a & c & c \\ b & b & m \end{bmatrix}\right) = \sum_{\substack{b<x<m \\ a \neq x \neq c}} \begin{bmatrix} a & a & c \\ b & x & x \end{bmatrix} - \sum_{\substack{y<b \\ a \neq y \neq c}} \begin{bmatrix} a & a & c \\ y & b & b \end{bmatrix}.$$

Proof By (43), the only 1–faces in the expression of $\overline{\Phi}(\delta)$ that can lead to a summand $\pm\left[\begin{smallmatrix} r & r & t \\ s & u & u \end{smallmatrix}\right]$ in the product $\overline{\Phi}(\gamma) \smile \overline{\Phi}(\delta)$ have the form $\pm\left[\begin{smallmatrix} r & t \\ u & u \end{smallmatrix}\right]$. From the expressions of $\overline{\Phi}$ in Proposition 4.13, this can hold only with $t < m$ and in fact $t < m-1$ whenever $u = m$, in view of the form of the basis elements of type (k.3). So $\left[\begin{smallmatrix} r & r & t \\ s & u & u \end{smallmatrix}\right]$ fits either (i) or (iii). Likewise, the only 1–faces in the expression of $\overline{\Phi}(\delta)$ that can lead to a summand $\pm\left[\begin{smallmatrix} r & t & t \\ s & s & u \end{smallmatrix}\right]$ in $\overline{\Phi}(\gamma) \smile \overline{\Phi}(\delta)$ have the form $\pm\left[\begin{smallmatrix} t & t \\ s & u \end{smallmatrix}\right]$, which can hold only under one of the following conditions:

- $u = m$ and $t \leq m-1$, as well as $s < m-2$ if $t = m-1$ (recall the form of basis elements of type (k.1)).
- $u < m-1$ (recall the form of basis elements of type (k.2)).

In the former possibility, $\left[\begin{smallmatrix} r & t & t \\ s & s & u \end{smallmatrix}\right]$ fits (iv). In the latter, $\left[\begin{smallmatrix} r & t & t \\ s & s & u \end{smallmatrix}\right]$ fits (ii) unless $t = m$, in which case

(44) the expression of $\overline{\Phi}(\gamma)$ should include a summand of the form $\pm\left[\begin{smallmatrix} r & m \\ s & s \end{smallmatrix}\right]$.

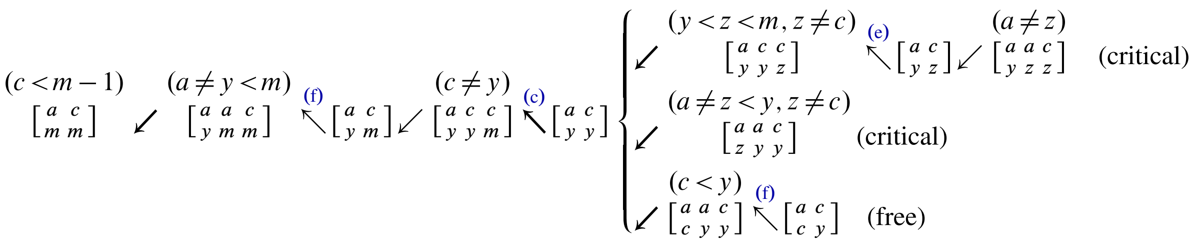


Figure 17: Gradient paths landing on a critical cell $\left[\begin{smallmatrix} a & c \\ m & m \end{smallmatrix}\right]$ of type (k.3).

	λ_{41}	λ_{42}	λ_{14}	λ_{24}	λ_{34}	λ_{32}	δ_{12}	δ_{13}	δ_{23}	υ_{12}	υ_{13}	υ_{23}
λ_{41}												$-g$
λ_{42}							g	$-g$	g	g		g
λ_{14}									$-g$			
λ_{24}								g				
λ_{34}								$-g$	g	g	$-g$	g
λ_{32}							$-g$	g	$-g$	$-g$	g	$-g$
δ_{12}		$-g$				g						
δ_{13}		g		$-g$	g	$-g$						
δ_{23}		$-g$	g		$-g$	g						
υ_{12}		$-g$			$-g$	g						
υ_{13}					g	$-g$						
υ_{23}	g	$-g$			$-g$	g						

Table 1: Here g stands for the generator of $H^2(\mathrm{Conf}(|K_m|,2))$, zeros are not shown, and brackets for cohomology classes are omitted.

But inspection of the expressions of $\overline{\Phi}$ in Proposition 4.13 rules out (44). Lastly, the four asserted expressions for the cochain map $\underline{\Phi}$ follow from (8) and the analysis of gradient paths in Figures 3–11. □

Example 4.15 For $m=5$, Corollaries 4.7 and 4.11 render the following list of cocycles representing a graded basis for $H^*(\mathrm{Conf}(|K_m|,2))$. In dimension 2 there is the single cocycle $\left[\begin{smallmatrix}3&3&4\\1&2&2\end{smallmatrix}\right]$, while in dimension 1 there are the twelve cocycles:

$$\delta_{12}:=\left[\begin{smallmatrix}5&5\\1&2\end{smallmatrix}\right],\quad \upsilon_{12}:=\left[\begin{smallmatrix}1&2\\5&5\end{smallmatrix}\right],\quad \delta_{13}:=\left[\begin{smallmatrix}5&5\\1&3\end{smallmatrix}\right],\quad \upsilon_{13}:=\left[\begin{smallmatrix}1&3\\5&5\end{smallmatrix}\right],\quad \delta_{23}:=\left[\begin{smallmatrix}5&5\\2&3\end{smallmatrix}\right],\quad \upsilon_{23}:=\left[\begin{smallmatrix}2&3\\5&5\end{smallmatrix}\right],$$

$$\lambda_{14}:=\left[\begin{smallmatrix}1&4\\4&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}1&4\\3&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}1&4\\2&5\end{smallmatrix}\right],\qquad \lambda_{24}:=\left[\begin{smallmatrix}2&4\\4&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}2&4\\3&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}2&4\\1&5\end{smallmatrix}\right],$$

$$\lambda_{41}:=\left[\begin{smallmatrix}4&4\\1&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}3&4\\1&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}2&4\\1&5\end{smallmatrix}\right],\qquad \lambda_{42}:=\left[\begin{smallmatrix}4&4\\2&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}3&4\\1&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}2&4\\3&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}2&4\\1&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}1&4\\3&5\end{smallmatrix}\right],$$

$$\lambda_{32}:=\left[\begin{smallmatrix}3&4\\2&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}3&4\\1&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}2&4\\3&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}2&4\\1&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}1&4\\3&5\end{smallmatrix}\right]+\left[\begin{smallmatrix}1&4\\2&5\end{smallmatrix}\right],\quad \lambda_{34}:=\left[\begin{smallmatrix}3&4\\4&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}2&4\\3&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}2&4\\1&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}1&4\\3&5\end{smallmatrix}\right]-\left[\begin{smallmatrix}1&4\\2&5\end{smallmatrix}\right].$$

Then the complete algebra structure of $H^*(\mathrm{Conf}(|K_m|,2))$ is spelled out by the matrix of cup products in Table 1. In particular, replacing λ_{42} by $\lambda'_{42}:=\lambda_{42}+\lambda_{41}+\lambda_{34}+\lambda_{32}+\lambda_{24}+\lambda_{14}$, λ_{34} by $\lambda'_{34}:=\lambda_{34}+\lambda_{32}$ and λ_{32} by $\lambda'_{32}:=\lambda_{32}+\lambda_{42}$, we get a cohomology basis whose only (up to anticommutativity) nontrivial products are

$$(45)\quad \lambda'_{42}\smile\upsilon_{12}=\lambda_{24}\smile\delta_{13}=\lambda'_{32}\smile\upsilon_{13}=g \quad\text{and}\quad \lambda_{14}\smile\delta_{23}=\lambda_{41}\smile\upsilon_{23}=\lambda'_{34}\smile\delta_{12}=-g.$$

The cohomology ring $H^*(\mathrm{Conf}(|K_m,2|))$ becomes richer as m increases (with $\mathrm{Conf}(|K_m|,2)$ no longer being a homotopy closed surface). Yet, some aspects of the particularly simple structure in (45) are kept for all $m>5$. Explicitly, let $\upsilon_{a,c}$, $\delta_{b,d}$ and $\lambda_{e,f}$ stand for the basis elements of $H^1(\mathrm{Conf}(|K_m,2|))$ represented, respectively, by the 1–cocycles $\left[\begin{smallmatrix}a&c\\m&m\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}m&m\\b&d\end{smallmatrix}\right]$ and $\left\langle\begin{smallmatrix}e\\f\end{smallmatrix}\right\rangle$ described in Corollary 4.7. Then:

$\delta_{b,d} \cdot \upsilon_{a,c}$, $a \in \{b,d\}$	$\upsilon_{a,c} \cdot \delta_{b,d}$, $c \in \{b,d\}$
$\lambda_{m-1,f_1} \cdot \lambda_{m-1,f_2}$, $1 \le f_i \le m-4$	$\lambda_{m-1,f} \cdot \lambda_{e,m-1}$, $1 \le f \le m-4$ and $1 \le e \le m-3$
$\lambda_{m-1,f} \cdot \lambda_{m-1,m-3}$, $1 \le f \le m-4$	
$\lambda_{m-1,f} \cdot \lambda_{m-2,m-3}$, $1 \le f \le m-4$	
$\lambda_{m-1,f} \cdot \lambda_{m-2,m-1}$, $1 \le f \le m-4$	
$\lambda_{m-1,m-3} \cdot \lambda_{m-2,m-3}$	$\lambda_{m-1,m-3} \cdot \lambda_{e,m-1}$, $1 \le e \le m-3$
$\lambda_{m-1,m-3} \cdot \lambda_{m-2,m-1}$	
	$\lambda_{e_1,m-1} \cdot \lambda_{e_2,m-1}$, $1 \le e_i \le m-3$ with $e_1 > e_2$
$\lambda_{m-2,m-3} \cdot \lambda_{m-2,m-1}$	$\lambda_{m-2,m-3} \cdot \lambda_{e,m-1}$, $1 \le e \le m-3$
	$\lambda_{m-2,m-1} \cdot \lambda_{e,m-1}$, $1 \le e \le m-3$

Table 2

Corollary 4.16 *Any cup product of the form $\delta_{b_1,d_1} \cdot \delta_{b_2,d_2}$, $\upsilon_{a_1,c_1} \cdot \upsilon_{a_2,c_2}$ or $\lambda_{e_1,f_1} \cdot \lambda_{e_2,f_2}$ vanishes. On the other hand, a cup product $\delta_{b,d} \cdot \upsilon_{a,c}$ is nonzero if and only if $\{a,b\} \cap \{c,d\} = \varnothing$, in which case $\delta_{b,d} \cdot \upsilon_{a,c}$ is represented by $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$.*

Proof This is a straightforward calculation using Proposition 4.13. We only indicate the two main checking steps for the reader's benefit. In what follows we assume $m \ge 6$. First, Proposition 4.13(i) is used to check that, modulo 1–faces not taking part on nonzero products (43), $\overline{\Phi}\left\langle\begin{smallmatrix} e \\ f \end{smallmatrix}\right\rangle$ is congruent to

- $\sum_{i<m}\left[\begin{smallmatrix} i & i \\ f & m \end{smallmatrix}\right]$ for $e = m-1$ and $1 \le f \le m-4$,
- $\sum_{j<m}\left(\left[\begin{smallmatrix} e & e \\ j & m \end{smallmatrix}\right] + \sum_{x<m}\left[\begin{smallmatrix} e & x \\ j & j \end{smallmatrix}\right] - \sum_{y<m}\left[\begin{smallmatrix} y & e \\ j & j \end{smallmatrix}\right]\right)$ for $1 \le e \le m-3$ and $f = m-1$,
- $\left[\begin{smallmatrix} m-1 & m-1 \\ m-3 & m \end{smallmatrix}\right] - \sum_{\substack{i<m-1>j \\ j \ne m-3}}\left[\begin{smallmatrix} i & i \\ j & m \end{smallmatrix}\right] - \sum_{i<m-1>j}\left[\begin{smallmatrix} i & m-1 \\ j & j \end{smallmatrix}\right]$ for $(e,f) = (m-1,m-3)$,
- $\left[\begin{smallmatrix} m-2 & m-2 \\ m-1 & m \end{smallmatrix}\right] + \left[\begin{smallmatrix} m-2 & m-1 \\ m-1 & m-1 \end{smallmatrix}\right] - \sum_y\left[\begin{smallmatrix} y & m-2 \\ m-1 & m-1 \end{smallmatrix}\right] - \sum_{\substack{i<m-1>j \\ i \ne m-2}}\left(\left[\begin{smallmatrix} i & i \\ j & m \end{smallmatrix}\right] + \sum_{\substack{x=m-\varepsilon \\ \varepsilon\in\{1,2\}}}\left[\begin{smallmatrix} i & x \\ j & j \end{smallmatrix}\right]\right)$ for $(e,f) = (m-2,m-1)$,
- $\sum_{i<m-1>j}\left(\left[\begin{smallmatrix} i & i \\ j & m \end{smallmatrix}\right] + \left[\begin{smallmatrix} i & m-1 \\ j & j \end{smallmatrix}\right]\right)$ for $(e,f) = (m-2,m-3)$.

The above congruences together with those in Proposition 4.13(ii) and (iii) are then used to check that each of the products asserted to vanish do so because there is no room for nonzero products (43) in the corresponding portion $\overline{\Phi}(x) \smile \overline{\Phi}(y)$ of (9). Such an assertion is easily seen for products $\delta_{b_1,d_1} \cdot \delta_{b_2,d_2}$ and $\upsilon_{a_1,c_1} \cdot \upsilon_{a_2,c_2}$, but the explicit details are not so direct for $\delta_{b,d} \cdot \upsilon_{a,c}$ and $\lambda_{e_1,f_1} \cdot \lambda_{e_2,f_2}$. In fact, in the latter two cases, a convenient order of factors needs to be chosen in order to ensure the vanishing of the corresponding $\overline{\Phi}(x) \smile \overline{\Phi}(y)$; see Table 2. The order chosen is immaterial for the trivial-product conclusion, as cohomology cup products are anticommutative. Keep in mind that $H^*(\mathrm{Conf}(|K_m|,2))$ is torsion-free, so cup squares of 1–dimensional classes are trivial for free.

Lastly, the fact that $\delta_{b,d} \cdot \upsilon_{a,c}$ is represented by $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ when $\{a,b\} \cap \{c,d\} = \varnothing$ follows by noticing that $\underline{\Phi}\left(\overline{\Phi}\left[\begin{smallmatrix} m & m \\ b & d \end{smallmatrix}\right] \smile \overline{\Phi}\left[\begin{smallmatrix} a & c \\ m & m \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$. Here $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ fails to represent one of our basis elements when

$(a,b)=(1,2)$, $(a,b,c)=(1,3,2)$, $(a,b,d)=(2,1,3)$ or $(a,b,c,d)=(2,1,3,4)$; recall (37)–(40). In each such case, one of Corollary 4.11(E_1)–(E_6) applies to write (the cohomology class of) $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ in terms of basis elements. Either way, inspection of (E_1)–(E_6) shows that $\left[\begin{smallmatrix} a & a & c \\ b & d & d \end{smallmatrix}\right]$ represents a nonzero cohomology class. □

5 Topological complexity

Fix a positive integer $s \geq 2$ and a path-connected space X. The s^{th} topological complexity $\mathrm{TC}_s(X)$ of X is the sectional category of the evaluation map $e_s \colon PX \to X^s$ which sends a (free) path on X, $\gamma \in PX$, to

$$e_s(\gamma) = \left(\gamma\left(\frac{0}{s-1}\right), \gamma\left(\frac{1}{s-1}\right), \dots, \gamma\left(\frac{s-1}{s-1}\right)\right).$$

The term "sectional category" is used in the reduced sense, so $\mathrm{TC}_s(X)+1$ stands for the smallest number of open sets covering X^s on each of which e_s admits a section. For instance, the (reduced) Lusternik–Schnirelmann category $\mathrm{cat}(X)$ of X is the sectional category of the evaluation map $e_1 \colon P_0X \to X$ sending a based path $\gamma \in P_0X$ (ie $\gamma(0)=\star$ for a fixed base point $\star \in X$) to $e_1(\gamma)=\gamma(1)$.

Proposition 5.1 [5, Theorem 3.9] *For a c–connected space X having the homotopy type of a CW complex,*

$$\mathrm{cl}(X) \leq \mathrm{cat}(X) \leq \mathrm{hdim}(X)/(c+1) \quad \textit{and} \quad \mathrm{zcl}_s(X) \leq \mathrm{TC}_s(X) \leq s\,\mathrm{cat}(X).$$

Here $\mathrm{hdim}(X)$ denotes the minimal dimension of cell complexes homotopy equivalent to X, while $\mathrm{cl}(X)$ and $\mathrm{zcl}_s(X)$ stand, respectively, for the cup length of X and the s^{th} zero-divisor cup length of X. Explicitly, $\mathrm{cl}(X)$ is the largest integer $l \geq 0$ such that there are classes[6] $c_j \in \widetilde{H}^*(X)$ for $1 \leq j \leq l$ with nonzero cup product. Likewise, $\mathrm{zcl}_s(X)$ is the largest integer $l \geq 0$ such that there are classes $z_j \in H^*(X^s)$ for $1 \leq j \leq l$ ("zero divisors") with nonzero cup product and such that each factor restricts trivially under the diagonal $X \hookrightarrow X^s$.

Let Γ be a 1–dimensional cell complex — a graph. While the fundamental group of $\mathrm{Conf}(\Gamma,n)$ is a central character in geometric group theory, the topological complexity of $\mathrm{Conf}(\Gamma,n)$ becomes relevant for the task of planning collision-free motion of n autonomous distinguishable agents moving on a Γ–shaped system of tracks. It is known that $\mathrm{hdim}(\mathrm{Conf}(\Gamma,n))$ is bounded from above by $m=m(\Gamma)$, the number of essential vertices of Γ; see for instance [12, Theorem 4.4]. Thus Proposition 5.1 yields

$$\mathrm{TC}_s(\mathrm{Conf}(\Gamma,n)) \leq sm. \tag{46}$$

For $s=2$, Farber proved in [8] that (46) is an equality when Γ is a tree and $n \geq 2m$, with the single (and well known) exception of $(n,m)=(2,1)$ where the (unique) essential vertex of Γ has valency 3 — which we call the "Y_2–exception". Farber also conjectured that the tree restriction is superfluous in obtaining equality in (46). The conjecture has recently been confirmed in [23] by Knudsen, who proved

[6] For the purposes of this section, cohomology will be taken with mod 2 coefficients.

equality in (46) for any $s \geq 2$ and any graph Γ, as long as the "stable" restriction $n \geq 2m$ is kept (and the Y_2–exception is avoided). Note that the stable condition forces $\mathrm{hdim}(\mathrm{Conf}(\Gamma, n)) = m$. More generally, it would be interesting to characterize the triples (s, Γ, n) for which the (in principle) improved bound

$$\mathrm{TC}_s(\mathrm{Conf}(\Gamma, n)) \leq s\, \mathrm{hdim}(\mathrm{Conf}(\Gamma, n)) \tag{47}$$

holds as an equality, preferably determining the value of $\mathrm{hdim}(\mathrm{Conf}(\Gamma, n))$. For instance, it is known from [2, Section 5] that, for any s and n (possibly with $n < 2m$),

$$\mathrm{hdim}(\mathrm{Conf}(\Gamma, n)) = \mathrm{cat}(\mathrm{Conf}(\Gamma, n)) = \min\{\lfloor n/2 \rfloor, m\} \tag{48}$$

when Γ is a tree, in which case (47) is an equality — the Y_2–exception still applies. The goal of this section is to prove Theorem 1.2, which adds a new and completely different family of instances where equality holds in (47) outside the stable regime $n \geq 2m$.

Note that $\mathrm{Conf}(|K_m|, 2)$ is empty for $m = 1$, and disconnected for $m = 2$, while $\mathrm{Conf}(|K_3|, 2)$ leads to the Y_2–exception. On the other hand, the cases $m = 4$ and $m = 5$ in Theorem 1.2 are well known in view of Corollary 4.2 and the last assertion in Example 4.15. We prove Theorem 1.2 for $m \geq 6$ by constructing $2s$ zero divisors in $\mathrm{Conf}(|K_m|, 2)$ with a nonzero cup product, and using Proposition 5.1 together with the obvious fact that $\mathrm{hdim}(\mathrm{Conf}(|K_m|, 2)) \leq 2$. It is natural to think that the expected richness of cup products in general graph configuration spaces might lead to many more instances where (47) would hold as an equality — even if $n < 2m$.

For integers $1 \leq i \leq s \geq 2$ and a cohomology class x in a space X, consider the exterior tensor product $x_{(i)} := 1 \otimes \cdots \otimes 1 \otimes x \otimes 1 \otimes \cdots \otimes 1 \in H^*(X)^{\otimes s} = H^*(X^s)$, where the tensor factor x appears in the i^{th} position. The following result is straightforward to check:

Lemma 5.2 *Let x, y, z and w be four elements in the mod 2 cohomology of a space X satisfying the relations $x^2 = y^2 = xz = yz = yw = 0$. Then*

$$\left(\prod_{i=2}^{s}(x_{(1)} + x_{(i)})\right)\left(\prod_{i=2}^{s}(y_{(1)} + y_{(i)})\right)(z_{(1)} + z_{(s)})(w_{(1)} + w_{(s)})$$
$$= zw \otimes xy \otimes xy \otimes \cdots \otimes xy + xy \otimes xy \otimes \cdots \otimes xy \otimes zw.$$

Proof of Theorem 1.2 for $m \geq 6$ In view of Corollary 4.16 and Lemma 5.2, the 1–dimensional basis elements $x := \delta_{1,2}$, $y := \upsilon_{3,4}$, $z := \upsilon_{1,3}$, $w := \delta_{2,4} \in H^* \mathrm{Conf}(|K_m|, 2)$ yield a product of $2s$ zero divisors, with product-representative

$$\begin{bmatrix}1&1&3\\2&4&4\end{bmatrix} \otimes \begin{bmatrix}3&3&4\\1&2&2\end{bmatrix} \otimes \begin{bmatrix}3&3&4\\1&2&2\end{bmatrix} \otimes \cdots \otimes \begin{bmatrix}3&3&4\\1&2&2\end{bmatrix} + \begin{bmatrix}3&3&4\\1&2&2\end{bmatrix} \otimes \begin{bmatrix}3&3&4\\1&2&2\end{bmatrix} \otimes \cdots \otimes \begin{bmatrix}3&3&4\\1&2&2\end{bmatrix} \otimes \begin{bmatrix}1&1&3\\2&4&4\end{bmatrix}. \tag{49}$$

The tensor factor $\left[\begin{smallmatrix}3&3&4\\1&2&2\end{smallmatrix}\right]$ represents one of the basis elements in the previous section. However, as indicated in Figure 13, we need to apply relation (E_2) in Corollary 4.11 in order to write the (cohomology class of the) tensor factor $\left[\begin{smallmatrix}1&1&3\\2&2&4\end{smallmatrix}\right]$ as a sum $\sum b_i$ of basis elements b_i (recall, we work mod 2). If $m \geq 6$, the basis element $\left[\begin{smallmatrix}3&3&5\\2&4&4\end{smallmatrix}\right]$ appears as a summand b_i, from which the nontriviality of the cohomology class represented by (49) follows. □

References

[1] **A D Abrams**, *Configuration spaces and braid groups of graphs*, PhD thesis, University of California, Berkeley (2000) Available at `https://www.proquest.com/docview/304583880`

[2] **J Aguilar-Guzmán**, **J González**, **T Hoekstra-Mendoza**, *Farley–Sabalka's Morse-theory model and the higher topological complexity of ordered configuration spaces on trees*, Discrete Comput. Geom. 67 (2022) 258–286 MR Zbl

[3] **B H An**, **G C Drummond-Cole**, **B Knudsen**, *Edge stabilization in the homology of graph braid groups*, Geom. Topol. 24 (2020) 421–469 MR Zbl

[4] **K Barnett**, **M Farber**, *Topology of configuration space of two particles on a graph, I*, Algebr. Geom. Topol. 9 (2009) 593–624 MR Zbl

[5] **I Basabe**, **J González**, **Y B Rudyak**, **D Tamaki**, *Higher topological complexity and its symmetrization*, Algebr. Geom. Topol. 14 (2014) 2103–2124 MR Zbl

[6] **U Bauer**, **C Lange**, **M Wardetzky**, *Optimal topological simplification of discrete functions on surfaces*, Discrete Comput. Geom. 47 (2012) 347–377 MR Zbl

[7] **S Chettih**, **D Lütgehetmann**, *The homology of configuration spaces of trees with loops*, Algebr. Geom. Topol. 18 (2018) 2443–2469 MR Zbl

[8] **M Farber**, *Collision free motion planning on graphs*, from "Algorithmic Foundations of Robotics, VI" (M Erdmann, M Overmars, D Hsu, F van der Stappen, editors), Springer Tracts Adv. Robot. 17, Springer (2005) 123–138

[9] **M Farber**, **E Hanbury**, *Topology of configuration space of two particles on a graph, II*, Algebr. Geom. Topol. 10 (2010) 2203–2227 MR Zbl

[10] **D Farley**, *Homology of tree braid groups*, from "Topological and asymptotic aspects of group theory" (R Grigorchuk, M Mihalik, M Sapir, Z Šunić, editors), Contemp. Math. 394, Amer. Math. Soc., Providence, RI (2006) 101–112 MR Zbl

[11] **D Farley**, *Presentations for the cohomology rings of tree braid groups*, from "Topology and robotics" (M Farber, R Ghrist, M Burger, D Koditschek, editors), Contemp. Math. 438, Amer. Math. Soc., Providence, RI (2007) 145–172 MR Zbl

[12] **D Farley**, **L Sabalka**, *Discrete Morse theory and graph braid groups*, Algebr. Geom. Topol. 5 (2005) 1075–1109 MR Zbl

[13] **D Farley**, **L Sabalka**, *On the cohomology rings of tree braid groups*, J. Pure Appl. Algebra 212 (2008) 53–71 MR Zbl

[14] **D Farley**, **L Sabalka**, *Presentations of graph braid groups*, Forum Math. 24 (2012) 827–859 MR Zbl

[15] **R Forman**, *A discrete Morse theory for cell complexes*, from "Geometry, topology, & physics" (S-T Yau, editor), Conf. Proc. Lect. Notes Geom. Topol. 4, International Press, Cambridge, MA (1995) 112–125 MR Zbl

[16] **R Forman**, *Discrete Morse theory and the cohomology ring*, Trans. Amer. Math. Soc. 354 (2002) 5063–5085 MR Zbl

[17] **R Ghrist**, *Configuration spaces and braid groups on graphs in robotics*, from "Knots, braids, and mapping class groups" (J Gilman, W W Menasco, X-S Lin, editors), AMS/IP Stud. Adv. Math. 24, Amer. Math. Soc., Providence, RI (2001) 29–40 MR Zbl

[18] **R Ghrist**, *Configuration spaces, braids, and robotics*, from "Braids" (A J Berrick, F R Cohen, E Hanbury, Y-L Wong, J Wu, editors), Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. 19, World Sci., Hackensack, NJ (2010) 263–304 MR Zbl

[19] **R W Ghrist**, **D E Koditschek**, *Safe cooperative robot dynamics on graphs*, SIAM J. Control Optim. 40 (2002) 1556–1575 MR Zbl

[20] **E J Gonzalez**, **J Gonzalez**, *An algorithmic discrete gradient field for non-colliding cell-like objects and the topology of pairs of points on skeleta of simplexes* (2023) arXiv 2307.14454

[21] **J González**, **T Hoekstra-Mendoza**, *Cohomology ring of tree braid groups and exterior face rings*, Trans. Amer. Math. Soc. Ser. B 9 (2022) 1065–1101 MR Zbl

[22] **S Harker**, **K Mischaikow**, **M Mrozek**, **V Nanda**, *Discrete Morse theoretic algorithms for computing homology of complexes and maps*, Found. Comput. Math. 14 (2014) 151–184 MR Zbl

[23] **B Knudsen**, *The topological complexity of pure graph braid groups is stably maximal*, Forum Math. Sigma 10 (2022) art. id. e93 MR Zbl

[24] **K H Ko**, **H W Park**, *Characteristics of graph braid groups*, Discrete Comput. Geom. 48 (2012) 915–963 MR Zbl

[25] **L Lampret**, *Chain complex reduction via fast digraph traversal*, preprint (2019) arXiv 1903.00783

[26] **T Lewiner**, **H Lopes**, **G Tavares**, *Applications of Forman's discrete Morse theory to topology visualization and mesh compression*, IEEE Trans. Visualiz. Comput. Graphics 10 (2004) 499–508

[27] **T Maciążek**, **A Sawicki**, *Homology groups for particles on one-connected graphs*, J. Math. Phys. 58 (2017) art. id. 062103 MR Zbl

[28] **F Mori**, **M Salvetti**, *(Discrete) Morse theory on configuration spaces*, Math. Res. Lett. 18 (2011) 39–57 MR Zbl

[29] **M Mrozek**, **B Batko**, *Coreduction homology algorithm*, Discrete Comput. Geom. 41 (2009) 96–118 MR Zbl

[30] **J R Munkres**, *Elements of algebraic topology*, Addison-Wesley, Menlo Park, CA (1984) MR Zbl

[31] **E Ramos**, *An application of the theory of FI-algebras to graph configuration spaces*, Math. Z. 294 (2020) 1–15 MR Zbl

[32] **M Salvetti**, **S Settepanella**, *Combinatorial Morse theory and minimality of hyperplane arrangements*, Geom. Topol. 11 (2007) 1733–1766 MR Zbl

[33] **C Severs**, **J A White**, *On the homology of the real complement of the k–parabolic subspace arrangement*, J. Combin. Theory Ser. A 119 (2012) 1336–1350 MR Zbl

[34] **J Shareshian**, *Discrete Morse theory for complexes of 2–connected graphs*, Topology 40 (2001) 681–701 MR Zbl

[35] **V A Vassiliev**, *Cohomology of knot spaces*, from "Theory of singularities and its applications" (V I Arnold, editor), Adv. Soviet Math. 1, Amer. Math. Soc., Providence, RI (1990) 23–69 MR Zbl

[36] **V A Vassiliev**, *Complexes of connected graphs*, from "The Gelfand Mathematical Seminars, 1990–1992" (L Corwin, I Gelfand, J Lepowsky, editors), Birkhäuser, Boston (1993) 223–235 MR Zbl

[37] **V A Vassiliev**, *Topology of two-connected graphs and homology of spaces of knots*, from "Differential and symplectic topology of knots and curves" (S Tabachnikov, editor), Amer. Math. Soc. Transl. Ser. 2 190, Amer. Math. Soc., Providence, RI (1999) 253–286 MR Zbl

Departamento de Matemáticas, Centro de Investigación y de Estudios Avanzados del IPN
Mexico City, Mexico

Departamento de Matemáticas, Centro de Investigación y de Estudios Avanzados del IPN
Mexico City, Mexico

jgonzalez@math.cinvestav.mx, jesus@math.cinvestav.mx

Received: 1 August 2022 Revised: 16 January 2023

Algebraic & Geometric Topology **24**:7 (2024) 3759–3799
DOI: 10.2140/agt.2024.24.3759
Published: 9 December 2024

Spectral diameter of Liouville domains

PIERRE-ALEXANDRE MAILHOT

The group of compactly supported Hamiltonian diffeomorphisms of a symplectic manifold is endowed with a natural bi-invariant distance, due to Viterbo, Schwarz, Oh, Frauenfelder and Schlenk, coming from spectral invariants in Hamiltonian Floer homology. This distance has found numerous applications in symplectic topology. However, its diameter is still unknown in general. In fact, for closed symplectic manifolds there is no unifying criterion for the diameter to be infinite. We prove that for any Liouville domain this diameter is infinite if and only if its symplectic cohomology does not vanish. This generalizes a result of Monzner, Vichery and Zapolsky and has applications in the setting of closed symplectic manifolds.

57R17, 51F99, 53D05, 57R58

1 Introduction and results

Liouville domains are a special kind of compact symplectic manifold with boundary. They are characterized by their exact symplectic form $\omega = \mathrm{d}\lambda$ and the fact that their boundary is of contact type. Liouville domains allow us to study under a common theoretical framework many important classes of symplectic manifolds. Examples of such manifolds include cotangent disk bundles over closed manifolds, complements of Donaldson divisors [Giroux 2017], preimages of some intervals under exhausting functions of Stein manifolds [Cieliebak and Eliashberg 2012], positive regions of convex hypersurfaces in contact manifolds [Giroux 1991] and total spaces of Lefschetz fibrations.

A key invariant of a Liouville domain D is its symplectic cohomology $\mathrm{SH}^*(D)$. It was first defined in [Floer and Hofer 1994; Cieliebak et al. 1995] and later developed in [Viterbo 1999]. Symplectic cohomology allows one to study the behavior of periodic Reeb orbits on the boundary of D. It is defined in terms of the Floer cohomology groups of a specific class of Hamiltonian functions on the completion $\widehat{D}$ of D which results from the gluing of the cylinder $[1,\infty)\times\partial D$ to ∂D.

The primary goal of this paper is to relate symplectic cohomology and spectral invariants, an important tool in Hamiltonian dynamics. When defined on a symplectic manifold (M,ω), spectral invariants associate to any pair $(\alpha, H)\in \mathrm{H}^*(M)\times C_c^\infty(S^1\times M)$ a real number $c(\alpha,H)$ that belongs to the spectrum of the action functional associated to H.[1] Spectral invariants were first defined in $\mathbb{R}^{2n}$ from the point of view of generating functions in [Viterbo 1992]. They were then constructed on closed symplectically aspherical manifolds in [Schwarz 2000] and general closed symplectic manifolds in [Oh 2005] (see also [Usher 2013]).

Frauenfelder and Schlenk [2007] constructed spectral invariants on Liouville domains. These spectral invariants are homotopy invariant in the Hamiltonian term in the following sense. If two compactly supported Hamiltonians H and F generate the same time-1 map, $\varphi_H=\varphi_F$, then $c(\alpha,H)=c(\alpha,F)$. Thus $c(\alpha,\cdot)$ descends to the group of compactly supported Hamiltonian diffeomorphisms $\mathrm{Ham}_c(D)$. This allows one to define a bi-invariant norm γ on $\mathrm{Ham}_c(D)$, called the spectral norm, by

$$\gamma(\varphi)=c(1,\varphi)+c(1,\varphi^{-1}).$$

One key feature of the spectral norm γ is the fact that it acts as a lower bound to the celebrated Hofer norm [1990] (see [Lalonde and McDuff 1995] and the book [Polterovich 2001] for further developments in the subject). It is thus natural to ask whether the spectral diameter

$$\mathrm{diam}_\gamma(M)=\sup\{\gamma(\varphi)\mid \varphi\in \mathrm{Ham}_c(M)\}$$

is finite or not. In particular, if $\mathrm{diam}_\gamma(M)=+\infty$ then the Hofer norm is assured to be unbounded. Further links between the spectral norm and Hofer geometry are discussed in Section 1.4.

1.1 Main results

We find a characterization of the finiteness of $\mathrm{diam}_\gamma(D)$ in the case of a Liouville domain $(D,\mathrm{d}\lambda)$ in terms of its symplectic cohomology.

Our main technical result shows that if $\mathrm{SH}^*(D)\neq 0$ then $c(1,H)$ can be made arbitrarily large. This, combined with the converse implication which was proved in [Benedetti and Kang 2022], implies:

Theorem A1 *Let (D,λ) be a Liouville domain. Then $\mathrm{diam}_\gamma(D)=+\infty$ if and only if $\mathrm{SH}^*(D)\neq 0$.*

As an intermediate step to proving Theorem A1, we show the following auxiliary result.

Lemma B *Let H be a compactly supported Hamiltonian on a Liouville domain (D,λ). Then,*

$$c(1,H)\geq 0.$$

[1] At least if the Hamiltonian satisfies certain technical conditions.

Lemma B is a cohomological adaptation for Liouville domains of a result of [Ganor and Tanny 2023, Lemma 4.1]. They show that $c([\mathrm{pt}], F) \leq 0$ for Hamiltonians F supported in certain incompressible domains of closed aspherical manifolds. It follows from [loc. cit., Section 5.1], that this inequality extends to Liouville domains. We remark that the inequality $c([M], F) \geq 0$ for the same class of Hamiltonians F was already shown by Humilière, Le Roux and Seyfaddini [Humilière et al. 2016]. It follows directly from [loc. cit., Theorem 45]. The main difference in Lemma B here is that we consider spectral invariants on Floer cohomology (instead of Floer homology) with respect to the unit (instead of the point class). Furthermore it applies to general Liouville domains without any ambient symplectic manifold. The proof uses an adaptation to Floer cohomology on Liouville domains of the barricade construction introduced in [Ganor and Tanny 2023].

In fact, when the symplectic cohomology of a Liouville domain is nonvanishing, the implication of Theorem A1 follows from a sharper result. Denote by $d_\gamma(\varphi, \psi) = \gamma(\varphi \circ \psi^{-1})$ the spectral distance on $\mathrm{Ham}_c(D)$ and by d_{st} the standard Euclidean distance on $\mathbb{R}$.

Theorem A2 *Let (D, λ) be a Liouville domain such that $\mathrm{SH}^*(D) \neq 0$. Then there exists an isometric group embedding $(\mathbb{R}, d_{\mathrm{st}}) \to (\mathrm{Ham}_c(D), d_\gamma)$.*

The proof of Theorem A2 uses an explicit construction of an isometric group embedding. This construction is a generalization of the procedure used by Monzner, Vichery and Zapolsky to prove Theorem 3 below. The construction of the aforementioned embedding relies primarily on the computation of spectral invariants of Hamiltonians which are constant on the skeleton of D, a special subset of Liouville domains which we define in Section 2.1.

Lemma C *Suppose (D, λ) is a Liouville domain such that $\mathrm{SH}^*(D) \neq 0$. Let H be a compactly supported autonomous Hamiltonian on D such that*

$$H|_{\mathrm{Sk}(D)} = -A \quad \text{and} \quad -A \leq H|_D \leq 0$$

for a constant $A > 0$. Then

$$c(1, H) = A.$$

1.2 What is already known for Liouville domains

Following the [Benedetti and Kang 2022], it is known that the spectral diameter of a Liouville domain D is bounded if its symplectic cohomology vanishes. This result was achieved using a special capacity derived from the filtered symplectic cohomology of D. To better understand how this is done, let us give an overview of the construction of $\mathrm{SH}^*(D)$ following [Viterbo 1999].

Consider the class of admissible Hamiltonians $H : \widehat{D} \to \mathbb{R}$ which are affine in the radial coordinate on the cylindrical part of $\widehat{D}$.[2] We can define the filtered Floer cohomology groups $\mathrm{HF}^*_{(a,b)}(H)$ of such

[2] See Definition 7 for the precise conditions.

Hamiltonians by considering only the 1–periodic orbits with action inside the interval (a,b).[3] Taking an increasing sequence of admissible Hamiltonians $\{H_i\}_i$ with corresponding slopes $\{\tau_i\}_i$ satisfying $\tau_i \to +\infty$, one can define the filtered symplectic cohomology $\mathrm{SH}^*_{(a,b)}(D)$ of D as

$$\mathrm{SH}^*_{(a,b)}(D) = \varinjlim_{H_i} \mathrm{HF}^*_{(a,b)}(H_i).$$

It follows from the above definition that for $a \le a'$ and $b \le b'$ there is a natural map

$$\iota^{b,b'}_{a,a'} \colon \mathrm{SH}^*_{(a,b)}(D) \longrightarrow \mathrm{SH}^*_{(a',b')}(D).$$

Moreover, the full symplectic cohomology $\mathrm{SH}^*(D) = \mathrm{SH}^*_{(-\infty,\infty)}(D)$ comes with a natural map

$$v^* \colon \mathrm{H}^*(D) \longrightarrow \mathrm{SH}^*(D)$$

called the Viterbo map. The failure of v^* to be an isomorphism signals the presence of Reeb orbits on the boundary of D. Thus, $\mathrm{SH}^*(D)$ is a useful tool to study the Weinstein conjecture [1979], which claims that on any closed contact manifold the Reeb vector field should admit at least one periodic orbit. For instance, Viterbo [1999] proved the Weinstein conjecture for the boundary of subcritical Stein manifolds.

We can extend any compactly supported Hamiltonian $H \in C^\infty_c(S^1 \times D)$ to an admissible Hamiltonian with small slope H^ε and define its Floer cohomology as $\mathrm{HF}^*(H) = \mathrm{HF}^*(H^\varepsilon)$. A key property of Floer cohomology on Liouville domains is that if an admissible Hamiltonian F has a slope close enough to zero, then we have an isomorphism $\Phi_F \colon \mathrm{H}^*(D) \to \mathrm{HF}^*(F)$. Thus, the Floer cohomology of compactly supported Hamiltonians on D is well-defined.

Let H be a compactly supported Hamiltonian. Following [Frauenfelder and Schlenk 2007], the spectral invariant associated to $(\alpha, H) \in \mathrm{H}^*(D) \times C^\infty_c(S^1 \times D)$ corresponds to the real number

$$c(\alpha, H) = \inf\{c \in \mathbb{R} \mid \Phi_H(\alpha) \in \operatorname{im} \iota^{<c}\},$$

where

$$\iota^{<c} = \iota^{c,+\infty}_{-\infty,-\infty} \colon \mathrm{HF}^*_{(-\infty,c)}(H) \longrightarrow \mathrm{HF}^*(H)$$

is the map induced by natural inclusion of subcomplexes.

Now, define the SH-capacity of D as

$$c_{\mathrm{SH}}(D) = \inf\{c > 0 \mid \iota^{\varepsilon,c}_{-\infty,-\infty} = 0\} \in (0,\infty],$$

where, for $\epsilon > 0$ sufficiently small,

$$\iota^{\varepsilon,c}_{-\infty,-\infty} \colon \mathrm{SH}^*_{(-\infty,\varepsilon)}(D) \longrightarrow \mathrm{SH}^*_{(-\infty,c)}(D).$$

It is known that $c_{\mathrm{SH}}(D)$ is finite if and only if $\mathrm{SH}^*(D)$ vanishes. Using this, Benedetti and Kang proved the following upper bound on spectral invariants of compactly supported Hamiltonians with respect to the unit.

[3] See Section 2.2.1 for details on the action convention we use in this paper.

Theorem 1 [Benedetti and Kang 2022] *Let $(D, \mathrm{d}\lambda)$ be a Liouville domain with $\mathrm{SH}^*(D) = 0$. Then,*

$$\sup\{c(1, H)\} \leq c_{\mathrm{SH}}(D) < +\infty,$$

where the supremum is taken over all compactly supported Hamiltonians in D.

In particular, by the definition of the spectral norm, if $\mathrm{SH}^*(D) = 0$, then for any compactly supported Hamiltonian H generating $\varphi_H \in \mathrm{Ham}_c(D)$, we have

$$\gamma(\varphi) = c(1, \varphi) + c(1, \varphi^{-1}) \leq 2c_{\mathrm{SH}}(D) < +\infty.$$

Therefore, Theorem 1 provides the *only if* part of Theorem A1.

On the other hand, symplectic cohomology is known to be nonzero in many cases [Seidel 2008, Section 5]. Since we will be using $\mathbb{Z}_2$ coefficients throughout this article, one case of particular interest to us is the following.

Proposition 2 [Viterbo 1999] *Suppose D contains a closed exact Lagrangian submanifold L. Then, $\mathrm{SH}^*(D) \neq 0$.*

This result of Viterbo can be used, in conjunction with Theorem A1, to prove that the spectral diameter is infinite for quite general classes of Liouville domains.

1.2.1 Cotangent bundles Monzner, Vichery and Zapolsky [Monzner et al. 2012] showed the following.

Theorem 3 *Let N be a closed manifold. There exists an isometric group embedding of $(\mathbb{R}, d_{\mathrm{st}})$ in $(\mathrm{Ham}_c(T^*N), d_\gamma)$.*

Note that Theorem 3 follows directly from Theorem A2 and Proposition 2. Indeed, since the zero section $N \subset DT^*N$ is an exact closed Lagrangian submanifold, Proposition 2 assures us that $\mathrm{SH}^*(DT^*N) \neq 0$. Therefore, Theorem A2 guarantees the existence of an isometric group embedding

$$(\mathbb{R}, d_{\mathrm{st}}) \longrightarrow (\mathrm{Ham}_c(T^*N), d_\gamma).$$

Theorem 3 immediately implies:

Corollary 4 *Let N be a closed manifold. Then $\mathrm{diam}_\gamma(DT^*N) = +\infty$.*

Similarly to Theorem 3, Corollary 4 follows directly from Proposition 2 and Theorem A1.

1.3 The spectral diameter of other symplectic manifolds

It has been known for a long time [Entov and Polterovich 2003] that for $(\mathbb{C}P^n, \omega\mathrm{FS})$,

$$\mathrm{diam}_\gamma(\mathbb{C}P^n) \leq \int_{\mathbb{C}P^1} \omega\mathrm{FS}.$$

The above upper bound was latter optimized in [Kislev and Shelukhin 2021, Theorem G] to

$$\mathrm{diam}_\gamma(\mathbb{C}P^n) = \frac{n}{n+1}\int_{\mathbb{C}P^1} \omega\mathrm{FS}.$$

However, for a surface Σ_g of genus $g \geq 1$, the spectral diameter is infinite. This case is covered by the following theorem [Kislev and Shelukhin 2021, Theorem D], which is a sharpening of [Usher 2013, Theorem 1.1].

Theorem 5 *Let (M,ω) be a closed symplectic manifold that admits an autonomous Hamiltonian $H \in C^\infty(M,\mathbb{R})$ such that*

(U1) *all the contractible periodic orbits of X_H are constant.*

Then $\mathrm{diam}_\gamma(M) = +\infty$.

Theorem 5 allows one to prove that the spectral diameter is infinite in many cases. A list of examples in which condition (U1) holds can be found in [Usher 2013, Section 1]. As mentioned above, surfaces of positive genus satisfy (U1). Also, if (N,ω_N) satisfies (U1) then so does $(M \times N, \omega_M \oplus \omega_N)$ for any other closed symplectic manifold (M,ω_M).

Kawamoto [2022b] proved that the spectral diameters of the quadrics Q^2 and Q^4 (of real dimensions 4 and 8 respectively), and certain stabilizations of them, are infinite.

1.3.1 Symplectically aspherical manifolds Recall that a symplectic manifold (M,ω_M) is symplectically aspherical if both ω_M and the first Chern class $c_1(M)$ of M vanish on $\pi_2(M)$; namely, for every continuous map $f\colon S^2 \to M$,

$$\langle[\omega_M], f_*[S^2]\rangle = 0 = \langle c_1(M), f_*[S^2]\rangle.$$

An open subset $U \subset M$ is said to be incompressible if the map $\pi_1(U) \to \pi_1(M)$ induced by the inclusion is injective.

As pointed out in [Buhovsky et al. 2021], it has been conjectured that $\mathrm{diam}_\gamma(M) = +\infty$ on all closed symplectically aspherical manifolds. Here, we prove that conjecture in the case of the twisted product $(M \times M, \omega \oplus -\omega)$ of a closed symplectically aspherical manifold (M,ω) with itself. But first, a more general result.

Proposition D *Let (M,ω) be a closed symplectically aspherical manifold of dimension $2n$. Suppose there exists an incompressible Liouville domain D of codimension 0 embedded inside M with $\mathrm{SH}^*(D) \neq 0$. Then*, $\mathrm{diam}_\gamma(M) = +\infty$.

Proof Let H be a compactly supported Hamiltonian in D and denote by $\iota\colon D \to M$ the embedding. By a cohomological analogue of [Ganor and Tanny 2023, Claim 5.2], we have that

$$c_D(\beta, H) = \max_{\substack{\alpha \in \mathrm{H}^*(M) \\ \iota^*(\alpha)=\beta}} c_M(\alpha, H)$$

for all $\beta \in \mathrm{H}^*(D)$, where c_D and c_M are the spectral invariants on D and M respectively. In particular, we know that the unit $1_M \in \mathrm{H}^*(M)$ is sent to the unit $1_D \in \mathrm{H}^*(D)$ under the map $\iota^*\colon \mathrm{H}^*(M) \to \mathrm{H}^*(D)$. Moreover, it is well known that the spectral invariant with respect to the unit can be implicitly written as

$$c_M(1_M, H) = \max_{\alpha \in \mathrm{H}^*(M)} c_M(\alpha, H)$$

(see Lemma 26). Therefore, fixing $\beta = 1_D$, we have

$$c_D(1_D, H) = c_M(1_M, H).$$

Using Theorem A1, the above equation thus yields the desired result. □

Corollary E *Let (M, ω) be a closed symplectically aspherical manifold. Then,*

$$\mathrm{diam}_\gamma(\mathrm{Ham}(M \times M, \omega \oplus -\omega)) = +\infty.$$

Proof Consider the closed Lagrangian given by the diagonal $L = \Delta$ inside $(M \times M, \omega_M \oplus -\omega_M)$. By virtue of the Weinstein neighborhood theorem, there exists an open neighborhood U of L and a symplectomorphism $\psi\colon U \to D_\varepsilon T^*L$ such that $\varphi(L)$ coincides with the zero section of an ϵ–radius codisk bundle $D_\varepsilon T^*L$ over L. The Liouville structure on $D_\varepsilon T^*L$ pulls back to a Liouville structure on U. Note that, inside $M \times M$, L is incompressible; ie the map $\pi_1(L) \to \pi_1(M \times M)$ of first homotopy groups induced by the inclusion $L \to M \times M$ is injective. Therefore, by homotopy equivalence, U and $D_\varepsilon T^*L$ are also incompressible. The desired result follows directly from Proposition D. □

1.4 Hofer geometry

As hinted at above, the finiteness of the spectral diameter plays a role in Hofer geometry. In particular, it can be used to study the following question posed in [Le Roux 2010]:

Question For any $A > 0$, let

$$E_A(M, \omega) := \{\varphi \in \mathrm{Ham}(M, \omega) \mid d_H(\mathrm{Id}, \varphi) > A\}$$

be the complement of the closed ball of radius A in Hofer's metric. For all $A > 0$, does $E_A(M, \omega)$ have nonempty C^0–interior?

Indeed, in the case of closed symplectically aspherical manifolds with infinite spectral diameter, a positive answer to the Question above was given by Buhovsky, Humilière and Seyfaddini (see also [Kawamoto 2022a; 2022b] for the positive and negative monotone cases).

Theorem 6 [Buhovsky et al. 2021] *Let (M,ω) be a closed, connected and symplectically aspherical manifold. If $\mathrm{diam}_\gamma(M) = +\infty$, then $E_A(M,\omega)$ has nonempty C^0–interior for all $A > 0$.*

Using Theorem 6 in conjunction with Corollary E, we directly obtain the following answer to the Question above in the specific setting of Corollary E.

Corollary F *Let (M,ω) be a closed symplectically aspherical manifold. Then, $E_A(M\times M,\omega\oplus -\omega)$ has a nonempty C^0–interior for all $A > 0$.*

Acknowledgements

This research is a part of my PhD thesis at the Université de Montréal under the supervision of Egor Shelukhin. I thank him for proposing this project and outlining the approach used to carry it out in this paper. I am deeply indebted to him for the countless valuable discussions we had regarding spectral invariants and symplectic cohomology. I would also like to thank Octav Cornea and François Lalonde for their comments on an early draft of this project. I thank Leonid Polterovich, Felix Schlenk and Shira Tanny for their comments which helped improve the exposition. Finally, I am grateful to Marcelo Atallah, Filip Brocic, François Charette, Jean-Philippe Chassé, Dustin Connery-Grigg, Jonathan Godin, Jordan Payette and Dominique Rathel-Fournier for fruitful conversations. This research was partially supported by Fondation Courtois.

2 Liouville domains and admissible Hamiltonians

In this subsection we recall the definition of Liouville domains, specify the class of Hamiltonians we will restrict our attention to and describe how their Floer trajectories behave at infinity.

2.1 Completion of Liouville domains

A Liouville domain $(D, \mathrm{d}\lambda, Y)$ is an exact symplectic manifold with boundary on which the vector field Y, defined by $Y \lrcorner\, \mathrm{d}\lambda = \lambda$ and called the Liouville vector field, points outwards along ∂D. Denote by $\widehat{D} = D \cup [1,\infty)\times\partial D$ the completion of D and (r,x) the coordinates on $[1,\infty)\times\partial D$. Here, we glue ∂D and $\{1\}\times\partial D$ with respect to the reparametrization $\psi_Y^{\ln r}$ of the Liouville flow generated by Y. Given $\delta > 0$, let

$$D^\delta = \psi_Y^{\ln\delta}(D) = \widehat{D}\setminus(\delta,\infty)\times\partial D.$$

We extend the Liouville form λ to $\widehat{D}$ by defining $\hat{\lambda}\colon T\widehat{D}\to\mathbb{R}$ as

$$\hat{\lambda}\,|_D = \lambda \quad\text{and}\quad \hat{\lambda}\,|_{\widehat{D}\setminus D} = r\alpha,$$

where $\alpha = \lambda|_{\partial D}$. The cylindrical portion $[1,\infty)\times\partial D$ of $\widehat{D}$ is thus equipped with the symplectic form $\omega = \mathrm{d}(r\alpha)$. See Figure 1.

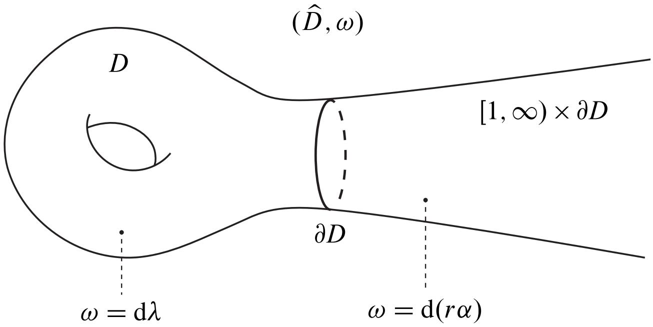


Figure 1: A Liouville domain with its completion.

The skeleton $\mathrm{Sk}(D)$ of $(D, \mathrm{d}\lambda, Y)$ is defined by

$$\mathrm{Sk}(D) = \bigcap_{0<r<1} \psi_Y^{\ln r}(D).$$

Denote by R_α the Reeb vector field on ∂D associated to α, meaning

$$R_\alpha \lrcorner \mathrm{d}\alpha = 0, \quad \alpha(R) = 1.$$

We define $\mathrm{Spec}(\partial D, \alpha)$ to be the set of periods of closed characteristics, the periodic orbits generated by R_α, on ∂D and put

$$T_0 = \min \mathrm{Spec}(\partial D, \lambda).$$

As a subset of $\mathbb{R}$, $\mathrm{Spec}(\partial D, \alpha)$ is known to be closed and nowhere dense. For any $A \in \mathbb{R}$, let η_A denote the distance between A and $\mathrm{Spec}(\partial D, \lambda)$.

2.2 Admissible Hamiltonians and almost-complex structures

2.2.1 Periodic orbits and action functional Given a Hamiltonian $H: S^1 \times \widehat{D} \to \mathbb{R}$, one defines its time-dependent Hamiltonian vector field $X_H^t: \widehat{D} \to T\widehat{D}$ by

$$X_H^t \lrcorner \omega = -dH_t,$$

where $H_t(p) = H(t, p)$. We denote by $\varphi_H^t: \widehat{D} \to \widehat{D}$ the flow generated by X_H^t. The set of all contractible 1–periodic orbits of φ_H^t is denoted by $\mathcal{P}(H)$. An orbit $x \in \mathcal{P}(H)$ is said to be *nondegenerate* if

$$\det(\mathrm{id} - d_{x(0)}\varphi_H^1) \neq 0$$

and *transversally nondegenerate* if the eigenspace associated to the eigenvalue 1 of the map $d_{x(0)}\varphi^1$ is of dimension 1.

Let $\mathcal{L}\widehat{D}$ be the space of contractible loops in $\widehat{D}$. For a Hamiltonian $H: S^1 \times \widehat{D} \to \mathbb{R}$, the *Hamiltonian action functional* $\mathcal{A}_H: \mathcal{L}\widehat{D} \to \mathbb{R}$ associated to H is defined as

$$\mathcal{A}_H(x) = \int_0^1 x^*\hat{\lambda} - \int_0^1 H_t(x(t))\,\mathrm{d}t.$$

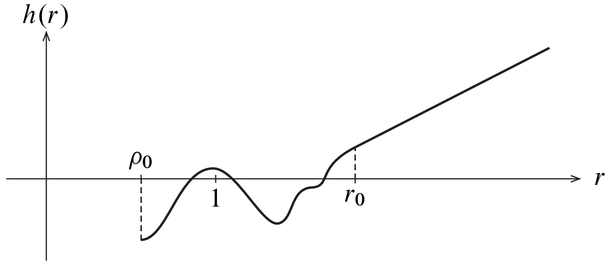


Figure 2: An r_0–admissible Hamiltonian.

It is well known that the elements of $\mathcal{P}(H)$ correspond to the critical points of $\mathcal{A}_H$; see [Audin and Damian 2014, Section 6]. The image of $\mathcal{P}(H)$ under the Hamiltonian action functional is called the *action spectrum of* H and is denoted by $\operatorname{Spec}(H)$. For an open set $U \subset \widehat{D}$ we define

$$\mathcal{P}_U(H) = \{x \in \mathcal{P}(H) \mid \operatorname{im} x \subset U\}.$$

2.2.2 Admissible Hamiltonians The completion of a Liouville domain is obviously noncompact. We thus need to control the behavior at infinity of Hamiltonians we use in order for them to have finitely many 1–periodic contractible orbits.

Definition 7 Let $r_0 > 1$. A Hamiltonian H is r_0–admissible if there exists $\rho_0 \in (0, r_0)$ such that

- $H(t, x, r) = h(r)$ on $\widehat{D} \setminus D^{\rho_0}$,
- $h(r) = \tau_H r + \eta_H$ on $(r_0, +\infty)$ for $\tau_H \in (0, \infty) \setminus \operatorname{Spec}(\partial D, \alpha)$,
- H is regular: every element of $\mathcal{P}_{D^\rho}(H)$ is nondegenerate and every element of $\mathcal{P}_{\widehat{D}\setminus D^\rho}(H)$ is transversally nondegenerate.

We denote the set of such Hamiltonians by $\mathcal{H}_{r_0}$. See Figure 2.

We will also consider the set $\mathcal{H}^0_{r_0} \subset \mathcal{H}_{r_0}$ of r_0–admissible Hamiltonians which are negative on D. In some cases, it is not necessary to specify r_0 as long as it is greater than 1. For that purpose, we define

$$\mathcal{H} = \bigcup_{r_0>1} \mathcal{H}_{r_0}, \quad \mathcal{H}^0 = \bigcup_{r_0>1} \mathcal{H}^0_{r_0}.$$

Remark 8 Suppose $H \in \mathcal{H}$. If $x \in \mathcal{P}_{\widehat{D}\setminus D^{\rho_0}}(H)$ is nonconstant, then it is necessarily transversally nondegenerate. Indeed, since H is time-independent there by definition, for any $c \in \mathbb{R}$, we know $x(t-c)$ is also a 1–periodic orbit of H.

Lemma 9 *If $H \in \mathcal{H}$, then $|\mathcal{P}_{D^{\rho_0}}(H)|$ is finite and $\mathcal{P}_{\widehat{D}\setminus D^{\rho_0}}(H)$ consists of a finite number of periodic orbits and S^1 families of periodic orbits.*

Proof Since $\overline{D^{\rho_0}}$ is compact and elements of $\mathcal{P}_{D^{\rho_0}}(H)$ are nondegenerate, there is a finite number of 1–periodic orbits of H inside it.

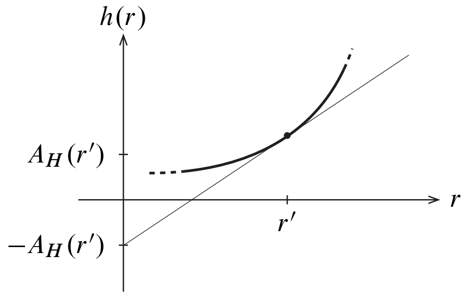


Figure 3: Action value of a periodic orbit contained in $\{r'\} \times \partial D$.

Next, we look at the elements of $\mathcal{P}_{\widehat{D} \setminus D^{\rho_0}}(H)$. On $\widehat{D} \setminus D^{\rho_0}$, we know that $H = h(r)$ and $\omega = \mathrm{d}\hat{\lambda}$. Therefore, on $\widehat{D} \setminus D^{\rho_0}$

$$\begin{aligned} X_H \lrcorner\, \omega &= X_H \lrcorner\, (\mathrm{d}r \wedge \alpha + r\,\mathrm{d}\alpha) \\ &= \mathrm{d}r(X_H)\alpha - \alpha(X_H)\mathrm{d}r + r X_H \lrcorner\, \mathrm{d}\alpha \end{aligned}$$

and $\mathrm{d}H = h'(r)\mathrm{d}r$. Hamilton's equation thus yields

$$\mathrm{d}r(X_H) = 0 = X_H \lrcorner\, \mathrm{d}\alpha, \quad \alpha(X_H) = h'(r).$$

The three equations above imply the following two facts:

- On $\widehat{D} \setminus D^{\rho_0}$, $X_H = h'(r) R_\alpha$.
- If $x \in \mathcal{P}(H)$ is such that $x \cap \widehat{D} \setminus D^{\rho_0} \neq \varnothing$, then $\operatorname{im} x \subset \{r\} \times \partial D$ for some $r > \rho_0$.

We conclude that a 1–periodic orbit x of H which lies inside $\{r\} \times \partial D$ corresponds to a Reeb orbit of period $h'(r)$. Notice that since $\tau_H \notin (0, +\infty) \cap \operatorname{Spec}(\partial D, \alpha)$, we have $\mathcal{P}_{\widehat{D} \setminus D^{\rho_0}}(H) = \mathcal{P}_{D^{r_0} \setminus D^{\rho_0}}(H)$. Therefore, since $\overline{D^{r_0} \setminus D^{\rho_0}}$ is compact and every element of $\mathcal{P}_{D^{r_0} \setminus D^{\rho_0}}(H)$ is transversally nondegenerate by definition, $\mathcal{P}_{D^{r_0} \setminus D^{\rho_0}}(H)$ is finite. □

Remark 10 The fact that admissible Hamiltonians are radial on the cylindrical part of $\widehat{D}$ allows us to express the action of the 1–periodic orbits inside $\widehat{D} \setminus D$ in terms of that radial function. To see this, we fix $H \in \mathcal{H}$ and compute the action of a nonconstant orbit $x \in \mathcal{P}(H) \cap (\widehat{D} \setminus D)$ which we suppose lies inside $\{r\} \times \partial D$ for $r > 1$:

$$\mathcal{A}_H(x) = \int_0^1 x^* \hat{\lambda} - \int_0^1 H \circ x \,\mathrm{d}t = \int_0^1 r\alpha(X_H)\,\mathrm{d}t - \int_0^1 h(r)\,\mathrm{d}t = r h'(r) - h(r).$$

The function $A_H(r) = r h'(r) - h(r)$ on the right-hand side of the above equation has a nice geometric interpretation. Looking at the graph of h, we notice that $A_H(r')$ corresponds to minus the y–coordinate of the intersection of the tangent at the point $(r', h(r'))$ and the y–axis. See Figure 3.

2.2.3 Monotone homotopies

We will need to also restrict the types of Hamiltonian homotopies we consider to the following class.

Definition 11 Let $H_s = \{H_s\}_{s\in\mathbb{R}}$ be a smooth homotopy from $H_+ \in \mathcal{H}_{r_0}$ to $H_- \in \mathcal{H}_{r_0'}$ We say that H_s is a *monotone homotopy* if the following conditions hold:

- There exists $S > 0$ such that $H_{s'} = H_-$ for $s' < -S$ and $H_{s'} = H_+$ for $s' > S$.
- $H_s = h_s(r)$ on $\widehat{D} \setminus D^\rho$ for $\rho = \max\{\rho_0, \rho_0'\}$.
- For $R = \max\{r_0, r_0'\}$, we have $h_s(r) = \tau_s r + \eta_s$ on $(R, +\infty)$ for smooth functions τ_s, η_s of s.
- $\partial_s H_s(t, p) \leq 0$ for $(t, p, s) \in S^1 \times \widehat{D} \times \mathbb{R}$.

For $H_+ \in \mathcal{H}_{r_0}$ and $H_- \in \mathcal{H}_{r_0'}$ such that $H_+ \leq H_-$ pointwise everywhere on $\widehat{D}$, we can explicitly construct a monotone homotopy in the following way. Fix a positive constant $S > 0$. Let $\beta\colon \mathbb{R} \to [0, 1]$ be a smooth function such that $\beta(s) = 0$ for $s \leq -S$, $\beta(s) = 1$ for $s \geq S$ and $\beta'(s) > 0$ for all $s \in (-S, S)$. Define

$$H_s = H_- + \beta(s)(H_+ - H_-).$$

Notice that, since $\beta'(s) \geq 0$ and $H_+ \leq H_-$, we have

$$\partial_s H_s = \beta'(s)(H_+ - H_-) \leq 0.$$

For $R = \max\{r_0, r_0'\}$ we have, on $\widehat{D} \setminus D^R$,

$$H_s(t, r, p) = (\beta(s)(\tau_+ - \tau_-) + \tau_-)r + \beta(s)(\eta_+ - \eta_-) + \eta_- = h_s(r)$$

as desired and

$$\partial_s \partial_r h_s(r) = \beta(s)'(\tau_+ - \tau_-) \leq 0. \tag{1}$$

This inequality will be needed for the maximum principle of Section 2.3.2. Equation (1) also holds for general monotone homotopies. Indeed, by definition, $H_s(r)$ decreases in s and $H_s(r)$ is affine for $r \geq R$.

2.2.4 Admissible almost-complex structures Let J be an almost-complex structure on $\widehat{D}$. Recall that J is ω–compatible if the map $g_J\colon TM \otimes TM \to \mathbb{R}$ defined by

$$g_J(v, w) = \omega(v, Jw)$$

is a Riemannian metric. To control the behavior of ω–compatible almost-complex structures at infinity, we make the following definition.

Definition 12 Let J be an ω–compatible almost-complex structure on $\widehat{D}$. We say that J is *admissible* if $J_1 = J|_{\widehat{D}\setminus D}$ is of *contact type*. Namely, we ask that

$$J_1^* \hat{\lambda} = \mathrm{d}r.$$

We denote the set of such almost-complex structures by $\mathcal{J}$. A pair (H, J), where $H \in \mathcal{H}_{r_0}$ and $J \in \mathcal{J}$, is called an r_0–*admissible pair*.

2.3 Floer trajectories and maximum principle

In this subsection, we recall some analytical aspects of Floer theory on Liouville domains. Issues regarding transversality will be dealt with in the next section.

2.3.1 Floer trajectories Consider a Hamiltonian $H: S^1 \times \widehat{D} \to \mathbb{R}$ and two 1–periodic orbits $x_\pm \in \mathcal{P}(H)$. Let J be an ω–compatible almost-complex structure on $\widehat{D}$. A *Floer trajectory* between x_- and x_+ is a solution $u: \mathbb{R} \times S^1 \to \widehat{D}$ to the *Floer equation*

$$\partial_s u + J(\partial_t u - X_H) = 0$$

that converges uniformly in t to x_- and x_+ as $s \to \pm\infty$:

$$\lim_{s \to \pm\infty} u(s,t) = x_\pm(t).$$

We denote the moduli space of such trajectories by $\mathcal{M}'(x_-, x_+; H)$. We may reparametrize a solution $u \in \mathcal{M}'(x_-, x_+; H)$ in the $\mathbb{R}$–coordinate by adding a constant. Thus, Floer trajectories occur in $\mathbb{R}$–families. The space of unparametrized solutions is denoted by $\mathcal{M}(x_-, x_+; H) = \mathcal{M}'(x_-, x_+; H)/\mathbb{R}$. When the context is clear, we will drop H from the notation and simply write $\mathcal{M}(x_-, x_+)$.

If we replace H with a monotone homotopy $H_\bullet = \{H_s\}_{s \in \mathbb{R}}$, then we can instead consider solutions $u: \mathbb{R} \times S^1 \to \widehat{D}$ to the *s-dependent Floer equation*

$$\partial_s u + J(\partial_t u - X_{H_s}) = 0$$

that converge uniformly in t to $x_\pm \in \mathcal{P}(H_\pm)$ as $s \to \pm\infty$. The moduli space of such trajectories is denoted by $\mathcal{M}(x_-, x_+; H_\bullet)$. Unlike the s–independent case, $\mathcal{M}(x_-, x_+; H_\bullet)$ does not admit a free $\mathbb{R}$–action by which we can quotient.

2.3.2 Maximum principle To define Floer cohomology of $\widehat{D}$, we need to control the behavior of the Floer trajectories. In particular, we have to make sure they do not escape to infinity. Admissible Hamiltonians and admissible complex structures allow us to achieve that requirement. The first result in that direction is the maximum principle for Floer trajectories. In what follows we say that v is a local Floer solution of (H, J) in $\widehat{D} \setminus D$ if

$$v = u|_{u^{-1}(\operatorname{im} u \cap \widehat{D} \setminus D)} : u^{-1}(\operatorname{im} u \cap \widehat{D} \setminus D) \longrightarrow \widehat{D} \setminus D$$

for some $u \in \mathcal{M}(x_-, x_+; H)$.

Lemma 13 (generalized maximum principle [Viterbo 1999]) *Let (H, J) be an r_0–admissible pair on $\widehat{D}$. Suppose v is a local Floer solution of (H, J) in $\widehat{D} \setminus D^{r_0}$. Then, the r–coordinate $r \circ v$ of v does not admit an interior maximum unless $r \circ v$ is constant.*

Remark 14 The generalized maximum principle still holds if we replace $H \in \mathcal{H}$ by a monotone homotopy H_s between $H_+ \in \mathcal{H}_{r_0}$ and $H_- \in \mathcal{H}_{r_0'}$ and if v is a local solution of the s–dependent Floer equation

$$\partial_s v + J(\partial_t v - X_{H_s}) = 0$$

inside $\widehat{D} \setminus D^R$, where $r = \max\{r_0, r_0'\}$. Here it is crucial that $\partial_s \partial_r h_s(r) \leq 0$ for large enough r. From the maximum principle above, we immediately obtain the following corollary which guarantees that Floer trajectories do not escape to infinity.

Corollary 15 *Let (H, J) be an r_0–admissible pair on $\widehat{D}$ and let $x_\pm \in \mathcal{P}(H)$. If $u \in \mathcal{M}(x_-, x_+)$, then*

$$\operatorname{im} u \subset D^R \quad \textit{for } R = \max\{r \circ x_-, r \circ x_+, r_0\}.$$

If H_s is a monotone homotopy between $H_- \in \mathcal{H}_{r_0}$ and $H_+ \in \mathcal{H}_{r_0'}$ and u is a solution to the s–dependent Floer equation between $x_- \in \mathcal{P}(H_-)$ and $x_+ \in \mathcal{P}(H_+)$, then

$$\operatorname{im} u \subset D^R, \quad \textit{for } R = \max\{r \circ x_-, r \circ x_+, r_0, r_0'\}.$$

2.3.3 Energy An important quantity which is associated to a Floer trajectory is its *energy*. It is defined as

$$E(u) = \frac{1}{2}\int_{\mathbb{R}\times S^1} (|\partial_s u|_J^2 + |\partial_t u - X_H|_J^2)\,\mathrm{d}s \wedge \mathrm{d}t,$$

where $|\cdot|_J$ is the norm corresponding to g_J. Using the Floer equation, we can write

$$|\partial_t u - X_H|_J^2 = \omega(J\partial_s u, -\partial_s u) = \omega(\partial_s u, J\partial_s u) = |\partial_s u|_J^2.$$

Thus, the energy can be written more compactly as

$$E(u) = \int_{\mathbb{R}\times S^1} |\partial_s u|_J^2 \,\mathrm{d}s \wedge \mathrm{d}t.$$

It is often useful to estimate the difference in Hamiltonian action of the ends of a Floer trajectory in terms of the energy of that trajectory. This can be achieved using the maximum principle and Stokes' theorem.

Lemma 16 *Let (H, J) be an r_0–admissible pair and let $u \in \mathcal{M}'(x_-, x_+; H)$ for $x_\pm \in \mathcal{P}(H)$. Then,*

$$0 \leq E(u) = \mathcal{A}_H(x_+) - \mathcal{A}_H(x_-).$$

If H_s is a monotone homotopy between $H_+ \in \mathcal{H}_{r_0}$ and $H_- \in \mathcal{H}_{r_0'}$ that is constant in the s–coordinate for $s > |S|$ then

$$0 \leq E(u) \leq \mathcal{A}_{H_+}(x_+) - \mathcal{A}_{H_-}(x_-) + \sup_{\substack{s\in[-S,S],\\ t\in S^1, p\in D^R}} \partial_s H_s(t, p),$$

where $R = \max\{r \circ x_-, r \circ x_+, r_0, r_0'\}$.

3 Filtered Floer and symplectic cohomology

We present in this section a brief overview of Floer cohomology for completions of Liouville domains and their symplectic cohomology. For more details we refer the reader to [Cieliebak et al. 1995; 1996; 2010; Viterbo 1999; Weber 2006; Ritter 2013].

3.1 Filtered Floer cohomology

3.1.1 The Floer cochain complex Let (H, J) be an admissible pair. As mentioned in Remark 8, the 1–periodic orbits of H on $\widehat{D} \setminus D^{\rho_0}$ come in a finite number of S^1–families, which we denote by $\hat{x}_i$. To break each $\hat{x}_i$ in a finite number of isolated periodic orbits, we first choose an open neighborhood U_i of each $\hat{x}_i$ such that $U_i \cap U_j = \varnothing$ for $i \neq j$. Then, we define on each $\hat{x}_i$ a Morse function f_i having exactly two critical points: one of index 0 and another of index 1. We extend each f_i to its corresponding U_i. When added to H, these perturbations, which can be chosen as small as we want, break each of the S^1–families into two critical points. By virtue of the action formula derived in Remark 10, the actions of the new critical points are as close as we want to the action of their original S^1–family. We denote by H_1 the Hamiltonian resulting from this procedure. By abuse of notation we will write $\mathcal{P}(H)$ for the set of 1–periodic orbits of H_1.

We define the *Floer cochain group of H* as the $\mathbb{Z}_2$–vector space[4]

$$\mathrm{CF}^*(H) = \bigoplus_{x \in \mathcal{P}(H)} \mathbb{Z}_2\langle x \rangle.$$

As the notation above suggests, $\mathrm{CF}^*(H)$ is in fact a *graded* $\mathbb{Z}_2$–vector space. Assuming that the first Chern class $c_1(\omega) \in \mathrm{H}^2(\widehat{D}; \mathbb{Z})$ of $(T\widehat{D}, J)$ vanishes on $\pi_2(\widehat{D})$, the Conley-Zehnder index $\mathrm{CZ}(x) \in \mathbb{Z}$ of a 1–periodic orbit $x \in \mathcal{P}(H)$ is well-defined [Salamon and Zehnder 1992]. We can therefore equip $\mathrm{CF}^*(H)$ with the degree

$$|x| = \tfrac{1}{2} \dim \widehat{D} - \mathrm{CZ}(x)$$

and define

$$\mathrm{CF}^k(H) = \bigoplus_{\substack{x \in \mathcal{P}(H) \\ |x| = k}} \mathbb{Z}_2\langle x \rangle.$$

Here, CZ is normalized such that for a C^2–small time-independent admissible Hamiltonian F,

$$\mathrm{CZ}(x) = \tfrac{1}{2} \dim \widehat{D} - \mathrm{ind}(x),$$

where $\mathrm{ind}(x)$ corresponds to the Morse index of $x \in \mathrm{Crit}(F) = \mathcal{P}(F)$. In particular, if x is a local minimum of F, then $|x| = 0$. This convention therefore ensures that the cohomological unit has degree 0.

[4] We use $\mathbb{Z}_2$ coefficients here for simplicity but the cohomological construction that follows can be carried out with any coefficient ring.

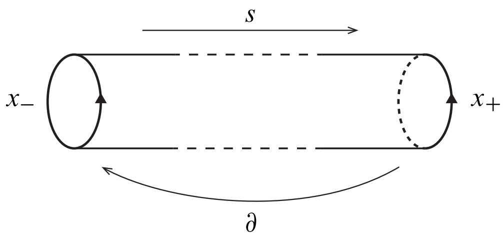


Figure 4: The differential in Floer cohomology goes from right to left.

For a generic perturbation of J, the space $\mathcal{M}(x_-, x_+; H)$ is a smooth manifold of dimension

$$\dim \mathcal{M}(x_-, x_+; H) = \mathrm{CZ}(x_+) - \mathrm{CZ}(x_-) - 1.$$

In the case where $|x_-| = |x_+| + 1$, Corollary 15 and Lemma 16 allow us to use the standard compactness arguments, as in [Audin and Damian 2014, Chapter 8], to show that $\mathcal{M}(x_-, x_+; H)$ is a compact manifold of dimension 0. Knowing that, we define the coboundary operator $\partial\colon \mathrm{CF}^k(H) \to \mathrm{CF}^{k+1}(H)$ by

$$\partial x_+ = \sum_{|x_-|=k+1} \#_2 \mathcal{M}(x_-, x_+; H) x_-,$$

where $\#_2 \mathcal{M}(x_-, x_+; H)$ is the count modulo 2 of components in $\mathcal{M}(x_-, x_+, H)$. See Figure 4.

Using once again Corollary 15, $\partial \circ \partial = 0$ holds by standard arguments which appear in [Audin and Damian 2014, Chapter 9]. The pair $(\mathrm{CF}^*(H), \partial)$ is thus a graded cochain complex that we call the *Floer cochain complex of H*.

3.1.2 Filtered Floer cochain complex The Hamiltonian action functional induces a filtration on the Floer cochain complex. For $a \in (\mathbb{R} \cup \{\pm\infty\}) \setminus \mathrm{Spec}(H)$, we define

$$\mathrm{CF}^k_{<a}(H) = \bigoplus_{\substack{x \in \mathcal{P}(H) \\ |x|=k,\ \mathcal{A}_H(x)<a}} \mathbb{Z}_2\langle x\rangle.$$

By definition, we have $\mathrm{CF}^*(H) = \mathrm{CF}^*_{<+\infty}(H)$. Lemma 16 assures that ∂ decreases the action. Thus, the restriction $\partial_{<a}\colon \mathrm{CF}^k_{<a}(H) \to \mathrm{CF}^{k+1}_{<a}(H)$ of the coboundary operator is well-defined and $(\mathrm{CF}^*_{<a}(H), \partial_{<a})$ is a subcomplex of $(\mathrm{CF}^*(H), \partial)$. Now, for $a, b \in (\mathbb{R} \cup \{\pm\infty\}) \setminus \mathrm{Spec}(H)$ such that $a < b$, we can define the Floer cochain complex in the action window (a, b) as the quotient

$$\mathrm{CF}^*_{(a,b)}(H) = \frac{\mathrm{CF}^*_{<b}(H)}{\mathrm{CF}^*_{<a}(H)},$$

on which we denote the projection of the coboundary operator by

$$\partial_{(a,b)}\colon \mathrm{CF}^k_{(a,b)}(H) \longrightarrow \mathrm{CF}^{k+1}_{(a,b)}(H).$$

Therefore, for $a, b, c \in (\mathbb{R} \cup \{\pm\infty\}) \setminus \mathrm{Spec}(H)$ such that $a < b < c$, we have an inclusion and a projection

$$\iota^{b,c}_{a,a} \colon \mathrm{CF}^*_{(a,b)}(H) \longrightarrow \mathrm{CF}^*_{(a,c)}(H), \quad \pi^{c,c}_{a,b} \colon \mathrm{CF}^*_{(a,c)}(H) \longrightarrow \mathrm{CF}^*_{(b,c)}(H)$$

that produce the short exact sequence

$$0 \longrightarrow \mathrm{CF}^*_{(a,b)}(H) \xrightarrow{\iota^{b,c}_{a,a}} \mathrm{CF}^*_{(a,c)}(H) \xrightarrow{\pi^{c,c}_{a,b}} \mathrm{CF}^*_{(b,c)}(H) \longrightarrow 0.$$

For simplicity, we define $\iota^{<c} = \iota^{+\infty,c}_{-\infty,-\infty}$ and $\pi_{>b} = \pi^{+\infty,+\infty}_{-\infty,b}$.

3.1.3 Filtered Floer cohomology Let $a, b \in (\mathbb{R} \cup \{\pm\infty\}) \setminus \mathrm{Spec}(H)$ such that $a < b$. The above filtered cochain complexes allow us to define the *Floer cohomology group of H* in the action window (a, b) *as*

$$\mathrm{HF}^*_{(a,b)}(H) = \frac{\ker \partial_{(a,b)}}{\mathrm{im}\, \partial_{(a,b)}}.$$

The full Floer cohomology group of H is defined as $\mathrm{HF}^*(H) = \mathrm{HF}^*_{(-\infty,+\infty)}(H)$. For $a, b, c \in (\mathbb{R} \cup \{\pm\infty\}) \setminus \mathrm{Spec}(H)$ such that $a < b < c$, the short exact sequence on the cochain level induces a long exact sequence in cohomology:

(2)
$$\begin{array}{ccc} \mathrm{HF}^*_{(a,b)}(H) & & \\ \uparrow {\scriptstyle [+1]} & \searrow {\scriptstyle [\iota^{b,c}_{a,a}]} & \\ & & \mathrm{HF}^*_{(a,c)}(H) \\ & \swarrow {\scriptstyle [\pi^{c,c}_{a,b}]} & \\ \mathrm{HF}^*_{(b,c)}(H) & & \end{array}$$

For C^2–small admissible Hamiltonians with small slope at infinity, the Floer cohomology recovers the standard cohomology of D.

Lemma 17 [Ritter 2013, Section 15.2] *Let $H \in \mathcal{H}$ be a C^2–small Hamiltonian with $\tau_H < T_0$ for* $T_0 = \min \mathrm{Spec}(\partial D, \lambda)$. *Then, we have an isomorphism*

$$\Phi_H \colon \mathrm{H}^*(D) \longrightarrow \mathrm{HF}^*(H).$$

Remark 18 We can endow $\mathrm{HF}^*(H)$ with a ring structure [Ritter 2013] where the product is given by the pair of pants product. The unit in $\mathrm{HF}^*(H)$, which we denote by 1_H, coincides with $\Phi_H(e_D)$, where e_D is the unit in $\mathrm{H}^*(D)$.

3.1.4 Compactly supported Hamiltonians We can define the Floer cohomology of compactly supported Hamiltonians on Liouville domains by first extending to affine functions on the cylindrical portion of $\widehat{D}$.

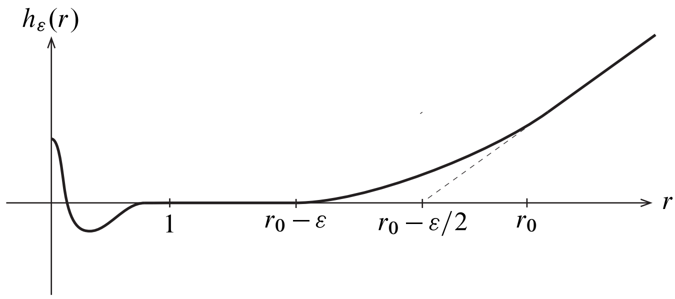


Figure 5: The τ–extension of a compactly supported Hamiltonian.

Definition 19 Denote by $\mathcal{C}(D)$ the set of Hamiltonians with support in $S^1 \times (D \setminus \partial D)$. Let $H \in \mathcal{C}(D)$. For $\tau \in (0,\infty) \setminus \mathrm{Spec}(\partial D, \lambda)$, we define the τ–extension $H^\tau \in \mathcal{H}_1$ of H as follows. Fix $0 < \varepsilon < 1$ and $r_0 > 1$ so that $1 < r_0 - \varepsilon$,

- $H^\tau = H$ on D and $H^\tau = 0$ on $D^{r_0-\varepsilon} \setminus D$,
- $H^\tau = h_\varepsilon(r)$ on $\widehat{D} \setminus D^{r_0-\varepsilon}$,
- $h_\varepsilon(r)$ is convex for $r \in [r_0-\varepsilon, r_0]$, with $h_\varepsilon^{(k)}(1) = 0$ for all $k \geq 0$, $h'_\varepsilon(1+\varepsilon) = \tau$ and $h_\varepsilon^{(\ell)}(1+\varepsilon) = 0$ for all $\ell > 1$,
- $h_\varepsilon(r) = \tau(r - (r_0 - \varepsilon/2))$ for $r \in [r_0, +\infty)$.

We perturb H^τ so that it is r_0–admissible. The Floer cohomology of H is defined as

$$\mathrm{HF}^*_{(a,b)}(H) = \mathrm{HF}^*_{(a,b)}(H^\tau),$$

where $0 < \tau < T_0$. See Figure 5.

Since we take a slope τ smaller than the minimum Reeb period to define $\mathrm{HF}^*_{(a,b)}(H)$, the above definition doesn't depend on the choice of τ, ε and r_0, as we will see in Lemma 20 below.

3.1.5 Continuation maps Let $K \in \mathcal{H}_{r_0}$ and $F \in \mathcal{H}_{r'_0}$ such that $F \leq K$. Consider a monotone homotopy $H_\bullet$ from F to K. Then from Corollary 15 and Lemma 16 in the case of homotopies, we can apply the techniques shown in [Audin and Damian 2014, Chapter 11] to show that, for $x_- \in \mathcal{P}(K)$ and $x_+ \in \mathcal{P}(F)$ with $|x_-| = |x_+|$, $\mathcal{M}(x_-, x_+; H_\bullet)$ is a smooth compact manifold of dimension 0. The *continuation map* $\Phi^{H_\bullet}\colon \mathrm{CF}^k(F) \to \mathrm{CF}^k(K)$ *induced by* H_s on the cochain level is defined as

$$\Phi^{H_\bullet}(x_+) = \sum_{|x_-|=k} \#_2\mathcal{M}(x_-, x_+; H_\bullet)x_-,$$

where $\#_2\mathcal{M}(x_-, x_+; H_\bullet)$ is the count modulo 2 of components in $\mathcal{M}(x_-, x_+; H_\bullet)$. The map

$$[\Phi^{H_\bullet}]\colon \mathrm{HF}^*(F) \longrightarrow \mathrm{HF}^*(K)$$

is independent of the chosen monotone homotopy and we can denote it by $[\Phi^{K,F}]$. Consider the monotone homotopy

$$H_s = K + \beta(s)(F - K)$$

described in Section 2.2. We note that $\partial_s H_s \le 0$ since $F \le K$ and $\beta' \ge 0$. Thus the action estimate given by Lemma 16 for homotopies yields

$$\mathcal{A}_K(x_-) \le \mathcal{A}_H(x_+) + \sup_{\substack{s\in[-S,S],\\ t\in S^1, p\in D^R}} \partial_s H_s(t,p) \le \mathcal{A}_H(x_+)$$

for $x_- \in \mathcal{P}(K)$ and $x_+ \in \mathcal{P}(F)$. Therefore, the continuation map decreases the action and hence induces maps

$$[\Phi^{K,F}_{(a,b)}]\colon \mathrm{HF}^*_{(a,b)}(F) \longrightarrow \mathrm{HF}^*_{(a,b)}(K)$$

that commute with the inclusion and restriction maps as follows [Ritter 2013, Section 8]:

(3)
$$\begin{CD}
\cdots @>>> \mathrm{HF}^*_{(a,b)}(F) @>[\iota^{b,c}_{a,a}]>> \mathrm{HF}^*_{(a,c)}(F) @>[\pi^{c,c}_{a,b}]>> \mathrm{HF}^*_{(b,c)}(F) @>>> \cdots \\
@. @V[\Phi^{K,F}_{(a,b)}]VV @V[\Phi^{K,F}_{(a,c)}]VV @V[\Phi^{K,F}_{(b,c)}]VV \\
\cdots @>>> \mathrm{HF}^*_{(a,b)}(K) @>>[\iota^{b,c}_{a,a}]> \mathrm{HF}^*_{(a,c)}(K) @>>[\pi^{c,c}_{a,b}]> \mathrm{HF}^*_{(b,c)}(K) @>>> \cdots
\end{CD}$$

Suppose we are given another Hamiltonian $H \ge K$. Then we have the commutative diagram

$$\mathrm{HF}^*_{(a,b)}(F) \xrightarrow{[\Phi^{K,F}_{(a,b)}]} \mathrm{HF}^*_{(a,b)}(K) \xrightarrow{[\Phi^{H,K}_{(a,b)}]} \mathrm{HF}^*_{(a,b)}(H), \qquad [\Phi^{H,F}_{(a,b)}]\colon \mathrm{HF}^*_{(a,b)}(F) \longrightarrow \mathrm{HF}^*_{(a,b)}(H).$$

As opposed to the closed case, for completion of Liouville domains, continuation maps do not necessarily yield isomorphisms. One case in which they do is when both Hamiltonians have the same slope.

Lemma 20 [Ritter 2009, Section 2.12] *Let $F, K \in \mathcal{H}$ and suppose τ_F and τ_K are both contained in an open interval that does not intersect* $\mathrm{Spec}(\partial D, \alpha)$. *Then, if $\tau_F \le \tau_K$,*

$$[\Phi^{K,F}]\colon \mathrm{HF}^*(F) \longrightarrow \mathrm{HF}^*(K)$$

is an isomorphism. Under $[\Phi^{K,F}]$, 1_F and 1_K are identified.

In action windows, we have the following isomorphisms.

Lemma 21 [Viterbo 1999, Proposition 1.1] *Let $H_\bullet$ be a monotone homotopy between $H_\pm \in \mathcal{H}$ that is constant in the s–coordinate for $|s| > S > 0$. Suppose $a_s, b_s \colon \mathbb{R} \to \mathbb{R}$ are functions which are constant* *outside* $[-S,S]$ *and* $a_s, b_s \notin \operatorname{Spec}(H_s)$ *for all* s. *Then,*

$$[\Phi^{H_-,H_+}]\colon \mathrm{HF}^*_{(a_+,b_+)}(H_+) \xrightarrow{\cong} \mathrm{HF}^*_{(a_-,b_-)}(H_-)$$

for $a_\pm = \lim_{s\to\pm\infty} a_s$ *and* $b_\pm = \lim_{s\to\pm\infty} b_s$.

3.2 Filtered symplectic cohomology

Equip the set of admissible Hamiltonians $\mathcal{H}^0$ negative on D with the partial order

$$H \preceq K \iff H(t,p) \le K(t,p) \quad \text{for all } (t,p) \in S^1 \times \widehat{D}.$$

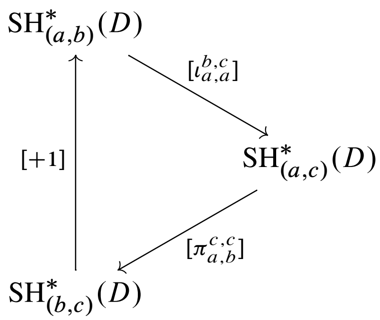
Let $\{H_i\}_{i\in I} \subset \mathcal{H}^0$ be a cofinal sequence with respect to $\preceq$. We define the *symplectic cohomology of* D as the direct limit

$$\mathrm{SH}^*_{(a,b)}(D) = \varinjlim_{H_i} \mathrm{HF}^*_{(a,b)}(H_i)$$

taken with respect to the continuation maps

$$[\Phi^{H_j,H_i}_{(a,b)}]\colon \mathrm{HF}^*_{(a,b)}(H_i) \longrightarrow \mathrm{HF}^*_{(a,b)}(H_j)$$

for $i < j$. We let $\mathrm{SH}^*(D) = \mathrm{SH}^*_{(-\infty,+\infty)}(D)$. The long exact sequence on Floer cohomology carries through the direct limit and we also have a long exact sequence on symplectic cohomology:



The Viterbo map Let $F \in \mathcal{H}$ and consider $H \in \mathcal{H}^0$ with $\tau_H = \tau_F$. Then, by Lemma 20, we have $\mathrm{HF}^*(F) \cong \mathrm{HF}^*(H)$ and there exist, by the definition of symplectic cohomology, a map

$$j_F\colon \mathrm{HF}^*(F) \cong \mathrm{HF}^*(H) \longrightarrow \mathrm{SH}^*(D) \tag{4}$$

sending each element of $\mathrm{HF}^*(H)$ to its equivalence class. Now, for $H \in \mathcal{H}^0$ with slope $\tau_H < T_0$ we can define, by Lemma 17, the map $v^*\colon \mathrm{H}^*(D) \to \mathrm{SH}^*(D)$ first introduced in [Viterbo 1999] by

$$\mathrm{H}^*(D) \xrightarrow{\Phi_H} \mathrm{HF}^*(H) \xrightarrow{j_H} \mathrm{SH}^*(D), \qquad v^* = j_H \circ \Phi_H.$$

This map induces a unit on symplectic cohomology. Recall that 1_H denotes the unit in $\mathrm{HF}^*(H)$ (see Remark 18).

Theorem 22 [Ritter 2013] *The ring structure on* $\mathrm{HF}^*(H)$ *induces a ring structure on* $\mathrm{SH}^*(D)$*. The unit on* $\mathrm{SH}^*(D)$ *is given by the image of the unit* $e_D \in \mathrm{H}^*(D)$ *under the map* v^**. Moreover,*

$$v^*(e_D) \in \operatorname{im}([\iota^{\varepsilon,\infty}_{-\infty,\infty}]\colon \mathrm{SH}^*_{(-\infty,\varepsilon)}(D) \longrightarrow \mathrm{SH}^*_{(-\infty,\infty)}(D)).$$

4 Spectral invariants and spectral norm

4.1 Spectral invariants

Denote by $\mathrm{Ham}_c(D, \mathrm{d}\lambda)$ the group of compactly supported Hamiltonian diffeomorphisms of $(D, \mathrm{d}\lambda)$ and by $\mathrm{Symp}_c(D, \mathrm{d}\lambda)$ the group of compactly supported symplectomorphisms of $(D, \mathrm{d}\lambda)$. The Hofer norm of a compactly supported Hamiltonian $H \in \mathcal{C}(D)$ is defined as

$$\|H\| = \int_0^1 \bigl(\sup_{p\in D} H(t,p) - \inf_{p\in D} H(t,p)\bigr)\,\mathrm{d}t.$$

Using the Hofer norm, we can define a bi-invariant metric [Hofer 1990; Lalonde and McDuff 1995] on $\mathrm{Ham}_c(D, \mathrm{d}\lambda)$ by

$$d_H(\varphi,\psi) = d_H(\varphi\psi^{-1}, \mathrm{id}), \quad d_H(\varphi,\mathrm{id}) = \inf\{\|H\| \mid \varphi = \varphi_H\}.$$

Recall that $\mathcal{C}(D)$ forms a group under the multiplication

$$H \mathbin{\#} K(t,p) = H(t,p) + K(t, (\varphi_H^t)^{-1}(p)),$$

with the inverse of some $H \in \mathcal{C}(D)$ given by $\overline{H}(t,p) = -H(t, \varphi_H^t(p))$.

From Lemma 17 and by the definition of $\mathrm{HF}^*(H)$ for $H \in \mathcal{C}(D)$, we know that $\mathrm{HF}^*(H) \cong \mathrm{H}^*(D)$. For $\beta \in \mathrm{H}^*(D)$, we define, following [Schwarz 2000], the *spectral invariant of* H *relative to* β *as*

$$c(\beta, H) = \inf\bigl\{\ell \in \mathbb{R} \mid \Phi_H(\beta) \in \operatorname{im}\bigl([\iota^{\ell,\infty}_{-\infty,-\infty}]\colon \mathrm{HF}^*_{(-\infty,\ell)}(H) \to \mathrm{HF}^*(H)\bigr)\bigr\},$$

which is, by exactness of the long exact sequence (2), equivalent to

$$c(\beta, H) = \inf\{\ell \in \mathbb{R} \mid [\pi^{\infty,\infty}_{-\infty,\ell}] \circ \Phi_H(\beta) = 0\}.$$

The following proposition gathers all the properties of spectral invariants we need for the rest of the text. Proofs of these properties can be found[5] in [Frauenfelder and Schlenk 2007, Section 5].

Proposition 23 *Let* $\beta, \eta \in H^*(D)$ *and let* $H, K \in \mathcal{C}(D)$*. Then:*

- ***Continuity*** $\int_0^1 \min_{x\in D}(K-H)\,\mathrm{d}t \le c(\beta,H) - c(\beta,K) \le \int_0^1 \max_{x\in D}(K-H)\,\mathrm{d}t$.
- ***Spectrality*** $c(\beta, H) \in \mathrm{Spec}(H)$.
- ***Triangle inequality*** $c(\beta \smile \eta, H \mathbin{\#} K) \le c(\beta,H) + c(\eta,K)$.
- ***Monotonicity*** *If* $H(t,x) \le K(t,x)$ *for all* $(t,x) \in [0,1]\times D$*, then* $c(\beta,H) \ge c(\beta,K)$*.*

[5] Note that the signs for continuity and monotonicity differ from [Frauenfelder and Schlenk 2007, Section 5] because of differences in sign conventions.

Remark 24 The continuity property of Proposition 23 allows us to define spectral invariants of compactly supported continuous Hamiltonians $H \in C_c^0([0,1] \times D)$. They satisfy continuity, the triangle inequality and monotonicity.

4.1.1 Additional properties of c The following lemma assures us that spectral invariants are well-defined on $\mathrm{Ham}_c(D, \mathrm{d}\lambda)$. The proof relies on the spectrality and the triangle inequality.

Lemma 25 *Let $H, K \in \mathcal{C}(D)$ such that $\varphi_H = \varphi_K$ and let $\beta \in \mathrm{H}^*(D)$. Then,*

$$c(\beta, H) = c(\beta, K).$$

Proof We have $\varphi_{H \# \overline{K}} = \varphi_0 = \mathrm{id}$ and in that case $\mathrm{Spec}(H \# \overline{K}) = \{0\}$. Now, by spectrality of spectral invariants, $c(\beta, H \# \overline{K}) = 0$. Thus, the triangle inequality yields

$$c(\beta, H) = c(\beta, H \# \overline{K} \# K) \le c(\beta, H \# \overline{K}) + c(\beta, K) = c(\beta, K).$$

Repeating the same argument with $K \# \overline{H}$ instead of $H \# \overline{K}$, we obtain $c(\beta, K) \le c(\beta, H)$, which concludes the proof. □

The spectral invariant with respect to the cohomological unit admits an implicit definition which depends on the spectral invariants with respect to all other cohomology classes in $\mathrm{H}^*(D)$. This follows directly from the triangle inequality.

Lemma 26 *Let $H \in \mathcal{C}(D)$. Then,*

$$c(1, H) = \max_{\beta \in \mathrm{H}^*(D)} c(\beta, H).$$

Proof Let $\beta \in H^*(D)$. By the definition of the unit and the concatenation of Hamiltonians, we have

$$c(\beta, H) = c(\beta \smile 1, H) = c(\beta \smile 1, 0 \# H).$$

Then, since $c(\beta, 0) = 0$, the triangle inequality guarantees that

$$c(\beta, H) = c(\beta \smile 1, 0 \# H) \le c(\beta, 0) + c(1, H) = c(1, H).$$

The choice of β being arbitrary, this concludes the proof. □

4.1.2 The symplectic contraction principle We conclude this section by recalling the symplectic contraction technique introduced in [Polterovich 2014, Section 5.4]. This principle allows one to describe the effect of the Liouville flow $\{\psi_Y^{\log r}\}_{0<r<1}$ on spectral invariants.

First, we need to describe how the Liouville flow acts on the symplectic form ω of D and on compactly supported Hamiltonians on D. Since $L_Y \omega = \omega$, we have that the Liouville flow contracts the symplectic form:

$$(\psi_Y^{\log r})^* \omega = r\omega.$$

Now, consider a Hamiltonian $H \in \mathcal{C}(D)$ supported in $U \subset D$. For fixed $0 < r < 1$ define the Hamiltonian

$$H_r(t,x) = \begin{cases} rH(t, (\psi_Y^{\log r})^{-1}(x)) & \text{if } x \in \psi_Y^{\log r}(U), \\ 0 & \text{if } x \notin \psi_Y^{\log r}(U). \end{cases} \tag{5}$$

It then follows from the two previous equations that $\mathrm{Spec}(H_r) = r\,\mathrm{Spec}(H)$. This allows one to prove:

Lemma 27 [Polterovich 2014] *Suppose $H \in \mathcal{C}(D)$ and let $H_r \in \mathcal{C}(D)$ be as in (5). Then,*

$$c(1, H_r) = rc(1, H).$$

4.2 Spectral norm

We define the *spectral norm $\gamma(H)$ of $H \in \mathcal{C}(D)$* as

$$\gamma(H) = c(1, H) + c(1, \overline{H}).$$

For $\varphi \in \mathrm{Ham}_c(D, \mathrm{d}\lambda)$ such that $\varphi = \varphi_H$, define

$$\gamma(\varphi) = \gamma(H).$$

By virtue of Lemma 25, this is well-defined.

From [Frauenfelder and Schlenk 2007, Section 7], we have the following theorem which justifies calling γ a norm.

Theorem 28 *Let $\varphi, \psi \in \mathrm{Ham}_c(D, \mathrm{d}\lambda)$ and let $\chi \in \mathrm{Symp}_c(D, \mathrm{d}\lambda)$. Then:*

- ***Nondegeneracy*** $\gamma(\mathrm{id}) = 0$ *and* $\gamma(\varphi) > 0$ *if* $\gamma \neq \mathrm{id}$.
- ***Triangle inequality*** $\gamma(\varphi\psi) \leq \gamma(\varphi) + \gamma(\psi)$.
- ***Symplectic invariance*** $\gamma(\chi \circ \varphi \circ \chi^{-1}) = \gamma(\varphi)$.
- ***Symmetry*** $\gamma(\varphi) = \gamma(\varphi^{-1})$.
- ***Hofer bound*** $\gamma(\varphi) \leq d_H(\varphi, \mathrm{id})$.

5 Cohomological barricades on Liouville domains

Ganor and Tanny [2023] introduced a particular perturbation of Hamiltonians compactly supported inside *contact incompressible boundary* domains (CIB) of closed aspherical symplectic manifolds. For instance, if $U \subset M$ is an incompressible open set which is a Liouville domain, then U is a CIB. In Floer homology, the aforementioned Hamiltonian perturbation, which is called a barricade, prohibits the existence of Floer trajectories exiting and entering the CIB. We consider barricades in the particular case of Liouville domains and adapt them to Floer cohomology.

In the present setting, we define barricades for a special class of admissible Hamiltonians.

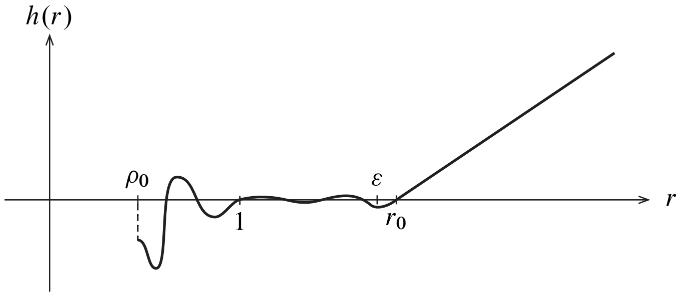


Figure 6: An r_0–barricade-admissible Hamiltonian.

Definition 29 A Hamiltonian H is said to be r_0-*barricade-admissible* if $H \in \mathcal{H}_{r_0}$ and the following conditions hold:

- $H(t,x,r) = h(r)$ on $\widehat{D} \setminus D^{\rho_0}$ for some $\rho_0 \in (0,1)$.
- $h(r)$ is C^2–small on $(1, r_0 - \varepsilon)$.
- $h(r)$ is strictly convex on $(r_0 - \varepsilon, r_0)$.

(See Figure 6.) Here $\varepsilon > 0$ is small enough so that $1 < r_0 - \varepsilon$. We denote the set of r_0–barricade-admissible Hamiltonians by $\overline{\mathcal{H}}_{r_0}$.

We say that $(F_\bullet, J)$ is an r_0-*barricade-admissible pair* if $F_\bullet$ is a monotone homotopy such that $F_s \in \overline{\mathcal{H}}_{r_0}$ for all s and J is an admissible almost-complex structure.

Remark 30 By Definition 19, the extension H^τ of any Hamiltonian H compactly supported in D can be chosen so that it is r_0–barricade-admissible.

Definition 31 Let $r_0 > 1$ and $0 < \varepsilon < r_0 - 1$. Define $B_{r_0,\varepsilon} = D^{r_0-\varepsilon} \setminus D$, where, for $\rho > 0$, $D^\rho = \Psi_Y^{\log \rho}(D)$. Suppose $(F_\bullet, J)$ is an r_0–barricade-admissible pair from F_+ to F_-. We say that $(F_\bullet, J)$ admits a barricade on $B_{r_0,\varepsilon}$ if for every $x_\pm \in \mathcal{P}(F_\pm)$ and every Floer trajectory $u\colon \mathbb{R} \times S^1 \to \widehat{D}$ connecting $x_\pm$, we have, for $D_{\mathrm{b}} := D^{r_0-\varepsilon} = D \cup B_{r_0,\varepsilon}$:

(1) If $x_- \in D$, then $\mathrm{im}(u) \subset D$.

(2) If $x_+ \in D_b$, then $\mathrm{im}(u) \subset D_{\mathrm{b}}$.

Remark 32 In the language of [Ganor and Tanny 2023], a barricade on $B_{r_0,\varepsilon}$ as described above would be called a barricade in $D^{r_0-\varepsilon}$ around D.

5.1 How to construct barricades

To construct barricades, we need to consider special classes of pairs of Hamiltonians and almost-complex structures. These are defined using a refinement of Definition 3.5 in [Ganor and Tanny 2023].

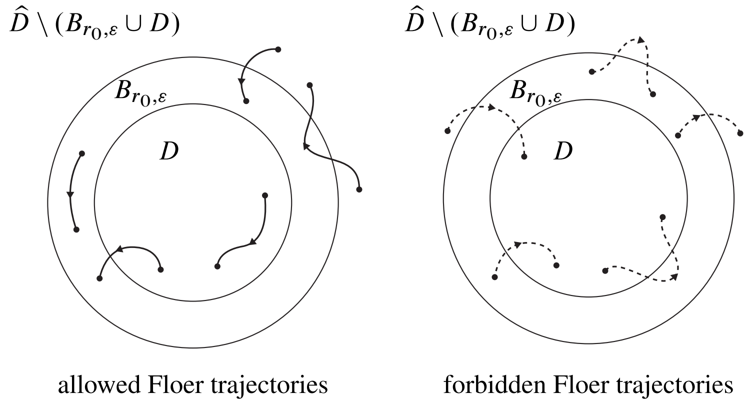


Figure 7: Floer cylinders in a barricade. The arrows follow the direction of the Floer differential and the continuation map: from x_+ to x_-.

Definition 33 Let $r_0 > 1$, $\sigma \in (0, +\infty) \setminus \mathrm{Spec}(\partial D, \lambda)$ and $0 < \varepsilon < r_0 - 1$. An r_0–barricade-admissible pair $(F_\bullet, J)$ admits a cylindrical bump of slope σ on $B_{r_0,\varepsilon}$ if:

(1) $F = 0$ on $\partial B_{r_0,\varepsilon} \times S^1 \times \mathbb{R}$.

(2) $JY = R_\alpha$ for Y the Liouville vector field on D, on a neighborhood of $\partial B_{r_0,\varepsilon}$; ie J is cylindrical near $\partial B_{r_0,\varepsilon} = \partial D \sqcup (\{r_0 - \varepsilon\} \times \partial D)$.

(3) $\nabla_J F = \sigma Y$ near $(\{1\} \times \partial D) \times S^1 \times \mathbb{R}$ and $\nabla_J F = -\sigma Y$ near $(\{r_0 - \varepsilon\} \times \partial D) \times S^1 \times \mathbb{R}$. Here, ∇_J denotes the gradient induced by the metric g_J.

(4) All 1–periodic orbits of $F_\pm$ contained in $B_{r_0,\varepsilon}$ are critical points with values in the interval $(-\sigma, \sigma)$. (In particular, $\sigma < T_0$.)

A cohomological adaptation of Lemma 3.3 in [Ganor and Tanny 2023] yields the following action estimates for pairs with cylindrical bumps.

Lemma 34 *Suppose that the r_0–barricade-admissible pair (F, J) admits a cylindrical bump of slope σ on $B_{r_0,\varepsilon}$. For every finite-energy solution u connecting $x_\pm \in \mathcal{P}(F_\pm)$:*

(1) $\mathrm{im}\, x_- \subset D$ *and* $\mathrm{im}\, x_+ \subset \widehat{D} \setminus D \Longrightarrow \mathcal{A}_{F_+}(x_+) > \sigma$.

(2) $\mathrm{im}\, x_+ \subset D_{\mathrm{b}}$ *and* $\mathrm{im}\, x_- \subset \widehat{D} \setminus D_{\mathrm{b}} \Longrightarrow \mathcal{A}_{F_-}(x_-) < -\sigma$.

See Figure 7.

Lemma 34 and the maximum principle are all we need to prove that every pair with a cylindrical bump admits a barricade. More precisely, we have:

Proposition 35 *Let (F, J) be an r_0–barricade-admissible pair with a cylindrical bump of slope σ on $B_{r_0,\varepsilon}$. Then, (F, J) admits a barricade on $B_{r_0,\varepsilon}$.*

Proof Suppose $u\colon \mathbb{R} \times S^1 \to \widehat{D}$ is a Floer trajectory between $x_\pm \in \mathcal{P}(F_\pm)$. We only need to study the case where $\mathrm{im}\, x_- \subset D$ and the case where $\mathrm{im}\, x_+ \subset D_{\mathrm{b}}$.

Suppose that $\operatorname{im} x_- \subset D$. We first establish that x_+ must lie inside D. Indeed, if $\operatorname{im} x_+ \subset \widehat{D} \setminus D$, part (1) of Lemma 34 assures us that $\mathcal{A}_{F_+}(x_+) > \sigma$, which contradicts the fact that orbits on $\widehat{D} \setminus D$ must have action in the interval $(-\sigma, \sigma)$ by the construction of the cylindrical bump. Therefore, $\operatorname{im} x_+ \subset D$ as desired. Now, since $\operatorname{im} x_\pm \subset D$, the maximum principle guarantees that $\operatorname{im} u \subset D$.

To finish the proof, we look at the case where $\operatorname{im} x_+ \subset D_{\mathrm{b}}$. Similarly to the previous case, we prove that x_- also lies inside D_{b}. If $\operatorname{im} x_- \subset \widehat{D} \setminus D_{\mathrm{b}}$, part (2) of Lemma 34 imposes $\mathcal{A}_{F_-}(x_-) < -\sigma$, which is again impossible by construction of the cylindrical bump. Therefore, $\operatorname{im} x_- \subset D_{\mathrm{b}}$ and the maximum principle implies $\operatorname{im} u \subset D_{\mathrm{b}}$. □

Given a pair (F, J) and $\sigma > 0$ small, we can add to F a $\mathcal{C}^\infty$–small radial bump function χ with support inside $B_{r_0,\varepsilon}$ such that $(F + \chi, J)$ has a cylindrical bump of slope σ on $B_{r_0,\varepsilon}$. By Proposition 35, the perturbed pair will also admit a barricade on $B_{r_0,\varepsilon}$. A second perturbation of the Hamiltonian term at its ends, under which the barricade survives, allows us to achieve Floer regularity for the pair. This procedure is carried out carefully in [Ganor and Tanny 2023, Section 9] and yields the following.

Theorem 36 [Ganor and Tanny 2023] *Let $F_\bullet$ be a monotone homotopy. Then, there exists a $\mathcal{C}^\infty$–small perturbation $f_\bullet$ of $F_\bullet$ and an almost-complex structure J such that the pairs $(f_\bullet, J)$ and $(f_\pm, J)$ are Floer-regular and have a barricade on $B_{r_0,\varepsilon}$.*

5.2 Decomposition of the Floer cochain complex

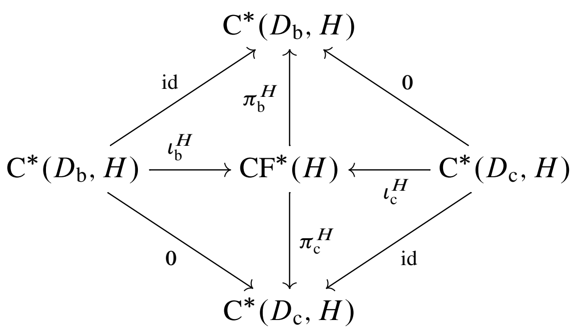
Let us investigate what structure barricades impose on the Floer cochain complex. Let $H \in \overline{\mathcal{H}}_{r_0}$ and suppose the pair (H, J) admits a barricade on $B_{r_0,\varepsilon}$. For an open subset $U \subset \widehat{D}$, denote by $\mathrm{C}^*(U, H)$ the set of 1–periodic orbits of H in U. By the definition of the differential ∂ on Floer cohomology, $\mathrm{C}^*(D_{\mathrm{b}}, H)$ is closed under ∂ and it therefore forms a subcomplex of $\mathrm{CF}^*(H)$. Moreover, for $D_{\mathrm{c}} = \widehat{D} \setminus D_{\mathrm{b}}$, we also have that

$$\mathrm{C}^*(D_{\mathrm{c}}, H) = \frac{\mathrm{CF}^*(H)}{\mathrm{C}^*(D_{\mathrm{b}}, H)}$$

is a well-defined cochain complex. In terms of vector spaces, we have the decomposition

$$\mathrm{CF}^*(H) \cong \mathrm{C}^*(D_{\mathrm{b}}, H) \oplus \mathrm{C}^*(D_{\mathrm{c}}, H).$$

The direct product gives us injections ι_{b}^H, ι_{c}^H and projections π_{b}^H, π_{c}^H for which the diagram



commutes and the equation

$$\iota_{\mathrm{b}}^H \circ \pi_{\mathrm{b}}^H(q) + \iota_{\mathrm{c}}^H \circ \pi_{\mathrm{c}}^H(q) = q$$

holds for any $q \in \mathrm{CF}^*(H)$. Here, the projection π_{c}^H coincides with the canonical projection

$$\mathrm{CF}^*(H) \twoheadrightarrow \frac{\mathrm{CF}^*(H)}{\mathrm{C}^*(D_{\mathrm{b}}, H)}.$$

The differential ∂_{b} on $\mathrm{C}^*(D_{\mathrm{b}})$ is simply the restriction of the differential ∂ of $\mathrm{CF}^*(H)$ on $\mathrm{C}^*(D_{\mathrm{b}})$. The differential ∂_{c} on $\mathrm{C}^*(D_{\mathrm{c}})$ is the quotient complex differential defined by

$$\partial_{\mathrm{c}} \pi_{\mathrm{c}}^H(p) = \pi_{\mathrm{c}}^H(\partial p).$$

5.2.1 Continuation maps Let $(F_\bullet, J)$ be an r_0–barricade-admissible pair that admits a barricade on $B_{r_0,\varepsilon}$. Then, since the continuation map $\Phi_{F_\bullet} : \mathrm{CF}^*(F_+) \to \mathrm{CF}^*(F_-)$ counts Floer trajectories of F connecting 1–periodic orbits of F_+ to 1–periodic orbits of F_-, it restricts, due to the barricade, to a chain map

$$\Phi_F^{\mathrm{b}} : \mathrm{C}^*(D_{\mathrm{b}}, F_+) \longrightarrow \mathrm{C}^*(D_{\mathrm{b}}, F_-).$$

Moreover, in virtue of Lemma 38 below, Φ_F projects to a chain map

$$\Phi_F^{\mathrm{c}} : \mathrm{C}^*(D_{\mathrm{c}}, F_+) \longrightarrow \mathrm{C}^*(D_{\mathrm{c}}, F_-)$$

such that the following diagram commutes

$$\begin{array}{ccc} \mathrm{CF}^*(F_+) & \xrightarrow{\ \Phi_{F_\bullet}\ } & \mathrm{CF}^*(F_-) \\ \Big\downarrow{\scriptstyle \pi_{\mathrm{b}}^+} & & \Big\downarrow{\scriptstyle \pi_{\mathrm{b}}^-} \\ \mathrm{C}^*(D_{\mathrm{c}}, F_+) & \xrightarrow[\ \Phi_{F_\bullet}^{\mathrm{c}}\ ]{} & \mathrm{C}^*(D_{\mathrm{c}}, F_-) \end{array}$$

where we write $\pi_{\mathrm{b}}^+ = \pi_{\mathrm{b}}^{F_+}$ and $\pi_{\mathrm{b}}^- = \pi_{\mathrm{b}}^{F_-}$.

5.2.2 Chain homotopies Let $(F_\pm, J)$ be r_0–barricade-admissible pairs that admit cylindrical bumps of slope σ on $B_{r_0,\varepsilon}$ such that F_+ and F_- have the same slope $\tau_+ = \tau_-$ at infinity. Consider the linear homotopy

$$F_s = F_- + \beta(s)(F_+ - F_-),$$

where $\beta : \mathbb{R} \to [0,1]$ is a smooth function such that $\beta(s) = 0$ for $s \le -1$, $\beta(s) = 1$ for $s \ge 1$ and $\beta'(s) > 0$ for all $s \in (-1, 1)$. Denote by $\overline{F}_\bullet$ the inverse homotopy defined by $\overline{F}_s = F_{-s}$. For $\rho > 1$ large, we define the concatenation $F \,\#\, \overline{F}_\bullet$ as

$$(F \,\#\, \overline{F})_s = \begin{cases} F_{s+\rho} & \text{for } s \le 0, \\ \overline{F}_{s-\rho} & \text{for } s \ge 0. \end{cases}$$

Using the definition of $F_\bullet$ and $\overline{F}_\bullet$, we can simply write

$$(F \,\#\, \overline{F})_s = F_- + \beta_\rho(s)(F_+ - F_-)$$

for $\beta_\rho(s)=\beta(-|s|+\rho)$. The homotopy $F\#\overline{F}_\bullet$ generates the composition of continuation homomorphisms $\Phi_F\circ\Phi_{\overline{F}}\colon \mathrm{CF}^*(F_-)\to\mathrm{CF}^*(F_-)$, which is chain homotopic to the identity on $\mathrm{CF}^*(F_-)$,

$$\Phi_{F_\bullet}\circ\Phi_{\overline{F}_\bullet}-\mathrm{id}_-=\partial_-\circ\Psi_--\Psi_-\circ\partial_-$$

for $\Psi_-\colon\mathrm{CF}^*(F_-)\to\mathrm{CF}^{*-1}(F_-)$ and ∂_- the differential on $\mathrm{CF}^*(F_-)$. The chain homotopy Ψ_- is built by counting Floer solutions of the homotopy $\{\Gamma^\kappa\}_{\kappa\in[0,1]}$ between $F\#\overline{F}_\bullet$ and the constant homotopy F_-, which is defined by

$$\Gamma^\kappa_s=F_-+\kappa\beta_\rho(s)(F_+-F_-).$$

For $x\in\mathcal{P}(F_-)$ and $y\in\mathcal{P}(F_+)$, define

$$\mathcal{M}^\Gamma(x,y)=\{(\kappa,u)\mid\kappa\in[0,1],\ u\in\mathcal{M}(x,y;\Gamma^k_\bullet)\}.$$

We can perturb Γ with a C^∞–small function in order to make it regular [Audin and Damian 2014, Chapter 11]. Now, since the pairs $(F_\pm,J)$ admit cylindrical bumps of slope σ on $B_{r_0,\varepsilon}$, and thus have barricades on $B_{r_0,\varepsilon}$, solutions to the parametric Floer equation for Γ^κ also admit cylindrical bumps of slope σ on $B_{r_0,\varepsilon}$ and have barricades on $B_{r_0,\varepsilon}$. To see this, first fix $\kappa\in[0,1]$, We need to show that Γ^κ satisfies conditions (1) through (4) of Definition 33. For (1), we have, on $\partial B_{r_0,\varepsilon}\times S^1\times\mathbb{R}$,

$$\Gamma^\kappa=F_-+\kappa\beta_\rho(s)(F_+-F_-)=0+\kappa\beta_\rho(s)(0-0)=0.$$

Condition (2) is automatically satisfied since J is fixed. For condition (3), we have on $(\{1\}\times\partial D)\times S^1\times\mathbb{R}$,

$$\nabla_J\Gamma^\kappa=\nabla_JF_-+\kappa\beta_\rho(s)(\nabla_JF_+-\nabla_JF_-)=\sigma Y+\kappa\beta_\rho(s)(\sigma Y-\sigma Y)=\sigma Y$$

and, by the same computation, $\nabla_J\Gamma^\kappa=-\sigma Y$ on $(\{r_0-\varepsilon\}\times\partial D)\times S^1\times\mathbb{R}$. Condition (4) is also satisfied since $\Gamma^\kappa_{\pm\infty}=F_-$. All of this still holds with regular perturbations of Γ.

Lemma 37 *Let $F_-,F_+\in\overline{\mathcal{H}}_{r_0}$ with same slope at infinity and suppose they both admit barricades on $B_{r_0,\varepsilon}$. Furthermore, suppose that solutions to the parametric Floer equation for Γ^κ also admit barricades on $B_{r_0,\varepsilon}$. Then, for any C^∞–small perturbation Γ' of Γ which satisfies $\mathcal{P}(F'_\pm)=\mathcal{P}(F_\pm)$, Floer trajectories in $\mathcal{M}^{\Gamma'}$ follow the rules of the barricade on $B_{r_0,\varepsilon}$.*

Proof The proof follows the same ideas as the proof of Proposition 9.21 in [Ganor and Tanny 2023]. By Gromov compactness, any sequence $(\kappa_n,u_n)\in\mathcal{M}^\Gamma(x_-,y_+)$ of solutions to the parametric Floer equation converges, up to taking a subsequence, to a broken trajectory $(\kappa,\bar v)$, where $\bar v=(v_1,\dots,v_k,w,v'_1,\dots,v'_\ell)$ connects two orbits $x_\pm\in\mathcal{P}(F_\pm)$. The fact that $F_\pm$ both admit a barricade on $B_{r_0,\varepsilon}$ assures us that

- $x_-\in D\implies\bar v\subset D$,
- $x_+\in D\implies\bar v\subset D_\flat$.

Now, consider a sequence of regular homotopies $\{\Gamma_n\}_n$ with ends $\lim_{s\to\pm\infty}\Gamma_{s,n}=F_{n\pm}$ converging to Γ such that $\mathcal{P}(F_{n\pm})=\mathcal{P}(F_\pm)$ for all n. Then, the above two implications regarding broken trajectories imply that every trajectory $(\kappa_n,u'_n)\in\mathcal{M}^{\Gamma'}(x_-,x_+)$, for $x_\pm\in\mathcal{P}(F_\pm)$, obey the rules of the barricade. □

Thus, Ψ_- restricts to a map $\Psi_-^b \colon \mathrm{C}^*(D_{\mathrm{b}}, F_-) \to \mathrm{C}^{*-1}(D_{\mathrm{b}}, F_-)$ and by Lemma 39 below, we can define its projection $\Psi_-^c \colon \mathrm{C}^*(D_{\mathrm{c}}, F_-) \to \mathrm{C}^{*-1}(D_{\mathrm{c}}, F_-)$.

Technical lemmas When adapting computations from homology to cohomology, we often have to rely on quotient complexes instead of subcomplexes. Here are a few simple results from homological algebra which will be useful in that regard. Let (A, d_A) and (C, d_C) be cochain complexes and let $B \subset A$ and $D \subset C$ be subcomplexes.

Lemma 38 *Suppose $f\colon (A, B) \to (C, D)$ is a chain map. Then, there exists a unique chain map $\bar{f}\colon A/B \to C/D$ such that the following diagram commutes*

$$\begin{array}{ccc} A & \xrightarrow{\;f\;} & C \\ {\scriptstyle \pi_B}\downarrow & & \downarrow{\scriptstyle \pi_D} \\ A/B & \xrightarrow[\;\bar{f}\;]{} & C/D \end{array}$$

for π_B and π_D the canonical projections. It follows that, on cohomology, we have the following commutative diagram:

$$\begin{array}{ccc} \mathrm{H}^*(A) & \xrightarrow{\;[f]\;} & \mathrm{H}^*(C) \\ {\scriptstyle [\pi_B]}\downarrow & & \downarrow{\scriptstyle [\pi_D]} \\ \mathrm{H}^*(A/B) & \xrightarrow[\;[\bar{f}]\;]{} & \mathrm{H}^*(C/D) \end{array}$$

Proof Define, for all $x \in A$,

$$\bar{f}(\pi_B(x)) = \pi_D(f(x)).$$

We first need to show that $\bar{f}$ is well-defined. Suppose $x' = x + b$ for $x \in A$ and $b \in B$. Then, since f restricts to a map from B to D, there exists $d \in D$ such that $f(b) = d$ and we have

$$\bar{f}(\pi_B(x')) = \pi_D(f(x+b)) = \pi_D(f(x)+d) = \pi_D(f(x)).$$

Thus, $\bar{f}$ is well-defined.

To prove uniqueness, we simply use the definition of $\bar{f}$. Suppose we have another map $\bar{g}\colon A/B \to C/D$ which makes the above diagram commute as well. Then, for all $x \in A$,

$$\bar{f}(\pi_B(x)) - \bar{g}(\pi_B(x)) = \pi_D(f(x)) - \pi_D(f(x)) = 0. \qquad \square$$

Lemma 39 *Suppose $f\colon (A, B) \to (C, D)$ and $g\colon (C, D) \to (A, B)$ are chain maps such that $f \circ g$ is chain homotopic to the identity*

$$f \circ g - \mathrm{id}_C = d_C \circ \psi - \psi \circ d_C,$$

where the chain homotopy is a map $\psi\colon (C, D) \to (C, D)$. Then, $\bar{f} \circ \bar{g}\colon C/D \to C/D$ is also chain homotopic to the identity.

Proof Since the chain homotopy $\psi\colon (C,D)\to(C,D)$ is a chain map of pairs, Lemma 38 allows us to define its projection $\bar\psi\colon C/D\to C/D$. Thus, for all $y\in C$,

$$\begin{aligned}\bar f\circ\bar g(\pi_D(y))-\mathrm{id}_{C/D}(\pi_D(y)) &= \bar f\circ\pi_B(g(y))-\pi_D(\mathrm{id}_C(y))\\ &= \pi_D(f\circ g(y))-\pi_D(\mathrm{id}_C(y))\\ &= \pi_D((d_C\circ\psi-\psi\circ d_C)(y))\\ &= (d_{C/D}\circ\pi_D\circ\psi-\pi_D\circ\psi\circ d_C)(y)\\ &= d_{C/D}\circ\bar\psi(\pi_D(y))-\bar\psi\circ d_{C/D}(\pi_D(y)),\end{aligned}$$

which proves that $\bar f\circ\bar g$ is chain homotopic to the identity on C/D since any $z\in C/D$ is of the form $z=\pi_D(y)$. □

6 Proofs of main results

6.1 Proof of Theorem A1

Fix $A\in(0,\infty)\setminus\mathrm{Spec}(\partial D,\lambda)$. The idea of the proof is to construct a special admissible Hamiltonian for which $c(1,\cdot)$ is bounded from below by $A-\varepsilon$ for ε a small constant which depends on A. This construction is inspired by [Cieliebak et al. 2010, Proposition 2.5]. Then, we use the fact that $c(1,\cdot)\geq 0$ to conclude.

6.1.1 Construction of the Hamiltonian Fix some $r_0>1$. For any $\delta\in(0,1)$ and $\sigma\in(0,T_0)$, we define the Hamiltonian $H_{\delta,A}$ as follows:

- $H_{\delta,A}$ is the constant function $A(\delta-1)$ on D^δ.
- $H_{\delta,A}(r,x)=A(r-1)$ on $D\setminus D^\delta$.
- $H_{\delta,A}(r,x)=0$ on $D^{r_0}\setminus D$.
- $H_{\delta,A}(r,x)=\sigma(r-r_0)$ on $\widehat D\setminus D^{r_0}$.

See Figure 8. We add a small perturbation to $H_{\delta,A}$ so that it lies in $\overline{\mathcal{H}}_{r_0}$. Denote by $h_{\delta,A}$ the function of one variable for which $H_{\delta,A}=h_{\delta,A}\circ r$ on D^c. If γ is a 1–periodic orbit of $h_{\delta,A}$ inside the level set $\{r\}\times\partial D$, its action can be written as

$$\mathcal{A}_{H_{\delta,A}}(\gamma)=\mathcal{A}_{H_{\delta,A}}(r)=rh'_{\delta,A}(r)-h_{\delta,A}(r).$$

The 1–periodic orbits of $H_{\delta,A}$ can be classified into three different categories. Recall that η_A denotes the distance between A and $\mathrm{Spec}(\partial D,\alpha)$.

(I) Critical points in D^δ with action close to $r_{\mathrm{I}}:=(1-\delta)A$.

(II) Nonconstant 1–periodic orbits near $\{\delta\}\times\partial D$ with action in a small neighborhood of the interval

$$I_{\mathrm{II}}=[\delta T_0+(1-\delta)A,\,A-\delta\eta_A].$$

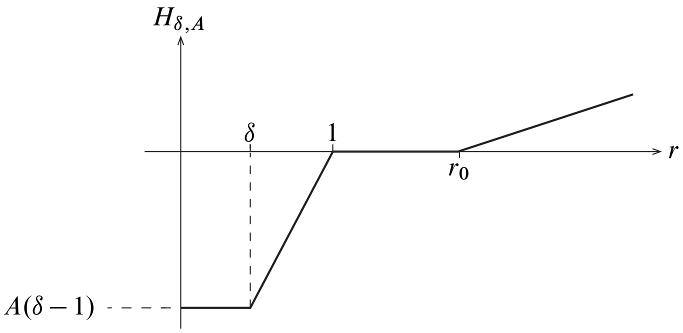


Figure 8: Radial portion of the Hamiltonian $H_{\delta,A}$.

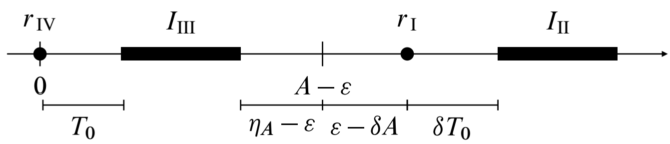


Figure 9: Distances that separate the action windows under consideration.

(III) Nonconstant 1–periodic orbits near $\{1\}\times\partial D$ with action in a small neighborhood of the interval

$$I_{\mathrm{III}} = [T_0, A-\eta_A].$$

(IV) Critical points in $D^{r_0}\setminus D$ with action close to $r_{\mathrm{IV}} := 0$.

Note that there are no nonconstant 1–periodic orbits near $\{r_0\}\times\partial D$, since the slope of the Hamiltonian there ranges from 0 to σ, which is less than T_0 by assumption.

We now want to construct a Floer complex $\mathrm{C}^*_{\mathrm{I,II}}$ which will contain the orbits of type (I) and (II) and another complex $\mathrm{C}^*_{\mathrm{III,IV}}$ containing orbits of type (III) and (IV). To that end, pick $0<\delta<1$ small enough so that $\delta A<\eta_A$. Now choose $\varepsilon>0$ such that

$$\delta A<\varepsilon<\eta_A.$$

Then, we have the inequalities

$$r_{\mathrm{IV}} < I_{\mathrm{III}} < A-\varepsilon < r_{\mathrm{I}} < I_{\mathrm{II}}.$$

As shown in Figure 9, r_{I}, I_{II}, I_{III} and r_{IV} are all separated by distances which depend only on T_0, A, η_A, δ and ε. Thus, we can choose the perturbation we add to $H_{\delta,A}$ to be small enough so that, in terms of action, we have

$$(\mathrm{IV}) < (\mathrm{III}) < A-\varepsilon < (\mathrm{I}) < (\mathrm{II}).$$

Therefore, since the Floer differential decreases the action, we can define the Floer cochain complexes as

$$\mathrm{C}^*_{\mathrm{III,IV}} = \mathrm{CF}^*_{(-\infty,A-\varepsilon)}(H_{\delta,A}), \quad \mathrm{C}^*_{\mathrm{I,II}} = \frac{\mathrm{CF}^*(H_{\delta,A})}{\mathrm{C}^*_{\mathrm{III,IV}}} = \mathrm{CF}^*_{(A-\varepsilon,\infty)}(H_{\delta,A})$$

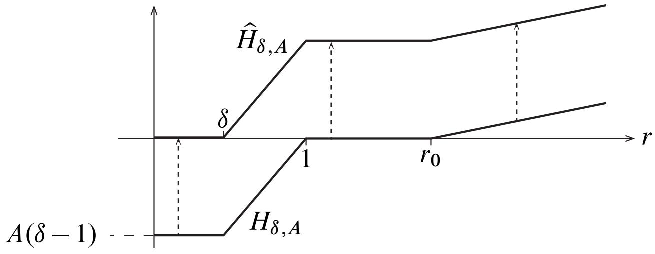


Figure 10: Homotopy from $H_{\delta,A}$ to $\widehat{H}_{\delta,A}$.

and they yield the Floer cohomology groups

$$\mathrm{H}^*(\mathrm{C}^*_{\mathrm{III,IV}}) = \mathrm{HF}^*_{(-\infty,A-\varepsilon)}(H_{\delta,A}), \quad \mathrm{H}^*(\mathrm{C}^*_{\mathrm{I,II}}) = \mathrm{HF}^*_{(A-\varepsilon,\infty)}(H_{\delta,A}).$$

A quick look at the action windows under consideration informs us that the above complexes fit into the short exact sequence

$$0 \longrightarrow \mathrm{C}^*_{\mathrm{III,IV}} \xrightarrow{\iota^{A-\varepsilon,+\infty}_{-\infty,-\infty}} \mathrm{CF}^*(H_{\delta,A}) \xrightarrow{\pi^{+\infty,+\infty}_{-\infty,A-\varepsilon}} \mathrm{C}^*_{\mathrm{I,\,II}} \longrightarrow 0,$$

which in turn yields an exact triangle in cohomology:

$$\begin{array}{ccc} \mathrm{H}^*(\mathrm{C}^*_{\mathrm{III,IV}}) & \xrightarrow{[\iota^{A-\varepsilon,+\infty}_{-\infty,-\infty}]} & \mathrm{HF}^*(H_{\delta,A}) \\ {\scriptstyle [+1]}\nwarrow & & \swarrow{\scriptstyle [\pi^{+\infty,+\infty}_{-\infty,A-\varepsilon}]} \\ & \mathrm{H}^*(\mathrm{C}^*_{\mathrm{I,II}}) & \end{array}$$

6.1.2 Factoring a map to $\mathbf{SH}^*(D)$ We now build maps Ψ and $\Psi_{\mathrm{I,II}}$ such that the diagram

$$\begin{array}{ccc} \mathrm{HF}^*(H_{\delta,A}) & \xrightarrow{[\pi^{\infty,\infty}_{-\infty,A-\varepsilon}]} & \mathrm{H}^*(\mathrm{C}^*_{\mathrm{I,II}}) \\ & {\scriptstyle \Psi}\searrow & \downarrow{\scriptstyle \Psi_{\mathrm{I,II}}} \\ & & \mathrm{SH}^*(D) \end{array} \tag{6}$$

commutes. We need to construct Ψ so that it coincides with the map $j_{H_{\delta,A}}\colon \mathrm{HF}^*(H_{\delta,A}) \to \mathrm{SH}^*(D)$ (see (4)). By virtue of Theorem 22, this assures us that Ψ is a map of unital algebras.

First, we construct $\Psi_{\mathrm{I,II}}$ in three steps.

Step 1 $[\Phi_1]\colon \mathrm{H}^*(\mathrm{C}^*_{\mathrm{I,II}}) \cong \mathrm{HF}^*_{(\delta A-\varepsilon,\infty)}(H_{\delta,A} + A(1-\delta))$. This isomorphism follows from a simple shift of $A(1-\delta)$ in the Hamiltonian term, which translates to a shift of $A(\delta-1)$ in action (see Figure 10). In what follows, we let $\widehat{H}_{\delta,A} := H_{\delta,A} + A(1-\delta)$.

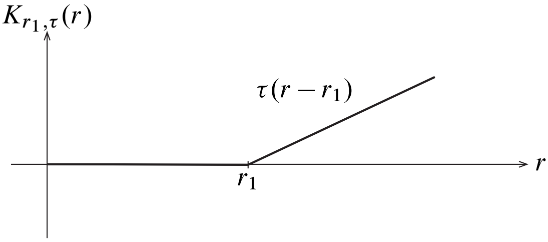


Figure 11: Radial portion of the Hamiltonian $K_{r_1,\tau}$.

For the next steps, we need to define another special family of Hamiltonians. Given $r_1 \in (0,+\infty)$ and $\tau \in (0,\infty) \setminus \mathrm{Spec}(\partial D,\lambda)$, define the Hamiltonian $K_{r_1,\tau}$ as follows (see Figure 11):

- $K_{r_1,\tau}$ is the constant zero function on D^{r_1}.
- $K_{r_1,\tau}(x,r) = \tau(r-r_1)$ on $\widehat{D} \setminus D^{r_1}$.

We add a small perturbation to $K_{r_1,\tau}$ so that it r_1–admissible. The 1–periodic orbits of $K_{r_1,\tau}$ fall in two categories:

(I′) Critical points in D^{r_1} with action near zero.

(II′) Nonconstant 1–periodic orbits near $\{r_1\} \times \partial D$ with action in a small neighborhood of the interval

$$[r_1 T_0, r_1\tau - r_1\eta_\tau].$$

By the same argument used for $H_{\delta,A}$, the action windows (I′) and (II′) are separated if we choose a small enough perturbation.

Step 2 $[\Phi_2]\colon \mathrm{HF}^*_{(\delta A-\varepsilon,\infty)}(\widehat{H}_{\delta,A}) \cong \mathrm{HF}^*(K_{\delta,A})$. Consider the homotopy

$$F_s = (1-\beta(s))K_{\delta,A} + \beta(s)\widehat{H}_{\delta,A},$$

where $\beta\colon \mathbb{R} \to [0,1]$ is a smooth function such that $\beta(s)=0$ for $s \le -1$, $\beta(s)=1$ for $s \ge 1$ and $\beta'(s) > 0$ for all $s \in (-1,1)$ (see Figure 12). Denote by

$$\Phi_{F_\bullet}\colon \mathrm{CF}^*(\widehat{H}_{\delta,A}) \longrightarrow \mathrm{CF}^*(K_{\delta,A})$$

the continuation map generated by $F_\bullet$.

Notice that since $H_{\delta,A} \preceq K_{\delta,A}$ we can restrict the continuation map on the action window $(\delta A-\varepsilon,\infty)$. Thus,

$$[\Phi_{F_\bullet}]\colon \mathrm{HF}^*_{(\delta A-\varepsilon,\infty)}(\widehat{H}_{\delta,A}) \longrightarrow \mathrm{HF}^*_{(\delta A-\varepsilon,\infty)}(K_{\delta,A})$$

is well-defined. Moreover, since $\delta A-\varepsilon<0$, $K_{\delta,A}$ has no orbits outside the action window $(\delta A-\varepsilon,\infty)$ and thus

$$[\iota^{-\infty,\infty}_{\delta A-\varepsilon,\infty}]\colon \mathrm{HF}^*_{(\delta A-\varepsilon,\infty)}(K_{\delta,A}) \longrightarrow \mathrm{HF}^*(K_{\delta,A})$$

is an isomorphism. We define $[\Phi_2]$ to be the composition $[\iota^{-\infty,\infty}_{\delta A-\varepsilon,\infty}] \circ [\Phi_{F_\bullet}]$.

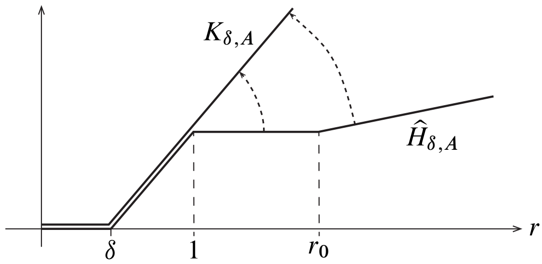


Figure 12: Homotopy from $\widehat{H}_{\delta,A}$ to $K_{\delta,A}$.

Step 3 Recall from (4), that we have a natural map

$$j_{K_{\delta,A}} : \mathrm{HF}^*(K_{\delta,A}) \longrightarrow \mathrm{SH}^*(D).$$

We define $\Psi_{\mathrm{I,II}} : \mathrm{H}^*(\mathrm{C}^*_{\mathrm{I,II}}) \to \mathrm{SH}^*(D)$ to be the composition

$$\Psi_{\mathrm{I,II}} = j_{K_{\delta,A}} \circ [\Phi_2] \circ [\Phi_1].$$

The morphism Ψ is built in a similar fashion. We define it as the following composition of maps:

$$\begin{array}{ccccc} \mathrm{HF}^*(H_{\delta,A}) & \xrightarrow[{[\Phi'_1]}]{\cong} & \mathrm{HF}^*(\widehat{H}_{\delta,A}) & & \\ & & \Big\downarrow {\scriptstyle [\Phi'_2]} & & \\ & & \mathrm{HF}^*(K_{\delta,A}) & \xrightarrow{j_{K_{\delta,A}}} & \mathrm{SH}^*(D) \end{array}$$

Here, the isomorphism $[\Phi'_1]$ follows from the fact that both $H_{\delta,A}$ and $\widehat{H}_{\delta,A}$ have the same slope at infinity. We defined $[\Phi'_2]$ to be the continuation map $[\Phi^{K_{\delta,A}\widehat{H}_{\delta,A}}]$. The last map is given, just as in Step 3, by $j_{K_{\delta,A}} : \mathrm{HF}^*(K_{\delta,A}) \to \mathrm{SH}^*(D)$. By construction, we therefore have

$$\Psi = j_{K_{\delta,A}} \circ [\Phi'_2] \circ [\Phi'_1] = j_{K_{\delta,A}} \circ [\Phi_{K_{\delta,A}\widehat{H}_{\delta,A}}] \circ [\Phi'_1] = j_{H_{\delta,A}}$$

as desired.

Now, we need to prove that diagram (6) commutes. Writing the maps Ψ and $\Psi_{\mathrm{I,II}}$ explicitly, we have the following diagram:

$$\begin{array}{ccc} \mathrm{HF}^*(H_{\delta,A}) & \xrightarrow{[\pi^{+\infty,+\infty}_{-\infty,A-\varepsilon}]} & \mathrm{H}^*(\mathrm{C}^*_{\mathrm{I,II}}) \\ \Big\downarrow {\scriptstyle [\Phi'_1]} & & {\scriptstyle [\Phi_1]} \Big\downarrow \\ \mathrm{HF}^*(\widehat{H}_{\delta,A}) & \xrightarrow{[\pi^{+\infty,+\infty}_{-\infty,\delta A-\varepsilon}]} & \mathrm{HF}^*_{(\delta A-\varepsilon,+\infty)}(\widehat{H}_{\delta,A}) \\ \Big\downarrow {\scriptstyle [\Phi_{K_{\delta,A}\widehat{H}_{\delta,A}}]} & & {\scriptstyle [\Phi_2]} \Big\downarrow \\ \mathrm{HF}^*(K_{\delta,A}) & \xrightarrow{\mathrm{id}} & \mathrm{HF}^*(K_{\delta,A}) \\ & & {\scriptstyle j_{K_{\delta,A}}} \Big\downarrow \\ & & \mathrm{SH}^*(D) \end{array} \tag{7}$$

The top square in diagram (7) commutes because, since $\widehat{H}_{\delta,A} \geq H_{\delta,A}$, there exists a continuation map from $\mathrm{HF}^*(H_{\delta,A}) \cong \mathrm{HF}^*_{(\delta A-\varepsilon,\infty)}(H_{\delta,A})$ to $\mathrm{HF}^*_{(\delta A-\varepsilon,+\infty)}(\widehat{H}_{\delta,A})$, where the isomorphism follows from the fact that $H_{\delta,A}$ has no orbits outside the action window $(\delta A-\varepsilon,\infty)$. Now, since the projection $[\pi^{+\infty,+\infty}_{-\infty,\delta A-\varepsilon}]$ commutes with continuation maps (see diagram (3)), the bottom square in diagram (7) also commutes. Therefore, we can conclude that diagram (6) commutes.

6.1.3 Spectral invariant and spectral norm of $H_{\delta,A}$ Recall that, by definition,

$$c(1,H_{\delta,A}) = \inf\{\ell \in \mathbb{R} \mid [\pi^{+\infty,+\infty}_{-\infty,\ell}] \circ [\iota^{\ell,+\infty}_{-\infty,-\infty}](1) = 0\}.$$

Since Ψ is a morphism of unital algebras, the commutative diagram (6) assures us that

$$[\pi^{+\infty,+\infty}_{-\infty,A-\varepsilon}](1_{H_{\delta,A}}) \neq 0$$

since we assume that $\mathrm{SH}^*(D) \neq 0$. Thus, from the exact triangle in cohomology induced by $[\iota^{A-\varepsilon,+\infty}_{-\infty,-\infty}]$ and $[\pi^{+\infty,+\infty}_{-\infty,A-\varepsilon}]$, we have $1 \notin \operatorname{im}[\iota^{A-\varepsilon,+\infty}_{-\infty,-\infty}]$ and therefore

$$c(1,H_{\delta,A}) \geq A-\varepsilon.$$

Now, we turn our attention to the spectral norm $\gamma(H_{\delta,A})$. We know from Lemma B that

$$c(1,H_{\delta,A}),\, c(1,\overline{H}_{\delta,A}) \geq 0.$$

It thus follows from the previous inequality that

$$\gamma(H_{\delta,A}) = c(1,H_{\delta,A}) + c(1,\overline{H}_{\delta,A}) \geq A-\varepsilon$$

as desired. This completes the proof.

6.2 Proof of Lemma B

We give a proof of Lemma B which relies on the decomposition of the Floer complex induced by the barricade. We expect that Lemma B could also be proven using Poincaré duality between filtered Floer cohomology and filtered Floer homology (as in [Cieliebak and Oancea 2018, Section 3]) and Lemma 4.1 of [Ganor and Tanny 2023].

Let $H \in \mathcal{H}_{r_0}$ with slope $0 < \tau_H < T_0$. Consider a linear homotopy $F_\bullet$ from $F_+ = K_{r_0,\tau_H}$ (see Figure 11) to $F_- = H$. There exists a small perturbation $f_\bullet$ of $F_\bullet$ and an almost-complex structure J such that the pairs $(f_\bullet, J)$ and $(f_\pm, J)$ admit a barricade on $B_{r_0,\varepsilon}$ for $\varepsilon > 0$ small enough. Fix $\delta > 0$. The construction of Theorem 36 allows us to choose J time independent [Ganor and Tanny 2023, Remark 3.7] and f such that

$$-\delta \leq \int_0^1 \min_{x \in \widehat{D}\setminus(r_0,+\infty)\times\partial D} (f_- - H)\,\mathrm{d}t \leq \delta.$$

We may assume further that f_+ has a local minimum point $p \in D_{\mathrm{c}} = \widehat{D} \setminus D_{\mathrm{b}}$, since f_+ is C^2–small there. It follows from Lemma 17 that $1_{f_+} = [p] \in \mathrm{HF}^*(f_+)$ is the image of the unit $e_D \in \mathrm{H}^*(D)$ under the isomorphism $\Phi_{f_+}\colon \mathrm{H}^*(D) \to \mathrm{HF}^*(f_+)$. Moreover, since f_+ and f_- have the same slope at infinity, Lemma 20 assures us that the isomorphism $[\Phi_{f_\bullet}]\colon \mathrm{HF}^*(f_+) \to \mathrm{HF}^*(f_-)$ induced by the continuation morphism $\Phi_{f_\bullet}\colon \mathrm{CF}^*(f_+)\to\mathrm{CF}^*(f_-)$ preserves the unit. To summarize, we have

$$\Phi_{f_+}(e_D)=[p]=1_{f_+}\quad\text{and}\quad[\Phi_{f_\bullet}(p)]=[\Phi_{f_\bullet}](1_{f_+})=1_{f_-}.$$

By the continuity of spectral invariants, we know that

$$c(1,H)-c(1,f_-)\geq\int_0^1\min_{x\in D^{r_0}}(f_--H)\,\mathrm{d}t.$$

Therefore, by our choice of f_-, we have $c(1,H)\geq-\delta+c(1,f_-)$. To complete the proof, it suffices to show that $c(1,f_-)\geq-k\delta$ for $k\geq 0$ independent of f_-. However, the definition of spectral invariants guarantees the existence of $q\in\mathrm{CF}^*(f_-)$ cohomologous to 1 for which $c(1,f_-)\geq\mathcal{A}_{f_-}(q)-\delta$. We thus only need to prove that $\mathcal{A}_{f_-}(q)\geq-\delta$. In the case where q is a combination $q_1+\cdots+q_k$ of orbits, the action of q is defined as

$$\mathcal{A}_{f_-}(q)=\max_i\mathcal{A}_{f_-}(q_i).$$

Recall from Section 5.2 that the barricade construction assures that we have, in terms of vector spaces, the decomposition

$$\mathrm{CF}^*(f_\pm)\cong\mathrm{C}^*(D_\mathrm{b},f_\pm)\oplus\mathrm{C}^*(D_\mathrm{c},f_\pm)$$

with inclusions and projections respectively given by

$$\iota_\heartsuit^\pm\colon\mathrm{C}^*(D_\heartsuit,f_\pm)\to\mathrm{CF}^*(f_\pm)\quad\text{and}\quad\pi_\heartsuit^\pm\colon\mathrm{CF}^*(f_\pm)\to\mathrm{C}^*(D_\heartsuit,f_\pm)$$

for $\heartsuit\in\{\mathrm{b},\mathrm{c}\}$. Moreover, Floer trajectories starting in D_b must have ends in D_b and Floer trajectories starting in D_c can have ends in D_b and D_c. Thus,

$$\Phi_{f_\bullet}(p)=p_\mathrm{b}+p_\mathrm{c}\quad\text{and}\quad q=p_\mathrm{b}+p_\mathrm{c}+\partial(r_\mathrm{b}+r_\mathrm{c})$$

for $p_\mathrm{b},r_\mathrm{b}\in\operatorname{im}i_\mathrm{b}^-$ and $p_\mathrm{c},r_\mathrm{c}\in\operatorname{im}\iota_\mathrm{c}^-$. Furthermore,

$$\partial(r_\mathrm{b})=r_\mathrm{bb}\quad\text{and}\quad\partial(r_\mathrm{c})=r_\mathrm{cb}+r_\mathrm{cc},$$

where $r_\mathrm{bb},r_\mathrm{cb}\in\operatorname{im}\iota_\mathrm{b}^-$ and $r_\mathrm{cc}\in\operatorname{im}\iota_\mathrm{c}^-$. See Figure 13 for an illustration of the Floer trajectories under consideration here.

Notice that since f_- is C^2–small on D_c, we have $\mathcal{A}_{f_-}(p_\mathrm{c}+r_\mathrm{cc})\geq-\delta$. Thus, if $p_\mathrm{c}+r_\mathrm{cc}\neq 0$, we have

$$\mathcal{A}_{f_-}(q)=\mathcal{A}_{f_-}(p_\mathrm{b}+p_\mathrm{c}+r_\mathrm{bb}+r_\mathrm{cb}+r_\mathrm{cc})\geq\mathcal{A}_{f_-}(p_\mathrm{c}+r_\mathrm{cc})\geq-\delta.$$

We now prove that $p_\mathrm{c}+r_\mathrm{cc}\neq 0$. This is equivalent to showing that the class $[\pi_\mathrm{c}^-(p_\mathrm{c})]$ in $\mathrm{H}^*(D_\mathrm{c},f_-)$ is nonzero. Indeed, if $p_\mathrm{c}+r_\mathrm{cc}=0$, we have, by the definition of r_cc, $p_\mathrm{c}=-\partial r_\mathrm{c}$ and thus

$$[\pi_\mathrm{c}^-(p_\mathrm{c})]=[\pi_\mathrm{c}^-(-\partial r_\mathrm{c})]=[-\partial_\mathrm{c}\pi_\mathrm{c}^-(r_\mathrm{c})]=0.$$

Denote by $\Phi_{\bar{f}_\bullet}\colon\mathrm{CF}^*(f_-)\to\mathrm{CF}^*(f_+)$ the continuation map generated by the inverse homotopy $\bar{f}_s=f_{-s}$. We know that both $\Phi_{\bar{f}_\bullet}\circ\Phi_{f_\bullet}$ and $\Phi_{f_\bullet}\circ\Phi_{\bar{f}_\bullet}$ are chain homotopic to the identity:

$$\Phi_{\bar{f}_\bullet}\circ\Phi_{f_\bullet}-\mathrm{id}_+=\partial_+\circ\Psi_+-\Psi_+\circ\partial_+,$$
$$\Phi_{f_\bullet}\circ\Phi_{\bar{f}_\bullet}-\mathrm{id}_-=\partial_-\circ\Psi_--\Psi_-\circ\partial_-$$

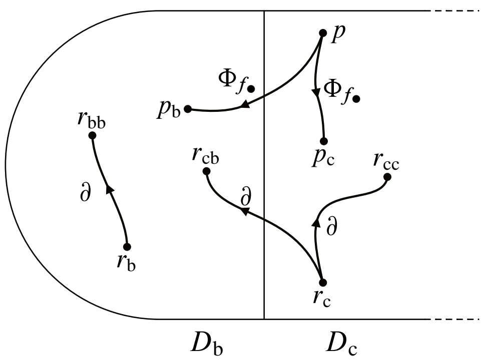


Figure 13: The possible trajectories for the differential of r_{b}, r_{c} and the continuation map applied to p according to the rules of the barricade.

for the differentials $\partial_\pm\colon \mathrm{CF}^*(f_\pm) \to \mathrm{CF}^{*+1}(f_\pm)$ and chain homotopies $\Psi_\pm\colon \mathrm{CF}^*(f_\pm) \to \mathrm{CF}^{*-1}(f_\pm)$. (In fact, for our purpose here, we only need the first homotopy relation.) Since $\Psi_\pm$ also obey the rules of the barricade by Lemma 37, the composition of the projections $\Phi^{\mathrm{c}}_{f_\bullet}\colon \mathrm{C}^*(D_{\mathrm{c}}, f_+) \to \mathrm{C}^*(D_{\mathrm{c}}, f_-)$ and $\Phi^{\mathrm{c}}_{\bar{f}_\bullet}\colon \mathrm{C}^*(D_{\mathrm{c}}, f_-) \to \mathrm{C}^*(D_{\mathrm{c}}, f_+)$ is chain homotopic to the identity on $\mathrm{C}^*(D_{\mathrm{c}}, f_+)$ by Lemma 39. Therefore, on cohomology, the morphism

$$[\Phi^{\mathrm{c}}_{\bar{f}_\bullet} \circ \Phi^{c}_{f_\bullet}]\colon \mathrm{H}^*(D_{\mathrm{c}}, f_+) \to \mathrm{H}^*(D_{\mathrm{c}}, f_+)$$

is given by the identity. Moreover, recall that by definition, $p \in D_{\mathrm{c}}$, which guarantees that, as a cycle, $p \in \operatorname{im} \iota^+_{\mathrm{c}}$ and since $[p] = 1_{f_+}$, we have $[\pi^+_{\mathrm{c}}(p)] \neq 0$. Therefore,

$$[\pi^-_{\mathrm{c}}(p_{\mathrm{c}})] = [\Phi^{\mathrm{c}}_{f_\bullet} \circ \pi^+_{\mathrm{c}}(p)] = [\Phi^{\mathrm{c}}_{f_\bullet}]([\pi^+_{\mathrm{c}}(p)]) \neq 0.$$

This concludes the proof.

6.3 Proof of Lemma C

Let $0 < \delta < 1$ be small enough so that

$$\delta A < \delta A + \delta \eta_A < \eta_A.$$

Then, following the proof of Theorem A1 with $\varepsilon = \delta(A + \eta_A)$, we have that

$$c(1, H_{\delta,A}) \geq A - \delta(A + \eta_A).$$

Notice that $H_{\delta,A}$ converges uniformly as $\delta \to 0$ to the continuous function $H_{0,A}$ (see Figure 14). Then, by continuity of spectral invariants and the previous equation, we have

$$c(1, H_{0,A}) = \lim_{\delta\to 0} c(1, H_{\delta,A}) \geq \lim_{\delta\to 0} (A - \delta(A + \eta_A)) = A.$$

Moreover, since $H_{0,A} \geq -A$, continuity of spectral invariants yields

$$c(1, H_{0,A}) \leq \max_{x\in D} -H_{0,A} = A,$$

which allows us to conclude that $c(1, H_{0,A}) = A$.

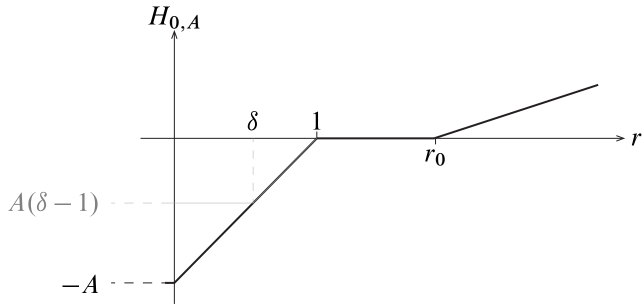


Figure 14: The continuous Hamiltonian $H_{0,A}$.

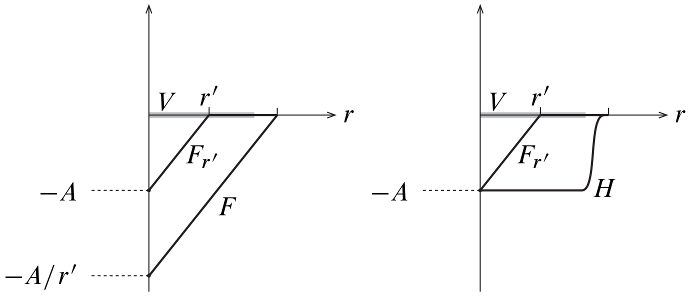


Figure 15: The Hamiltonians F, $F_{r'}$ and H.

First, we prove the lemma for Hamiltonians which are constant on an open neighborhood of the skeleton of D. Consider an autonomous Hamiltonian $H \in \mathcal{C}(D)$ such that $H|_V = -A$ and $-A \le H \le 0$ for an open neighborhood V of $\mathrm{Sk}(D)$ and a constant $A > 0$. The last condition on H allows us to use continuity of spectral invariance to conclude that

$$c(1, H) \le A. \tag{8}$$

All we need to do now is prove that A bounds $c(1, H)$ from below.

Define $F \in \mathcal{C}(D)$ to be the continuous autonomous Hamiltonian that agrees with $H_{0,A/r'}$ on D for some $0 < r' < 1$. Since $H|_V = -A$, we can choose r' so that the r'–contraction $F_{r'}$ of F under the Liouville flow (see (5) and Figure 15) has support in V and $-A \le F_{r'} \le 0$. Therefore,

$$F_{r'}(x) \ge H(x) \quad \text{for all } x \in D. \tag{9}$$

From the contraction principle stated in Lemma 27 and the computation of $c(1, H_{0,A})$ above, we have

$$c(1, F_{r'}) = r'c(1, F) = r'c(1, H_{0,A/r'}) = A.$$

This computation and (9) yield, by virtue of the monotonicity of spectral invariants, the lower bound $A = c(1, F_{r'}) \le c(1, H)$ as desired. In conjunction with (8), we conclude that $c(1, H) = A$.

Now, we prove the lemma in general. Suppose $H|_{\mathrm{Sk}(D)} = -A$ and $-A \le H \le 0$. For any $\varepsilon \in (0,1)$, there exists a compactly supported Hamiltonian H_ε such that $H_\varepsilon|_{V_\varepsilon} = -A$ for an open neighborhood V_ε of $\mathrm{Sk}(D)$ and $H_\varepsilon \le H$ everywhere. Indeed, define H_ε as follows: $H_\varepsilon|_{\mathrm{Sk}(D)} = -A$,

$$H_\varepsilon|_{D^\varepsilon \setminus \mathrm{Sk}(D)} = \beta_\varepsilon(r)H + (1-\beta_\varepsilon(r))(-A),$$

where $\beta_\varepsilon\colon (0,1) \to \mathbb{R}$ is such that

- $\beta_\varepsilon|_{(0,\varepsilon]} \equiv 0$,
- $\beta'_\varepsilon|_{(\varepsilon,2\varepsilon/3)} > 0$,
- $\beta_\varepsilon|_{(2\varepsilon/3,1)} \equiv 1$.

Then, H_ε satisfies the required conditions and converges uniformly to H as $\varepsilon \to 0$. We have $c(1, H_\varepsilon) = A$ by the previous computation, and by continuity of spectral invariants, we can conclude that

$$c(1, H) = c(1, H_\varepsilon) = A.$$

This completes the proof.

6.4 Proof of Theorem A2

Let $H \in \mathcal{C}(D)$ be an autonomous Hamiltonian such that $H|_V = -1$ and $-1 \le H \le 0$ everywhere for an open neighborhood V of $\mathrm{Sk}(D)$.

Define $\iota\colon \mathbb{R} \to \mathrm{Ham}_c(D)$ as

$$\iota(s) = \varphi_{sH},$$

where $\varphi_{sH} \in \mathrm{Ham}_c(D)$ is the time-1 map associated to sH. We claim that ι is the desired embedding.

We first bound $d_\gamma(\iota(s), \iota(s'))$ from above. If $F \in \mathcal{C}(D)$, then $\gamma(\varphi_F) \le \|F\|$. Moreover, since H is autonomous, $sH \,\#\, \overline{s'H} = (s-s')H$. Therefore,

$$d_\gamma(\iota(s), \iota(s')) = \gamma(\iota(s)\iota(s')^{-1}) \le \|(s-s')H\| = |s-s'|.$$

Now, we bound $d_\gamma(\iota(s), \iota(s'))$ from below. Since d_γ is symmetric, we can assume that $s \ge s'$. Then, by Lemmas B and C, we have

$$d_\gamma(\iota(s), \iota(s')) \ge c(1, (s-s')H) = s - s',$$

which completes the proof.

References

[Audin and Damian 2014] **M Audin**, **M Damian**, *Morse theory and Floer homology*, Springer (2014) MR Zbl

[Benedetti and Kang 2022] **G Benedetti**, **J Kang**, *Relative Hofer–Zehnder capacity and positive symplectic homology*, J. Fixed Point Theory Appl. 24 (2022) art. id. 44 MR Zbl

[Buhovsky et al. 2021] **L Buhovsky**, **V Humilière**, **S Seyfaddini**, *The action spectrum and C^0 symplectic topology*, Math. Ann. 380 (2021) 293–316 MR Zbl

[Cieliebak and Eliashberg 2012] **K Cieliebak**, **Y Eliashberg**, *From Stein to Weinstein and back: symplectic geometry of affine complex manifolds*, Amer. Math. Soc. Colloq. Publ. 59, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[Cieliebak and Oancea 2018] **K Cieliebak**, **A Oancea**, *Symplectic homology and the Eilenberg–Steenrod axioms*, Algebr. Geom. Topol. 18 (2018) 1953–2130 MR Zbl

[Cieliebak et al. 1995] **K Cieliebak**, **A Floer**, **H Hofer**, *Symplectic homology, II: A general construction*, Math. Z. 218 (1995) 103–122 MR Zbl

[Cieliebak et al. 1996] **K Cieliebak**, **A Floer**, **H Hofer**, **K Wysocki**, *Applications of symplectic homology, II: Stability of the action spectrum*, Math. Z. 223 (1996) 27–45 MR Zbl

[Cieliebak et al. 2010] **K Cieliebak**, **U Frauenfelder**, **A Oancea**, *Rabinowitz Floer homology and symplectic homology*, Ann. Sci. École Norm. Sup. 43 (2010) 957–1015 MR Zbl

[Entov and Polterovich 2003] **M Entov**, **L Polterovich**, *Calabi quasimorphism and quantum homology*, Int. Math. Res. Not. 2003 (2003) 1635–1676 MR Zbl

[Floer and Hofer 1994] **A Floer**, **H Hofer**, *Symplectic homology, I: Open sets in $\mathbb{C}^n$*, Math. Z. 215 (1994) 37–88 MR Zbl

[Frauenfelder and Schlenk 2007] **U Frauenfelder**, **F Schlenk**, *Hamiltonian dynamics on convex symplectic manifolds*, Israel J. Math. 159 (2007) 1–56 MR Zbl

[Ganor and Tanny 2023] **Y Ganor**, **S Tanny**, *Floer theory of disjointly supported Hamiltonians on symplectically aspherical manifolds*, Algebr. Geom. Topol. 23 (2023) 645–732 MR Zbl

[Giroux 1991] **E Giroux**, *Convexité en topologie de contact*, Comment. Math. Helv. 66 (1991) 637–677 MR Zbl

[Giroux 2017] **E Giroux**, *Remarks on Donaldson's symplectic submanifolds*, Pure Appl. Math. Q. 13 (2017) 369–388 MR Zbl

[Hofer 1990] **H Hofer**, *On the topological properties of symplectic maps*, Proc. Roy. Soc. Edinburgh Sect. A 115 (1990) 25–38 MR Zbl

[Humilière et al. 2016] **V Humilière**, **F Le Roux**, **S Seyfaddini**, *Towards a dynamical interpretation of Hamiltonian spectral invariants on surfaces*, Geom. Topol. 20 (2016) 2253–2334 MR Zbl

[Kawamoto 2022a] **Y Kawamoto**, *Homogeneous quasimorphisms, C^0–topology and Lagrangian intersection*, Comment. Math. Helv. 97 (2022) 209–254 MR Zbl

[Kawamoto 2022b] **Y Kawamoto**, *On C^0–continuity of the spectral norm for symplectically non-aspherical manifolds*, Int. Math. Res. Not. 2022 (2022) 17187–17230 MR Zbl

[Kislev and Shelukhin 2021] **A Kislev**, **E Shelukhin**, *Bounds on spectral norms and barcodes*, Geom. Topol. 25 (2021) 3257–3350 MR Zbl

[Lalonde and McDuff 1995] **F Lalonde**, **D McDuff**, *The geometry of symplectic energy*, Ann. of Math. 141 (1995) 349–371 MR Zbl

[Le Roux 2010] **F Le Roux**, *Six questions, a proposition and two pictures on Hofer distance for Hamiltonian diffeomorphisms on surfaces*, from "Symplectic topology and measure preserving dynamical systems", Contemp. Math. 512, Amer. Math. Soc., Providence, RI (2010) 33–40 MR Zbl

[Monzner et al. 2012] **A Monzner**, **N Vichery**, **F Zapolsky**, *Partial quasimorphisms and quasistates on cotangent bundles, and symplectic homogenization*, J. Mod. Dyn. 6 (2012) 205–249 MR Zbl

[Oh 2005] **Y-G Oh**, *Construction of spectral invariants of Hamiltonian paths on closed symplectic manifolds*, from "The breadth of symplectic and Poisson geometry", Progr. Math. 232, Birkhäuser, Boston, MA (2005) 525–570 MR Zbl

[Polterovich 2001] **L Polterovich**, *The geometry of the group of symplectic diffeomorphisms*, Birkhäuser, Basel (2001) MR Zbl

[Polterovich 2014] **L Polterovich**, *Symplectic geometry of quantum noise*, Comm. Math. Phys. 327 (2014) 481–519 MR Zbl

[Ritter 2009] **A F Ritter**, *Novikov-symplectic cohomology and exact Lagrangian embeddings*, Geom. Topol. 13 (2009) 943–978 MR Zbl

[Ritter 2013] **A F Ritter**, *Topological quantum field theory structure on symplectic cohomology*, J. Topol. 6 (2013) 391–489 MR Zbl

[Salamon and Zehnder 1992] **D Salamon**, **E Zehnder**, *Morse theory for periodic solutions of Hamiltonian systems and the Maslov index*, Comm. Pure Appl. Math. 45 (1992) 1303–1360 MR Zbl

[Schwarz 2000] **M Schwarz**, *On the action spectrum for closed symplectically aspherical manifolds*, Pacific J. Math. 193 (2000) 419–461 MR Zbl

[Seidel 2008] **P Seidel**, *A biased view of symplectic cohomology*, from "Current developments in mathematics, 2006", International, Somerville, MA (2008) 211–253 MR Zbl

[Usher 2013] **M Usher**, *Hofer's metrics and boundary depth*, Ann. Sci. École Norm. Sup. 46 (2013) 57–128 MR Zbl

[Viterbo 1992] **C Viterbo**, *Symplectic topology as the geometry of generating functions*, Math. Ann. 292 (1992) 685–710 MR Zbl

[Viterbo 1999] **C Viterbo**, *Functors and computations in Floer homology with applications, I*, Geom. Funct. Anal. 9 (1999) 985–1033 MR Zbl

[Weber 2006] **J Weber**, *Noncontractible periodic orbits in cotangent bundles and Floer homology*, Duke Math. J. 133 (2006) 527–568 MR Zbl

[Weinstein 1979] **A Weinstein**, *On the hypotheses of Rabinowitz' periodic orbit theorems*, J. Differential Equations 33 (1979) 353–358 MR Zbl

Département de mathématiques et de statistique, Université de Montréal
Montréal, QC, Canada

pierre-alexandre.mailhot@umontreal.ca

Received: 2 August 2022 Revised: 8 April 2023

Algebraic & Geometric Topology **24**:7 (2024) 3801–3825
DOI: 10.2140/agt.2024.24.3801
Published: 9 December 2024

Classifying rational G–spectra for profinite G

DAVID BARNES
DANNY SUGRUE

For G an arbitrary profinite group, we construct an algebraic model for rational G–spectra in terms of G–equivariant sheaves over the space of subgroups of G. This generalises the known case of finite groups to a much wider class of topological groups. It improves upon earlier work of the first author on the case where G is the p–adic integers.

As the purpose of an algebraic model is to allow one to use homological algebra to study questions of homotopy theory, we prove that the homological dimension (injective dimension) of the algebraic model is determined by the Cantor–Bendixson rank of the space of closed subgroups of the profinite group G. This also provides a calculation of the homological dimension of the category of rational Mackey functors.

55P91; 54B40, 55P42, 55Q91

1 Introduction

The usefulness of equivariant cohomology theories in equivariant (stable) homotopy theory has long been proven. Examples of equivariant cohomology theories include the equivariant K–theory of Segal [1968] and the equivariant cobordism spectra used in Hill, Hopkins and Ravenel, [Hill et al. 2016]. To effectively study equivariant cohomology theories, one studies the category of their representing objects: equivariant spectra. That is, Brown representability holds equivariantly.

The study of equivariant spectra up to homotopy is even more demanding than the nonequivariant case, so one often works with rational equivariant spectra. Under Brown representability, rational equivariant spectra correspond to equivariant cohomology theories that take values in rational vector spaces. Rationalising preserves most of the interesting behaviour coming from the group, while removing much of the topological complexity.

A major goal in the study of rational equivariant stable homotopy theory is to find a more tractable model category that has the same homotopy theory as rational equivariant spectra. That is, one chooses a group G of interest, constructs an abelian category $\mathcal{A}(G)$ and a Quillen equivalence between the *algebraic model* $\mathrm{Ch}(\mathcal{A}(G))$ and the model category of rational G–spectra. The Quillen equivalence induces an equivalence of categories between the homotopy category of rational G–spectra and the homotopy category of the algebraic model. The primary advantage of having an algebraic model is that one can use the simplicity of the abelian category $\mathcal{A}(G)$ and the tools of homological algebra to construct objects and calculate sets of maps in the rational G–equivariant stable homotopy category. For an introduction to algebraic models and summary of the known cases see [Barnes and Kędziorek 2022].

In this paper, the authors generalise the known case of algebraic models for finite groups (see [Schwede and Shipley 2003, Example 5.1.2]) to profinite groups. Profinite groups are a commonly encountered class of compact topological groups, appearing most often as Galois groups or when one has a diagram of finite groups. They are defined as the compact Hausdorff totally disconnected topological groups. It can be shown that a group is profinite if and only if it is the limit of a filtered system of finite groups. For example, the Morava stabiliser group $\mathbb{S}_n$ from chromatic homotopy theory is profinite. Profinite groups occur in many other mathematical fields: number theory makes substantial use of profinite groups, as seen in [Bley and Boltje 2004], and the étale fundamental groups of algebraic geometry are profinite. This ubiquity drives our interest in rational G–spectra for profinite G and hence our interest in finding algebraic models in the profinite case.

1.1 Main results

Let G be a profinite group and let $\mathcal{S}G$ be the space of closed subgroups of G, topologised as the inverse limit of the finite discrete spaces $\mathcal{S}G/N$, for N open and normal in G. There is an abelian category of rational G–equivariant sheaves over $\mathcal{S}G$. Consider the full subcategory of those equivariant sheaves E such that the stalk E_K over the closed subgroup K is K–fixed. These are called rational Weyl-G–sheaves and are introduced in earlier work of the authors, [Barnes and Sugrue 2023; 2022]. In this paper, the category of rational Weyl-G–sheaves occur as the abelian category $\mathcal{A}(G)$ used to make the algebraic model $\mathrm{Ch}(\mathcal{A}(G))$ for rational G–spectra.

Theorem A (Corollary 4.10) *For G a profinite group, there is a zigzag of Quillen equivalences between the model category of rational G–spectra and the model category of chain complexes of rational Weyl-G–sheaves.*

This result generalises work of the first author [Barnes 2011] on the case where $G = \mathbb{Z}_p^\wedge$ is the p–adic integers. As with the current work, that paper uses the tilting theory (Morita theory) of Schwede and Shipley [2003, Theorem 5.1] to obtain a Quillen equivalence between rational G–spectra and a category of chain complexes in an abelian category. The p–adic case uses a hand-crafted abelian category designed for that specific group. Contrastingly, the current work uses rational G–Mackey functors in the tilting theory step, then applies the equivalence between rational G–Mackey functors and rational Weyl-G–sheaves of [Barnes and Sugrue 2023, Theorem A] to obtain the final result.

There are two reasons why the sheaf description of the algebraic model is important. The first is that Greenlees' conjecture [2006] on algebraic models (for compact Lie groups) is described in terms of sheaves over $\mathcal{S}G$, the space of closed subgroups of G. This result is the first realisation of that conjecture, albeit for profinite groups.

The second is that using the sheaf description we can calculate the homological dimension (also called the injective dimension) of the abelian category. This dimension is a measure of the complexity of the abelian model. In this case, the result is phrased in terms of the Cantor–Bendixson rank of the space of subgroups $\mathcal{S}G$; see Section 5 for details. When G is profinite, $\mathcal{S}G$ is a profinite space (one that is compact, Hausdorff and totally disconnected). The Cantor–Bendixson rank of a profinite space can be thought of as a measure of how far the space is from being discrete. To illustrate, a discrete space has rank 1 and $\mathcal{S}\mathbb{Z}_p^\wedge$, consisting of countably many points with one accumulation point, has rank 2.

Theorem B (Corollary 5.16) *Let G be a profinite group whose space of subgroups $\mathcal{S}G$ is scattered of Cantor–Bendixson rank n. The homological dimension of the category of rational Weyl-G–sheaves is $n-1$.*

If $\mathcal{S}G$ has infinite Cantor–Bendixson rank, then the homological dimension of the category of rational Weyl-G–sheaves is infinite.

Using the equivalence between rational G–Mackey functors and rational Weyl-G–sheaves, [Barnes and Sugrue 2023, Theorem A], this result gives the homological dimension of categories of rational G–Mackey functors.

Corollary C *Let G be a profinite group whose space of subgroups $\mathcal{S}G$ is scattered of Cantor–Bendixson rank n. The homological dimension of the category of rational G–Mackey functors is $n-1$.*

If $\mathcal{S}G$ has infinite Cantor–Bendixson rank, then the homological dimension of the category of rational G–Mackey functors is infinite.

As well as the two cases in the previous results there is a third possibility, that $\mathcal{S}G$ is of finite rank, but not scattered. We make the following conjecture for this case, which occurs as Conjecture 5.17 in the main body.

Conjecture D *Let G be a profinite group. If the G–space X has finite Cantor–Bendixson rank and nonempty perfect hull, then the homological dimension of rational G–sheaves over X is infinite.*

If $X=\mathcal{S}G$, then the homological dimension of the category of rational Weyl-G–sheaves is infinite.

1.2 Future questions

The question of how the change of groups functors on spectra compare with functors relating algebraic models for varying G is surprisingly involved. The currently known cases are for (co)free equivariant spectra; see [Williamson 2022]. For the case of (pro)finite groups, the authors expect that having the sheaf description and the Mackey functor description will be vital.

The Quillen equivalences as given are not monoidal. There are two sources of difficulty, firstly that the Quillen equivalences of the tilting theory of Schwede and Shipley are not monoidal. Resolving this would require a fundamentally different approach to the classification. Secondly, it is not known how the equivalence of [Barnes and Sugrue 2023] interacts with the tensor product of sheaves and the two known monoidal structures on G–Mackey functors: the box product and the equivariant tensor product of Hill and Mazur [2019].

A further question is whether one can construct an Adams spectral sequence which takes values in the abelian category. The difficulty here is expected to be around constructing a suitable set of geometric fixed point functors for the closed subgroups of G, which should be the topological equivalent of taking the stalk of an equivariant sheaf and hence detect equivalences in the homotopy category of rational G–spectra. The geometric fixed point functors and the Adams spectral sequence are needed to give a good set of examples of rational G–spectra and their image in the algebraic model.

1.3 Strategy of the classification

The following diagram gives the major steps in the classification of rational G–spectra, for profinite G. The tilting theorem (Morita theory) of Schwede and Shipley [2003, Theorem 5.1] is used in Section 4.2 to create a Quillen equivalence between rational G–spectra and a category of (chain complexes of) "topological Mackey functors". This gives the upper horizontal functor. The key input to apply the tilting theorem is Theorem 2.14, which proves that the homotopical information of rational G–spectra is concentrated in degree zero.

In Section 3 we study spans and the stable orbit category. The aim is to prove Theorem 3.11, which gives an equivalence between $\pi_0(\mathfrak{O}_G)$, the G–equivariant stable orbit category and $\mathrm{Span}(G_{df}\text{–sets})$, a category of spans of finite discrete G–sets. That equivalence provides the upper vertical functor of the diagram. The lower vertical functor is an equivalence of categories describing Mackey functors in terms of spans, as detailed in Section 4.1. Theorem 4.4 is where these results are combined to give the zigzag of Quillen equivalences between rational G–spectra and chain complexes of rational G–Mackey functors.

The lower horizontal functor is an equivalence by earlier work of the authors, [Barnes and Sugrue 2023, Theorem A], which proves that the category of rational G–Mackey functors is equivalent to the category of rational Weyl-G–sheaves over the space of closed subgroups of G.

$$
\begin{array}{ccc}
\text{rational } G\text{–spectra} & \overset{\simeq}{\longleftrightarrow} & \mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathfrak{O}_G^{\mathbb{Q}}), \mathbb{Q}\text{–mod})\big) \\
 & & \big\updownarrow \cong \\
 & & \mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\mathrm{Span}(G_{df}\text{–sets}), \mathbb{Q}\text{–mod})\big) \\
 & & \big\updownarrow \cong \\
\mathrm{Ch}(\text{Weyl-}G\text{–sheaf}_{\mathbb{Q}}(\mathcal{S}G)) & \overset{\cong}{\longleftrightarrow} & \mathrm{Ch}(\mathrm{Mackey}_{\mathbb{Q}}(G))
\end{array}
$$

In Section 5 we look at the homological dimension of the algebraic model and relate it to the Cantor–Bendixson rank of $\mathcal{S}G$.

Acknowledgements

Sugrue gratefully acknowledges support from the Engineering and Physical Sciences Research Council under Grant 1631308. No data were created or analysed in this study.

2 Basics on equivariant spectra for profinite G

We recap the construction of the model category of rational orthogonal G–spectra, for G a profinite group. We then give various properties of this homotopy theory that will be used in the classification. Along the way we will need some facts about profinite groups, profinite sets, and topological spaces with a profinite group action. The constructions will all be generalisations of the finite group case. The expert reader may like to skip to Section 3.

2.1 Profinite groups

We give a few reminders of useful facts on profinite groups. More details can be found in [Wilson 1998] or [Ribes and Zalesskii 2000].

A *profinite group* is a compact, Hausdorff, totally disconnected topological group. A profinite group G is homeomorphic to the inverse limit of its finite quotients:

$$G \cong \lim_{\substack{N \trianglelefteq G \\ \text{open}}} G/N \subseteq \prod_{\substack{N \trianglelefteq G \\ \text{open}}} G/N.$$

The limit has the canonical topology which can either be described as the subspace topology on the product or as the topology generated by the preimages of the open sets in G/N under the projection map $G \to G/N$, as N runs over the open normal subgroups of G.

Closed subgroups and quotients by closed subgroups of profinite groups are also profinite. A subgroup of a profinite group is open if and only if it is finite index and closed. The trivial subgroup $\{e\}$ is open if and only if the group is finite. The intersection of all open normal subgroups is $\{e\}$. Any open subgroup H contains an open normal subgroup, the *core* of H in G, which is defined as the finite intersection

$$\mathrm{Core}_G(H) = \bigcap_{g \in G} gHg^{-1}.$$

2.2 Equivariant orthogonal spectra

Recall that an orthogonal spectrum is a sequence of based spaces indexed by finite dimensional inner product spaces, related by suspension maps that are suitably compatible with linear isometries of those vector spaces. For details, see the work of Mandell, May Schwede and Shipley [Mandell et al. 2001] or work of Barnes and Roitzheim [2020].

Equivariantly, the picture is similar, starting from based spaces with G–action indexed by finite dimensional G–inner product spaces, with suspension maps that are compatible with both the linear isometries and G–actions. This construction was first given in work of Mandell and May [2002] and adapted to the profinite setting by Fausk [2008].

The starting point is to describe the model category of based topological G–spaces that we will use to create our G–spectra. We focus on those model structures built using the *open* subgroups H of G as then G/H is a finite set. This ensures that the forgetful functor from G–spaces to spaces is a left Quillen functor. Throughout this section G will be a profinite group.

Proposition 2.1 *There is a cofibrantly generated proper model structure on the category of based topological G–spaces with weak equivalences those maps f such that f^H is a weak equivalence of spaces, for all open subgroups H of G. Similarly, fibrations are those maps f such that f^H is a fibration of based spaces for all open subgroups H of G. This model structure is denoted $G\mathrm{Top}_*$.*

The generating cofibrations are the standard inclusions

$$G/H_+ \wedge S^{n-1}_+ \to G/H_+ \wedge D^n_+, \quad G/H_+ \wedge D^n_+ \to G/H_+ \wedge (D^n \times [0,1])_+$$

for H an open subgroup of G and $n \geqslant 0$.

Just as CW–complexes are built from iteratively attaching cells by taking the pushout over the inclusion $S^{n-1} \to D^n$, one can define G–CW–complexes using the inclusions

$$G/H \times S^{n-1} \to G/H \times D^n.$$

That is, take X_0 to be a disjoint union of copies of G/H_+, then attach cells of the above form with $n = 1$ and H varying to obtain X_1. Continuing inductively gives X_n and X is defined as the union of the X_n. We see that if X is built using finitely many cells, the stabiliser of X (the intersection of all open subgroups H used in the cells) is also open. The evident pointed analogue gives the definition of pointed G–CW–complexes. Since the spaces G/H are finite, we see that every G–CW–complex is indeed a CW–complex, after forgetting the group action. Note that our choices mean that G itself is not a G–CW–complex. Indeed, as it has no fixed points, the space G is weakly equivalent to the empty set.

Our category of G–spectra will be indexed by the finite dimensional sub-G–inner product spaces of a complete G–universe $\mathcal{U}$, as defined below.

Definition 2.2 A *G–universe* $\mathcal{U}$ is a countably infinite direct sum $\mathcal{U} = \bigoplus_{i=1}^{\infty} U$ of a real G–inner product space U, such that

(1) there is a canonical choice of trivial representation $\mathbb{R} \subset U$,

(2) $\mathcal{U}$ is topologised as the union of all finite dimensional G–subspaces of U (each equipped with the norm topology).

A G–universe is said to be *complete* if every finite dimensional irreducible representation is contained (up to isomorphism) within $\mathcal{U}$.

A complete G–universe always exists, one can be obtained by setting U to be direct sum of a representative of each isomorphism class of irreducible representations of G, then defining $\mathcal{U}$ to be the direct sum of countably many copies of U.

Remark 2.3 Since a profinite group G is compact and Hausdorff, the action of G on a finite dimensional G–inner product space factors through a Lie group quotient by [Fausk 2008, Lemma A.1]. The only such quotients of a profinite group are finite, hence if V is a finite dimensional G–inner product space, there is an open normal subgroup N of G such that $V^N = V$.

In particular, the one-point compactification of V, denoted S^V is fixed by some open normal subgroup N. Thus, S^V can be given the structure of a finite (pointed) G/N–CW–complex, by [Illman 1983], and hence is a finite (pointed) G–CW–complex.

For brevity we define equivariant orthogonal spectra in terms of enriched functors from a particular enriched category made using Thom spaces. Recall that $G\mathrm{Top}_*$ is enriched over itself, via the space of (not-necessarily equivariant) maps, where G acts by conjugation

$$(f : X \to Y) \mapsto (g_Y \circ f \circ g_X^{-1} : X \to Y).$$

Definition 2.4 Define an *indexing space* to be a finite dimensional sub-G–inner product space of $\mathcal{U}$. We define $\mathcal{L}$ to be the category of all real G–inner product spaces that are isomorphic to indexing G-spaces in $\mathcal{U}$ and morphisms the (not-necessarily equivariant) linear isometries.

Definition 2.5 For each $V \subseteq W$ there is a vector bundle (a subset of the product bundle)

$$\gamma(V, W) = \{(f, x) \subseteq \mathcal{L}(V, W) \times W \mid x \in W - f(V)\}$$

over $\mathcal{L}(V, W)$, where $W - f(V)$ is the orthogonal complement of $f(V)$ in W.

Let $\mathcal{J}(V, W)$ be the Thom space of $\gamma(V, W)$, with G–action given by $g(f, x) = (gfg^{-1}, gx)$.

Lemma 2.6 *For inclusions of indexing spaces $U \subseteq V \subseteq W$ the G–equivariant map*

$$\gamma(V, W) \times \gamma(U, V) \to \gamma(U, W), \quad ((f, x), (k, y)) \mapsto (f \circ k, x + f(y))$$

induces a composition for the $G\mathrm{Top}_*$–enriched category $\mathcal{J}$ whose objects are the objects of $\mathcal{L}$ and morphism G–spaces are given by $\mathcal{J}(V, W)$.

Definition 2.7 An *orthogonal G–spectrum* X on a universe $\mathcal{U}$ is an $G\mathrm{Top}_*$–enriched functor from $\mathcal{J}$ to $G\mathrm{Top}_*$. A map of orthogonal G–spectra is a $G\mathrm{Top}_*$–enriched natural transformation. The category of orthogonal G–spectra is denoted $G\mathrm{Sp}^O$.

In particular, an orthogonal G–spectrum X defines based G–spaces $X(V)$ for each indexing space $V \subset \mathcal{U}$ and based G–maps

$$\sigma_{V,W} : S^{W-V} \wedge X(V) \to X(W)$$

for each $V \subseteq W$. A map $f\colon X \to Y$ of orthogonal G–spectra defines a set of G–maps $f\colon X(V) \to Y(V)$, which commute with the maps $\sigma_{V,W}$ for each $V \subseteq W$.

2.3 Model categories of spectra

With the category of G–spectra defined, the next step is to give model structures. This follows the usual path, a levelwise model structure, then a stable model structure and finally a rational model structure (as a Bousfield localisation of the stable model structure). In each case, the weak equivalences are defined using the *open* subgroups of G. For brevity, we state the existence of the rational model structure as a theorem, giving the essential properties afterwards. These results are standard, details can be found in [Barnes 2008, Section 2.2].

Definition 2.8 Let X be a G–spectrum and n a nonnegative integer. We define the *H-homotopy groups* of X as

$$\pi_n^H(X) = \operatorname*{colim}_V \pi_n^H(\Omega^V X(V)), \quad \pi_{-n}^H(X) = \operatorname*{colim}_{V \supset \mathbb{R}^n} \pi_0^H(\Omega^{V-\mathbb{R}^n} X(V)).$$

In the first case the colimit runs over the indexing spaces of $\mathcal{U}$, in the second case over the indexing spaces of $\mathcal{U}$ that contain $\mathbb{R}^n$.

A map $f\colon X \to Y$ of orthogonal G–spectra is called a *π_*–isomorphism* if $\pi_k^H(f)$ is an isomorphism for all open subgroups H of G and all integers k. We call f a *rational π_*–isomorphism* if $\pi_k^H(f) \otimes \mathbb{Q}$ is an isomorphism for all open subgroups H of G and all integers k.

The rational model structure on G–spectra is made using the rational sphere spectrum $S^0\mathbb{Q}$; see [Barnes 2008, Definition 1.5.2] for a construction as an equivariant Moore spectrum for $\mathbb{Q}$.

Theorem 2.9 *There is a cofibrantly generated proper stable model structure on the category of orthogonal G–spectra whose weak equivalences are the class of rational π_*–isomorphisms and whose cofibrations are the class of q–cofibrations. This model structure is called the* ***rational model structure*** *and we write $G\mathrm{Sp}^O_{\mathbb{Q}}$ for the model category of orthogonal G–spectra equipped with the rational model structure.*

Proof The stable model structure on orthogonal G–spectra is cofibrantly generated, proper and stable. Thus [Barnes and Roitzheim 2020, Theorem 7.2.17] implies that any left Bousfield localisation of the stable model structure at a set of maps which are closed under desuspension is also cofibrantly generated, proper and stable.

We construct the rational model structure by localising the stable model structure at the set of all suspensions and desuspensions of the map from the sphere spectrum to the rational sphere spectrum

$$S^0 \to S^0\mathbb{Q}.$$

That this gives the correct weak equivalences is a consequence of Proposition 2.10. □

Proposition 2.10 *There is a natural isomorphism*

$$\pi_*^H(X \wedge S^0\mathbb{Q}) \cong \pi_*^H(X) \otimes \mathbb{Q}.$$

A G–spectrum X is fibrant in $G\mathrm{Sp}^O_{\mathbb{Q}}$ if and only if X is an Ω–spectrum and has rational homotopy groups.

The last statement gives the following zigzag of π_*–isomorphisms for any G–spectrum X, where $\hat{f}_{\mathbb{Q}}X$ is the fibrant replacement of X in $G\mathrm{Sp}^O_{\mathbb{Q}}$:

$$\hat{f}_{\mathbb{Q}}X \to \hat{f}_{\mathbb{Q}}X \wedge S^0\mathbb{Q} \leftarrow X \wedge S^0\mathbb{Q}.$$

This shows that our localisation is a smashing localisation.

Corollary 2.11 *If the G–spectrum A is compact in the homotopy category of $G\mathrm{Sp}^O$, then there is a natural isomorphism*

$$[A, X]_{\mathbb{Q}} \cong [A, X] \otimes \mathbb{Q}.$$

In particular, for $\hat{f}_{\mathbb{Q}}X$ the fibrant replacement of X in $G\mathrm{Sp}^O_{\mathbb{Q}}$, there is a natural isomorphism

$$\pi_*^H(\hat{f}_{\mathbb{Q}}X) \cong \pi_*^H(X) \otimes \mathbb{Q}.$$

Since the weak equivalences of the stable model structure are defined in terms of π_*–isomorphisms, the triangulated category $\mathrm{Ho}(G\mathrm{Sp}^O)$ has a set of compact generators: the suspension spectra $\Sigma^\infty G/H_+$, for H an open subgroup of G. Similarly, this set is also a set of compact generators for $\mathrm{Ho}(G\mathrm{Sp}^O_{\mathbb{Q}})$, as the weak equivalences of $G\mathrm{Sp}^O_{\mathbb{Q}}$ are defined in terms of rational π_*–isomorphisms.

Corollary 2.12 *For G a profinite group, the homotopy category of $G\mathrm{Sp}^O_{\mathbb{Q}}$ is generated by the set of compact objects $\Sigma^\infty G/H_+$, for H an open subgroup of G.*

This completes the construction of the model category we wish to model with algebra. Our next task is to study maps between objects like $\Sigma^\infty G/H_+$. We use the profinite version of tom Dieck splitting to show these maps are concentrated in degree zero.

Proposition 2.13 [Fausk 2008, Proposition 7.10] *For G a profinite group and X a based G–space, there is an isomorphism of abelian groups*

$$\bigoplus_{(H) \underset{\text{open}}{\leqslant} G} \pi_*(\Sigma^\infty EW_G H_+ \wedge_{W_G H} X^H) \to \pi_*^G(\Sigma^\infty X).$$

The sum runs over the conjugacy classes of open subgroups of G.

Theorem 2.14 [Barnes 2011, Theorem 2.9] *For G a profinite group, the graded $\mathbb{Q}$–module*

$$[\Sigma^\infty G/H_+, \Sigma^\infty G/K_+]_*^G \otimes \mathbb{Q}$$

is concentrated in degree zero.

3 Spans and the stable orbit category

The aim for this section is Theorem 3.11, which provides a combinatorial description of the stable orbit category for G. That is, the full triangulated subcategory of $\mathrm{Ho}(G\mathrm{Sp}^{O})$ (the homotopy category of G–spectra) defined by the suspension spectra of finite pointed G–sets is shown to be equivalent to the category of spans of finite G–sets. This result is well-known in the case of finite groups but is seemingly new in the case of profinite groups. The method of proof is to relate the profinite case to the finite case by describing maps in $\mathrm{Ho}(G\mathrm{Sp}^{O})$ as a colimit of maps in $\mathrm{Ho}((G/N)\mathrm{Sp}^{O})$, as N varies over the open normal subgroups of G, see Lemma 3.9.

3.1 The Burnside category

In the case of a finite group G, the Burnside category is the category of G–sets with (equivalence classes) of spans as morphisms. In this subsection we generalise this construction to profinite groups and show how it relates to the finite group case.

Definition 3.1 A set X with an action of G is said to be *discrete* if the canonical map

$$\operatorname*{colim}_{\substack{H \leqslant G \\ \text{open}}} X^H \to X$$

is an isomorphism. The category of finite discrete G–sets and equivariant maps is denoted G_{df}–sets.

Equally, one can define a discrete G–set as a G–set such that the stabiliser of each point is open. We then see that a G–set X is discrete if and only if the action on G is continuous, when X is equipped with the discrete topology. If X is a finite discrete G–set, then the stabiliser of each point is open, as is the intersection of all the stabilisers. Thus a finite G–set A is discrete if and only if there is an open subgroup H of G such that A is H–fixed.

The class of discrete G–sets is closed under arbitrary coproducts and finite products. The set of finite coproducts of G–sets of the form G/H, for H an open subgroup of G, is a skeleton for the class of finite discrete G–sets. We also note that the empty set is a finite discrete G–set.

Definition 3.2 The *Burnside ring* of G, written as $\mathsf{A}(G)$ is the Grothendieck ring of finite discrete G–sets. We further define the *rational Burnside ring* of G as $\mathsf{A}_{\mathbb{Q}}(G) = \mathsf{A}(G) \otimes \mathbb{Q}$.

Lemma 3.3 *For G a profinite group there is a natural isomorphism*

$$\varepsilon^*\colon \operatorname*{colim}_{\substack{N \trianglelefteqslant G \\ \text{open}}} \mathsf{A}(G/N) \to \mathsf{A}(G).$$

Proof A finite G/N–set A can be considered as a G–set via inflation, $\varepsilon^* A$. Since each point of $\varepsilon^* A$ is fixed by N, this is a finite discrete G–set. Inflation gives an injective ring map

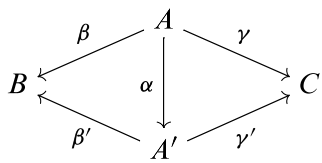
$$\varepsilon_N^*\colon \mathsf{A}(G/N) \to \mathsf{A}(G).$$

The maps ε_N^* are compatible with the maps forming the colimit (as these are also inflation maps), giving the map ε^*. As the colimit is filtered, ε^* is also injective.

Any finite discrete G–set A is fixed by some open subgroup H, which must contain an open normal subgroup N. Thus A is in the image of ε_N^* and so ε^* is surjective. □

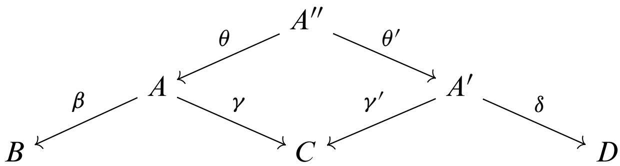
We can generalise the construction of the Burnside ring to make a category.

Definition 3.4 Let G be a profinite group. A *span* of finite discrete G–sets is a pair of equivariant maps $B \xleftarrow{\beta} A \xrightarrow{\gamma} C$, sometimes shortened to (β, γ). Two spans $B \xleftarrow{\beta} A \xrightarrow{\gamma} C$ and $B \xleftarrow{\beta'} A' \xrightarrow{\gamma'} C$ are *equivalent* if there is an equivariant isomorphism $A \to A'$ such that the following diagram commutes.



We write $[\beta, \gamma]$ for the equivalence class of (β, γ).

We recall the notion of composition of spans. Take two spans $B \xleftarrow{\beta} A \xrightarrow{\gamma} C$ and $C \xleftarrow{\gamma'} A' \xrightarrow{\delta} D$, then construct A'' as the pullback of γ and γ'



The composite of (β, γ) and (γ', δ) is the span $(\beta \circ \theta, \delta \circ \theta')$. This composition is well-defined under equivalence of spans.

Further, there is an addition on (equivalence classes of) spans with the same codomains. Consider two spans $B \xleftarrow{\beta} A \xrightarrow{\gamma} C$ and $B \xleftarrow{\beta'} A' \xrightarrow{\gamma'} C$. Their sum is the span

$$B \xleftarrow{\beta,\beta'} A \coprod A' \xrightarrow{\gamma,\gamma'} C.$$

This addition rule is associative, commutative, compatible with equivalence of spans and the unit is the span $B \leftarrow \varnothing \to C$.

Definition 3.5 The *Burnside category* of G is the category with objects the finite discrete G–sets and morphisms given by Grothendieck construction on sets of equivalence classes of spans of finite discrete G–sets, denoted $\mathrm{Span}(G_{df}\text{–sets})$.

Thus, a map $A \to B$ in the Burnside category is a formal difference of equivalence classes of spans. We also see that the Burnside ring of G is the ring $\mathrm{Span}(G_{df}\text{–sets})(*, *)$.

Just as for the Burnside ring, we can relate the Burnside category for G to the Burnside categories for the groups G/N, where N is an open normal subgroup of G.

Lemma 3.6 *If A and B are finite discrete G–sets, for G a profinite group, then there is an isomorphism*

$$\operatorname*{colim}_{\substack{N \trianglelefteq G \\ \text{open}}} \mathrm{Span}(G/N_{df}\text{–sets})(A, B) \cong \mathrm{Span}(G_{df}\text{–sets})(A, B).$$

Proof The maps of the colimit are given by inflation functors and inflation from G/N to G induces the map to the codomain. For the inverse, take a span

$$B \leftarrow A \rightarrow C,$$

and choose an open normal subgroup N that fixes each of element of A, B and C. Then this span appears in term N of the colimit. □

3.2 The stable orbit category

In this subsection we study maps in the G–equivariant stable homotopy category between spectra of the form $\Sigma^\infty A_+$, where A is a finite discrete G–set. Our aim is to relate this to unstable homotopy classes of maps of G/N–spaces, where N runs over the open normal subgroups of G. We start by comparing maps in the G–equivariant stable homotopy category to the G–equivariant unstable homotopy category.

We need two categories, one defined via the G–equivariant stable category and one via the rational analogue. It is important to note that these categories are not graded, we use π_0 in the notation of the categories as a reminder of this fact. The last sentence of the definition holds due to Corollary 2.11.

Definition 3.7 We define a category $\pi_0(\mathfrak{O}_G)$, called the *G–equivariant stable orbit category*. The objects are the class of G–spectra of the form $\Sigma^\infty A_+$, for A a finite discrete G–set. The morphisms are given by

$$\pi_0(\mathfrak{O}_G)(\Sigma^\infty A_+, \Sigma^\infty B_+) = [\Sigma^\infty A_+, \Sigma^\infty B_+]^G,$$

the set of maps in the homotopy category of $G\mathrm{Sp}^O$.

Similarly, we define a category $\pi_0(\mathfrak{O}_G^{\mathbb{Q}})$, called the *rational G–equivariant stable orbit category*. The objects are the same as for $\pi_0(\mathfrak{O}_G)$, but the morphisms are given by

$$\pi_0(\mathfrak{O}_G^{\mathbb{Q}})(\Sigma^\infty A_+, \Sigma^\infty B_+) = [\Sigma^\infty A_+, \Sigma^\infty B_+]^G \otimes \mathbb{Q},$$

the set of maps in the homotopy category of $G\mathrm{Sp}^O_{\mathbb{Q}}$.

Recall that a finite discrete G–set is a disjoint union of homogeneous spaces G/H for H open, hence every finite discrete G–set A is a finite G–CW–complex and A_+ is finite pointed G–CW–complex.

Lemma 3.8 [Fausk 2008, Corollary 7.2] *For G a profinite group, there is an isomorphism of abelian groups*

$$[\Sigma^\infty A, \Sigma^\infty B]^G \cong \operatorname*{colim}_{W \in \mathcal{U}} [A \wedge S^W, B \wedge S^W]^G,$$

where the right-hand terms indicate maps in the homotopy category of pointed G–spaces and A and B are finite pointed G–CW–complexes.

Now we show how the equivariant stable homotopy category for G relates to the equivariant stable homotopy category for finite quotients G/N. Whenever we talk about G/N–spectra, we will use the G/N–universe $\mathcal{U}^N$. This universe is a complete G/N–universe as any finite-dimensional G/N–inner product space V can be written as $(\varepsilon^* V)^N$ and $\varepsilon^* V$ is isomorphic to an indexing space of $\mathcal{U}$, as $\mathcal{U}$ is complete.

Let A and B be finite pointed G–CW–complexes and let $N_1 \leqslant N_2$ be open normal subgroups of G, which fix all of A and B. Then A can be considered as either a G/N_2–CW–complex or a G/N_1–CW–complex and the inflation functor ε^* from G/N_2–spaces to G/N_1–spaces sends the G/N_2–version of A to the G/N_1–version. Hence, the inflation functor ε^* induces a natural map

$$[\Sigma^\infty A, \Sigma^\infty B]^{G/N_2} \to [\Sigma^\infty A, \Sigma^\infty B]^{G/N_1}.$$

We can find a more direct description of this map using Lemma 3.8. An element of the domain can be represented as a map of pointed G/N_2–spaces

$$f : A \wedge S^V \to B \wedge S^V$$

for some indexing space $V \subset \mathcal{U}^{N_2}$. Applying the inflation functor gives a map of pointed G/N_1–spaces $\varepsilon^* f$ and since V is N_2–fixed, this defines a element of

$$\operatorname*{colim}_{W \in \mathcal{U}^{N_1}} [A \wedge S^W, B \wedge S^W]^{G/N_1}.$$

This assignment is compatible with taking colimits over $V^{N_2} \subset \mathcal{U}^{N_2}$ and taking G/N_2–equivariant homotopy classes. Hence, we may construct the colimit term of the following result.

Lemma 3.9 *If A and B are finite pointed G–CW–complexes, for G a profinite group, then there is an isomorphism of abelian groups*

$$\operatorname*{colim}_{\substack{N \trianglelefteqslant G \\ \text{open}}} [\Sigma^\infty A, \Sigma^\infty B]^{G/N} \cong [\Sigma^\infty A, \Sigma^\infty B]^G.$$

Proof We have seen that there are isomorphisms

$$\operatorname*{colim}_{\substack{N \trianglelefteqslant G \\ \text{open}}} [\Sigma^\infty A, \Sigma^\infty B]^{G/N} \cong \operatorname*{colim}_{\substack{N \trianglelefteqslant G \\ \text{open}}} \Big(\operatorname*{colim}_{W \in \mathcal{U}^N} [A \wedge S^W, B \wedge S^W]^{G/N} \Big)$$

The right-hand term maps to

$$\operatorname*{colim}_{W \in \mathcal{U}} [A \wedge S^W, B \wedge S^W]^G \cong [\Sigma^\infty A, \Sigma^\infty B]^G$$

via the inflation functors.

We construct an inverse. Take a representative $f: A \wedge S^V \to B \wedge S^V$. The finite pointed G–CW–complexes A and B are fixed by some open normal subgroup. By Remark 2.3, the G–inner product space V must also be fixed by some open normal subgroup of G. By taking intersections there is an open normal subgroup N of G, which also fixes all of A, B and V. It follows that f defines an element of $[A \wedge S^V, B \wedge S^V]^{G/N}$. One can verify that this is an inverse to the map of the statement. □

Recall the grading convention $[\Sigma^n X, \Sigma^m Y]^G = [X, Y]^G_{n-m}$. Since suspension by S^1 preserves finite (pointed) G–CW–complexes, we have the following extension.

Corollary 3.10 *For G a profinite group, there is an isomorphism of graded abelian groups*

$$\operatorname*{colim}_{\substack{N \trianglelefteq G\\ \text{open}}} [\Sigma^\infty A, \Sigma^\infty B]^{G/N}_* \cong [\Sigma^\infty A, \Sigma^\infty B]^G_*$$

for A and B finite G–CW–complexes.

Theorem 3.11 *For G a profinite group, there is an equivalence of categories*

$$\psi_G : \mathrm{Span}(G_{df}\text{–sets}) \to \pi_0(\mathfrak{O}_G),$$
$$A \mapsto \Sigma^\infty A_+ \qquad \text{on objects},$$
$$[B \xleftarrow{\beta} A \xrightarrow{\alpha} C] \mapsto \Sigma^\infty \beta \circ \tau(\alpha) \qquad \text{on morphisms},$$

where $\tau(\alpha)$ is the transfer map construction associated to α; see [Lewis et al. 1986, Construction II.5.1] *or work of the second author* [Sugrue 2019b, Construction 3.1.11].

Proof That ψ_G is an equivalence for finite groups G is well-known; see [Lewis et al. 1986, Proposition V.9.6]. We will use that result to extend from the finite case to the profinite case.

For an inclusion $N \to N'$ of open normal subgroups, the inflation functor from $\mathrm{Span}(G/N_{df}\text{–sets})$ to $\mathrm{Span}(G/N'_{df}\text{–sets})$ commutes with $\psi_{G/N}$ and $\psi_{G/N'}$, hence the following diagram commutes:

$$\begin{array}{ccc}
\mathrm{Span}(G_{df}\text{–sets})(A, B) & \xrightarrow{\quad\psi_G\quad} & [\Sigma^\infty A_+, \Sigma^\infty B_+]^G \\
\cong \uparrow & & \uparrow \cong \\
\operatorname*{colim}_{\substack{N \trianglelefteq G\\ \text{open}}} \mathrm{Span}(G/N_{df}\text{–sets})(A, B) & \xrightarrow[\cong]{\psi_{G/N}} & \operatorname*{colim}_{\substack{N \trianglelefteq G\\ \text{open}}} [\Sigma^\infty A_+, \Sigma^\infty B_+]^{G/N}
\end{array}$$

This proves that ψ_G is full and faithful; essential surjectivity is immediate. □

4 The classification

We give the main result, the classification (in terms of Quillen equivalences) of rational G–equivariant spectra, for profinite G, in terms of a simple algebraic model. In fact, by previous work of the authors [Barnes and Sugrue 2023] we give two equivalent algebraic models. The first is the category of (chain complexes of) rational G–Mackey functors, the second is the category of Weyl G–sheaves over the space of closed subgroups of G. The relative advantages of the two descriptions are explained in that reference, though we will need the sheaf description in Section 5.

4.1 Mackey functors

There are several equivalent definitions of Mackey functors, we briefly describe the three most common variations, leaving the axioms of the first two versions for references such as [Lindner 1976] or [Thiel 2011]. These three definitions are shown to be equivalent in [Sugrue 2019b, Section 2.1], which follows the work of Lindner [1976].

(1) A set of abelian groups $M(G/H)$, for H open in G, with induction, restriction and conjugation maps relating these groups that satisfy a list of axioms (unital, transitivity, associativity, equivariance and the Mackey axiom).

(2) A pair of functors from the category of finite discrete G–sets to abelian groups, one covariant, one contravariant that agree on objects and satisfy a pullback axiom and a coproduct axiom. These are sometimes known as categorical Mackey functors.

(3) An additive functor from the Burnside category Span(G_{df}–sets) to abelian groups.

The choice of focussing on the open subgroups (or equally the discrete finite G–sets) matches with the "finite natural Mackey system" of [Bley and Boltje 2004, Definition 2.1 and Examples 2.2; Thiel 2011, Definition 2.2.12]. In the case of a finite group, this choice restricts to the usual definitions. For our purposes we use the last definition.

Definition 4.1 A *Mackey functor* for a profinite group G is an additive functor from the Burnside category Span(G_{df}–sets) to abelian groups. We will write Mackey(G) for the category of Mackey functors and additive natural transformations between them.

A *rational Mackey functor* is an additive functor from the Burnside category Span(G_{df}–sets) to $\mathbb{Q}$–modules. We will write $\text{Mackey}_{\mathbb{Q}}(G)$ for the category of rational Mackey functors and additive natural transformations between them.

General examples of Mackey functors can be found in the references given at the start of the section. However, there is one class of rational Mackey functors of particular relevance to this paper.

Example 4.2 By Theorem 3.11 and Corollary 2.11, if X is a G–spectrum then we have a rational G–Mackey functor,

$$\pi_0(\mathfrak{O}_G^{\mathbb{Q}}) \to \mathbb{Q}\text{–mod}, \quad G/H_+ \mapsto \pi_0^H(X) \otimes \mathbb{Q} \cong [G/H_+, X]_{\mathbb{Q}}^G,$$

called the *homotopy group Mackey functor* of X.

We write out the definition of Mackey functors in terms of our notation and apply Theorem 3.11. We use $\text{Func}_{\text{Ab}}(-,-)$ to denote the category of enriched functors and natural transformations over abelian groups. This is the same category as taking additive functors, as an additive functor between additive categories is precisely the data of a functor enriched over abelian groups:

$$\mathrm{Mackey}(G) = \mathrm{Func}_{\mathrm{Ab}}(\mathrm{Span}(G_{df}\text{–sets}), \mathrm{Ab}) \cong \mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathfrak{O}_G), \mathrm{Ab}),$$
$$\mathrm{Mackey}_{\mathbb{Q}}(G) = \mathrm{Func}_{\mathrm{Ab}}(\mathrm{Span}(G_{df}\text{–sets}), \mathbb{Q}\text{–mod}) \cong \mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathfrak{O}_G), \mathbb{Q}\text{–mod}).$$

It remains to relate these categories of G–Mackey functors to the model category of rational G–spectra. For this we will need a model structure on chain complexes of G–Mackey functors.

Lemma 4.3 *There is a cofibrantly generated model structure on the category of chain complexes of* (*rational*) *G–Mackey functors where a map is a fibration if and only if it is a surjection and the class of weak equivalences is the class of homology isomorphisms.*

Proof Since $\mathrm{Ch}(\mathrm{Mackey}(G)) = \mathrm{Func}_{\mathrm{Ab}}(\mathrm{Span}(G_{df}\text{–sets}), \mathrm{Ch}(\mathbb{Z}))$, we use the cofibrant generation of $\mathrm{Ch}(\mathbb{Z})$ to obtain generating sets for $\mathrm{Ch}(\mathrm{Mackey}(G))$ in terms of the representable functors. □

4.2 Tilting theory

We give our classification theorem for rational G–spectra, where G is a profinite group.

Theorem 4.4 *For G a profinite group*, *there is a zigzag of Quillen equivalences between the model category of rational G–spectra and the model category of chain complexes of rational G–Mackey functors.*

Proof Choose a skeleton $\mathcal{G}$ of $\pi_0(\mathfrak{O}_G^{\mathbb{Q}})$ (such as the set of finite coproducts of the G–sets G/H for H an open subgroup of G). Define $\pi_0(\mathcal{G})$ to be the category whose objects are the elements of $\mathcal{G}$ and whose morphisms are given by the abelian groups

$$[\Sigma^\infty A_+, \Sigma^\infty B_+]^G \otimes \mathbb{Q}.$$

The objects of $\mathcal{G}$ define a set of compact generators for $G\mathrm{Sp}_{\mathbb{Q}}^O$ by Corollary 2.12 and the set of graded maps between them in $\mathrm{Ho}(G\mathrm{Sp}_{\mathbb{Q}}^O)$ is concentrated in degree zero by Theorem 2.14. Thus, we can use [Schwede and Shipley 2003, Theorem 5.1.1 and Proposition B.2.1] to obtain a zigzag of Quillen equivalences,

$$G\mathrm{Sp}_{\mathbb{Q}}^O \simeq \mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathcal{G}), \mathrm{Ab})\big).$$

As $\pi_0(\mathcal{G})$ is a skeleton of $\pi_0(\mathfrak{O}_G^{\mathbb{Q}})$, we have the first equivalence of categories below. The second is Lemma 4.5, which applies as Corollary 2.11 shows that $\pi_0(\mathfrak{O}_G^{\mathbb{Q}}) = \pi_0(\mathfrak{O}_G) \otimes \mathbb{Q}$:

$$\mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathcal{G}), \mathrm{Ab})\big) \cong \mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathfrak{O}_G^{\mathbb{Q}}), \mathrm{Ab})\big) \cong \mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathfrak{O}_G), \mathbb{Q}\text{–mod})\big).$$

Applying Theorem 3.11 gives the final step,

$$\mathrm{Ch}(\mathrm{Mackey}_{\mathbb{Q}}(G)) = \mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\mathrm{Span}(G_{df}\text{–sets}), \mathbb{Q}\text{–mod})\big) \cong \mathrm{Ch}\big(\mathrm{Func}_{\mathrm{Ab}}(\pi_0(\mathfrak{O}_G), \mathbb{Q}\text{–mod})\big). \quad □$$

Lemma 4.5 *Let $\mathcal{C}$ be a small additive category. The rationalisation functor $i:\mathcal{C}\to\mathcal{C}\otimes\mathbb{Q}$ (defined in the proof) and the forgetful functor $U:\mathbb{Q}\text{–mod}\to\mathrm{Ab}$ induce equivalences*

$$\mathrm{Func}_{\mathrm{Ab}}(\mathcal{C},\mathbb{Q}\text{–mod}) \xleftarrow{\;i_*\;} \mathrm{Func}_{\mathrm{Ab}}(\mathcal{C}\otimes\mathbb{Q},\mathbb{Q}\text{–mod}) \xrightarrow{\;U^*\;} \mathrm{Func}_{\mathrm{Ab}}(\mathcal{C}\otimes\mathbb{Q},\mathrm{Ab}).$$

Proof A small additive category $\mathcal{C}$ has a rationalisation $\mathcal{C}\otimes\mathbb{Q}$, this category has the same objects and morphisms given by

$$(\mathcal{C}\otimes\mathbb{Q})(c,c')=\mathcal{C}(c,c')\otimes\mathbb{Q}.$$

Composition is induced from that of $\mathcal{C}$ and $\mathcal{C}\otimes\mathbb{Q}$ is an additive category. Moreover, there is an additive functor from $\mathcal{C}$ to its rationalisation $i:\mathcal{C}\to\mathcal{C}\otimes\mathbb{Q}$.

The functors are equivalences as in each case the functor must take values in $\mathbb{Q}$–modules. □

Example 4.6 Let A be a finite discrete G–set. Let M_A be the homotopy group Mackey functor of $\Sigma^\infty A_+$ from Example 4.2. Then M_A is the representable functor given by

$$\mathrm{Span}(G_{df}\text{–sets})(-,A)\otimes\mathbb{Q}.$$

We recall the notion of Weyl-G–sheaves over the space of closed subgroups of G from [Barnes and Sugrue 2023, Section 2; 2022, Section 10]. For G a profinite group, let $\mathcal{S}G$ denote the set of closed subgroups of G. We topologise this set as the limit of finite discrete spaces

$$\mathcal{S}G:=\lim_{\substack{N\trianglelefteq G\\ \text{open}}}\mathcal{S}(G/N)$$

using the maps which send $K\in\mathcal{S}G$ to $KN/N\in\mathcal{S}(G/N)$.

Definition 4.7 A *G–equivariant sheaf* of $\mathbb{Q}$–modules over $\mathcal{S}G$ is a map $p\colon E\to\mathcal{S}G$ such that

(1) p is a G–equivariant map $p\colon E\to\mathcal{S}G$ of spaces with continuous G–actions,

(2) (E,p) is a sheaf space of $\mathbb{Q}$–modules,

(3) each map $g\colon p^{-1}(x)\to p^{-1}(gx)$ is a map of $\mathbb{Q}$–modules for every $x\in\mathcal{S}G,\ g\in G$.

We will write this as either the pair (E,p) or simply as E. A map $f\colon(E,p)\to(E',p')$ of G–sheaves of $\mathbb{Q}$–modules over $\mathcal{S}G$ is a G–equivariant map $f\colon E\to E'$ such that $p'f=p$ and $f_x\colon E_x\to E'_x$ is a map of $\mathbb{Q}$–modules for each $x\in\mathcal{S}G$.

Definition 4.8 A *rational Weyl-G–sheaf* E is a G–sheaf of $\mathbb{Q}$–modules over $\mathcal{S}G$ such that E_K is K–fixed and hence is a discrete $\mathbb{Q}[W_GK]$–module. A map of Weyl-G–sheaves is a map of G–sheaves of $\mathbb{Q}$ modules over $\mathcal{S}G$. We write this category as Weyl-G–$\mathrm{sheaf}_{\mathbb{Q}}(\mathcal{S}G)$

Theorem 4.9 [Barnes and Sugrue 2023, Theorem A] *If G is a profinite group then the category of rational G–Mackey functors is equivalent to the category of rational Weyl-G–sheaves over $\mathcal{S}G$. Furthermore, this is an exact equivalence.*

Corollary 4.10 *For G a profinite group, there is a zigzag of Quillen equivalences between the model category of rational G–spectra and the model category of chain complexes of rational Weyl-G–sheaves.*

5 Homological dimension of equivariant sheaves

Using the Weyl-G–sheaf description of rational G–spectra we can calculate the homological dimension (also known as the injective dimension) of the algebraic model. This gives an indication of the homological complexity of the algebraic model. It is already known that the algebraic model in the case of finite groups has homological dimension zero (we will recover this result). The algebraic model in the case of an r–torus $(S^1)^{\times r}$ is r, as shown in [Greenlees 2012, Theorem 8.1].

We prove that the Cantor–Bendixson rank (Definition 5.4) of the space $\mathcal{S}G$ will determine the homological dimension of rational (Weyl) G–sheaves on $\mathcal{S}G$. This is an equivariant generalisation of the results of the second author [Sugrue 2019a].

5.1 Cantor–Bendixson rank

We start with the basic definitions; see [Gartside and Smith 2010a; 2010b].

Definition 5.1 For a topological space X we define the *Cantor–Bendixson process* on X. Denote by X' the set of all isolated points of X.

(1) Let $X^{(0)} = X$ and $X^{(1)} = X \setminus X'$ have the subspace topology with respect to X.

(2) For successor ordinals suppose we have $X^{(\alpha)}$ for an ordinal α, we define
$$X^{(\alpha+1)} = X^{(\alpha)} \setminus X^{(\alpha)\prime}.$$

(3) If λ is a limit ordinal we define
$$X^{(\lambda)} = \bigcap_{\alpha<\lambda} X^{(\alpha)}.$$

Every Hausdorff topological space X has a minimal ordinal α such that $X^{(\alpha)} = X^{(\lambda)}$ for all $\lambda \geq \alpha$; see [Gartside and Smith 2010a, Lemma 2.7].

Definition 5.2 For X a Hausdorff topological space, the *Cantor–Bendixson rank* of X, written $\mathrm{rank}_{CB}(X)$, is the minimal ordinal α such that $X^{(\alpha)} = X^{(\lambda)}$ for all $\lambda \geq \alpha$.

A topological space X is called *perfect* if it has no isolated points, whereupon $\mathrm{rank}_{CB}(X) = 0$.

Remark 5.3 The definition given above agrees with that of [Gartside and Smith 2010b]. The convention of Dickmann, Schwartz and Tressl [Dickmann et al. 2019, Definition 4.3.1] is to take one less than the rank as defined above.

There are two ways that the Cantor–Bendixson process can stabilise, by reaching the empty set or a perfect subspace.

Definition 5.4 If X is a Hausdorff space with Cantor–Bendixson rank λ, then we define the *perfect hull* of X to be the subspace $X^{(\lambda)}$, written X_H.

We write X_S for the complement $X \setminus X_H$ and call it the *scattered part* of X. The space X is said to be *scattered* if $X_H = \varnothing$.

Example 5.5 The Cantor–Bendixson rank of the empty set is zero and the Cantor–Bendixson rank of a nonempty discrete space is 1 as every point is isolated.

The space $\mathcal{S}\mathbb{Z}_p^\wedge$ of closed subgroups of the p–adics is the subspace of $\mathbb{R}$ consisting of the points

$$\{1/n \mid n \in \mathbb{N}\} \cup \{0\}.$$

The isolated points are those of the form $1/n$, which are removed in the first stage of the Cantor–Bendixson process, thus $(\mathcal{S}\mathbb{Z}_p^\wedge)^{(1)} = \{0\}$. For $k > 1$, $(\mathcal{S}\mathbb{Z}_p^\wedge)^{(k)} = \varnothing$, so $\mathrm{rank}_{CB}(\mathcal{S}\mathbb{Z}_p^\wedge) = 2$.

Definition 5.6 If X is a space and $x \in X_S$, we define the *height* of x denoted $\mathrm{ht}(X, x)$, to be the ordinal κ such that $x \in X^{(\kappa)}$ but $x \notin X^{(\kappa+1)}$. We denote this by $\mathrm{ht}(x)$ when the background space X is understood.

We may rephrase the definitions to see that a point x of height $k > 0$ is a limit of points of height $k-1$. Consequently, an open neighbourhood of x contains infinitely many points of height j for each $j < k$.

Lemma 5.7 *Let X be a topological space with an action of a topological group G. The height of points of X is invariant under the action of G.*

If $x \in X$ has height $k > 0$, then every neighbourhood of x contains infinitely many points from orbits other than Gx.

Proof As G acts through homeomorphisms, the first statement holds. The second statement follows from our preceding discussion and the first statement. □

5.2 Equivariant Godement resolutions

We recap equivariant Godement resolutions from [Barnes and Sugrue 2022, Section 9]. The key change from the nonequivariant case is the use orbits in place of points.

Definition 5.8 Let $p\colon E \to X$ be a G–equivariant sheaf of $\mathbb{Q}$–modules over a G–space X, for G a profinite group. Define a G–sheaf

$$I^0(E) = \prod_A i_A^* E|_A$$

with the product taken over the G–orbits of X and i_A^* is the extension by zero functor induced by the map $A \to X$.

The restriction–extension adjunction induces morphisms $E \to i_A^* E|_A$, which combine to a monomorphism

$$\delta_E\colon E \to I^0(E).$$

Lemma 5.9 *Let $p\colon E \to X$ be a G–equivariant sheaf of $\mathbb{Q}$–modules over a G–space X, for G a profinite group. The G–sheaf $I^0(E)$ is injective.*

Proof We prove that $I^0(E)$ is the product of injective sheaves. For each orbit A of X, pick an element $x_A \in A$. By [Barnes and Sugrue 2022, Lemma 8.3] there is an isomorphism of sheaves over A

$$G \underset{\operatorname{stab}_G(x_A)}{\times} E_{x_A} \cong E|_A.$$

The left-hand sheaf is known as the equivariant skyscraper sheaf of E_{x_0} at x_0. It can be viewed as part of an adjunction with the left adjoint being taking the stalk at x_0. As this left adjoint preserves monomorphisms, the equivariant skyscraper sheaf construction preserves injective objects.

The result is completed by [Castellano and Weigel 2016, Proposition 3.1], which states that every object of the category of discrete $\mathbb{Q}[\operatorname{stab}_G(x_0)]$–modules is injective. □

Remark 5.10 The result [Castellano and Weigel 2016, Proposition 3.1] also fixes an omission in work of the first author [Barnes 2011, Lemma 6.2] which implicitly assumes that all discrete $\mathbb{Q}[\mathbb{Z}_p^\wedge]$–modules are injective.

Iterating this construction I^0 gives an injective resolution.

Definition 5.11 Let G be a profinite group. If E is a G–sheaf of R–modules over a profinite G–space we define the *equivariant Godement resolution* as follows:

$$\begin{array}{ccccccc} 0 & \longrightarrow & E & \xrightarrow{\delta_E} & I^0(E) & \longrightarrow & I^0(\operatorname{Coker}\delta_E) = I^1(E) \longrightarrow \cdots \\ & & & & \downarrow p & \nearrow \delta_{\operatorname{Coker}\delta_E} & \\ & & & & \operatorname{Coker}\delta_E & & \end{array}$$

We now connect the Cantor–Bendixson rank of a G–space X to the length of the equivariant Godement resolution of rational G–sheaves over X.

Theorem 5.12 *Let E be a rational G–sheaf over a profinite G–space X, for G a profinite group. For $n \in \mathbb{N}$ and $x \in X$, the stalk $I^n(E)_x$ is zero unless the height of x is at least n.*

Proof The proof is by induction. The base case follows from the fact that $(\delta_E)_x$ is an isomorphism for any isolated point of X. Similarly, if x has a open neighbourhood where the only nontrivial stalk is at x, then $(\delta_E)_x$ is an isomorphism. The equivariant Godement resolution with the Cantor–Bendixson process gives the inductive step using Lemma 5.7. Further details are given in the nonequivariant case of [Sugrue 2019a, Lemma 3.7]. □

If we restrict ourselves to the case of scattered spaces of finite rank (the Cantor–Bendixson process ends in the empty set after finitely many steps) this theorem gives an upper bound for the length of the equivariant Godement resolutions.

Corollary 5.13 *Let G be a profinite group. Let X be a scattered profinite G–space of Cantor–Bendixson rank $n \in \mathbb{N}$. The category of rational G–sheaves over X has homological dimension at most $n-1$.*

Proof If the rank is n, then $X^{(n)} = \varnothing$ and every point has height at most $n-1$. Hence, every stalk of $I^n(X)$ is zero, so the sheaf is itself zero. □

5.3 The case of the constant sheaf

It remains to prove that the upper bound on homological dimension is in fact an equality. To that end, we take the equivariant Godement resolution of the constant sheaf at $\mathbb{Q}$, $\operatorname{Const}\mathbb{Q}$,

$$0 \longrightarrow \operatorname{Const}\mathbb{Q} \xrightarrow{\delta_0} I^0 \xrightarrow{\delta_1} I^1 \xrightarrow{\delta_2} \cdots \longrightarrow I^{n-1}.$$

Adding a small assumption on X we can prove this resolution has length exactly $n-1$.

Proposition 5.14 *Let G be a profinite group, let X be a profinite scattered G–space of Cantor–Bendixson rank n and assume that each $x \in X$ has a neighbourhood basis $\mathcal{B}_x$ of $\operatorname{stab}_G(x)$–invariant sets. In the equivariant Godement resolution of the constant sheaf, the cokernel of $\delta_i(U)$ has a nonzero $\operatorname{stab}_G(x)$–equivariant section for each $U \in \mathcal{B}_x$ whenever i is smaller than the height of x.*

Proof We start with the case of δ_0 with x of height at least 1. Take a $\operatorname{stab}_G(x)$–invariant open neighbourhood of x. By Lemma 5.7, U contains infinitely many points of other orbits which are of lower height than x.

Choose a nonzero element of $\operatorname{Const}\mathbb{Q}(U)_x = \mathbb{Q}$ represented by a section $s \in \operatorname{Const}\mathbb{Q}(U)$. We define a section t of $I^0 = \prod_A i_A^* E|_A$ by the sequence

$$A \mapsto \begin{cases} 0 & \operatorname{ht}(A) \text{ has the same parity as } \operatorname{ht}(x), \\ s|_A & \text{otherwise.} \end{cases}$$

Since s is nonzero at infinitely many points near x, t_x is nonzero. If $t = \delta_0(s')$ for some $s' \in \operatorname{Const}\mathbb{Q}(U)$ then $s'|_{Gx} = 0$. This implies that there is an open neighbourhood of x where s' restricts to zero and hence $s'_y = 0$ for all y in that open neighbourhood, which implies that $t_x = 0$, a contradiction.

The rest follows inductively, as with the nonequivariant case of [Sugrue 2019a, Lemma 4.3], with two changes. The first is that when we need to construct a new nonzero section we use the alternating process described previously. The second is that since our sets U are $\operatorname{stab}_G(x)$–invariant and we begin with a $\operatorname{stab}_G(x)$–equivariant section, all sections constructed in the proof are $\operatorname{stab}_G(x)$–equivariant.

Note that the assumption on i and the height of x is required to ensure that the new section we construct is nonzero at infinitely many orbits. □

We can now give the equivariant analogue of [Sugrue 2019a, Theorem 4.4].

Theorem 5.15 *Let X be profinite G–space such that each $x \in X$ has a neighbourhood basis $\mathcal{B}_x$ of $\operatorname{stab}_G(x)$–invariant sets, for G a profinite group.*

If X is a scattered G–space of Cantor–Bendixson rank n and x has height $n-1$, then

$$\mathrm{Ext}^{n-1}\big(G \underset{\mathrm{stab}_G(x)}{\times} \mathbb{Q}, \mathrm{Const}\,\mathbb{Q}\big) \neq 0.$$

Hence, the homological dimension of the category of rational G–sheaves over X is $n-1$.

If X has infinite Cantor–Bendixson rank then the homological dimension of the category of rational G–sheaves over X is infinite.

Proof We begin with a general calculation of maps out of an equivariant skyscraper sheaf into $I^0(E)$ for E some rational G–sheaf over X,

$$\begin{aligned}\hom\big(G \underset{\mathrm{stab}_G(x)}{\times} \mathbb{Q}, I^0(E)\big) &\cong \prod_A \hom\big(G \underset{\mathrm{stab}_G(x)}{\times} \mathbb{Q}, i_A^* E|_A\big) \\ &\cong \hom\big(G \underset{\mathrm{stab}_G(x)}{\times} \mathbb{Q}, E|_{Gx}\big) \\ &\cong \hom\big(G \underset{\mathrm{stab}_G(x)}{\times} \mathbb{Q}, G \underset{\mathrm{stab}_G(x)}{\times} E_x\big) \\ &\cong E_x^{\mathrm{stab}_G x}.\end{aligned}$$

The final term has fixed points as $\mathbb{Q}$ has the trivial $\mathrm{stab}_G\, x$–action.

Assume that X is a scattered G–space of Cantor–Bendixson rank n and x has height $n-1$. Applying our calculation to our resolution of $\mathrm{Const}\,\mathbb{Q}$ we see that our Ext groups are the homology of the chain complex

$$\mathbb{Q} \xrightarrow{\alpha_0} (\mathrm{Coker}\,\delta_0)_x^{\mathrm{stab}_G(x)} \xrightarrow{\alpha_1} (\mathrm{Coker}\,\delta_1)_x^{\mathrm{stab}_G(x)} \xrightarrow{\alpha_2} \cdots \xrightarrow{\alpha_{n-1}} (\mathrm{Coker}\,\delta_{n-2})_x^{\mathrm{stab}_G(x)}.$$

By Proposition 5.14 we have a nonzero $\mathrm{stab}_G(x)$–equivariant section which implies that

$$(\delta_i)_x^{\mathrm{stab}_G(x)} \neq 0$$

whenever i is smaller than $n-1$. By a similar argument to Proposition 5.14 we see that α_{n-1} is not surjective, hence the n^{th} Ext group is nonzero. This calculation and Corollary 5.13 show that the homological dimension is $n-1$.

In the case of infinite Cantor–Bendixson rank we see that for each n there is a point of height n, hence our earlier work shows that the homological dimension is infinite. □

The space of closed subgroups $\mathcal{S}G$ of a profinite group G always satisfies the condition on the invariant neighbourhood basis; see [Barnes and Sugrue 2023, Section 2].

Corollary 5.16 *Let G be a profinite group whose space of subgroups $\mathcal{S}G$ is scattered of Cantor–Bendixson rank n. The homological dimension of the category of rational Weyl-G–sheaves is $n-1$.*

If $\mathcal{S}G$ has infinite Cantor–Bendixson rank, then the homological dimension of the category of rational Weyl-G–sheaves is infinite.

Proof All the sheaves used in the theorem proof were stalkwise fixed and hence were Weyl-G–sheaves. □

There is a remaining case of spaces which have finite Cantor–Bendixson rank and nonempty perfect hull. We make an equivariant version of [Sugrue 2019a, Conjecture 4.6].

Conjecture 5.17 *Let G be a profinite group. If the G–space X has finite Cantor–Bendixson rank and nonempty perfect hull, then the homological dimension of rational G–sheaves over X is infinite.*

Using the equivalence between Weyl-G–sheaves and rational G–Mackey functors [Barnes and Sugrue 2023, Theorem A] we obtain the following calculation of homological dimensions for categories of rational G–Mackey functors.

Corollary 5.18 *Let G be a profinite group whose space of subgroups $\mathcal{S}G$ is scattered of Cantor–Bendixson rank n. The homological dimension of the category of rational G–Mackey functors is $n-1$.*

If $\mathcal{S}G$ has infinite Cantor–Bendixson rank, then the homological dimension of the category of rational G–Mackey functors is infinite.

Example 5.19 Using our calculations from Example 5.5 we can give the homological dimension of some algebraic models.

For G a finite group, the homological dimension of rational Weyl-G–sheaves is zero. In terms of Mackey functors, this says that every rational G–Mackey functor is injective, as was proven independently by two sources [Greenlees and May 1995, Appendix A; Thévenaz and Webb 1995, Theorems 8.3 and 9.1]. See also [Barnes and Kędziorek 2022, Theorem 4.28].

For $G = \mathbb{Z}_p^\wedge$, the p–adic integers, the homological dimension of rational Weyl-G–sheaves is one, which agrees with [Barnes 2011, Lemma 6.2].

References

[Barnes 2008] **D Barnes**, *Rational equivariant spectra*, PhD thesis, University of Sheffield (2008) arXiv 0802.0954

[Barnes 2011] **D Barnes**, *Rational $\mathbb{Z}_p$–equivariant spectra*, Algebr. Geom. Topol. 11 (2011) 2107–2135 MR Zbl

[Barnes and Kędziorek 2022] **D Barnes**, **M Kędziorek**, *An introduction to algebraic models for rational G–spectra*, from "Equivariant topology and derived algebra", Lond. Math. Soc. Lect. Note Ser. 474, Cambridge Univ. Press (2022) 119–179 MR Zbl

[Barnes and Roitzheim 2020] **D Barnes**, **C Roitzheim**, *Foundations of stable homotopy theory*, Cambridge Stud. Adv. Math. 185, Cambridge Univ. Press (2020) MR Zbl

[Barnes and Sugrue 2022] **D Barnes**, **D Sugrue**, *Equivariant sheaves for profinite groups*, Topology Appl. 319 (2022) art. id. 108215 MR Zbl

[Barnes and Sugrue 2023] **D Barnes**, **D Sugrue**, *The equivalence between rational G–sheaves and rational G–Mackey functors for profinite G*, Math. Proc. Cambridge Philos. Soc. 174 (2023) 345–368 MR Zbl

[Bley and Boltje 2004] **W Bley**, **R Boltje**, *Cohomological Mackey functors in number theory*, J. Number Theory 105 (2004) 1–37 MR Zbl

[Castellano and Weigel 2016] **I Castellano**, **T Weigel**, *Rational discrete cohomology for totally disconnected locally compact groups*, J. Algebra 453 (2016) 101–159 MR Zbl

[Dickmann et al. 2019] **M Dickmann**, **N Schwartz**, **M Tressl**, *Spectral spaces*, New Math. Monogr. 35, Cambridge Univ. Press (2019) MR Zbl

[Fausk 2008] **H Fausk**, *Equivariant homotopy theory for pro-spectra*, Geom. Topol. 12 (2008) 103–176 MR Zbl

[Gartside and Smith 2010a] **P Gartside**, **M Smith**, *Classifying spaces of subgroups of profinite groups*, J. Group Theory 13 (2010) 315–336 MR Zbl

[Gartside and Smith 2010b] **P Gartside**, **M Smith**, *Counting the closed subgroups of profinite groups*, J. Group Theory 13 (2010) 41–61 MR Zbl

[Greenlees 2006] **J P C Greenlees**, *Triangulated categories of rational equivariant cohomology theories*, Oberwolfach Reports 3 (2006) 480–489 Part of the conference report "Thick subcategories: classifications and applications"

[Greenlees 2012] **J P C Greenlees**, *Rational torus-equivariant stable homotopy, II: Algebra of the standard model*, J. Pure Appl. Algebra 216 (2012) 2141–2158 MR Zbl

[Greenlees and May 1995] **J P C Greenlees**, **J P May**, *Generalized Tate cohomology*, Mem. Amer. Math. Soc. 543, Amer. Math. Soc., Providence, RI (1995) MR Zbl

[Hill and Mazur 2019] **M A Hill**, **K Mazur**, *An equivariant tensor product on Mackey functors*, J. Pure Appl. Algebra 223 (2019) 5310–5345 MR Zbl

[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

[Illman 1983] **S Illman**, *The equivariant triangulation theorem for actions of compact Lie groups*, Math. Ann. 262 (1983) 487–501 MR Zbl

[Lewis et al. 1986] **L G Lewis, Jr**, **J P May**, **M Steinberger**, **J E McClure**, *Equivariant stable homotopy theory*, Lecture Notes in Math. 1213, Springer (1986) MR Zbl

[Lindner 1976] **H Lindner**, *A remark on Mackey-functors*, Manuscripta Math. 18 (1976) 273–278 MR Zbl

[Mandell and May 2002] **M A Mandell**, **J P May**, *Equivariant orthogonal spectra and S–modules*, Mem. Amer. Math. Soc. 755, Amer. Math. Soc., Providence, RI (2002) MR Zbl

[Mandell et al. 2001] **M A Mandell**, **J P May**, **S Schwede**, **B Shipley**, *Model categories of diagram spectra*, Proc. Lond. Math. Soc. 82 (2001) 441–512 MR Zbl

[Ribes and Zalesskii 2000] **L Ribes**, **P Zalesskii**, *Profinite groups*, Ergebnisse der Math. 40, Springer (2000) MR Zbl

[Schwede and Shipley 2003] **S Schwede**, **B Shipley**, *Stable model categories are categories of modules*, Topology 42 (2003) 103–153 MR Zbl

[Segal 1968] **G Segal**, *Equivariant K–theory*, Inst. Hautes Études Sci. Publ. Math. 34 (1968) 129–151 MR Zbl

[Sugrue 2019a] **D Sugrue**, *Injective dimension of sheaves of rational vector spaces*, J. Pure Appl. Algebra 223 (2019) 3112–3128 MR Zbl

[Sugrue 2019b] **D Sugrue**, *Rational G–spectra for profinite G*, PhD thesis, Queen's University Belfast (2019) arXiv 1910.12951

[Thévenaz and Webb 1995] **J Thévenaz**, **P Webb**, *The structure of Mackey functors*, Trans. Amer. Math. Soc. 347 (1995) 1865–1961 MR Zbl

[Thiel 2011] **U Thiel**, *Mackey functors and abelian class field theories*, preprint (2011) arXiv 1111.0942

[Williamson 2022] **J Williamson**, *Algebraic models of change of groups functors in (co)free rational equivariant spectra*, J. Pure Appl. Algebra 226 (2022) art. id. 107108 MR

[Wilson 1998] **J S Wilson**, *Profinite groups*, Lond. Math. Soc. Monogr. 19, Oxford Univ. Press (1998) MR Zbl

Mathematical Sciences Research Centre, Queen's University Belfast
Belfast, Northern Ireland, United Kingdom

Mathematical Sciences Research Centre, Queen's University Belfast
Belfast, Northern Ireland, United Kingdom

d.barnes@qub.ac.uk, dannysugrue@rocketmail.com

Received: 10 September 2022 Revised: 4 October 2023

Algebraic & Geometric Topology **24**:7 (2024) 3827–3873
DOI: 10.2140/agt.2024.24.3827
Published: 9 December 2024

An explicit comparison between 2–complicial sets and Θ_2–spaces

Julia E Bergner
Viktoriya Ozornova
Martina Rovelli

We produce a direct Quillen equivalence between two models of $(\infty, 2)$–categories: the complete Segal Θ_2–spaces due to Rezk and the 2–complicial sets due to Verity.

18N10, 18N65, 55U35; 18N50, 55U10

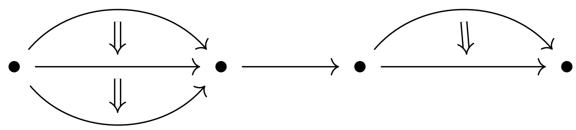
Introduction

The language of higher categories provides a way to describe many phenomena in areas of mathematics as diverse as topology, algebra, geometry, and mathematical physics. In a higher categorical structure, we not only have functions between objects, but functions between those functions and possibly further iterations of this idea, encoded by the notion of a k–morphism between $(k-1)$–morphisms. One might initially assume that these higher morphisms should satisfy conditions like associativity in the usual way, but for many natural examples they only hold up to isomorphism or, in topological settings, up to homotopy. In the latter situation, it is convenient to work in the setting of (∞, n)–categories, in which we have k–morphisms for arbitrarily large k, but they are all weakly invertible for $k > n$. These higher invertible morphisms provide a means for conveniently encoding the "up to isomorphism" data in the lower morphisms.

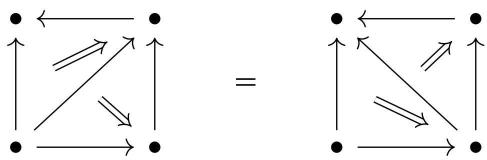
There have been many different approaches to realizing (∞, n)–categories as concrete mathematical objects; such realizations are often called *models* for (∞, n)–categories. A natural question, then, is whether these different models really do encode the same information, namely, whether we can establish an appropriate equivalence between them. Much work has been done in this direction, but there are still proposed models for which we do not have such comparisons. In some other cases, we know by general results that models must be equivalent, but do not have an explicit equivalence.

The motivation for this paper is the desire for an explicit comparison between two of these models, the complete Segal Θ_n–spaces as defined by [Rezk 2010] and the n–complicial sets as defined by [Verity 2008b] (see also [Ozornova and Rovelli 2020; Riehl 2018]); we give such a comparison when $n = 2$, for which more tools are available. Let us give a brief description of these two models.

A complete Segal Θ_2–space is described by a diagram of spaces indexed by 2–categories freely generated by pasting diagrams such as



which the expert reader may recognize as the generic element of Joyal's *cell category* Θ_2. In contrast, a 2–complicial set is given by a simplicial set with a suitable marking in which a k–simplex represents a diagram indexed by a truncated *oriental*, which is a free 2–category generated by a standard simplex, such as



A common way to show that two models are equivalent is to show that appropriate model categories for each are Quillen equivalent to each other. In this paper, we seek to establish such a Quillen equivalence between the model structure $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$ for complete Segal Θ_2–spaces and the model structure $ms\mathcal{S}et_{(\infty,2)}$ for 2–complicial sets.

Combining several prior results by different groups of authors, we already know that the two model categories are Quillen equivalent via a rather lengthy zigzag of Quillen equivalences between different models. Although we do not expect the reader to be familiar with all these models of $(\infty, 2)$–categories, to give an idea of the complexity of the comparison Figure 1 shows a diagram of an essentially optimal zigzag of Quillen equivalences, extracted from [Gagna et al. 2022].

To simplify the comparison, the goal of this paper is to produce the following direct Quillen equivalence.

Theorem *There is a Quillen equivalence between complete Segal Θ_2–spaces, presented by the model category $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$, and 2–complicial sets, presented by the model category $ms\mathcal{S}et_{(\infty,2)}$.*

$$\begin{array}{ccccc}
s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)} & & & & ms\mathcal{S}et_{(\infty,2)} \\
\wr\Vert \ \text{[Rezk 2001]} & & & & \text{[Gagna et al. 2022]}\ \Vert\wr \\
s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)} & & & & s\mathcal{S}et^{sc}_{(\infty,2)} \\
\wr\Vert \ \text{[Bergner and Rezk 2020]} & & & & \text{[Lurie 2009b]}\ \Vert\wr \\
s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)} & & & & \mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}} \\
\wr\Vert \ \text{[Bergner and Rezk 2020]} & & & & \text{[Lurie 2009a]}\ \Vert\wr \\
P\mathcal{C}at(s\mathcal{S}et^{\Delta^{op}})_{p,(\infty,2)} & \simeq & \mathcal{C}at_{s\mathcal{S}et^{\Delta^{op}}_{(\infty,1)}} & \simeq & \mathcal{C}at_{\mathcal{S}et^{\Delta^{op}}_{(\infty,1)}} \\
 & \text{[Bergner and Rezk 2013]} & & \text{[Joyal and Tierney 2007]} &
\end{array}$$

Figure 1

In addition to providing a more transparent comparison between the two models, this direct comparison facilitates the transport of constructions between these model structures. We now briefly illustrate the advantages of each model structure via the examples of duals and joins, and we refer the reader to Section 4 for a more detailed treatment of these cases, as well as other applications.

The structure of the category Θ_2 makes the description of duals straightforward in Θ_2–spaces, thanks to the globular shape of the objects. We can think of 1–dimensional duals as given by reversing the direction of the arrows, and 2–dimensional duals as given by similarly reversing the direction of 2–cells. Describing such 2–dimensional duals in the simplicial framework is more complicated, due to the triangular shape of the cells.

On the other hand, the join construction has been described for 2–complicial sets by [Verity 2008b] and is similar to familiar join constructions for simplicial sets. One can adjoin 1–simplices connecting the vertices of the two simplicial sets being joined, and higher-dimensional simplices analogously. In this case, working in a simplicial framework is much more straightforward than that of Θ_2.

While the existence of such a direct Quillen equivalence follows formally, for example using methods of [Dugger 2001], we find it valuable to have an explicit description.

Let us now describe the main ingredients of the proof of our main theorem.

(i) We use the compatibility of the 2–categorical nerve valued in marked simplicial sets, established by [Ozornova and Rovelli 2022], to construct a left Quillen functor

$$L\colon s\mathcal{S}et^{\Theta_2^{op}}_{(\infty,2)} \to ms\mathcal{S}et_{(\infty,2)}.$$

(ii) To show that this left Quillen functor is in fact a Quillen equivalence, we use a result of [Barwick and Schommer-Pries 2021] to reduce the problem to showing that it preserves cells in dimensions 0, 1, and 2. In Section 3 we use the intermediate comparisons of models from the diagram above to identify these cells in each model and thereby show that L does indeed preserve cells.

The outline of the paper is as follows. In Section 1 we recall some necessary results about model structures for 2–categories, Θ_2–spaces, and simplicial sets with marking, as well as functors between them, such as suspension and nerve functors. In Section 2 we construct the adjunction between Θ_2–spaces and simplicial sets with marking and we show that it is a Quillen pair. We then describe how it follows from [Barwick and Schommer-Pries 2021] that this adjunction is indeed a Quillen equivalence, modulo an explicit identification of the cells in the two models. In Section 3 we then provide the desired identification of the cells in the two models. In Section 4 we discuss some applications of the main theorem.

Acknowledgements

The authors would like to thank Lennart Meier and Lyne Moser for helpful conversations and feedback on this project. Bergner was partially supported by NSF grant DMS-1906281. Ozornova thankfully acknowledges the financial support by the DFG grant OZ 91/2-1 with the project 442418934. Rovelli was partially supported by NSF grant DMS-2203915.

This material is based upon work supported by the National Science Foundation under grant DMS-1440140 while Ozornova and Rovelli were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2020 semester.

1 Models of $(\infty, 2)$–categories

We assume the reader is familiar with the basics of strict 2–category theory (see eg [Borceux 1994]) and with the language of model categories (see eg [Hirschhorn 2003; Hovey 1999]), and we now recall some further preliminary material that we need in this paper.

1.1 Strict 2–categories

The category $2\mathscr{C}at$ of 2–categories is defined as the category whose objects are (small) categories enriched over the category $\mathscr{C}at$ of 1–categories. In particular, a *2–category* $\mathscr{D}$ consists of a set of objects and, for any objects x and x', a 1–category $\mathrm{Hom}_{\mathscr{D}}(x, x')$ together with a horizontal composition that defines a functor of hom-categories $\circ\colon \mathrm{Hom}_{\mathscr{D}}(x, x') \times \mathrm{Hom}_{\mathscr{D}}(x', x'') \to \mathrm{Hom}_{\mathscr{D}}(x, x'')$.

We consider the following model structure on $2\mathscr{C}at$ that was constructed by [Lack 2002, Theorem 3.3] (with a correction in [Lack 2004, Theorem 4]).

Theorem 1.1 *The category $2\mathscr{C}at$ of 2–categories supports a model structure in which*

- *all 2–categories are fibrant, and*
- *the weak equivalences are precisely the biequivalences of 2–categories.*

An important source of examples of 2–categories is given by suspending 1–categories, as follows.

Definition 1.2 Let $\mathscr{D}$ be a 1–category. The *suspension* of $\mathscr{D}$ is the 2–category $\Sigma\mathscr{D}$ in which

(a) there are two objects $x_\perp$ and $x_\top$;

(b) the hom-1–categories are given by

$$\mathrm{Hom}_{\Sigma\mathscr{D}}(a,b) := \begin{cases} \mathscr{D} & \text{if } a = x_\perp \text{ and } b = x_\top, \\ [0] & \text{if } a = b, \\ \varnothing & \text{if } a = x_\top \text{ and } b = x_\perp; \end{cases}$$

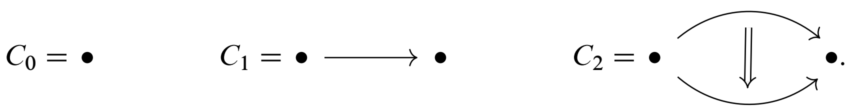
(c) there is no nontrivial horizontal composition.

This construction extends to a functor $\Sigma\colon \mathscr{C}at \to 2\mathscr{C}at_{*,*}$ valued in the category of bipointed categories, namely categories endowed with a pair of (possibly equal) specified objects, and basepoint-preserving functors.

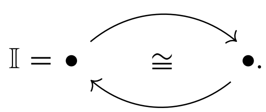
The 2–categorical suspension $\Sigma\mathscr{D}$ appears in [Barwick and Schommer-Pries 2021], where it is denoted by $\sigma(\mathscr{D})$. It is also often described in the literature as a special case of a simplicial suspension. For instance, applying the nerve to hom-categories of the suspension $\Sigma\mathscr{D}$ gives a simplicial category $N_*(\Sigma\mathscr{D})$ that agrees with what was denoted by $U(N\mathscr{D})$ in [Bergner 2007b], as $S(N\mathscr{D})$ in [Joyal 2007], as $[1]_{N\mathscr{D}}$ in [Lurie 2009a], and as $\mathbb{2}[N\mathscr{D}]$ in [Riehl and Verity 2020].

Notation 1.3 We record some notation for the following (nondisjoint) families of 2–categories.

- For $m \geq -1$, we denote by $[m]$ the finite ordinal with $m+1$ elements.
- For $j = 0, 1, 2$, we denote by C_j the free j–cell. These 2–categories can be pictured as



- For $m \geq 0$ and $k_1, \dots, k_m \geq 0$, we denote by $[m|k_1, \dots, k_m]$ a generic object of Joyal's cell category Θ_2, namely the full subcategory Θ_2 of $2\mathscr{C}at$ from [Joyal 1997].
- We denote by $\mathbb{I}$ the free-living isomorphism category. This category can be pictured as



1.2 Complete Segal spaces as a model for $(\infty, 1)$–categories

We briefly recall the theory of complete Segal spaces, as first defined by [Rezk 2001], of which the next model we discuss for $(\infty, 2)$–categories is a generalization.

First, consider functors $X\colon \Delta^{\mathrm{op}} \to s\mathcal{S}et$. For any $n \geq 1$, consider the *Segal map*

$$X_n \to \underbrace{X_1 \times_{X_0} \cdots \times_{X_0} X_1}_{n}$$

induced by the inclusion

$$\underbrace{\Delta[1] \amalg_{\Delta[0]} \Delta[1] \amalg_{\Delta[0]} \cdots \amalg_{\Delta[0]} \Delta[1]}_{n} \to \Delta[n]$$

of the spine of the n–simplex into the n–simplex $\Delta[n]$.

Definition 1.4 A *Segal space* is a functor $X\colon \Delta^{\mathrm{op}} \to s\mathcal{S}et$ such that the Segal maps are weak equivalences of simplicial sets for all $n \geq 1$.

The idea is that a Segal space behaves something like a category, with simplicial sets of objects and morphisms, but with composition defined only up to homotopy.

However, to have a model for $(\infty,1)$–categories, we do not want a simplicial set of objects, as in an internal category, but instead a discrete set of objects. The most straightforward way to get such a model is to ask for the simplicial set X_0 to be discrete.

Definition 1.5 A *Segal precategory* is a functor $X\colon \Delta^{\mathrm{op}} \to s\mathcal{S}et$ such that X_0 is a discrete simplicial set. We denote by $P\mathcal{C}at$ the full subcategory of $s\mathcal{S}et^{\Delta^{\mathrm{op}}}$ spanned by all Segal precategories. A *Segal category* is a Segal precategory that is also a Segal space.

There are two model structures for Segal precategories, the first of which has all objects cofibrant and is originally due to [Pellissier 2002, Theorem 6.4.4]; another proof is given in [Bergner 2007a, Theorem 5.1]. However, in this paper we make use of the following model structure that has cofibrations defined similarly to those in the projective model structure.

Theorem 1.6 [Bergner 2007a, Theorem 7.1; Bergner 2007c, Theorem 4.2] *The category $P\mathcal{C}at$ of Segal precategories admits a model structure in which*

- *the fibrant objects are the projectively fibrant Segal categories, and*
- *the cofibrations are projective cofibrations.*

We denote this model structure by $P\mathcal{C}at_{(\infty,1)}$.

However, from the point of view of homotopy theory, asking for discreteness is awkward. The completeness condition that we now describe can be more convenient from this perspective.

Let $N\mathbb{I}$ denote the nerve of the groupoid $\mathbb{I}$, and denote by X_{heq} the simplicial set $\mathrm{Hom}(N\mathbb{I}, X)$, which is sometimes called the space of *homotopy equivalences* of X. The unique map $N\mathbb{I} \to \Delta[0]$ induces a map

$$X_0 \to X_{\mathrm{heq}}.$$

Definition 1.7 A Segal space is *complete* if this map $X_0 \to X_{\mathrm{heq}}$ is a weak equivalence of simplicial sets.

Rezk builds a supporting model structure for the homotopy theory of complete Segal spaces.

Theorem 1.8 [Rezk 2001, Theorem 7.2] *The category $s\mathcal{S}et^{\Delta^{\mathrm{op}}}$ of simplicial spaces admits a model structure in which*

- *the fibrant objects are the injectively fibrant complete Segal spaces, and*
- *the cofibrations are the monomorphisms.*

We denote this model structure by $s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}$.

This model structure can be obtained by taking the left Bousfield localization of the injective model structure on $s\mathcal{S}et^{\Delta^{\mathrm{op}}}$ with respect to the following set of maps:

(1) the *Segal acyclic cofibrations*

$$\underbrace{\Delta[1] \amalg_{\Delta[0]} \Delta[1] \amalg_{\Delta[0]} \cdots \amalg_{\Delta[0]} \Delta[1]}_{n} \to \Delta[n]$$

for $n \geq 1$, and

(2) the *completeness cofibration*, given by either inclusion of the form

$$\Delta[0] \to N\mathbb{I}.$$

Complete Segal spaces, the fibrant objects in $s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}$, are then precisely the injectively fibrant simplicial spaces that are local with respect to the maps of type (1) and (2).

Remark 1.9 As briefly addressed in [Rezk 2010, Section 10], in presence of the maps of type (1), for the purpose of the localization one could replace the map of type (2) as completeness acyclic cofibration with

(2′) either inclusion of the form

$$\Delta[0] \to \Delta[0] \amalg_{\Delta[1]} \Delta[3] \amalg_{\Delta[1]} \Delta[0],$$

where the right-hand side is the colimit of the diagram

$$\Delta[0] \leftarrow \Delta[1] \xrightarrow{02} \Delta[3] \xleftarrow{13} \Delta[1] \to \Delta[0].$$

The following theorem establishes that the homotopy theories of Segal categories and complete Segal spaces are equivalent.

Theorem 1.10 [Bergner 2007a, Theorems 6.3 and 7.5] *The inclusion functor from the category of Segal precategories to the category of simplicial spaces induces a left Quillen equivalence*

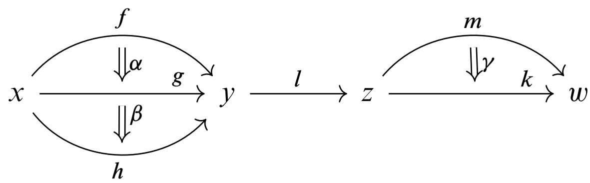
$$I\colon P\mathcal{C}at_{(\infty,1)} \to s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}.$$

1.3 Complete Segal Θ_2–spaces as a model of $(\infty, 2)$–categories

We now recall the notion of complete Segal Θ_2–spaces, which give a model for $(\infty, 2)$–categories.

Let Θ_2 be Joyal's cell category. For a precise account on how Θ_2 is defined, we refer the reader to the original source [Joyal 1997], or to [Berger 2007, Definition 3.3] or [Rezk 2010, Section 1.1] for an inductive approach; we give a brief review here.

Recall that Θ_2 is a full subcategory of $2\mathscr{C}at$ and that a generic object of Θ_2 is a 2–category $[m|k_1, \dots, k_m]$ generated by gluing horizontally the suspensions of $[k_i]$ for $i = 1, \dots, m$. An example is the 2–category $[3|2, 0, 1]$, which is generated by the following data:



Definition 1.11 A Θ_2–*set* is a presheaf $A\colon \Theta_2^{op} \to \mathscr{S}et$, and we denote the category of Θ_2–sets and natural transformations by $\mathscr{S}et^{\Theta_2^{op}}$. Similarly, a Θ_2–*space* is a simplicial presheaf $A\colon \Theta_2^{op} \to s\mathscr{S}et$, and we denote the category of Θ_2–spaces by $s\mathscr{S}et^{\Theta_2^{op}}$.

Remark 1.12 The reader familiar with [Rezk 2010] might observe that we are using the term "Θ_2–space" in a more general sense than he does. His Θ_2–spaces satisfy additional Segal and completeness conditions that we discuss below; we further specify such objects by calling them "complete Segal Θ_2–spaces".

Remark 1.13 The canonical inclusion $\mathscr{S}et \hookrightarrow s\mathscr{S}et$ of sets as discrete simplicial sets induces a canonical inclusion $\mathscr{S}et^{\Theta_2^{op}} \hookrightarrow s\mathscr{S}et^{\Theta_2^{op}}$, which is both a left and right adjoint. In particular, we often regard Θ_2–sets as discrete Θ_2–spaces without further specification.

Notation 1.14 For any object θ of Θ_2, we denote by $\Theta_2[\theta]$ the Θ_2–set represented by θ.

Remark 1.15 As a special case of [Ara 2014, Section 3.1], given any Θ_2–set A and any space B one can consider the Θ_2–space $A \boxtimes B$, which is defined levelwise as the simplicial set

$$(A \boxtimes B)_\theta := A_\theta \times B.$$

The construction extends to a bifunctor

$$\boxtimes\colon \mathscr{S}et^{\Theta_2^{op}} \times s\mathscr{S}et \to s\mathscr{S}et^{\Theta_2^{op}}$$

that preserves colimits in each variable.

In preparation for a localization on the category $s\mathcal{S}et^{\Theta_2^{op}}$, we introduce the following class of maps. The reader may notice the analogy with the maps treated in Section 1.2.

Definition 1.16 An *elementary acyclic cofibration* is a map of discrete Θ_2–spaces of the following kinds.

(1) A *vertical Segal acyclic cofibration* is given by, for some $k \geq 1$, the canonical map
$$\Theta_2[1|1] \amalg_{\Theta_2[1|0]} \cdots \amalg_{\Theta_2[1|0]} \Theta_2[1|1] \hookrightarrow \Theta_2[1|k].$$

(2) A *horizontal Segal acyclic cofibration* is given by, for some $m \geq 1$ and $k_i \geq 0$, where $0 \leq i \leq m$, the canonical map
$$\Theta_2[1|k_1] \amalg_{\Theta_2[0]} \cdots \amalg_{\Theta_2[0]} \Theta_2[1|k_m] \hookrightarrow \Theta_2[m|k_1, \ldots, k_m].$$

(3) The *horizontal completeness acyclic cofibration* is either of the inclusions of the form
$$\Theta_2[0] \to \Theta_2[0] \amalg_{\Theta_2[1|0]} \Theta_2[3|0,0,0] \amalg_{\Theta_2[1|0]} \Theta_2[0],$$
where the right-hand side is the colimit of the diagram
$$\Theta_2[0] \leftarrow \Theta_2[1|0] \xrightarrow{02} \Theta_2[3|0,0,0] \xleftarrow{13} \Theta_2[1|0] \to \Theta_2[0].$$

(4) The *vertical completeness acyclic cofibration* is the canonical map
$$\Theta_2[1|0] \to \Theta_2[1|0] \amalg_{\Theta_2[1|1]} \Theta_2[1|3] \amalg_{\Theta_2[1|1]} \Theta_2[1|0],$$
induced by suspending the previous one.

We now describe two model structures on the category $s\mathcal{S}et^{\Theta_2^{op}}$, both established by [Rezk 2010, Section 2.13, Proposition 11.5]. Our description, in terms of the elementary acyclic cofibrations defined above, differs slightly from his, but is designed to facilitate some of our proofs in the next section. We explain in Remark 1.21 why the two approaches give the same model structures.

Theorem 1.17 *The category $s\mathcal{S}et^{\Theta_2^{op}}$ of Θ_2–spaces supports the following two cofibrantly generated model structures*:

- *the model structure*
$$s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$$
obtained by taking the left Bousfield localization of the injective model structure $s\mathcal{S}et^{\Theta_2^{op}}_{inj}$ with respect to the set of elementary acyclic cofibrations from Definition 1.16, and
- *the model structure*
$$s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}$$
obtained by taking the left Bousfield localization of the projective model structure $s\mathcal{S}et^{\Theta_2^{op}}_{proj}$ with respect to the set of elementary acyclic cofibrations from Definition 1.16.

Although the model structure $s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$ is more common in the literature, for technical reasons that we discuss in Remark 2.3, in this paper we focus more on $s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}$. In this model structure

- the projectively fibrant objects, which we call *complete Segal Θ_2–spaces*, are precisely the projectively fibrant Θ_2–spaces that are local with respect to the elementary acyclic cofibrations from Definition 1.16, and
- the cofibrations are precisely the projective cofibrations.

Remark 1.18 Combining [Hirschhorn 2003, Theorem 11.6.1 and Definitions 11.5.33 and 11.5.25], we can obtain an explicit description of the generating cofibrations and generating acyclic cofibrations of $s\mathcal{S}et^{\Theta_2^{op}}_{\text{proj}}$. In particular,

(1) a set of generating cofibrations for the projective model structure on $s\mathcal{S}et^{\Theta_2^{op}}$ is given by all maps of the form

$$\Theta_2[\theta] \boxtimes \partial\Delta[\ell] \to \Theta_2[\theta] \boxtimes \Delta[\ell] \quad \text{for } \theta \in \text{Ob}(\Theta_2) \text{ and } \ell \geq 0;$$

(2) a set of generating acyclic cofibrations for the projective model structure on $s\mathcal{S}et^{\Theta_2^{op}}$ is given by all maps of the form

$$\Theta_2[\theta] \boxtimes \Lambda^k[\ell] \to \Theta_2[\theta] \boxtimes \Delta[\ell] \quad \text{for } \theta \in \text{Ob}(\Theta_2) \text{ and } 0 \leq k \leq \ell.$$

The following equivalence between the two model structures can alternatively also be seen as a direct application of [Hirschhorn 2003, Theorem 3.3.20].

Theorem 1.19 [Rezk 2010, Sections 2.5–2.13] *The identity functor defines a Quillen equivalence*

$$s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)} \rightleftarrows s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}.$$

We want to consider the suspension of a simplicial space to a Θ_2–space. In [Rezk 2010, Section 4.4], the notation $V[1](X)$ is used for what we denote here by ΣX to emphasize the analogy with similar constructions we have discussed.

Definition 1.20 The *suspension* ΣX of a simplicial space X is the Θ_2–space obtained by applying the cocontinuous functor $\Sigma\colon s\mathcal{S}et^{\Delta^{op}} \to s\mathcal{S}et^{\Theta_2^{op}}_{*,*}$ defined on representable simplicial spaces as

$$\Sigma(\Delta[k] \boxtimes \Delta[\ell]) := \Theta_2[1|k] \boxtimes \Delta[\ell].$$

This construction extends to a functor $\Sigma\colon s\mathcal{S}et^{\Delta^{op}} \to s\mathcal{S}et^{\Theta_2^{op}}_{*,*}$ valued in bipointed Θ_2–spaces.

Remark 1.21 In the original construction from [Rezk 2010, Section 2.13, Proposition 11.5], two model structures on $s\mathcal{S}et^{\Theta_2^{op}}$ are obtained by localizing the injective and projective model structure with respect to the set of maps of the following kinds:

(1′) a family of maps that can be recognized to be precisely the family of vertical Segal acyclic cofibrations, using [Rezk 2010, Proposition 11.7];

(2′) a family of maps that can be recognized to be precisely the family of horizontal Segal acyclic cofibrations, using [Rezk 2010, Proposition 11.7];

(3′) the unique map
$$N^{\Theta_2}\mathbb{I} \to \Theta_2[0];$$

(4′) the map
$$\Sigma N^{\Theta_2}\mathbb{I} \to \Theta_2[1|0]$$
obtained by suspending the map from (3′).

However, in presence of the maps of type (1) and (2), it is shown in [Rezk 2010, Section 10] and also in [Barwick and Schommer-Pries 2021, Section 13] that for the purpose of the localization the maps of type (3) and (4) are equivalent to the maps of type (3′) and (4′), respectively. It follows that, although presented differently, these two model structures in fact agree with the model structures $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,2)}$ and $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$ from Theorem 1.17.

1.4 Complicial sets as a model of $(\infty, 2)$–categories

The next model of $(\infty,2)$–categories that we consider is based on the following structure, originally referred to as a *simplicial set with hollowness* in [Street 1987] and later as a *stratified simplicial set* in [Verity 2008a].

Definition 1.22 A *simplicial set with marking* is a simplicial set endowed with a subset of simplices of strictly positive dimensions that contains all degenerate simplices, called *thin* or *marked*. We denote by $ms\mathcal{S}et$ the category of simplicial sets with marking and marking-preserving simplicial maps.

We want to consider a model structure on the category of simplicial sets with marking, in which the fibrant objects, called 2*–complicial sets*, provide a model for $(\infty,2)$–categories. The idea is that, in a 2–complicial set, the marked k–simplices are precisely the k–equivalences. We refer the reader to [Riehl 2018] for further elaboration on this viewpoint.

Remark 1.23 As discussed in [Verity 2008a, Observation 97], the underlying simplicial set functor $ms\mathcal{S}et \to s\mathcal{S}et$ fits into an adjoint triple
$$ms\mathcal{S}et \; \underset{\underset{(-)^\sharp}{\longleftarrow}}{\overset{\overset{(-)^\flat}{\longleftarrow}}{\underset{\perp}{\overset{\perp}{\longrightarrow}}}} \; s\mathcal{S}et.$$

For any simplicial set X, the left adjoint $X^\flat$ (sometimes also denoted simply by X) is obtained by marking only the degenerate simplices of X, and the right adjoint $X^\sharp$ is obtained by marking all simplices in positive dimensions.

Remark 1.24 As described in detail in [Verity 2008a, Observation 109], the category $ms\mathcal{S}et$ of simplicial sets with marking is complete and cocomplete, with limits and colimits constructed as follows.

- The underlying simplicial set of a limit $\lim_{i\in I} X_i$ of simplicial sets with marking is the limit of the corresponding underlying simplicial sets of X_i, and a simplex is marked in a limit of simplicial sets with marking $\lim_{i\in I} X_i$ if and only if it is marked in each component X_i for $i \in I$.
- The underlying simplicial set of a colimit $\mathrm{colim}_{i\in I} X_i$ of simplicial sets with marking is the colimit of the corresponding underlying simplicial sets of X_i, and a simplex is marked in a colimit of simplicial sets with marking $\mathrm{colim}_{i\in I} X_i$ if and only if it admits a marked representative in X_i for some $i \in I$.

The following model structure is one instance of the family of model structures constructed by [Verity 2008b, Theorem 100], and is described in more detail in [Riehl 2018, Section 3.3].

Theorem 1.25 [Ozornova and Rovelli 2020, Theorem 1.25] *The category $ms\mathcal{S}et$ of simplicial sets with marking supports a cofibrantly generated cartesian closed model structure in which*

- *the fibrant objects are the 2–complicial sets, as recalled in* [Ozornova and Rovelli 2020, Definition 1.21], *and*
- *the cofibrations are precisely the monomorphisms on underlying simplicial sets.*

We denote this model structure by $ms\mathcal{S}et_{(\infty,2)}$.

We warn the reader that the fibrant objects in this model structure have been given different names in the literature, and could perhaps more accurately be called "2–trivial saturated weak complicial sets". We have chosen to call them "2–complicial sets" for the sake of brevity; in what follows we do not make explicit use of their definition. We recall the key results we need, in particular about the weak equivalences in this model structure, in the remainder of this section.

Remark 1.26 Because of the way the model structure $ms\mathcal{S}et_{(\infty,2)}$ is constructed, if $\Delta[3]_{\mathrm{eq}}$ denotes the 3–simplex $\Delta[3]$ in which the nondegenerate marked 1–simplices are precisely the one between the vertices 0 and 2 and the one between the vertices 1 and 3, and all simplices in dimension 2 or higher are marked, the canonical map $\Delta[3]_{\mathrm{eq}} \to \Delta[3]^\sharp$ is a weak equivalence. Indeed, the model structure $ms\mathcal{S}et_{(\infty,2)}$ is a Cisinski–Olschok model structure (in the sense of [Olschok 2011]) for which the map $\Delta[3]_{\mathrm{eq}} \to \Delta[3]^\sharp$ is an anodyne extension.

Lemma 1.27 *The functor*

$$(-)^\sharp : s\mathcal{S}et_{(\infty,0)} \to ms\mathcal{S}et_{(\infty,2)}$$

is a left Quillen functor, where $s\mathcal{S}et_{(\infty,0)}$ denotes the Kan–Quillen model structure on the category $s\mathcal{S}et$.

Proof The fact that the functor admits a right adjoint, often called the *core functor*, is discussed in [Riehl and Verity 2022, Definition D.1.2]. It is straightforward from its description that the functor $(-)^\sharp$

preserves cofibrations, and it is shown in [Ozornova and Rovelli 2020, Lemma 2.16] that it also sends acyclic cofibrations of $s\mathcal{S}et_{(\infty,0)}$ to weak equivalences of $ms\mathcal{S}et_{(\infty,2)}$. It follows that $(-)^\sharp$ defines indeed a left Quillen functor between the desired model categories. □

For $n = 2$, the Street nerve was studied in detail by [Duskin 2001], and can be described explicitly as follows.

Definition 1.28 The *nerve* $N\mathfrak{D}$ of a 2–category $\mathfrak{D}$ is the 3–coskeletal simplicial set in which

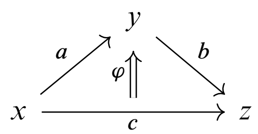
(0) a 0–simplex consists of an object of $\mathfrak{D}$:

$$x;$$

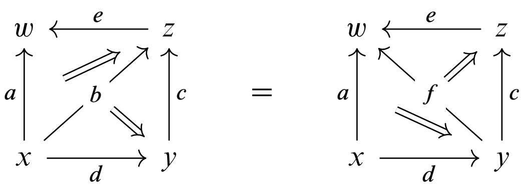
(1) a 1–simplex consists of a 1–morphism of $\mathfrak{D}$:

$$x \xrightarrow[a]{} y;$$

(2) a 2–simplex consists of a 2–cell of $\mathfrak{D}$ of the form $c \Rightarrow b \circ a$:



(3) a 3–simplex consists of four 2–cells of $\mathfrak{D}$ that satisfy the relation



and in which the simplicial structure is as indicated in the pictures.

Definition 1.29 [Verity 2008a, Chapter 10] The *Roberts–Street nerve* of a 2–category $\mathfrak{D}$ is the simplicial set with marking $N^{\mathrm{RS}}\mathfrak{D}$, defined by the following properties.

(0) The underlying simplicial set is the Duskin nerve $N\mathfrak{D}$.

(1) Only degenerate 1–simplices are marked.

(2) A 2–simplex of $N\mathfrak{D}$ is marked in $N^{\mathrm{RS}}\mathfrak{D}$ if and only if corresponding 2–morphism $\varphi\colon c \Rightarrow b \circ a$ is an identity.

(3) Any m–simplex of $N\mathfrak{D}$ for $m \geq 3$ is marked in $N^{\mathrm{RS}}\mathfrak{D}$.

This construction extends to a functor $N^{\mathrm{RS}}\colon 2\mathcal{C}at \to ms\mathcal{S}et$.

The Roberts–Street nerve is a right adjoint functor, but, as proved by the second- and third-named authors, does not preserve fibrant objects on the model structures we want to consider. However, it is a *homotopical* functor between model categories, in the sense that it preserves weak equivalences.

Proposition 1.30 [Ozornova and Rovelli 2022, Proposition 1.18] *The Roberts–Street nerve defines a homotopical functor of model categories*

$$N^{\mathrm{RS}} \colon 2\mathcal{C}at \to ms\mathcal{S}et_{(\infty,2)}.$$

The following two technical results essentially tell us that horizontal and vertical composition of 2–cells can be encoded via Segal-type maps that are acyclic cofibrations in the model structure for 2–complicial sets.

Theorem 1.31 [Ozornova and Rovelli 2022, Corollary 2.10] *For any $m \geq 0$ and $k_i \geq 0$ for $i = 1, \ldots, m$ there is a canonical map of simplicial sets with marking*

$$N^{\mathrm{RS}}[1|k_1] \coprod_{N^{\mathrm{RS}}[0]} \cdots \coprod_{N^{\mathrm{RS}}[0]} N^{\mathrm{RS}}[1|k_m] \hookrightarrow N^{\mathrm{RS}}[m|k_1, \ldots, k_m]$$

that is an acyclic cofibration, and in particular a weak equivalence, in $ms\mathcal{S}et_{(\infty,2)}$.

Theorem 1.32 [Ozornova and Rovelli 2022, Corollary 2.11] *For any $k \geq 0$ there is a canonical map of simplicial sets with marking*

$$N^{\mathrm{RS}}[1|1] \coprod_{N^{\mathrm{RS}}[1]} \cdots \coprod_{N^{\mathrm{RS}}[1]} N^{\mathrm{RS}}[1|1] \hookrightarrow N^{\mathrm{RS}}[1|k]$$

that is an acyclic cofibration, and in particular a weak equivalence, in $ms\mathcal{S}et_{(\infty,2)}$.

Note that, when taking the nerve we simply write $N^{\mathrm{RS}}[1]$ rather than $N^{\mathrm{RS}}[1|0]$, since the 2–category $[1|0]$ is just the category $[1]$ thought of as a 2–category.

An important construction in this paper is the suspension of a simplicial set with marking. We conclude this section with the definition and some key results about it.

Definition 1.33 [Ozornova and Rovelli 2022, Definition 2.6] The *suspension* ΣX of a simplicial set with marking X is the simplicial set with marking defined as follows.

- It has precisely two 0–simplices that we denote by $x_\perp$ and $x_\top$.
- The set of m–simplices for $m > 0$ is given by all k–simplices of X for $0 \leq k \leq m-1$, as well as the m–fold degeneracies of the two 0–simplices $x_\perp$ and $x_\top$, namely
$$(\Sigma X)_m \cong \{s_0^m x_\perp\} \amalg X_{m-1} \amalg \cdots \amalg X_0 \amalg \{s_0^m x_\top\}.$$
- The simplicial structure can be read off from [Ozornova and Rovelli 2022, Definition 2.6].
- The set of nondegenerate m–simplices for $m > 0$ is given by the nondegenerate $(m-1)$–simplices of X.
- A nondegenerate m–simplex σ is marked in ΣX if and only if it is marked as an $(m-1)$–simplex of X.

This construction extends to a functor $\Sigma \colon ms\mathcal{S}et \to ms\mathcal{S}et_{*,*}$ valued in bipointed marked simplicial sets.

We now recall that this functor can be upgraded to a left Quillen functor of model categories. Recall from [Hirschhorn 2021] that, given any cofibrantly model category $\mathcal{M}$, there is a model structure on the category $\mathcal{M}_{*,*}$ of bipointed objects in $\mathcal{M}$, in which cofibrations, fibrations, and weak equivalences are created in $\mathcal{M}$.

Lemma 1.34 [Ozornova and Rovelli 2022, Lemma 2.7] *Regarding ΣX as a simplicial set with marking bipointed on $x_\perp$ and $x_\top$, the marked suspension defines a left Quillen functor*

$$\Sigma\colon ms\mathcal{S}et_{(\infty,2)} \to (ms\mathcal{S}et_{(\infty,2)})_{*,*}.$$

In particular, it is homotopical and it respects connected colimits as a functor $\Sigma\colon ms\mathcal{S}et \to ms\mathcal{S}et$.

Finally, we recall that the suspension of a marked simplicial set is homotopically compatible with the Roberts–Street nerve, as one would expect.

Theorem 1.35 [Ozornova and Rovelli 2022, Theorem 2.9] *For any 1–category $\mathcal{D}$ there is a canonical map*

$$\Sigma N^{\mathrm{RS}}\mathcal{D} \to N^{\mathrm{RS}}\Sigma\mathcal{D}$$

that is a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$.

2 The comparison of models of $(\infty, 2)$–categories

In this section, we set up our explicit comparison between the two models for $(\infty, 2)$–categories that we are considering. We first establish the desired Quillen pair of functors between the unlocalized model structure on the category of Θ_2–spaces and the model structure on simplicial sets with marking, then show that it is still a Quillen pair after localization of the former model category. We then show that it is a Quillen equivalence, deferring some steps in the proof to later sections.

2.1 The Quillen pair before localizing

Let us begin by defining the functor that we use to make our comparison.

Construction 2.1 The functor $\Theta_2 \times \Delta \subseteq s\mathcal{S}et^{\Theta_2^{\mathrm{op}}} \to ms\mathcal{S}et$ given by

$$(\theta, [\ell]) \mapsto (\Theta_2 \times \Delta)[\theta, \ell] = \Theta_2[\theta] \boxtimes \Delta[\ell] \mapsto N^{\mathrm{RS}}\theta \times \Delta[\ell]^{\sharp}$$

induces an adjunction

$$L\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}} \rightleftarrows ms\mathcal{S}et : R.$$

Roughly speaking, for any Θ_2–space W, the simplicial set with marking LW is obtained by gluing together a copy of the Roberts–Street nerve of θ, for any θ in Θ_2 that maps to W. While describing this gluing explicitly is complicated, it is essentially specified by the definition of left Kan extension.

We now show that these adjoint functors define Quillen pair on unlocalized model categories.

Proposition 2.2 *The adjunction*

$$L\colon s\mathcal{S}et_{\mathrm{proj}}^{\Theta_2^{\mathrm{op}}} \rightleftarrows ms\mathcal{S}et_{(\infty,2)} : R$$

is a Quillen pair.

Proof We want to show that the functor L preserves cofibrations and acyclic cofibrations. From Remark 1.18 we know that

(1) a set of generating cofibrations for the projective model structure on $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}$ is given by all maps of the form

$$\Theta_2[\theta] \boxtimes \partial\Delta[\ell] \to \Theta_2[\theta] \boxtimes \Delta[\ell] \quad \text{for } \theta \in \Theta_2 \text{ and } \ell \geq 0;$$

(2) a set of generating acyclic cofibrations for the projective model structure on $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}$ is given by all maps of the form

$$\Theta_2[\theta] \boxtimes \Lambda^k[\ell] \to \Theta_2[\theta] \boxtimes \Delta[\ell] \quad \text{for } \theta \in \Theta_2 \text{ and } 0 \leq k \leq \ell.$$

Using the facts that $(-)^\sharp$ commutes with colimits, which is a consequence of Lemma 1.27, and that the box product $\boxtimes$ preserves colimits in each variable, which was recalled in Remark 1.15, we see that

(1) the image of the generating cofibration via L is the map

$$N^{\mathrm{RS}}\theta \times \partial\Delta[\ell]^\sharp \to N^{\mathrm{RS}}\theta \times \Delta[\ell]^\sharp \quad \text{for } \theta \in \Theta_2 \text{ and } \ell \geq 0;$$

(2) the image of the generating acyclic cofibration via L is the map

$$N^{\mathrm{RS}}\theta \times \Lambda^k[\ell]^\sharp \to N^{\mathrm{RS}}\theta \times \Delta[\ell]^\sharp \quad \text{for } \theta \in \Theta_2 \text{ and } 0 \leq k \leq \ell.$$

Since the model structure $ms\mathcal{S}et_{(\infty,2)}$ is cartesian closed by Theorem 1.25 and $(-)^\sharp$ is a left Quillen functor by Lemma 1.27, we conclude that

(1) the map $N^{\mathrm{RS}}\theta \times \partial\Delta[\ell]^\sharp \to N^{\mathrm{RS}}\theta \times \Delta[\ell]^\sharp$ is a cofibration, and
(2) the map $N^{\mathrm{RS}}\theta \times \Lambda^k[\ell]^\sharp \to N^{\mathrm{RS}}\theta \times \Delta[\ell]^\sharp$ is an acyclic cofibration.

It follows that L preserves cofibrations and acyclic cofibrations, so it is a left Quillen functor, as desired. □

Remark 2.3 One might wonder, in contrast with much of the literature on the subject, why we have chosen to use the projective, rather than the injective, model structure on $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}$. However, it is not clear whether the functor

$$L\colon s\mathcal{S}et_{\mathrm{inj}}^{\Theta_2^{\mathrm{op}}} \to ms\mathcal{S}et_{(\infty,2)}$$

is a left Quillen functor, since we do not know whether it preserves cofibrations. More precisely, it is unclear whether L sends the injective cofibration

$$\partial\Theta_2[3|1,0,1] \to \Theta_2[3|1,0,1]$$

to a cofibration of $ms\mathcal{S}et_{(\infty,2)}$.

2.2 The Quillen pair after localizing

We now show that we still have a Quillen pair after localizing the projective model structure on $s\mathcal{S}et^{\Theta_2^{op}}$.

Theorem 2.4 *The adjunction*

$$L: s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)} \rightleftarrows ms\mathcal{S}et_{(\infty,2)} : R$$

is a Quillen pair.

Since cofibrations are unchanged by localization, it suffices to prove that L preserves acyclic cofibrations. We do so by proving that L preserves all elementary acyclic cofibrations, in the following sequence of propositions.

Proposition 2.5 *The functor L sends the vertical Segal acyclic cofibrations*

$$\Theta_2[1|1] \coprod_{\Theta_2[1|0]} \cdots \coprod_{\Theta_2[1|0]} \Theta_2[1|1] \hookrightarrow \Theta_2[1|k] \text{ for } k \geq 0$$

from Definition 1.16 to weak equivalences in $ms\mathcal{S}et_{(\infty,2)}$.

Proof The functor L sends the elementary acyclic cofibration

$$\Theta_2[1|1] \coprod_{\Theta_2[1|0]} \cdots \coprod_{\Theta_2[1|0]} \Theta_2[1|1] \hookrightarrow \Theta_2[1|k]$$

to the canonical inclusion

$$N^{RS}[1|1] \coprod_{N^{RS}[1]} \cdots \coprod_{N^{RS}[1]} N^{RS}[1|1] \hookrightarrow N^{RS}[1|k],$$

which is an acyclic cofibration by Theorem 1.32. □

Proposition 2.6 *The functor L sends the horizontal Segal acyclic cofibrations*

$$\Theta_2[1|k_1] \coprod_{\Theta_2[0]} \cdots \coprod_{\Theta_2[0]} \Theta_2[1|k_m] \hookrightarrow \Theta_2[m|k_1,\dots,k_m] \text{ for } m \geq 0 \text{ and } k_i \geq 0$$

from Definition 1.16 to weak equivalences in $ms\mathcal{S}et_{(\infty,2)}$.

Proof The functor L sends the elementary acyclic cofibration

$$\Theta_2[1|k_1] \coprod_{\Theta_2[0]} \cdots \coprod_{\Theta_2[0]} \Theta_2[1|k_m] \hookrightarrow \Theta_2[m|k_1,\dots,k_m]$$

to the canonical inclusion

$$N^{RS}[1|k_1] \coprod_{N^{RS}[0]} \cdots \coprod_{N^{RS}[0]} N^{RS}[1|k_m] \hookrightarrow N^{RS}[m|k_1,\dots,k_m],$$

which is an acyclic cofibration by Theorem 1.31. □

Proposition 2.7 *The functor L sends the horizontal completeness acyclic cofibration*

$$\Theta_2[0] \to \Theta_2[0] \coprod_{\Theta_2[1|0]} \Theta_2[3|0,0,0] \coprod_{\Theta_2[1|0]} \Theta_2[0]$$

from Definition 1.16 to a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$.

To prove this proposition, we need the following preliminary lemma.

Lemma 2.8 *The unique map*

$$\Delta[0] \amalg_{\Delta[1]} N^{\mathrm{RS}}[3] \amalg_{\Delta[1]} \Delta[0] \to \Delta[0]$$

is a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$.

Proof We observe that this map fits into a commutative diagram of simplicial sets with marking

$$\begin{array}{ccc} \Delta[0] \amalg_{\Delta[1]} N^{\mathrm{RS}}[3] \amalg_{\Delta[1]} \Delta[0] & \longrightarrow & \Delta[0] \\ \uparrow & & \uparrow \\ \Delta[1]^\sharp \amalg_{\Delta[1]} N^{\mathrm{RS}}[3] \amalg_{\Delta[1]} \Delta[1]^\sharp \cong \Delta[3]_{\mathrm{eq}} & \longrightarrow & \Delta[3]^\sharp \end{array}$$

where $\Delta[1]^\sharp$ denotes the standard 1–simplex with the maximal marking. In this diagram, we observe that

- the bottom horizontal map is an acyclic cofibration by Remark 1.26;
- the left vertical map is a map between (homotopy) pushouts induced by the identity of $N^{\mathrm{RS}}[3]$ and two copies of the weak equivalence $\Delta[1]^\sharp \to \Delta[0]$; and
- the right vertical map is a weak equivalence since $(-)^\sharp$ preserves weak equivalences by Lemma 1.27.

It follows by two-out-of-three that the top horizontal map is a weak equivalence, as desired. □

We can now use this lemma to prove Proposition 2.7.

Proof of Proposition 2.7 The functor L sends the map

$$\Theta_2[0] \to \Theta_2[0] \amalg_{\Theta_2[1|0]} \Theta_2[3|0,0,0] \amalg_{\Theta_2[1|0]} \Theta_2[0]$$

to a map

$$\Delta[0] \to \Delta[0] \amalg_{\Delta[1]} N^{\mathrm{RS}}[3] \amalg_{\Delta[1]} \Delta[0]$$

that we want to show is a weak equivalence. However, we can conclude this fact by the two-out-of-three property, since we know from Lemma 2.8 that the unique map

$$\Delta[0] \amalg_{\Delta[1]} N^{\mathrm{RS}}[3] \amalg_{\Delta[1]} \Delta[0] \to \Delta[0]$$

is a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$. □

To complete the proof of Theorem 2.4, it remains to show that L preserves one more acyclic cofibration.

Proposition 2.9 *The functor L sends the vertical completeness acyclic cofibration*

$$\Theta_2[1|0] \to \Theta_2[1|0] \amalg_{\Theta_2[1|1]} \Theta_2[1|3] \amalg_{\Theta_2[1|1]} \Theta_2[1|0]$$

from Definition 1.16 to a weak equivalence of $ms\mathcal{S}et_{(\infty,2)}$.

Proof The functor L sends the map

$$\Theta_2[1|0] \to \Theta_2[1|0] \coprod_{\Theta_2[1|1]} \Theta_2[1|3] \coprod_{\Theta_2[1|1]} \Theta_2[1|0]$$

to a map

$$N^{\mathrm{RS}}[1|0] \to N^{\mathrm{RS}}[1|0] \coprod_{N^{\mathrm{RS}}[1|1]} N^{\mathrm{RS}}[1|3] \coprod_{N^{\mathrm{RS}}[1|1]} N^{\mathrm{RS}}[1|0],$$

which we want to show is a weak equivalence. By the two-out-of-three property, it suffices to show that the map

$$N^{\mathrm{RS}}[1|0] \coprod_{N^{\mathrm{RS}}[1|1]} N^{\mathrm{RS}}[1|3] \coprod_{N^{\mathrm{RS}}[1|1]} N^{\mathrm{RS}}[1|0] \to N^{\mathrm{RS}}[1|0],$$

induced by the unique map $[1|3] \to [1|0]$ in Θ_2 that is bijective on objects, is a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$. This map can be rewritten in terms of suspensions of 1–categories, as in Definition 1.2, as

$$N^{\mathrm{RS}}\Sigma[0] \coprod_{N^{\mathrm{RS}}\Sigma[1]} N^{\mathrm{RS}}\Sigma[3] \coprod_{N^{\mathrm{RS}}\Sigma[1]} N^{\mathrm{RS}}\Sigma[0] \to N^{\mathrm{RS}}\Sigma[0].$$

By Theorem 1.35, this map fits into a commutative diagram of simplicial sets with marking

$$\begin{array}{ccc} N^{\mathrm{RS}}\Sigma[0] \coprod_{N^{\mathrm{RS}}\Sigma[1]} N^{\mathrm{RS}}\Sigma[3] \coprod_{N^{\mathrm{RS}}\Sigma[1]} N^{\mathrm{RS}}\Sigma[0] & \longrightarrow & N^{\mathrm{RS}}\Sigma[0] \\ \simeq\uparrow & & \uparrow\cong \\ \Sigma N^{\mathrm{RS}}[0] \coprod_{\Sigma N^{\mathrm{RS}}[1]} \Sigma N^{\mathrm{RS}}[3] \coprod_{\Sigma N^{\mathrm{RS}}[1]} \Sigma N^{\mathrm{RS}}[0] & \longrightarrow & \Sigma N^{\mathrm{RS}}[0] \end{array}$$

in which the two vertical maps are weak equivalences. Note that for the left-hand map we are using the fact that these pushouts are actually homotopy pushouts. In particular, by the two-out-of-three property, to prove the theorem it is enough to prove that the bottom map is a weak equivalence. Using the fact that suspension commutes with pushouts by Lemma 1.34, this map can be rewritten as

$$\Sigma\left(\Delta[0] \coprod_{\Delta[1]} N^{\mathrm{RS}}[3] \coprod_{\Delta[1]} \Delta[0]\right) \to \Sigma\Delta[0],$$

namely the suspension of the map

$$\Delta[0] \coprod_{\Delta[1]} N^{\mathrm{RS}}[3] \coprod_{\Delta[1]} \Delta[0] \to \Delta[0],$$

which was shown in Lemma 2.8 to be a weak equivalence. Since suspension is a left Quillen functor by Lemma 1.34, we are done. □

2.3 The Quillen equivalence

It remains to show that this Quillen pair is in fact a Quillen equivalence. Our proof, however, is not done directly via the definition, but instead uses some machinery due to [Barwick and Schommer-Pries 2021] that we now briefly recall.

The first thing we need to consider is their criterion for when a model category is a "model of $(\infty, 2)$–categories". We begin with some notation.

Notation 2.10 Given a model category $\mathcal{M}$, we denote by $\mathcal{M}_\infty$ the underlying $(\infty,1)$–category of $\mathcal{M}$. While we do not need one here, for explicit (different but equivalent) constructions of $\mathcal{M}_\infty$ in the model of quasicategories, we refer the reader to [Hinich 2016] or [Lurie 2009a, Section A.2].

Notation 2.11 Given a Quillen pair $F\colon \mathcal{M} \rightleftarrows \mathcal{M}' : G$ between model categories $\mathcal{M}$ and $\mathcal{M}'$, we denote by

$$F_\infty\colon \mathcal{M}_\infty \rightleftarrows \mathcal{M}'_\infty : G_\infty$$

the adjunction that (F,G) induces at the level of underlying $(\infty,1)$–categories.

On objects, the value of F_∞ on any object of $\mathcal{M}$ can be computed up to equivalence in $\mathcal{M}'_\infty$ by applying F to any cofibrant replacement of the given object. Similarly the value of G_∞ on any object of $\mathcal{M}'$ can be computed up to equivalence in $\mathcal{M}_\infty$ by applying G to any fibrant replacement of the given object. Moreover, if (F,G) is a Quillen equivalence, the induced adjunction (F_∞, G_∞) is an equivalence of $(\infty,1)$–categories. For more details on how to obtain this adjunction of $(\infty,1)$–categories in the model of quasicategories we refer the reader to [Hinich 2016, Proposition 1.5.1].

Definition 2.12 (Barwick–Schommer-Pries) A model category $\mathcal{M}$ is a *model of* $(\infty,2)$*–categories* if the underlying $(\infty,1)$–category is equivalent to the colossal model $\mathcal{K}$ from [Barwick and Schommer-Pries 2021, Section 8], namely if there exists an equivalence of $(\infty,1)$–categories

$$\mathcal{M}_\infty \simeq \mathcal{K}.$$

The colossal model is constructed as an $(\infty,1)$–category in [Barwick and Schommer-Pries 2021, Section 8]. As we discuss in the appendix, with standard techniques one can also present the colossal model as the underlying $(\infty,1)$–category of a model category. More precisely, we show as Theorem A.3 that it is the underlying $(\infty,1)$–category $\big(s\mathcal{S}et^{\Upsilon_2^{op}}_{(\infty,2)}\big)_\infty$ of a model category $s\mathcal{S}et^{\Upsilon_2^{op}}_{(\infty,2)}$.

In any case, for the main purpose of this paper the arguments are packaged in a way that no explicit construction for the colossal model is needed.

Theorem 2.13 *The model categories* $s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}$ *and* $s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$ *are models of* $(\infty,2)$*–categories.*

Proof The fact that $s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$ is a model of $(\infty,2)$–categories is shown in [Barwick and Schommer-Pries 2021, Section 13] and there is an equivalence

$$\big(s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}\big)_\infty \simeq \big(s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}\big)_\infty$$

induced by the Quillen equivalence from Theorem 1.19. □

Theorem 2.14 *The model category* $ms\mathcal{S}et_{(\infty,2)}$ *is a model of* $(\infty,2)$*–categories.*

Proof An equivalence of $(\infty,1)$–categories

$$(ms\mathcal{S}et_{(\infty,2)})_\infty \simeq \big(s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}\big)_\infty$$

can be obtained combining several equivalences of $(\infty,1)$–categories induced by Quillen equivalences due to [Bergner and Rezk 2013; 2020; Gagna et al. 2022; Joyal and Tierney 2007; Lurie 2009b], as we recalled in Figure 1. □

Next, we recall the definition of a j–cell in a model of $(\infty,2)$–categories.

Definition 2.15 Let $\mathcal{M}$ be a model category that is a model for $(\infty,2)$–categories. An object of $\mathcal{M}$ is a representative of the *j–cell* for $j=0,1,2$ if it corresponds to the j–cell of the colossal model through any equivalence of $(\infty,1)$–categories $\mathcal{M}_\infty \simeq \big(s\mathcal{S}et^{\Upsilon_2^{op}}_{(\infty,2)}\big)_\infty$.

For completeness, the definitions of the 0–, 1– and 2–cells in the colossal model are recalled in the appendix, but will not be needed explicitly.

Remark 2.16 The object in $\mathcal{M}$ that represents the j–cell is unique up to equivalence in $\mathcal{M}_\infty$ and also up to isomorphism in $\operatorname{Ho}\mathcal{M}$, the homotopy category of $\mathcal{M}$. The definition makes sense in particular because any auto-equivalence of $\big(s\mathcal{S}et^{\Upsilon_2^{op}}_{(\infty,2)}\big)_\infty$ preserves j–cells for $j=0,1,2$, as shown in [Barwick and Schommer-Pries 2021, Theorem 7.3].

The following statements describe j–cells in $s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}$ and $ms\mathcal{S}et_{(\infty,2)}$.

Proposition 2.17 *In $s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}$ the object $\Theta_2[C_j]$ is a representative of the j–cell for $j=0,1,2$.*

Proposition 2.18 *In $ms\mathcal{S}et_{(\infty,2)}$ the object $N^{\mathrm{RS}}C_j$ is a representative of the j–cell for $j=0,1,2$.*

Although the two statements are not surprising, the argument to identify cells in $ms\mathcal{S}et_{(\infty,2)}$ requires significant work and makes use of many external results. We therefore postpone both proofs to Section 3.

Finally, the following theorem is the key ingredient to prove that the functor L is a Quillen equivalence.

Theorem 2.19 [Barwick and Schommer-Pries 2021, Proposition 15.10] *Let $\mathcal{M}$ and $\mathcal{N}$ be model categories that are models for $(\infty,2)$–categories, and $L\colon \mathcal{M} \rightleftarrows \mathcal{N} : R$ a Quillen pair between them. Then the Quillen pair (L,R) is a Quillen equivalence if and only if the derived functor of L sends j–cells to j–cells for $j=0,1,2$.*

Once the proofs of Propositions 2.17 and 2.18 are provided in Section 3, we can then apply Theorem 2.19 to the Quillen pair from Theorem 2.4 to conclude the desired Quillen equivalence.

Theorem 2.20 *The adjunction*

$$L\colon s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)} \rightleftarrows ms\mathcal{S}et_{(\infty,2)} : R$$

is a Quillen equivalence, and in particular induces an equivalence of $(\infty,1)$–categories

$$L_\infty\colon \big(s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}\big)_\infty \rightleftarrows (ms\mathcal{S}et_{(\infty,2)})_\infty : R_\infty.$$

3 Recognizing cells in models of $(\infty, 2)$–categories

The goal of this section is to identify the j–cells in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$, and most importantly the j–cells in $ms\mathcal{S}et_{(\infty,2)}$, as defined in Definition 2.15. The structure of the argument involves the identification of the j–cells in several established model categories that are models of $(\infty, 2)$–categories.

In Figure 2, we display the equivalences used to identify the cells in the marked simplicial sets, and the propositions displayed show how the cells behave under the corresponding equivalence.

While it is impractical to make this section completely self-contained, we have included precise references for all relevant constructions and definitions.

Lemma 3.1 *Suppose that a functor $F\colon \mathcal{M} \to \mathcal{M}'$ is a left (resp. right) Quillen equivalence between models of $(\infty, 2)$–categories, and an object X is cofibrant (resp. fibrant) in $\mathcal{M}$. Then X is a j–cell in $\mathcal{M}$ for some $0 \le j \le 2$ if and only if $F(X)$ is a j–cell in $\mathcal{M}'$.*

Proof Consider the induced functor $F_\infty\colon \mathcal{M}_\infty \to \mathcal{M}'_\infty$, which is an equivalence of $(\infty, 1)$–categories. It follows that, for any $j = 0, 1, 2$ and j–cell X_j of $\mathcal{M}$, the object $F_\infty(X_j)$ is a cell in $\mathcal{M}'_\infty$, either by direct verification, or by appealing to Theorem 2.19. Now, an object X is a j–cell in $\mathcal{M}$ if and only if there is an isomorphism $X \cong X_j$ in $\operatorname{Ho}\mathcal{M}$. Again using the fact that F_∞ is an equivalence, this statement is equivalent to saying that there is an isomorphism $F_\infty(X) \cong F_\infty(X_j)$ in $\operatorname{Ho}(\mathcal{M}')$. But the existence of such an isomorphism is equivalent to having $F_\infty(X)$ being a j–cell of $\mathcal{M}'_\infty$ because $F_\infty(X_j)$ is one. Since $F(X)$ computes $F_\infty(X)$, the result follows. □

3.1 Recognizing cells in Θ_2–models of $(\infty, 2)$–categories

We now begin the work of identifying j–cells in different models for $(\infty, 2)$–categories. We begin with the j–cells in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,2)}$, which have been identified by Barwick and Schommer-Pries.

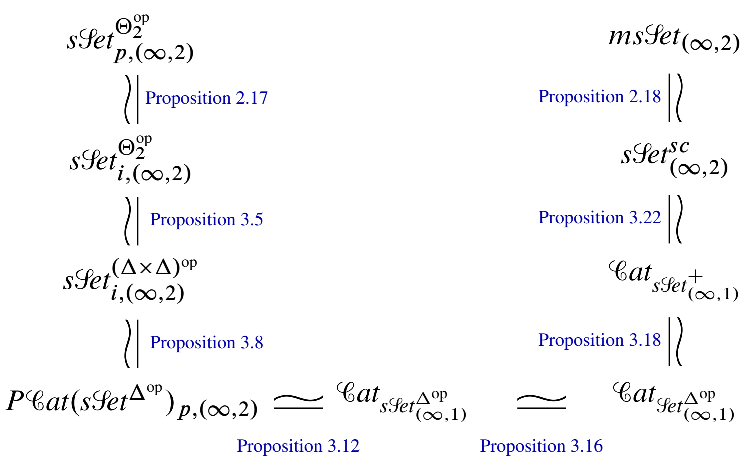


Figure 2

Proposition 3.2 [Barwick and Schommer-Pries 2021, Section 13] *In $s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$ the object $\Theta_2[C_j]$ is a representative of the j–cell for $j=0,1,2$.*

We can now prove Proposition 2.17, which identifies the cells in $s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}$.

Proof of Proposition 2.17 Consider the identity functor on $s\mathcal{S}et^{\Theta_2^{op}}$, which by Theorem 1.19 defines a left Quillen equivalence

$$\mathrm{id}\colon s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)} \to s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}.$$

By Proposition 3.2 we know that $\Theta_2[C_j]$ is a j–cell in $s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$ for $j=0,1,2$. Moreover, the object $\Theta_2[C_j]$ is projectively cofibrant by [Hirschhorn 2003, Proposition 11.6.2], since it is representable. It then follows from Lemma 3.1 that $\Theta_2[C_j]$ is a j–cell in $s\mathcal{S}et^{\Theta_2^{op}}_{p,(\infty,2)}$. □

3.2 Recognizing cells in multisimplicial models of $(\infty,2)$–categories

We now turn to identifying j–cells in multisimplicial models of $(\infty,2)$–categories. Because we have not yet considered these models in this paper, we describe them briefly as we go.

Theorem 3.3 [Barwick 2005, Chapter 2] *The category $s\mathcal{S}et^{(\Delta\times\Delta)^{op}}$ of bisimplicial spaces admits a model structure in which*

- *the fibrant objects are the injectively fibrant complete Segal objects in complete Segal spaces*; *and*
- *the cofibrations are the monomorphisms*, *and in particular every object is cofibrant.*

We denote this model structure by $s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)}$.

The idea behind complete Segal objects in complete Segal spaces is that we apply similar Segal and completeness conditions to functors $\Delta^{op}\to s\mathcal{S}et^{\Delta^{op}}$, where the target category is equipped with the complete Segal space model structure. Thus, the Segal and completeness maps are now equivalences in this model structure, rather than equivalences of simplicial sets. For more details on the definition of complete Segal objects in complete Segal spaces, see [Barwick 2005, Chapter 2; Bergner and Rezk 2020, Definition 5.3; Lurie 2009b].

There is an explicit equivalence of model categories between this model structure and the one for complete Segal Θ_2–spaces. See also [Barwick and Schommer-Pries 2021] for a different proof of this equivalence.

Theorem 3.4 [Bergner and Rezk 2020, Corollary 7.1] *The functor $d\colon \Delta\times\Delta\to\Theta_2$ given by*

$$[m,k]\mapsto[m|k,\dots,k]$$

induces a left Quillen equivalence

$$d^*\colon s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)} \to s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)}.$$

In particular, $s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)}$ is a model for $(\infty,2)$–categories. We now characterize the j–cells in this model.

Proposition 3.5 *In $s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)}$ the object $d^*\Theta_2[C_j]$ is a representative of the j–cell for $j=0,1,2$.*

Proof We consider the functor d^*, which defines a left Quillen equivalence

$$d^*\colon s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)} \to s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)}.$$

By Proposition 3.2, for $j=0,1,2$, the object $\Theta_2[C_j]$ is a j–cell in $s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$. Moreover, every object is cofibrant in $s\mathcal{S}et^{\Theta_2^{op}}_{i,(\infty,2)}$. It follows from Lemma 3.1 that $d^*\Theta_2[C_j]$ is a j–cell in $s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)}$. □

We can now generalize the notion of Segal precategory to this context; in analogy with the notion of complete Segal objects described above, we can define *Segal precategory objects* in complete Segal spaces, given by functors $X\colon \Delta^{op}\to s\mathcal{S}et^{\Delta^{op}}$ such that X_0 is a discrete object and the Segal maps are weak equivalences in the complete Segal space model structure. See [Bergner and Rezk 2013, Section 6] for more details.

Let us briefly describe the comparison with complete Segal objects, which is analogous to Theorem 1.10. We denote by $P\mathcal{C}at(\mathcal{S}et^{\Delta^{op}})$ the full subcategory of $s\mathcal{S}et^{(\Delta\times\Delta)^{op}}$ given by Segal precategory objects in simplicial spaces, namely those bisimplicial spaces $X\colon \Delta^{op}\to s\mathcal{S}et^{\Delta^{op}}$ for which X_0 is discrete, and we denote by $I\colon P\mathcal{C}at(\mathcal{S}et^{\Delta^{op}})\to s\mathcal{S}et^{(\Delta\times\Delta)^{op}}$ the inclusion functor.

Theorem 3.6 [Bergner and Rezk 2013, 6.12] *The category $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{op}})$ of precategories in simplicial spaces admits a model structure in which*

- *the fibrant objects are the projectively fibrant Segal category objects, and*
- *the cofibrations are the projective cofibrations.*

We denote this model structure by $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{op}})_{p,(\infty,2)}$.

This model was compared to the previous ones by [Bergner and Rezk 2020].

Theorem 3.7 [Bergner and Rezk 2020, Theorem 9.6 and Propositions 7.1 and 9.5] *The natural inclusion functor from* [Bergner and Rezk 2020, Section 9] *induces a left Quillen equivalence*

$$I\colon P\mathcal{C}at(s\mathcal{S}et^{\Delta^{op}})_{p,(\infty,2)} \to s\mathcal{S}et^{(\Delta\times\Delta)^{op}}_{i,(\infty,2)}.$$

In particular, $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{op}})_{p,(\infty,2)}$ is a model for $(\infty,2)$–categories.

For each $j=0,1,2$, the bisimplicial space $d^*\Theta_2[C_j]$, a priori an object of $s\mathcal{S}et^{(\Delta\times\Delta)^{op}}$, is actually a precategory, so it can be regarded as an object of $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{op}})$.

Proposition 3.8 *In $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{op}})_{p,(\infty,2)}$ the object $d^*\Theta_2[C_j]$ is a representative of the j–cell for $j=0,1,2$.*

To prove this result, we make use of the following lemma.

Lemma 3.9 *In $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}$ the object $d^*\Theta_2[C_j]$ is a cofibrant for $j = 0, 1, 2$.*

Proof If $\Delta[\varnothing]$ denotes the initial bisimplicial space, we show that the canonical map $\Delta[\varnothing] \to d^*\Theta_2[C_j]$ is a cofibration in $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}$ for $j = 0, 1, 2$.

For $j = 0$ and $j = 1$, the object $d^*\Theta_2[C_j]$ is representable as an object of $s\mathcal{S}et^{(\Delta\times\Delta)^{\mathrm{op}}}$, so by [Hirschhorn 2003, Proposition 11.6.2] the map $\Delta[\varnothing] \to d^*\Theta_2[C_j]$ is a projective cofibration and hence $d^*\Theta_2[C_j]$ is cofibrant in $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}$.

For $j = 2$, we recall from [Bergner and Rezk 2013, Section 6.2] that any map of the form $A_{[p]} \to B_{[p]}$, where $p \geq 0$ and $A \to B$ is a cofibration of $s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}$, is a cofibration in $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}$. Recall that $A_{[p]}$ is defined as the pushout in $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})$

$$\begin{array}{ccc} A \boxtimes (\Delta[p])_0 & \hookrightarrow & A \boxtimes \Delta[p] \\ \downarrow & & \downarrow \\ \Delta[0] \boxtimes (\Delta[p])_0 & \hookrightarrow & A_{[p]} \end{array}$$

for any $p \geq 0$ and any simplicial space A. We can now write the map $\Delta[\varnothing] \to d^*\Theta_2[C_2]$ as the following composite of three cofibrations:

$$\Delta[\varnothing] \to d^*\Theta_2[C_0] \cong \Delta[0]_{[0]} \to (\Delta[0] \amalg \Delta[0])_{[0]} \cong \Delta[\varnothing]_{[1]} \to \Delta[1]_{[1]} \cong d^*\Theta_2[C_2],$$

concluding the proof. □

We can now prove the proposition.

Proof of Proposition 3.8 We consider the inclusion functor, which defines a left Quillen equivalence

$$I \colon P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)} \to s\mathcal{S}et^{(\Delta\times\Delta)^{\mathrm{op}}}_{i,(\infty,2)}.$$

For $j = 0, 1, 2$, by Proposition 3.5 we know that $I(d^*\Theta_2[C_j])$ is a j–cell in $s\mathcal{S}et^{(\Delta\times\Delta)^{\mathrm{op}}}_{i,(\infty,2)}$. Moreover, the object is cofibrant in $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}$ by Lemma 3.9. It follows from Lemma 3.1 that $d^*\Theta_2[C_i]$ is a j–cell in $P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}$. □

3.3 Recognizing cells in enriched models of $(\infty, 2)$–categories

We now turn to recognizing cells in models that are given by enriched categories. Many model structures on enriched categories can be obtained by the following general result of Lurie.

Theorem 3.10 [Lurie 2009a, Theorem A.3.2.24] *Let $\mathcal{V}$ be an excellent monoidal model category, in the sense of* [Lurie 2009a, Definition A.3.2.16]. *The category of small categories enriched over $\mathcal{V}$ admits a model structure in which*

- *the fibrant objects are the* ***locally fibrant categories****, ie the enriched categories C such that for any pair of objects c, c' in C, the mapping object $\mathrm{Hom}_C(c, c')$ is fibrant in $\mathcal{V}$;*

- *the weak equivalences, which are described in* [Lurie 2009a, Definition A.3.2.1] *and* [Lawson 2017], *are enriched functors* $F\colon C \to D$ *such that*
  (1) *for every pair of objects* c, c' *of* C, *the map induced by* F *of mapping objects*
  $$F_{c,d}\colon \mathrm{Hom}_C(c,c') \to \mathrm{Hom}_D(Fc, Fc'),$$
  is a weak equivalence in $\mathcal{V}$, *and*
  (2) *the functor induced by* F *on (underlying categories) of* $\mathrm{Ho}\,\mathcal{V}$*–categories is essentially surjective*;
- *the cofibrations are those described in* [Lurie 2009a, Proposition A.3.2.4].

We denote this model structure by $\mathcal{C}at_{\mathcal{V}}$.

To give an idea, the technical condition for a combinatorial monoidal model category to be *excellent* requests a closure property for cofibrations and weak equivalences, in addition to compatibility of the model structure with the monoidal structure. Lurie's original definition also requires a further condition, known as "invertibility hypothesis", which was shown to follow from the other conditions by [Lawson 2017, Theorem 0.1].

We specialize this construction to the following situations.

- Let $\mathcal{V} = \mathcal{C}at$ be the canonical model structure on the category $\mathcal{C}at$ of small categories from [Rezk 1996], which is seen to be excellent using the fact that the nerve functor creates weak equivalences and commutes with filtered colimits. We then obtain precisely the model category $\mathcal{C}at_{\mathcal{C}at} = 2\mathcal{C}at$ as discussed in [Berger and Moerdijk 2013, Example 1.8].
- Let $\mathcal{V} = s\mathcal{S}et_{(\infty,1)}$ be the Joyal model structure on the category $s\mathcal{S}et$ of simplicial sets from [Joyal 2008, Theorem 6.12], which is excellent by [Lurie 2009a, Example A.3.2.23]. We then obtain the model category $\mathcal{C}at_{s\mathcal{S}et_{(\infty,1)}}$.
- Let $\mathcal{V} = s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}$ be Rezk's model structure from Theorem 1.8 on the category $s\mathcal{S}et^{\Delta^{\mathrm{op}}}$ of simplicial spaces, which is discussed to be excellent [Bergner and Rezk 2013, Theorem 3.11]. We then obtain the model category $\mathcal{C}at_{s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$.
- Let $\mathcal{V} = s\mathcal{S}et^{+}_{(\infty,1)}$ be Lurie's model structure on the category $s\mathcal{S}et^{+}$ of marked simplicial sets from [Lurie 2009a, Proposition 3.1.3.7], which is excellent by [Lurie 2009a, Example A.3.2.22]. We then obtain the model category $\mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}}$.

We now turn to an explicit Quillen equivalence between one of these enriched models and one of the models we have already discussed.

Theorem 3.11 [Bergner and Rezk 2013, 7.1–7.6] *The enriched nerve functor from* [Bergner and Rezk 2013, Definition 7.3], *obtained by regarding a bisimplicial category as a simplicial object in simplicial spaces, defines a right Quillen equivalence*
$$R\colon \mathcal{C}at_{s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}} \to P\mathcal{C}at(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}.$$

In particular, $\mathcal{C}at_{s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$ is a model for $(\infty,2)$–categories.

Now, we would like to identify the j–cells in the model structure $\mathcal{C}at_{s\mathcal{S}et^{\Delta^{op}}_{(\infty,1)}}$, for which we make use of the discrete nerve functor $N^{\mathrm{disc}}\colon \mathcal{C}at \to s\mathcal{S}et^{\Delta^{op}}$ considered in [Rezk 2001]. Since it preserves products, being a right adjoint functor, it induces a functor $N^{\mathrm{disc}}_*\colon \mathcal{C}at_{\mathcal{C}at} \to \mathcal{C}at_{s\mathcal{S}et^{\Delta^{op}}}$, given by applying N^{disc} to each mapping category.

Proposition 3.12 *In $\mathcal{C}at_{s\mathcal{S}et^{\Delta^{op}}_{(\infty,1)}}$ the object $N^{\mathrm{disc}}_* C_j$ is a representative of the j–cell for $j = 0, 1, 2$.*

Before proving this proposition, we establish two lemmas that tell us more about the structure of these discrete nerves.

Lemma 3.13 *For any $j = 0, 1, 2$, the $s\mathcal{S}et^{\Delta^{op}}$–enriched category $N^{\mathrm{disc}}_*\Theta_2[C_j]$ is fibrant in $\mathcal{C}at_{s\mathcal{S}et^{\Delta^{op}}_{(\infty,1)}}$.*

Proof For $j = 0, 1, 2$, all hom-categories of C_j are of the form $\varnothing = [-1]$, $[0]$, or $[1]$. So all hom-bisimplicial sets of $N^{\mathrm{disc}}_*\Theta_2[C_j]$ are of the form $N^{\mathrm{disc}}[-1]$, $N^{\mathrm{disc}}[0]$ or $N^{\mathrm{disc}}[1]$, which are all complete Segal spaces, namely fibrant in $s\mathcal{S}et^{\Delta^{op}}_{(\infty,1)}$, since the categories $[-1]$, $[0]$, and $[1]$ do not have any nontrivial isomorphisms. □

Lemma 3.14 *For any θ in Θ_2, there is an isomorphism of precategories*

$$R(N^{\mathrm{disc}}_*\theta) \cong d^*\Theta_2[\theta].$$

Proof For $i, j, k \geq 0$, we first compute the set $(RN^{\mathrm{disc}}_*\theta)_{[i],[j],[k]}$. If $\mathcal{D}$ is a bisimplicial category with object set $\mathcal{D}_0$, and $\mathcal{D}_1$ denotes the bisimplicial space

$$\mathcal{D}_1 = \coprod_{a,b\in\mathcal{D}_0} \mathrm{Hom}_{\mathcal{D}}(a,b),$$

by definition of R (as given in [Bergner and Rezk 2013, Definition 7.3]) for any $i \geq 0$ there is an isomorphism of bisimplicial sets

$$(R\mathcal{D})_{[i]} \cong \underbrace{\mathcal{D}_1 \times_{\mathcal{D}_0} \mathcal{D}_1 \times_{\mathcal{D}_0} \cdots \times_{\mathcal{D}_0} \mathcal{D}_1}_{i}$$

that is natural in i. When specializing to the case $\mathcal{D} = N^{\mathrm{disc}}_*\theta$, we obtain a natural isomorphism

$$(RN^{\mathrm{disc}}_*\theta)_{[i]} \cong \underbrace{(N^{\mathrm{disc}}_*\theta)_1 \times_{(N^{\mathrm{disc}}_*\theta)_0} (N^{\mathrm{disc}}_*\theta)_1 \times_{(N^{\mathrm{disc}}_*\theta)_0} \cdots \times_{(N^{\mathrm{disc}}_*\theta)_0} (N^{\mathrm{disc}}_*\theta)_1}_{i}.$$

In particular, if θ_0 denotes the set of objects of θ and θ_1 denotes the category

$$\theta_1 := \coprod_{a,b\in\theta_0} \mathrm{Hom}_\theta(a,b),$$

for any $j, k \geq 0$ we have a bijection

$$(RN^{\mathrm{disc}}_*\theta)_{[i],[j],[k]} \cong \underbrace{N_j\theta_1 \times_{\theta_0} N_j\theta_1 \times_{\theta_0} \cdots \times_{\theta_0} N_j\theta_1}_{i},$$

that is natural in i, j, k.

Next, for $i, j, k \geq 0$, we compute the set $(d^*\Theta_2[\theta])_{[i],[j],[k]}$. By definition of d^*, and using the fact that Θ_2 is a full subcategory of $2\mathscr{C}at$, we have bijections

$$\begin{aligned}
(d^*\Theta_2[\theta])_{[i],[j],[k]} &\cong \mathrm{Hom}_{\Theta_2}([i\,|\,\underbrace{j, j, \dots, j}_{i}], \theta) \\
&\cong \mathrm{Hom}_{2\mathscr{C}at}([i\,|\,\underbrace{j, j, \dots, j}_{i}], \theta) \\
&\cong \mathrm{Hom}_{2\mathscr{C}at}(\underbrace{[1|j]\amalg_{[0]}[1|j]\amalg_{[0]}\cdots\amalg_{[0]}[1|j]}_{i}, \theta) \\
&\cong \underbrace{\mathrm{Hom}_{2\mathscr{C}at}([1|j], \theta) \times_{\mathrm{Hom}_{2\mathscr{C}at}([0],\theta)} \cdots \times_{\mathrm{Hom}_{2\mathscr{C}at}([0],\theta)} \mathrm{Hom}_{2\mathscr{C}at}([1|j], \theta)}_{i} \\
&\cong \underbrace{\mathrm{Hom}_{2\mathscr{C}at}([1|j], \theta) \times_{\theta_0} \cdots \times_{\theta_0} \mathrm{Hom}_{2\mathscr{C}at}([1|j], \theta)}_{i}
\end{aligned}$$

that are natural in i, j, k.

Finally, we show that there is a bijection

$$\mathrm{Hom}_{2\mathscr{C}at}([1|j], \theta) \cong N_j\theta_1$$

that is natural in j, from which the lemma follows. To do so, we observe that there are natural bijections

$$\begin{aligned}
\mathrm{Hom}_{2\mathscr{C}at}([1|j], \theta) &\cong \coprod_{a,b\in\theta_0} \mathrm{Hom}_{2\mathscr{C}at_{*,*}}([1|j], (\theta, a, b)) \\
&\cong \coprod_{a,b\in\theta_0} \mathrm{Hom}_{\mathscr{C}at}([j], \mathrm{Hom}_\theta(a,b)) \\
&\cong \mathrm{Hom}_{\mathscr{C}at}\Big([j], \coprod_{a,b\in\theta_0} \mathrm{Hom}_\theta(a,b)\Big) \\
&\cong \mathrm{Hom}_{\mathscr{C}at}([j], \theta_1) \cong N_j\theta_1,
\end{aligned}$$

as desired. □

Proof of Proposition 3.12 Consider the right Quillen equivalence

$$R\colon \mathscr{C}at_{s\mathscr{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}} \to P\mathscr{C}at(s\mathscr{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}.$$

By Proposition 3.5 and Lemma 3.14, we know that for any $j = 0, 1, 2$, the object $d^*\Theta_2[C_j] \cong R(N_*^{\mathrm{disc}}C_j)$ is a j–cell in $P\mathscr{C}at(s\mathscr{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,2)}$. Moreover, by Lemma 3.13 the object $N_*^{\mathrm{disc}}C_j$ is fibrant in $\mathscr{C}at_{s\mathscr{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$. It follows from Lemma 3.1 that $N_*^{\mathrm{disc}}C_j$ is a j–cell in $\mathscr{C}at_{s\mathscr{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$, as desired. □

We now compare the model structure for categories enriched in complete Segal spaces to the model structure for categories enriched in quasicategories.

Theorem 3.15 *The functor induced by taking $(-)_0$ on each hom-simplicial space defines a right Quillen equivalence*

$$\mathcal{C}at_{s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}} \to \mathcal{C}at_{\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}.$$

In particular, $\mathcal{C}at_{\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$ is a model for $(\infty,2)$–categories.

Proof The functor $p\colon \Delta\times\Delta\to\Delta$, defined by $[m,n]\mapsto[m]$, induces an adjoint triple

$$\mathcal{S}et^{\Delta^{\mathrm{op}}} \underset{p_*}{\overset{p_!}{\leftrightarrows}} \xrightarrow{\;p^*\;} \mathcal{S}et^{(\Delta\times\Delta)^{\mathrm{op}}} = s\mathcal{S}et^{\Delta^{\mathrm{op}}},$$

where p^* is given by precomposition with p, while $p_!$ and p_* are the left and right Kan extensions along p, respectively. In particular, the functor p^* is (strong) monoidal with respect to cartesian product because it is a right adjoint. Moreover, it is shown as [Joyal and Tierney 2007, Theorem 4.11] that the adjunction

$$p^*\colon \mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)} \rightleftarrows s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)} : p_*$$

is a Quillen equivalence. One can then apply [Lurie 2009a, Remark A.3.2.6] to obtain the desired Quillen equivalence, observing that p_* is the functor $(-)_0$. □

We can now use this equivalence to identify the j–cells in $\mathcal{C}at_{\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$.

Proposition 3.16 *In $\mathcal{C}at_{\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$ the object N_*C_j is a representative of the j–cell for $j=0,1,2$.*

Proof Consider the right Quillen equivalence from Theorem 3.15

$$\mathcal{C}at_{s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}} \to \mathcal{C}at_{\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}.$$

By Proposition 3.12 and Lemma 3.13 we know that for each $j=0,1,2$, the object $N_*^{\mathrm{disc}}C_j$ is a j–cell in $\mathcal{C}at_{s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$ and is fibrant. It follows from Lemma 3.1 that N_*C_j, the image of $N_*^{\mathrm{disc}}C_j$ under the above Quillen equivalence, is a j–cell in $\mathcal{C}at_{\mathcal{S}et^{\Delta^{\mathrm{op}}}_{(\infty,1)}}$. □

We now make a similar comparison between categories enriched in quasicategories and categories enriched in marked simplicial sets.

Theorem 3.17 *The functor induced by taking the underlying simplicial set U on each mapping object defines a right Quillen equivalence*

$$U_*\colon \mathcal{C}at_{s\mathcal{S}et^+_{(\infty,1)}} \to \mathcal{C}at_{s\mathcal{S}et_{(\infty,1)}}.$$

In particular, $\mathcal{C}at_{s\mathcal{S}et^+_{(\infty,1)}}$ is a model for $(\infty,2)$–categories.

Proof The desired right Quillen equivalence is an instance of [Lurie 2009a, Remark A.3.2.6] applied to the right Quillen equivalence

$$U: s\mathcal{S}et^{+}_{(\infty,1)} \to \mathcal{S}et^{\Delta^{op}}_{(\infty,1)}$$

from [Lurie 2009a, Theorem 3.1.5.1]. □

Once again, our goal is to identify the j–cells in this model structure. To do so, consider the flat nerve functor $N^\flat: \mathcal{C}at \to s\mathcal{S}et^+$, obtained by regarding the nerve of a category in which the marked 1–simplices are precisely those corresponding to identity morphisms in the category. One can check that the functor $N^\flat$ preserves finite cartesian products, from which we obtain an induced functor $N^\flat_*: \mathcal{C}at_{\mathcal{C}at} \to \mathcal{C}at_{s\mathcal{S}et^+}$, given by applying $N^\flat$ to each mapping category.

Proposition 3.18 *In $\mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}}$ the object $N^\flat_* C_j$ is a representative of the j–cell for $j = 0, 1, 2$.*

We begin with a lemma establishing that these objects are fibrant.

Lemma 3.19 *For $j = 0, 1, 2$, the object $N^\flat_* C_j$ is fibrant in $\mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}}$.*

Proof For $j = 0, 1, 2$, all hom-marked simplicial sets of $N^\flat_* C_j$ are of the form $N^\flat[-1]$, $N^\flat[0]$ or $N^\flat[1]$, which are naturally marked quasicategories, and therefore fibrant in $\mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}}$, since the categories $[-1]$, $[0]$ and $[1]$ have no nontrivial isomorphisms. □

Proof of Proposition 3.18 We consider the right Quillen equivalence

$$U_*: \mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}} \to \mathcal{C}at_{s\mathcal{S}et_{(\infty,1)}}.$$

By Proposition 3.16 we know that for each $j = 0, 1, 2$, the object $U_* N^\flat_* C_j \cong N_* C_j$ is a j–cell in $\mathcal{C}at_{s\mathcal{S}et_{(\infty,1)}}$, and $N^\flat_* C_j$ is fibrant in $\mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}}$ by Lemma 3.19. It follows from Lemma 3.1 that $N^\flat_* C_j$ is a j–cell in $\mathcal{C}at_{s\mathcal{S}et^{+}_{(\infty,1)}}$. □

3.4 Recognizing cells in simplicial models of $(\infty, 2)$–categories

Finally, we want to identify the j–cells in the model of marked simplicial sets. To aid in doing so, we look first at the related model of scaled simplicial sets. A *scaled simplicial set* is a simplicial set with a collection of marked 2–simplices including degenerate 2–simplices.

Theorem 3.20 [Lurie 2009b, Theorem 4.2.7] *The category $s\mathcal{S}et^{sc}$ of scaled simplicial sets admits a model structure in which*

- *the fibrant objects are the ∞–bicategories from* [Lurie 2009b, Definition 4.2.8], *and*
- *the cofibrations are the monomorphisms* (*and in particular every object is cofibrant*).

We denote this model structure by $s\mathcal{S}et^{sc}_{(\infty,2)}$.

Lurie enhances the classical homotopy coherent nerve functor $\mathfrak{N}\colon \mathscr{C}at_{s\mathscr{S}et} \to s\mathscr{S}et$ to the context of scaled simplicial sets by taking into account the marking, obtaining a scaled homotopy coherent nerve functor $\mathfrak{N}\colon \mathscr{C}at_{s\mathscr{S}et^+} \to s\mathscr{S}et^{sc}$.

Theorem 3.21 [Lurie 2009b, Theorem 0.0.3] *The scaled homotopy coherent nerve functor from* [Lurie 2009b, Definition 3.1.10] *defines a right Quillen equivalence*

$$\mathfrak{N}^{sc}\colon \mathscr{C}at_{s\mathscr{S}et^+(\infty,1)} \to s\mathscr{S}et^{sc}_{(\infty,2)}.$$

In particular, $s\mathscr{S}et^{sc}_{(\infty,2)}$ is a model for $(\infty,2)$–categories. We now describe the j–cells in this model structure.

The description of the j–cells in this model structure makes use of a similar scaled nerve construction, in the form of a functor $N^{sc}\colon 2\mathscr{C}at \to s\mathscr{S}et^{sc}$, as described in [Gagna et al. 2022, Definition 8.1]. Given any 2–category $\mathfrak{D}$, the scaled nerve $N^{sc}\mathfrak{D}$ is given by the Duskin nerve of $\mathfrak{D}$ together with the marking of all 2–simplices arising from 2–isomorphisms.

Proposition 3.22 *In $s\mathscr{S}et^{sc}_{(\infty,2)}$ the object $N^{sc}C_j$ is a representative of the j–cell for $j=0,1,2$.*

Proof As a preliminary observation, we mention that there is an isomorphism of scaled simplicial sets

$$N^{sc}C_j \cong \mathfrak{N}^{sc}N^{\flat}_*C_j$$

for $j=0,1,2$. This fact can be deduced combining [Gagna et al. 2022, Definition 8.1] together with [Gagna et al. 2022, Proposition 8.2].

Consider now the right Quillen equivalence

$$\mathfrak{N}^{sc}\colon \mathscr{C}at_{s\mathscr{S}et^+(\infty,1)} \to s\mathscr{S}et^{sc}_{(\infty,2)}.$$

By Proposition 3.18 we know that for each $j=0,1,2$, the object $N^{\flat}_*C_j$ is a j–cell in $\mathscr{C}at_{s\mathscr{S}et^+_{(\infty,1)}}$. We have also proved that $N^{\flat}_*C_j$ is fibrant in $\mathscr{C}at_{s\mathscr{S}et^+_{(\infty,1)}}$ in Lemma 3.19. It follows from Lemma 3.1 that $N^{sc}C_j \cong \mathfrak{N}^{sc}N^{\flat}_*C_j$ is a j–cell in $s\mathscr{S}et^{sc}_{(\infty,2)}$. □

Finally, we can compare the models of scaled simplicial sets and marked simplicial sets.

Theorem 3.23 [Gagna et al. 2022, Theorem 7.7] *The forgetful functor defines a right Quillen equivalence*

$$U\colon ms\mathscr{S}et_{(\infty,2)} \to s\mathscr{S}et^{sc}_{(\infty,2)}.$$

We can now prove Proposition 2.18, which characterizes the cells in $ms\mathscr{S}et_{(\infty,2)}$.

Proof of Proposition 2.18 We consider the right Quillen equivalence

$$U\colon ms\mathscr{S}et_{(\infty,2)} \to s\mathscr{S}et^{sc}_{(\infty,2)}.$$

By [Gagna et al. 2022, Definition 8.1] we know that $UN^{\mathrm{RS}}C_j \cong N^{sc}C_j$ for each $j = 0, 1, 2$. Moreover, we know that $N^{sc}C_j$ is a j–cell in $s\mathcal{S}et^{sc}_{(\infty,2)}$ by Proposition 3.18 and that it is fibrant in $ms\mathcal{S}et_{(\infty,2)}$ by [Ozornova and Rovelli 2020, Theorem 5.1(1)]. It follows from Lemma 3.1 that $N^{\mathrm{RS}}C_j$ is a j–cell in $ms\mathcal{S}et_{(\infty,2)}$. □

4 Applications

Here we discuss four situations in which one can exploit the explicit Quillen equivalence

$$L\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)} \rightleftarrows ms\mathcal{S}et_{(\infty,2)} : R \tag{4-1}$$

from Theorem 2.20 to produce new theorems, new proofs or export constructions given some existing ones. Precisely, we show the following.

(1) The nerve construction for 2–categories is compatible with the suspension construction and the wedge constructions in an appropriate sense in the globular setting, using the analogous statement proven in the complicial setting in [Ozornova and Rovelli 2022].

(2) The nerve construction for 2–categories is compatible with the cone construction in an appropriate sense in the globular setting, using the analogous statement proven in the complicial setting in [Gagna et al. 2023].

(3) The nerve construction for 2–categories is compatible with the co-dual construction in an appropriate sense in the complicial setting, using the analogous statement that is formal in the globular setting.

(4) Weak equivalences can be tested on homotopy categories and homs in the complicial setting, using the analogous statement for the globular setting from [Bergner and Rezk 2020].

We expect that similar techniques can be applied to translate any new results from the setting of complicial sets to that of Θ_2–spaces, and vice versa.

As a preliminary preparation that is common to many of the applications, we define the *complicial nerve* to be the homotopical functor

$$N^{\mathrm{cmp}}\colon 2\mathcal{C}at \to \mathcal{S}et^{t\Delta^{\mathrm{op}}}_{(\infty,2)} \to ms\mathcal{S}et_{(\infty,2)}$$

obtained as a composite of the right Quillen functor $N^{\natural}\colon 2\mathcal{C}at \to \mathcal{S}et^{t\Delta^{\mathrm{op}}}_{(\infty,2)}$ from [Ozornova and Rovelli 2021, Theorem 4.12] with the left Quillen functor $\mathrm{Refl}\colon \mathcal{S}et^{t\Delta^{\mathrm{op}}}_{(\infty,2)} \to ms\mathcal{S}et_{(\infty,2)}$. Similarly, we consider the *globular nerve* construction to be the right Quillen functor

$$N^{\mathrm{gl}}\colon 2\mathcal{C}at \to \mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{(\infty,2)} \to s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,2)} \to s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$$

obtained by composing the right Quillen nerve $N\colon 2\mathcal{C}at \to \mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{(\infty,2)}$ from [Campbell 2020, Theorem 5.10] with the right Quillen equivalence $\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{(\infty,2)} \to s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,2)}$ from [Ara 2014, Corollary 8.8] and the identity viewed as a right Quillen equivalence $\mathrm{id}\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,2)} \to s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$. The two nerve constructions N^{gl} and N^{cmp} induce functors at the level of $(\infty,1)$–categories that have the "correct" universal property, namely they realize the $(\infty,1)$–category of strict 2–categories as a localization of the $(\infty,1)$–category of $(\infty,2)$–categories, as in [Moser et al. 2022, Remark 6.37].

The globular and complicial nerves are compatible in the following sense.

Proposition 4.1 *For every 2–category $\mathcal{D}$ there is a natural weak equivalence*

$$L((N^{\mathrm{gl}}\mathcal{D})^{\mathrm{cof}}) \xrightarrow{\simeq} N^{\mathrm{cmp}}\mathcal{D}$$

in $ms\mathcal{S}et_{(\infty,2)}$, where $(N^{\mathrm{gl}}\mathcal{D})^{\mathrm{cof}}$ denotes a functorial cofibrant replacement of $N^{\mathrm{gl}}\mathcal{D}$ in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$.

Proof As a preliminary fact, we observe that with techniques analogous to the ones employed to construct the Quillen equivalence (4-1), one could also show that there is a Quillen equivalence

$$L'\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)} \rightleftarrows \mathcal{S}et^{t\,\Delta^{\mathrm{op}}}_{(\infty,2)} : R'$$

by setting, for all θ in Θ_2 and $k \geq 0$,

$$L'(\Theta_2[\theta] \boxtimes \Delta[k]) := N^{\natural}(\theta) \times \Delta[k]^{\sharp}.$$

Here, $\mathcal{S}et^{t\,\Delta^{\mathrm{op}}}_{(\infty,2)}$ denotes the model structure from [Ozornova and Rovelli 2020, Theorem 1.28], and it is useful to recall that, by [Ozornova and Rovelli 2020, Proposition 1.35], there is a Quillen equivalence

$$\mathrm{Refl}\colon \mathcal{S}et^{t\,\Delta^{\mathrm{op}}}_{(\infty,2)} \rightleftarrows ms\mathcal{S}et_{(\infty,2)}.$$

By construction, one then has $L = \mathrm{Refl}\, L'$. Now, for all θ in Θ_2 and $k \geq 0$, we have a commutative diagram in $2\mathcal{C}at$

$$\begin{array}{ccc} c^{\mathrm{gl}}N^{\mathrm{gl}}(\theta) & \longleftarrow & c^{\natural}N^{\natural}(\theta) \\ & {\scriptstyle \epsilon,\ \simeq}\searrow \quad \swarrow{\scriptstyle \epsilon,\ \simeq} & \\ & \theta & \end{array}$$

Here, $c^{\mathrm{gl}}\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}} \to 2\mathcal{C}at$ and $c^{\natural}\colon \mathcal{S}et^{t\,\Delta^{\mathrm{op}}} \to 2\mathcal{C}at$ denote the left adjoint functors to N^{gl} and $N^{\natural}$, respectively, and the top map is adjoint to $N^{\natural}(\epsilon_\theta\colon c^{\mathrm{gl}}N^{\mathrm{gl}}(\theta) \to \theta)$ with respect to the adjunction $c^{\natural} \dashv N^{\natural}$, and the vertical maps can be seen to be weak equivalences in $2\mathcal{C}at$ combining [Ara 2014, Corollary 8.8; Campbell 2020, Section 5.1; Ozornova and Rovelli 2021, Theorem 4.10]. So there is a weak equivalence

$$c^{\mathrm{gl}}\Theta[\theta] \cong c^{\mathrm{gl}}N^{\mathrm{gl}}(\theta) \xleftarrow{\simeq} c^{\natural}N^{\natural}(\theta) = c^{\natural}L'\Theta[\theta]$$

in $2\mathcal{C}at$. Applying [Dugger 2001, Lemma 9.7] on the left Quillen functors

$$c^{\mathrm{gl}}, c^{\natural}L'\colon s\mathcal{S}et_p^{\Theta_2^{\mathrm{op}}} \to s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)} \to 2\mathcal{C}at,$$

it follows that, for all W cofibrant in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}})_{p,(\infty,2)}$, there is a natural weak equivalence

$$c^{\mathrm{gl}}W \xleftarrow{\simeq} c^{\natural}L'W$$

in $2\mathscr{C}at$. By [Hovey 1999, Corollary 1.4.4], it follows that for every (necessarily fibrant) 2–category $\mathscr{D}$ we obtain a natural equivalence

$$N^{\mathrm{gl}}\mathscr{D} \xrightarrow{\simeq} R' N^{\natural}\mathscr{D}$$

in $s\mathscr{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$. Hence, using that the left Quillen functor L' preserves weak equivalences between cofibrant objects in $s\mathscr{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$ and that the derived counit of $L' \dashv R'$ at $N^{\mathrm{gl}}\mathscr{D}$ is a weak equivalence in $\mathscr{S}et^{t\,\Delta^{\mathrm{op}}}_{(\infty,2)}$, we obtain that there are weak equivalences

$$L'((N^{\mathrm{gl}}\mathscr{D})^{\mathrm{cof}}) \xrightarrow{\simeq} L'((R' N^{\natural}\mathscr{D})^{\mathrm{cof}}) \xrightarrow{\simeq} N^{\natural}\mathscr{D}$$

in $\mathscr{S}et^{t\,\Delta^{\mathrm{op}}}_{(\infty,2)}$. Finally, using that the functor Refl is homotopical we obtain a weak equivalence

$$L((N^{\mathrm{gl}}\mathscr{D})^{\mathrm{cof}}) = \mathrm{Refl}\, L'((N^{\mathrm{gl}}\mathscr{D})^{\mathrm{cof}}) \xrightarrow{\simeq} \mathrm{Refl}\, N^{\natural}\mathscr{D} = N^{\mathrm{cmp}}\mathscr{D}$$

in $ms\mathscr{S}et_{(\infty,2)}$, as desired. □

4.1 Compatibility of the suspension and wedge constructions with the nerve

We consider the following two constructions in the globular setting:

(1) the *globular suspension* construction from Definition 1.20

$$\Sigma^{\mathrm{gl}} \colon s\mathscr{S}et^{\Delta^{\mathrm{op}}}_{p,(\infty,1)} \to (s\mathscr{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)})_{*,*},$$

which is a left Quillen functor; and

(2) the *globular wedge* construction

$$\vee^{\mathrm{gl}} \colon (s\mathscr{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)})_* \times (s\mathscr{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)})_* \to (s\mathscr{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)})_*,$$

defined by

$$W \vee^{\mathrm{gl}} Z := W \underset{\Theta_2[0]}{\amalg} Z,$$

which is a left Quillen bifunctor.

These constructions induce functors at the level of underlying $(\infty,1)$–categories that have the following recognized universal properties.

(1) The suspension construction induces precisely the construction studied in [Gepner and Haugseng 2015, Definition 4.3.21].

(2) The wedge construction induces the $(\infty,1)$–categorical coproduct in the $(\infty,1)$–category of pointed $(\infty,2)$–categories.

The analogous constructions can also be implemented in the complicial context as well, as

(1) the *complicial suspension* construction from [Ozornova and Rovelli 2022, Definition 2.6]

$$\Sigma^{\mathrm{cmp}} \colon ms\mathscr{S}et_{(\infty,1)} \to (ms\mathscr{S}et_{(\infty,2)})_{*,*};$$

(2) the *complicial wedge* construction from [Ozornova and Rovelli 2022, Definition 4.7]

$$\vee^{\mathrm{cmp}}\colon (ms\mathcal{S}et_{(\infty,2)})_* \times (ms\mathcal{S}et_{(\infty,2)})_* \to (ms\mathcal{S}et_{(\infty,2)})_*.$$

We record a precise relation between the globular and complicial suspension and wedge.

As a preliminary fact, we observe that with techniques analogous to the ones employed to construct the Quillen equivalence (4-1), we could also show that there is a Quillen equivalence

$$L_1\colon s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{p,(\infty,1)} \rightleftarrows \mathcal{S}et^{t\,\Delta^{\mathrm{op}}}_{(\infty,1)} : R_1$$

by setting, for all $m,k\geq 0$,

$$L_1(\Delta[m]\boxtimes\Delta[k]) := N^{\mathrm{RS}}([m])\times\Delta[k]^\sharp = \mathrm{th}_1(\Delta[m])\times\Delta[k]^\sharp.$$

Here, $s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{p,(\infty,1)}$ denotes the model structure for complete Segal spaces obtained by localizing the projective model structure, $ms\mathcal{S}et_{(\infty,1)}$ denotes the model structure for saturated 1–complicial sets, and $\mathrm{th}_1\colon ms\mathcal{S}et_{(\infty,1)}\to ms\mathcal{S}et_{(\infty,2)}$ is the left Quillen functor from [Verity 2008b, Notation 13].

Proposition 4.2 (a) *For all W cofibrant in $(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,1)}$ there is a weak equivalence*

$$\Sigma^{\mathrm{cmp}}L_1W \xrightarrow{\simeq} L\Sigma^{\mathrm{gl}}W$$

in $(ms\mathcal{S}et_{(\infty,2)})_{*,*}$.

(b) *Given any W and Z in $s\mathcal{S}et_*^{\Theta_2^{\mathrm{op}}}$, there is an isomorphism in $ms\mathcal{S}et$*

$$L(W\vee^{\mathrm{gl}}Z)\cong LW\vee^{\mathrm{cmp}}LZ.$$

Proof (a) By [Ozornova and Rovelli 2022, Theorem 2.9], there is a natural weak equivalence in $(ms\mathcal{S}et_{(\infty,2)})_{*,*}$ given by the composite

$$L\Sigma^{\mathrm{gl}}\Theta_2[m]\cong L\Theta_2[1|m] = N^{\mathrm{cmp}}[1|m] = N^{\mathrm{cmp}}\Sigma[m] \xleftarrow{\simeq} \Sigma^{\mathrm{cmp}}N^{\mathrm{cmp}}[m] = \Sigma^{\mathrm{cmp}}L_1\Delta[m].$$

The first three isomorphisms are given by the definitions of Σ^{gl}, L, and Σ, respectively, and the last is given by the definition L_1. The weak equivalence was established in [Ozornova and Rovelli 2022, Theorem 2.9]. Now, using [Dugger 2001, Lemma 9.7] on the functors

$$L\Sigma^{\mathrm{gl}},\ \Sigma^{\mathrm{cmp}}L_1\colon s\mathcal{S}et^{\Delta^{\mathrm{op}}}_p \to s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{p,(\infty,1)} \to (ms\mathcal{S}et_{(\infty,2)})_{*,*}$$

it follows that for all W cofibrant in $(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,1)}$ there is a weak equivalence

$$\Sigma^{\mathrm{cmp}}L_1W \xrightarrow{\simeq} L\Sigma^{\mathrm{gl}}W$$

in $(ms\mathcal{S}et_{(\infty,2)})_{*,*}$.

(b) Given any W and Z in $s\mathcal{S}et_*^{\Theta_2^{\mathrm{op}}}$, there is an isomorphism in $ms\mathcal{S}et$

$$L(W\vee^{\mathrm{gl}}Z)\cong L\Big(W\amalg_{\Theta_2[0]}Z\Big)\cong LW\amalg_{\Delta[0]}LZ\cong LW\vee^{\mathrm{cmp}}LZ,$$

concluding the proof. □

Remark 4.3 By [Hovey 1999, Corollary 1.4.4], for all (Y, c, d) fibrant in $(ms\mathcal{S}et_{(\infty,2)})_{*,*}$ there is a weak equivalence

$$\mathrm{Hom}^{\mathrm{gl}}_{RY}(c,d) \xrightarrow{\simeq} R_1 \mathrm{Hom}^{\mathrm{cmp}}_{Y}(c,d) \tag{4-2}$$

in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,1)}$ (and hence also in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,1)}$), where

$$\mathrm{Hom}^{\mathrm{gl}}\colon (s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)})_{*,*} \to s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{p,(\infty,1)}$$

denotes the right Quillen adjoint functor to Σ^{gl}, and similarly

$$\mathrm{Hom}^{\mathrm{cmp}}\colon (ms\mathcal{S}et_{(\infty,2)})_{*,*} \to ms\mathcal{S}et_{(\infty,1)}$$

denotes the right Quillen adjoint to Σ^{cmp}.

Recall that there are functors implementing the strict suspension construction

$$\Sigma\colon \mathcal{C}at \to 2\mathcal{C}at_{*,*}$$

and the strict wedge construction

$$\vee\colon 2\mathcal{C}at_* \times 2\mathcal{C}at_* \to 2\mathcal{C}at_*.$$

As an application of Theorem 2.20, combined with results from [Ozornova and Rovelli 2022], we can prove the following corollary, asserting that the suspension and wedge construction along a sieve/cosieve object are both compatible with the globular nerve of 2–categories. Recall from [Ozornova and Rovelli 2022, Definition 4.3] the definition of sieve and cosieve object in a 2–category, which are used to determine which 2–categories can be wedged together.

Corollary 4.4 (a) *Given any* 1*–category* $\mathcal{D}$, *there is a weak equivalence*

$$\Sigma^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{D} \xrightarrow{\simeq} N^{\mathrm{gl}}\Sigma\mathcal{D}$$

in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{(\infty,2)})_{*,*}$.

(b) *Given* 2*–categories* $\mathcal{A}$ *and* $\mathcal{B}$ *endowed with a sieve and a cosieve object*, *respectively*, *there is a weak equivalence in* $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{(\infty,2)})_*$,

$$N^{\mathrm{gl}}\mathcal{A} \vee^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{B} \xrightarrow{\simeq} N^{\mathrm{gl}}(\mathcal{A} \vee \mathcal{B}).$$

Proof We prove (b) and leave (a) to the interested reader.

First, in the commutative square

$$\begin{array}{ccc} N^{\mathrm{RS}}\mathcal{A} \vee^{\mathrm{cmp}} N^{\mathrm{RS}}\mathcal{B} & \xrightarrow{\simeq} & N^{\mathrm{cmp}}\mathcal{A} \vee^{\mathrm{cmp}} N^{\mathrm{cmp}}\mathcal{B} \\ \simeq\downarrow & & \downarrow \\ N^{\mathrm{RS}}(\mathcal{A} \vee \mathcal{B}) & \xrightarrow[\simeq]{} & N^{\mathrm{cmp}}(\mathcal{A} \vee \mathcal{B}) \end{array}$$

the left vertical map is a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$ by [Ozornova and Rovelli 2022, Theorem 4.9]. The two horizontal maps are weak equivalences in $ms\mathcal{S}et_{(\infty,2)}$, which can be seen by combining [Ozornova and Rovelli 2021, Theorem 5.2] and [Ozornova and Rovelli 2020, Proposition 1.31]. By the two-out-of-three property, the right-hand map is then a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$.

Next, in the commutative square

$$\begin{array}{ccc} L((N^{\mathrm{gl}}\mathcal{A})^{\mathrm{cof}}) \vee^{\mathrm{cmp}} L((N^{\mathrm{gl}}\mathcal{B})^{\mathrm{cof}}) & \xrightarrow{\simeq} & N^{\mathrm{cmp}}\mathcal{A} \vee^{\mathrm{cmp}} N^{\mathrm{cmp}}\mathcal{B} \\ \downarrow & & \downarrow \simeq \\ L((N^{\mathrm{gl}}(\mathcal{A}\vee\mathcal{B}))^{\mathrm{cof}}) & \xrightarrow[\simeq]{} & N^{\mathrm{cmp}}(\mathcal{A}\vee\mathcal{B}) \end{array}$$

the two horizontal maps are weak equivalences in $ms\mathcal{S}et_{(\infty,2)}$ by Proposition 4.1. By the two-out-of-three property, the left vertical map is a weak equivalence.

We have the commutative triangle

$$\begin{array}{ccc} L((N^{\mathrm{gl}}\mathcal{A})^{\mathrm{cof}} \vee^{\mathrm{gl}} (N^{\mathrm{gl}}\mathcal{B})^{\mathrm{cof}}) & \xrightarrow{\simeq} & L((N^{\mathrm{gl}}\mathcal{A} \vee^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{B})^{\mathrm{cof}}) \\ & \searrow^{\simeq} & \downarrow \\ & & L((N^{\mathrm{gl}}(\mathcal{A}\vee\mathcal{B}))^{\mathrm{cof}}) \end{array}$$

By [Hovey 1999, Corollary 1.3.16], the left Quillen equivalence L creates weak equivalences between cofibrant objects in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$, so we obtain a commutative square

$$\begin{array}{ccc} (N^{\mathrm{gl}}\mathcal{A} \vee^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{B})^{\mathrm{cof}} & \xrightarrow{\simeq} & (N^{\mathrm{gl}}(\mathcal{A}\vee\mathcal{B}))^{\mathrm{cof}} \\ \downarrow \simeq & & \downarrow \simeq \\ N^{\mathrm{gl}}\mathcal{A} \vee^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{B} & \longrightarrow & N^{\mathrm{gl}}(\mathcal{A}\vee\mathcal{B}) \end{array}$$

as desired. By the two-out-of-three property, the bottom map in the square is a weak equivalence in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$, as desired. □

While (a) can be essentially read off from [Rezk 2010], tackling directly (b) within the globular setting would require significant combinatorial work.

4.2 The cone construction and compatibility with the nerve

In the globular setting there does not seem to be a straightforward way to define join constructions, or even cone constructions, which play an important role in the development of the theory of limits and colimits. By contrast, the complicial setting is well-suited to implementing formal join constructions in general, and cones in particular. A *cone construction*

$$\mathrm{Cone}^{\mathrm{cmp}} := \Delta[0] \star -\colon ms\mathcal{S}et_{(\infty,1)} \to (ms\mathcal{S}et_{(\infty,2)})_*$$

is defined in [Gagna et al. 2023] in the form of a left Quillen functor.

Given any W in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}})_{p,(\infty,1)}$, taking advantage of the explicit Quillen equivalence (4-1), it is possible to define the cone construction for W in terms of the one for $L_1 W$, by setting

$$\mathrm{Cone}^{\mathrm{gl}} W := R((\mathrm{Cone}^{\mathrm{cmp}}(L_1(W^{\mathrm{cof}})))^{\mathrm{fib}}).$$

While the formula is fairly complicated, there is currently no competing way of treating cones in Θ_2–spaces.

Remark 4.5 For every W cofibrant in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}})_{p,(\infty,2)}$ there is a zigzag of weak equivalences

$$\begin{aligned} L((\mathrm{Cone}^{\mathrm{gl}} W)^{\mathrm{cof}}) = L((R((\mathrm{Cone}^{\mathrm{cmp}}(L_1(W^{\mathrm{cof}})))^{\mathrm{fib}}))^{\mathrm{cof}}) &\xrightarrow{\simeq} (\mathrm{Cone}^{\mathrm{cmp}}(L_1(W^{\mathrm{cof}})))^{\mathrm{fib}} \\ &\xleftarrow{\simeq} \mathrm{Cone}^{\mathrm{cmp}}(L_1(W^{\mathrm{cof}})) \\ &\xrightarrow{\simeq} \mathrm{Cone}^{\mathrm{cmp}}(L_1 W) \end{aligned}$$

in $ms\mathcal{S}et_{(\infty,2)}$.

Recall that there is a functor implementing the strict cone construction $\mathrm{Cone}\colon \mathcal{C}at \to 2\mathcal{C}at_*$. As an application of Theorem 2.20, combined with results from [Gagna et al. 2023], we can prove the following corollary, asserting that the cone construction is compatible with the nerve construction in a suitable sense in the globular setting for 1–categories that are freely generated by a loop-free graph. Such 1–categories are called *strong Steiner* in [Ara and Maltsiniotis 2020, Section 2.15].

Corollary 4.6 *Given any 1–category $\mathcal{D}$ that is freely generated by a loop-free graph, there is a zigzag of weak equivalences in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}})_{p,(\infty,2)}$,*

$$N^{\mathrm{gl}}\mathrm{Cone}\mathcal{D} \simeq \mathrm{Cone}^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{D}.$$

Proof There is a zigzag of weak equivalences

$$\begin{aligned} L((N^{\mathrm{gl}}\mathrm{Cone}\mathcal{D})^{\mathrm{cof}}) &\xrightarrow{\simeq} N^{\mathrm{cmp}}\mathrm{Cone}\mathcal{D} \\ &\xleftarrow{\simeq} \mathrm{Cone}^{\mathrm{cmp}} N^{\mathrm{cmp}}\mathcal{D} \\ &\xleftarrow{\simeq} \mathrm{Cone}^{\mathrm{cmp}} L_1((N^{\mathrm{gl}}\mathcal{D})^{\mathrm{cof}}) \simeq L(\mathrm{Cone}^{\mathrm{gl}}(N^{\mathrm{gl}}\mathcal{D}^{\mathrm{cof}})^{\mathrm{cof}}) \end{aligned}$$

in $ms\mathcal{S}et_{(\infty,2)}$ given by Proposition 4.1, [Gagna et al. 2023, Theorem 5.5], Proposition 4.1 for L_1, and Remark 4.5, respectively. Given that the left Quillen equivalence L creates weak equivalences between cofibrant objects, we obtain a zigzag of weak equivalences in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}})_{p,(\infty,2)}$

$$N^{\mathrm{gl}}\mathrm{Cone}\mathcal{D} \simeq (\mathrm{Cone}^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{D})^{\mathrm{cof}} \simeq (\mathrm{Cone}^{\mathrm{gl}}((N^{\mathrm{gl}}\mathcal{D})^{\mathrm{cof}}))^{\mathrm{cof}} \xrightarrow{\simeq} \mathrm{Cone}^{\mathrm{gl}}((N^{\mathrm{gl}}\mathcal{D})^{\mathrm{cof}}) \xrightarrow{\simeq} \mathrm{Cone}^{\mathrm{gl}} N^{\mathrm{gl}}\mathcal{D},$$

as desired. □

4.3 Dual constructions and compatibility with the nerve

It is determined in [Barwick and Schommer-Pries 2021, Theorem 7.3] that there are four types of dualities for $(\infty,2)$–categories: identity; op-dual, which reverses the direction of the 1–morphisms; co-dual, which reverses the direction of the 2–morphisms; and co-op-dual, which reverses both.

In the complicial setting, one can implement the op-dual construction $(-)^{\mathrm{op}}\colon ms\mathcal{S}et_{(\infty,2)} \to ms\mathcal{S}et_{(\infty,2)}$ in a straightforward way. However, there is no formal way to define the co-dual construction in $ms\mathcal{S}et_{(\infty,2)}$, and a co-dual construction $(-)^{\mathrm{co}}$ was only proposed recently in [Loubaton 2022, Proposition 4.2.7] in the form of a (highly nontrivial) left Quillen functor $(-)^{\mathrm{co}}\colon ms\mathcal{S}et_{(\infty,2)} \to ms\mathcal{S}et_{(\infty,2)}$.

By contrast, the globular setting is well-suited to implementing all four dualities; see [Haugseng 2021]. In particular, the co-dual construction can be realized as an isomorphism that is both a left and right Quillen equivalence for both the projectively-based and injectively-based model structures,

$$(-)^{\mathrm{co}}\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)} \to s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)} \quad \text{and} \quad (-)^{\mathrm{co}}\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,2)} \to s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{i,(\infty,2)}.$$

Given the Quillen equivalence (4-1), for all Y fibrant in $ms\mathcal{S}et_{(\infty,2)}$, there is a zigzag of weak equivalences

$$Y^{\mathrm{co}} \simeq L(((RY)^{\mathrm{co}})^{\mathrm{cof}}), \tag{4-3}$$

in $ms\mathcal{S}et_{(\infty,2)}$, allowing one to express the co-dual construction of Y in terms of the one for RY.

Remark 4.7 For every fibrant object Y in $ms\mathcal{S}et_{(\infty,2)}$, by (4-3) there is zigzag of weak equivalences

$$L(((RY)^{\mathrm{co}})^{\mathrm{cof}}) \simeq Y^{\mathrm{co}}$$

in $ms\mathcal{S}et_{(\infty,2)}$. By taking a functorial fibrant replacement in $ms\mathcal{S}et_{(\infty,2)}$, we obtain a zigzag of weak equivalences between fibrant objects in $ms\mathcal{S}et_{(\infty,2)}$

$$(L(((RY)^{\mathrm{co}})^{\mathrm{cof}}))^{\mathrm{fib}} \simeq (Y^{\mathrm{co}})^{\mathrm{fib}}.$$

Applying R then gives a zigzag of weak equivalences

$$R(L(((RY)^{\mathrm{co}})^{\mathrm{cof}}))^{\mathrm{fib}}) \simeq R((Y^{\mathrm{co}})^{\mathrm{fib}})$$

in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$. By composing with the component of the derived unit of $L \dashv R$ on $((RY)^{\mathrm{co}})^{\mathrm{cof}}$, we obtain a zigzag of weak equivalences

$$(RY)^{\mathrm{co}} \simeq ((RY)^{\mathrm{co}})^{\mathrm{cof}} \simeq R((Y^{\mathrm{co}})^{\mathrm{fib}})$$

in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$.

As an application of Theorem 2.20, we can prove the following corollary, asserting that the co-dual construction is compatible with the nerve of 2–categories in the complicial setting.

Corollary 4.8 *Given any 2–category $\mathcal{D}$, there is a zigzag of weak equivalences in $ms\mathcal{S}et_{(\infty,2)}$*

$$N^{\mathrm{cmp}}\mathcal{D}^{\mathrm{co}} \simeq (N^{\mathrm{cmp}}\mathcal{D})^{\mathrm{co}}.$$

Proof There a zigzag of weak equivalences

$$RN^{\mathrm{cmp}}(\mathcal{D}^{\mathrm{co}}) \simeq N^{\mathrm{gl}}(\mathcal{D}^{\mathrm{co}}) \simeq (N^{\mathrm{gl}}\mathcal{D})^{\mathrm{co}} \simeq (RN^{\mathrm{cmp}}\mathcal{D})^{\mathrm{co}} \simeq R((N^{\mathrm{cmp}}\mathcal{D})^{\mathrm{co}})^{\mathrm{fib}})$$

in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}})_{p,(\infty,2)}$, where the weak equivalences are given by Proposition 4.1, inspection, Proposition 4.1, and Remark 4.7, respectively. By [Hovey 1999, Corollary 1.3.16], the right Quillen equivalence R creates weak equivalences between fibrant objects in $(s\mathcal{S}et^{\Theta_2^{\mathrm{op}}})_{p,(\infty,2)}$, so we obtain a zigzag of weak equivalences

$$N^{\mathrm{cmp}}\mathcal{D}^{\mathrm{co}} \simeq ((N^{\mathrm{cmp}}\mathcal{D})^{\mathrm{co}})^{\mathrm{fib}} \simeq (N^{\mathrm{cmp}}\mathcal{D})^{\mathrm{co}}$$

in $ms\mathcal{S}et_{(\infty,2)}$, as desired. □

4.4 A fundamental theorem for 2–complicial sets

We define the *globular hom* construction

$$\mathrm{Hom}^{\mathrm{gl}}\colon (s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)})_{*,*} \to s\mathcal{S}et^{\Theta_1^{\mathrm{op}}}_{p,(\infty,1)}$$

as the right (Quillen) adjoint functor of the suspension Σ^{gl}. Similarly, we define the *globular homotopy category* construction

$$\mathrm{Ho}^{\mathrm{gl}}\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)} \to \mathcal{C}at$$

to be given by

$$\mathrm{Ho}^{\mathrm{gl}}\, X = h\tau_\Theta^* X,$$

where $h\colon s\mathcal{S}et^{\Delta^{\mathrm{op}}} \to \mathcal{C}at$ is the homotopy category functor from [Rezk 2010, Section 7.3] and

$$\tau_\Theta\colon \Delta = \Theta_1 \to \Theta_2$$

is the functor defined by $\tau_\Theta[m] = [m]([0],\dots,[0])$ from [Bergner and Rezk 2020, Section 3.2]; it is also defined in [Rezk 2010, Section 4.1] in more generality.

The following statement, which is essentially in [Bergner and Rezk 2020], can be thought of as a *fundamental theorem for* $(\infty,2)$*–categories*, referring to the terminology from [Rezk 2022], where the $(\infty,1)$–categorical case is treated in the model of quasicategories.

Proposition 4.9 *Let W and Z be fibrant in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$. A map $f\colon W\to Z$ in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$ is a weak equivalence if and only if*

(1) *the map f is* ***essentially surjective****, meaning that it induces an essentially surjective functor of homotopy categories*

$$\mathrm{Ho}^{\mathrm{gl}}\, f\colon \mathrm{Ho}^{\mathrm{gl}}\, W \to \mathrm{Ho}^{\mathrm{gl}}\, Z;$$

(2) *the map f is* ***homotopically fully faithful****, namely that it induces a weak equivalence*

$$f_{c,d}\colon \mathrm{Hom}^{\mathrm{gl}}_W(c,d) \simeq \mathrm{Hom}^{\mathrm{gl}}_Z(fc,fd)$$

in $s\mathcal{S}et^{\Delta^{\mathrm{op}}}_{p,(\infty,1)}$.

Proof As a consequence of [Bergner and Rezk 2020, Theorem 6.4], the functor

$$d^*\colon s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)} \to s\mathcal{S}et^{(\Delta\times\Delta)^{\mathrm{op}}}_{p,(\infty,2)}$$

creates weak equivalences, so saying that $f\colon W\to Z$ is a weak equivalence in $s\mathcal{S}et_{p,(\infty,2)}^{\Theta_2^{\mathrm{op}}}$ is equivalent to saying that

$$d^*f\colon d^*W\to d^*Z$$

is a weak equivalence in $s\mathcal{S}et_{p,(\infty,2)}^{(\Delta\times\Delta)^{\mathrm{op}}}$. Let us take a functorial fibrant replacement of d^*f in $s\mathcal{S}et_p^{\Theta_2^{\mathrm{op}}}$, and denote it by $(d^*f)^{\mathrm{pf}}\colon (d^*W)^{\mathrm{pf}}\to(d^*Z)^{\mathrm{pf}}$. Using the fact that d^* preserves fibrant objects [Bergner and Rezk 2020, Proposition 6.3], we can deduce that $d^*f\colon d^*W\to d^*Z$ is a weak equivalence in $s\mathcal{S}et_{p,(\infty,2)}^{\Theta_2^{\mathrm{op}}}\to s\mathcal{S}et_{p,(\infty,2)}^{(\Delta\times\Delta)^{\mathrm{op}}}$ if and only if

$$(d^*f)^{\mathrm{pf}}\colon (d^*W)^{\mathrm{pf}}\to(d^*Z)^{\mathrm{pf}}$$

is a weak equivalence in $s\mathcal{S}et_{p,(\infty,2)}^{(\Delta\times\Delta)^{\mathrm{op}}}$. Since $(d^*W)^{\mathrm{pf}}$ and $(d^*Z)^{\mathrm{pf}}$ are fibrant in $s\mathcal{S}et_{p,(\infty,2)}^{(\Delta\times\Delta)^{\mathrm{op}}}$, we know that $(d^*f)^{\mathrm{pf}}$ is a weak equivalence in $s\mathcal{S}et_{p,(\infty,2)}^{(\Delta\times\Delta)^{\mathrm{op}}}$ if and only if the same map is a weak equivalence in $s\mathcal{S}et_p^{(\Delta\times\Delta)^{\mathrm{op}}}$. Using the fact that Dwyer–Kan equivalences between complete Segal objects are precisely levelwise weak equivalences [Bergner and Rezk 2020, Proposition 8.17 and Definition 8.2], the map $(d^*f)^{\mathrm{pf}}$ is a weak equivalence in $s\mathcal{S}et_{p,(\infty,2)}^{(\Delta\times\Delta)^{\mathrm{op}}}$ if and only if

(1′) the map

$$\mathrm{Ho}((d^*f)^{\mathrm{pf}})\colon \mathrm{Ho}((d^*W)^{\mathrm{pf}})\to\mathrm{Ho}((d^*Z)^{\mathrm{pf}})$$

is an essentially surjective functor on homotopy categories, where Ho denotes the homotopy category from [Bergner and Rezk 2020, Section 8.1]; and

(2′) the map

$$((d^*f)^{\mathrm{pf}})_{a,b}\colon M^{\Delta}_{(d^*W)^{\mathrm{pf}}}((d^*f)^{\mathrm{pf}}(a),(d^*f)^{\mathrm{pf}}(b))\to M^{\Delta}_{(d^*Z)^{\mathrm{pf}}}((d^*f)^{\mathrm{pf}}(a),(d^*f)^{\mathrm{pf}}(b))$$

is a weak equivalence in $s\mathcal{S}et_{p,(\infty,1)}^{\Delta^{\mathrm{op}}}$, where M^{Δ} denotes the mapping object from [Bergner and Rezk 2020, Section 8.1].

Given the natural equivalence of categories

$$\mathrm{Ho}((d^*W)^{\mathrm{pf}})\simeq h\tau_{\Theta}^*W=\mathrm{Ho}^{\mathrm{gl}}\,W$$

and the natural weak equivalence

$$M^{\Delta}_{(d^*W)^{\mathrm{pf}}}(a,b)\simeq M^{\Delta}_{d^*W}(a,b)\cong\mathrm{Hom}^{\mathrm{gl}}_W(a,b)$$

from [Bergner and Rezk 2020, Proposition 3.10] in $s\mathcal{S}et_{p,(\infty,1)}^{\Delta^{\mathrm{op}}}$, we then obtain an equivalence to the conditions (1) and (2) from the statement, as desired. □

Aiming at providing a proof of the fundamental theorem for $(\infty,2)$–categories in the complicial context, recall the *complicial hom* construction

$$\mathrm{Hom}^{\mathrm{cmp}}\colon (ms\mathcal{S}et_{(\infty,2)})_{*,*}\to ms\mathcal{S}et_{(\infty,1)},$$

as the right (Quillen) adjoint functor of the suspension Σ^{cmp}, and the *complicial homotopy category* construction

$$\mathrm{Ho}^{\mathrm{cmp}}\colon ms\mathcal{S}et_{(\infty,2)} \to \mathcal{C}at$$

given by

$$\mathrm{Ho}^{\mathrm{cmp}} X = c_1^{\natural}\mathrm{sp}_1 X,$$

where $\mathrm{sp}_1\colon ms\mathcal{S}et \to ms\mathcal{S}et$ is the right Quillen functor of [Verity 2008b, Notation 13] and $c_1^{\natural}\colon ms\mathcal{S}et \to \mathcal{C}at$ is the left adjoint functor to the 1–dimensional natural nerve functor.

The homotopy category and hom constructions for the globular and complicial setting are compatible as follows:

Lemma 4.10 *Given a fibrant object X in $ms\mathcal{S}et_{(\infty,2)}$, there is an isomorphism of categories*

$$\mathrm{Ho}^{\mathrm{gl}} RX \cong \mathrm{Ho}^{\mathrm{cmp}} X.$$

Proof Given a fibrant object X in $ms\mathcal{S}et_{(\infty,2)}$, there is an isomorphism of categories

$$\mathrm{Ho}^{\mathrm{gl}} RX = h\tau_{\Theta}^{*} RX \cong hR_1\mathrm{sp}_1 X \cong c_1^{\natural}\mathrm{sp}_1 X = \mathrm{Ho}^{\mathrm{cmp}} X,$$

as desired. □

Using the Theorem 2.20, combined with Proposition 4.9, we can prove the following fundamental theorem for $(\infty,2)$–categories in the complicial setting.

Theorem 4.11 *Let X and Y be fibrant in $ms\mathcal{S}et_{(\infty,2)}$. A map $f\colon X \to Y$ in $ms\mathcal{S}et_{(\infty,2)}$ is a weak equivalence if and only if*

(1) *the map f is* ***essentially surjective****, meaning f induces an essentially surjective functor of homotopy categories*

$$\mathrm{Ho}^{\mathrm{cmp}} f\colon \mathrm{Ho}^{\mathrm{cmp}} X \simeq \mathrm{Ho}^{\mathrm{cmp}} Y;$$

(2) *the map f is* ***homotopically fully faithful****, namely f induces a weak equivalence in $ms\mathcal{S}et_{(\infty,1)}$,*

$$f_{c,d}\colon \mathrm{Hom}_X^{\mathrm{cmp}}(c,d) \simeq \mathrm{Hom}_Y^{\mathrm{cmp}}(fc,fd).$$

Proof A map $f\colon X \to Y$ is a weak equivalence in $ms\mathcal{S}et_{(\infty,2)}$ if and only if, by (4-1), the map $Rf\colon RX \to RY$ is a weak equivalence in $s\mathcal{S}et^{\Theta_2^{\mathrm{op}}}_{p,(\infty,2)}$. By Proposition 4.9, we can equivalently say that Rf is a Dwyer–Kan equivalence, in that the map

$$(Rf)^{\mathrm{gl}}_{c,d}\colon \mathrm{Hom}^{\mathrm{gl}}_{RX}(c,d) \to \mathrm{Hom}^{\mathrm{gl}}_{RY}(fc,fd)$$

is a weak equivalence in $(s\mathcal{S}et^{\Delta^{\mathrm{op}}})_{p,(\infty,1)}$ for all $c,d \in (RX)_{0,[0]}$ and that

$$\mathrm{Ho}^{\mathrm{gl}} Rf\colon \mathrm{Ho}^{\mathrm{gl}} RX \simeq \mathrm{Ho}^{\mathrm{gl}} RY$$

is an essentially surjective functor of homotopy categories. By Remark 4.3, these conditions are equivalent to having the analogous ones for $R_1(f^{\mathrm{cmp}})$, namely, that the map

$$R_1(f^{\mathrm{cmp}}_{c,d})\colon R_1 \operatorname{Hom}^{\mathrm{cmp}}_X(c,d) \to R_1 \operatorname{Hom}^{\mathrm{cmp}}_Y(fc,fd)$$

is a weak equivalence in $ms\mathcal{S}et_{(\infty,1)}$ for all $c,d \in X_0$ and that the map

$$\operatorname{Ho}^{\mathrm{gl}} Rf \colon \operatorname{Ho}^{\mathrm{gl}} RX \to \operatorname{Ho}^{\mathrm{gl}} RY$$

is essentially surjective. Then applying Lemma 4.10, we can equivalently say that f^{cmp} is a Dwyer–Kan equivalence, in that the map

$$f^{\mathrm{cmp}}_{c,d}\colon \operatorname{Hom}^{\mathrm{cmp}}_X(c,d) \to \operatorname{Hom}^{\mathrm{cmp}}_Y(fc,fd)$$

is a weak equivalence in $ms\mathcal{S}et_{(\infty,1)}$ for all $c,d \in X_0$ and that

$$\operatorname{Ho}^{\mathrm{cmp}} f \colon \operatorname{Ho}^{\mathrm{cmp}} X \to \operatorname{Ho}^{\mathrm{cmp}} Y$$

is an essentially surjective functor of homotopy categories, as desired. □

A proof of this fact internal to the complicial setting was outlined in [Campbell 2019], and provided recently in [Loubaton 2022, Corollary 3.2.11], but it relies on highly nontrivial combinatorics. Using our comparison with the Θ_2–model gives a much less technical proof.

Appendix The colossal model of $(\infty,2)$–categories

In this section, we give a model categorical variant of the colossal model by Barwick–Schommer-Pries.

In order to recall the original definition of the colossal model, we fix the following notations. We denote by Υ_2 the indexing category for the colossal model, namely the full subcategory of $2\mathcal{C}at$ as described by [Barwick and Schommer-Pries 2021, Definition 6.2]. In particular, $(\Upsilon_2^{\mathrm{op}})_\infty$ is the $(\infty,1)$–category obtained by regarding the category Υ_2^{op} as an $(\infty,1)$–category. We denote by $\mathcal{S}_\infty$ the $(\infty,1)$–category of spaces, namely $\mathcal{S}_\infty = (s\mathcal{S}et_{(\infty,0)})_\infty$.

Definition A.1 [Barwick and Schommer-Pries 2021] The *colossal model* is the $(\infty,1)$–category

$$\mathcal{L}_\infty(\mathcal{S}_\infty^{(\Upsilon_2^{\mathrm{op}})_\infty}),$$

obtained by localizing the presheaf $(\infty,1)$–category $\mathcal{S}_\infty^{(\Upsilon_2^{\mathrm{op}})_\infty}$ at the set of maps from [Barwick and Schommer-Pries 2021, Notation 8.3].

From the definition, we see that the colossal model is obtained by considering the $(\infty,1)$–category of spaces, taking a presheaf $(\infty,1)$–category valued in it, and then localizing. By contrast, one could instead present the same $(\infty,1)$–category by considering the Quillen model structure, which presents the $(\infty,1)$–category of spaces, taking the injective model structure on a presheaf category of functors valued in the Quillen model structure, and then a left Bousfield localization of it. More precisely, one can consider the following model structure.

Proposition A.2 *The category $s\mathcal{S}et^{\Upsilon_2^{op}}$ of Υ_2–spaces supports a cofibrantly generated model structure obtained by taking the left Bousfield localization of the injective model structure $s\mathcal{S}et_{inj}^{\Upsilon_2^{op}}$ with respect to the set of elementary acyclic cofibrations from* [Barwick and Schommer-Pries 2021, Notation 6.5]. *We denote this model structure by $s\mathcal{S}et_{(\infty,2)}^{\Upsilon_2^{op}}$.*

We want to prove that this model structure does present the colossal model, in the sense of the following theorem.

Theorem A.3 *There is an equivalence of $(\infty,1)$–categories*

$$\mathcal{L}_\infty\big((s\mathcal{S}et_{(\infty,0)})_\infty^{(\Upsilon_2^{op})_\infty}\big) \simeq \big(s\mathcal{S}et_{(\infty,2)}^{\Upsilon_2^{op}}\big)_\infty.$$

The proof is an application of the following result, which guarantees that one can build localizations of presheaf categories either as model categories or directly as $(\infty,1)$–categories.

Proposition A.4 *Let $\mathcal{A}$ be a category, $\mathcal{M}$ a left proper combinatorial simplicial model category, and Λ a set of maps in $\mathcal{M}^{\mathcal{A}}$. There is an equivalence of $(\infty,1)$–categories*

$$\mathcal{L}_\infty(\mathcal{M}_\infty^{\mathcal{A}_\infty}) \simeq (\mathcal{L}(\mathcal{M}_{inj}^{\mathcal{A}}))_\infty,$$

where $\mathcal{L}(\mathcal{M}_{inj}^{\mathcal{A}})$ denotes the Bousfield localization of the injective model structure $\mathcal{M}_{inj}^{\mathcal{A}}$ at Λ, and $\mathcal{L}_\infty(\mathcal{M}_\infty^{\mathcal{A}_\infty})$ denotes the localization of the $(\infty,1)$–category $\mathcal{M}_\infty^{\mathcal{A}_\infty}$ at Λ.

The proof of the proposition requires the following two ingredients.

Theorem A.5 [Lurie 2009a, Proof of Proposition A.3.7.8] *Let $\mathcal{N}$ be a left proper combinatorial simplicial model category, and Λ be a set of maps in $\mathcal{N}$. There is an equivalence of $(\infty,1)$–categories*

$$\mathcal{L}_\infty\mathcal{N}_\infty \simeq (\mathcal{L}\mathcal{N})_\infty,$$

where $\mathcal{L}\mathcal{N}$ denotes the Bousfield localization of the model structure $\mathcal{N}$ at Λ, and $\mathcal{L}_\infty\mathcal{N}_\infty$ denotes the localization of the $(\infty,1)$–category $\mathcal{N}_\infty$ at Λ_∞.

Theorem A.6 [Lurie 2009a, Proposition 4.2.4.4] *Let $\mathcal{A}$ be a category and $\mathcal{M}$ a combinatorial simplicial model category. There is an equivalence of $(\infty,1)$–categories*

$$\mathcal{M}_\infty^{\mathcal{A}_\infty} \simeq (\mathcal{M}_{inj}^{\mathcal{A}})_\infty,$$

where $\mathcal{M}_{inj}^{\mathcal{A}}$ denotes the injective model structure on $\mathcal{M}^{\mathcal{A}}$.

We can now prove the proposition.

Proof of Proposition A.4 Combining Theorems A.5 and A.6, we obtain an equivalence of $(\infty,1)$–categories

$$\mathscr{L}_\infty(\mathscr{M}_\infty^{\mathscr{A}_\infty}) \simeq \mathscr{L}_\infty(\mathscr{M}^{\mathscr{A}})_\infty \simeq (\mathscr{L}(\mathscr{M}^{\mathscr{A}}))_\infty,$$

as desired. □

We can now prove the theorem.

Proof of Theorem A.3 Applying Proposition A.4 with $\mathscr{M}=\mathscr{S}=s\mathscr{S}et_{(\infty,0)}$, $\mathscr{A}=\Upsilon_2^{\mathrm{op}}$ and $\Lambda=S$, the set of maps in $s\mathscr{S}et^{\Upsilon_2^{\mathrm{op}}}$ from [Barwick and Schommer-Pries 2021, Notation 8.3], we obtain the equivalence of $(\infty,1)$–categories

$$\mathscr{L}_\infty\big(\mathscr{S}_\infty^{(\Upsilon_2^{\mathrm{op}})_\infty}\big) = \mathscr{L}_\infty\big((s\mathscr{S}et_{(\infty,0)})_\infty^{(\Upsilon_2^{\mathrm{op}})_\infty}\big) \simeq \big(\mathscr{L}\big(s\mathscr{S}et^{\Upsilon_2^{\mathrm{op}}}_{(\infty,0)}\big)\big)_\infty = \big(s\mathscr{S}et^{\Upsilon_2^{\mathrm{op}}}_{(\infty,2)}\big)_\infty,$$

as desired. □

References

[Ara 2014] **D Ara**, *Higher quasi-categories vs higher Rezk spaces*, J. K–Theory 14 (2014) 701–749 MR Zbl

[Ara and Maltsiniotis 2020] **D Ara**, **G Maltsiniotis**, *Joint et tranches pour les ∞–catégories strictes*, Mém. Soc. Math. France 165, Soc. Math. France, Paris (2020) MR Zbl

[Barwick 2005] **C Barwick**, *(∞,n)–Cat as a closed model category*, PhD thesis, University of Pennsylvania (2005) Available at `https://www.proquest.com/docview/305445747`

[Barwick and Schommer-Pries 2021] **C Barwick**, **C Schommer-Pries**, *On the unicity of the theory of higher categories*, J. Amer. Math. Soc. 34 (2021) 1011–1058 MR Zbl

[Berger 2007] **C Berger**, *Iterated wreath product of the simplex category and iterated loop spaces*, Adv. Math. 213 (2007) 230–270 MR Zbl

[Berger and Moerdijk 2013] **C Berger**, **I Moerdijk**, *On the homotopy theory of enriched categories*, Q. J. Math. 64 (2013) 805–846 MR Zbl

[Bergner 2007a] **J E Bergner**, *Three models for the homotopy theory of homotopy theories*, Topology 46 (2007) 397–436 MR Zbl

[Bergner 2007b] **J E Bergner**, *A model category structure on the category of simplicial categories*, Trans. Amer. Math. Soc. 359 (2007) 2043–2058 MR Zbl

[Bergner 2007c] **J E Bergner**, *A characterization of fibrant Segal categories*, Proc. Amer. Math. Soc. 135 (2007) 4031–4037 MR Zbl

[Bergner and Rezk 2013] **J E Bergner**, **C Rezk**, *Comparison of models for (∞,n)–categories, I*, Geom. Topol. 17 (2013) 2163–2202 MR Zbl

[Bergner and Rezk 2020] **J E Bergner**, **C Rezk**, *Comparison of models for (∞,n)–categories, II*, J. Topol. 13 (2020) 1554–1581 MR Zbl

[Borceux 1994] **F Borceux**, *Handbook of categorical algebra, I: Basic category theory*, Encycl. Math. Appl. 50, Cambridge Univ. Press (1994) MR Zbl

[Campbell 2019] **A Campbell**, *Equivalences of complicial sets*, handwritten notes (2019) Available at https://acmbl.github.io/equivcompnotes.pdf

[Campbell 2020] **A Campbell**, *A homotopy coherent cellular nerve for bicategories*, Adv. Math. 368 (2020) art. id. 107138 MR Zbl

[Dugger 2001] **D Dugger**, *Universal homotopy theories*, Adv. Math. 164 (2001) 144–176 MR Zbl

[Duskin 2001] **J W Duskin**, *Simplicial matrices and the nerves of weak n–categories, I: Nerves of bicategories*, Theory Appl. Categ. 9 (2001) 198–308 MR Zbl

[Gagna et al. 2022] **A Gagna**, **Y Harpaz**, **E Lanari**, *On the equivalence of all models for $(\infty, 2)$–categories*, J. Lond. Math. Soc. 106 (2022) 1920–1982 MR Zbl

[Gagna et al. 2023] **A Gagna**, **V Ozornova**, **M Rovelli**, *Nerves and cones of free loop-free ω–categories*, Tunis. J. Math. 5 (2023) 273–326 MR Zbl

[Gepner and Haugseng 2015] **D Gepner**, **R Haugseng**, *Enriched ∞–categories via non-symmetric ∞–operads*, Adv. Math. 279 (2015) 575–716 MR Zbl

[Haugseng 2021] **R Haugseng**, *On lax transformations, adjunctions, and monads in $(\infty, 2)$–categories*, High. Struct. 5 (2021) 244–281 MR Zbl

[Hinich 2016] **V Hinich**, *Dwyer–Kan localization revisited*, Homology Homotopy Appl. 18 (2016) 27–48 MR Zbl

[Hirschhorn 2003] **P S Hirschhorn**, *Model categories and their localizations*, Math. Surv. Monogr. 99, Amer. Math. Soc., Providence, RI (2003) MR Zbl

[Hirschhorn 2021] **P S Hirschhorn**, *Overcategories and undercategories of cofibrantly generated model categories*, J. Homotopy Relat. Struct. 16 (2021) 753–768 MR Zbl

[Hovey 1999] **M Hovey**, *Model categories*, Math. Surv. Monogr. 63, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[Joyal 1997] **A Joyal**, *Disks, duality and Θ–categories*, preprint (1997) Available at https://ncatlab.org/nlab/files/JoyalThetaCategories.pdf

[Joyal 2007] **A Joyal**, *Quasi-categories vs simplicial categories*, preprint (2007) Available at http://www.math.uchicago.edu/~may/IMA/Incoming/Joyal/QvsDJan9(2007).pdf

[Joyal 2008] **A Joyal**, *The theory of quasi-categories and its applications*, preprint (2008) Available at http://mat.uab.cat/~kock/crm/hocat/advanced-course/Quadern45-2.pdf

[Joyal and Tierney 2007] **A Joyal**, **M Tierney**, *Quasi-categories vs Segal spaces*, from "Categories in algebra, geometry and mathematical physics", Contemp. Math. 431, Amer. Math. Soc., Providence, RI (2007) 277–326 MR Zbl

[Lack 2002] **S Lack**, *A Quillen model structure for 2–categories*, K–Theory 26 (2002) 171–205 MR Zbl

[Lack 2004] **S Lack**, *A Quillen model structure for bicategories*, K–Theory 33 (2004) 185–197 MR Zbl

[Lawson 2017] **T Lawson**, *Localization of enriched categories and cubical sets*, Theory Appl. Categ. 32 (2017) art. id. 35 MR Zbl

[Loubaton 2022] **F Loubaton**, *Dualities in the complicial model of ∞–categories*, preprint (2022) arXiv 2203.11845

[Lurie 2009a] **J Lurie**, *Higher topos theory*, Ann. of Math. Stud. 170, Princeton Univ. Press (2009) MR Zbl

[Lurie 2009b] **J Lurie**, *$(\infty, 2)$–categories and the Goodwillie calculus, I*, preprint (2009) arXiv 0905.0462

[Moser et al. 2022] **L Moser**, **V Ozornova**, **M Rovelli**, *Model independence of $(\infty, 2)$–categorical nerves*, preprint (2022) arXiv 2206.00660

[Olschok 2011] **M Olschok**, *Left determined model structures for locally presentable categories*, Appl. Categ. Structures 19 (2011) 901–938 MR Zbl

[Ozornova and Rovelli 2020] **V Ozornova**, **M Rovelli**, *Model structures for (∞, n)–categories on (pre)stratified simplicial sets and prestratified simplicial spaces*, Algebr. Geom. Topol. 20 (2020) 1543–1600 MR Zbl

[Ozornova and Rovelli 2021] **V Ozornova**, **M Rovelli**, *Nerves of 2–categories and 2–categorification of $(\infty, 2)$–categories*, Adv. Math. 391 (2021) art. id. 107948 MR Zbl

[Ozornova and Rovelli 2022] **V Ozornova**, **M Rovelli**, *Fundamental pushouts of n–complicial sets*, High. Struct. 6 (2022) 403–438 MR Zbl

[Pellissier 2002] **R Pellissier**, *Categories enrichies faibles*, PhD thesis, Université de Nice Sophia Antipolis (2002) arXiv math/0308246

[Rezk 1996] **C Rezk**, *A model category for categories*, preprint (1996) Available at `https://ncatlab.org/nlab/files/Rezk_ModelCategoryForCategories.pdf`

[Rezk 2001] **C Rezk**, *A model for the homotopy theory of homotopy theory*, Trans. Amer. Math. Soc. 353 (2001) 973–1007 MR Zbl

[Rezk 2010] **C Rezk**, *A Cartesian presentation of weak n–categories*, Geom. Topol. 14 (2010) 521–571 MR Zbl

[Rezk 2022] **C Rezk**, *Introduction to quasicategories* (2022) Available at `https://rezk.web.illinois.edu/quasicats.pdf`

[Riehl 2018] **E Riehl**, *Complicial sets: an overture*, from "2016 MATRIX annals", MATRIX Book Ser. 1, Springer (2018) 49–76 MR Zbl

[Riehl and Verity 2020] **E Riehl**, **D Verity**, *Recognizing quasi-categorical limits and colimits in homotopy coherent nerves*, Appl. Categ. Structures 28 (2020) 669–716 MR Zbl

[Riehl and Verity 2022] **E Riehl**, **D Verity**, *Elements of ∞–category theory*, Cambridge Stud. Adv. Math. 194, Cambridge Univ. Press (2022) MR Zbl

[Street 1987] **R Street**, *The algebra of oriented simplexes*, J. Pure Appl. Algebra 49 (1987) 283–335 MR Zbl

[Verity 2008a] **D Verity**, *Complicial sets characterising the simplicial nerves of strict ω–categories*, Mem. Amer. Math. Soc. 905, Amer. Math. Soc., Providence, RI (2008) MR Zbl

[Verity 2008b] **D R B Verity**, *Weak complicial sets, I: Basic homotopy theory*, Adv. Math. 219 (2008) 1081–1149 MR Zbl

Department of Mathematics, University of Virginia
Charlottesville, VA, United States

Max Planck Institute for Mathematics
Bonn, Germany

Department of Mathematics and Statistics, University of Massachusetts at Amherst
Amherst, MA, United States

`jeb2md@virginia.edu`, `viktoriya.ozornova@mpim-bonn.mpg.de`, `mrovelli@umass.edu`

Received: 26 September 2022 Revised: 2 August 2023

Algebraic & Geometric Topology **24**:7 (2024) 3875–3896
DOI: 10.2140/agt.2024.24.3875
Published: 9 December 2024

On products of beta and gamma elements in the homotopy of the first Smith–Toda spectrum

KATSUMI SHIMOMURA
MAO-NO-SUKE SHIMOMURA

We determine the first cohomology of the monochromatic comodule M_2^1 at an odd prime, and apply the results to show nontrivialities of some products of beta and gamma elements in the homotopy groups of the Smith–Toda spectrum $V(1)$. The cohomology for a prime greater than three was previously determined by the first author. Here we verify them and determine the cohomology at the prime 3 by elementary calculation. The cohomology will be a stepping stone for computing the cohomology of the monochromatic comodule M_0^3, which we hope to determine for a long time.

55Q45; 55Q51, 55T15

1 Introduction

Let p be an odd prime number and $\mathcal{S}_{(p)}$ denote the stable homotopy category of p–local spectra. Let $S \in \mathcal{S}_{(p)}$ denote the sphere spectrum. Then the mod p Moore spectrum M and the first Smith–Toda spectrum $V(1)$ are given by the cofiber sequences

(1.1) $$S \xrightarrow{p} S \xrightarrow{i} M \xrightarrow{j} \Sigma S \quad \text{and} \quad \Sigma^q M \xrightarrow{\alpha} M \xrightarrow{i_1} V(1) \xrightarrow{j_1} \Sigma^{q+1} M.$$

Here $p \in \pi_0(S) \cong \mathbb{Z}_{(p)}$, and $\alpha \in [M, M]_q$ denotes the Adams map. Hereafter we put

$$q = 2p - 2 \in \mathbb{Z}.$$

To study the homotopy groups $\pi_*(X)$ of a spectrum X, we adopt the Adams–Novikov spectral sequence

(1.2) $$E_2^{s,t}(X) = H^{s,t}\,\mathrm{BP}_*(X) \Rightarrow \pi_{t-s}(X).$$

Hereafter we abbreviate as

$$H^{s,t} M = \mathrm{Ext}^{s,t}_{\mathrm{BP}_*(\mathrm{BP})}(\mathrm{BP}_*, M)$$

for a $\mathrm{BP}_*(\mathrm{BP})$–comodule M over the Hopf algebroid

(1.3) $$(\mathrm{BP}_*, \mathrm{BP}_*(\mathrm{BP})) = (\mathbb{Z}_{(p)}[v_1, v_2, \dots], \mathrm{BP}_*[t_1, t_2, \dots])$$

based on the Brown–Peterson spectrum $\mathrm{BP} \in \mathcal{S}_{(p)}$. We note that the v_i are Hazewinkel's generators and the degrees of v_i and t_i are $|v_i| = 2p^i - 2 = |t_i|$; see Miller, Ravenel and Wilson [2, (1.1)].

Let

(1.4) $$I_n = (p, v_1, \dots, v_{n-1}) \quad \text{and} \quad J_j = (p, v_1, v_2^j)$$

for $v_0 = p$ denote the invariant ideals of BP_*. Since $\mathrm{BP}_*(\alpha) = v_1$, the cofiber sequences (1.1) induce the short exact sequences

$$(1.5)\qquad 0 \to \mathrm{BP}_* \xrightarrow{p} \mathrm{BP}_* \xrightarrow{i_*} \mathrm{BP}_*/I_1 \to 0 \quad\text{and}\quad 0 \to \mathrm{BP}_*/I_1 \xrightarrow{v_1} \mathrm{BP}_*/I_1 \xrightarrow{(i_1)_*} \mathrm{BP}_*/I_2 \to 0$$

along with the isomorphisms

$$\mathrm{BP}_*(S) = \mathrm{BP}_*, \quad \mathrm{BP}_*(M) = \mathrm{BP}_*/I_1 \quad\text{and}\quad \mathrm{BP}_*(V(1)) = \mathrm{BP}_*/I_2.$$

Furthermore, we have a short exact sequence

$$(1.6)\qquad 0 \to \mathrm{BP}_*/I_2 \xrightarrow{v_2^j} \mathrm{BP}_*/I_2 \xrightarrow{\bar{i}_j} \mathrm{BP}_*/J_j \to 0$$

for $j \ge 1$. We denote by

$$\delta_0 \colon H^s\mathrm{BP}_*/I_1 \to H^{s+1}\mathrm{BP}_*, \quad \delta_1 \colon H^s\mathrm{BP}_*/I_2 \to H^{s+1}\mathrm{BP}_*/I_1, \quad \bar{\delta}_j \colon H^s\mathrm{BP}_*/J_j \to H^{s+1}\mathrm{BP}_*/I_2$$

the connecting homomorphisms associated to the short exact sequences (1.5) and (1.6). We define the Greek letter elements by

$$\begin{aligned}
\bar{\beta}'_s &= \delta_1(v_2^s) \in E_2^1(M) = H^1\,\mathrm{BP}_*/I_1 && \text{for } v_2^s \in H^0\,\mathrm{BP}_*/I_2,\\
\bar{\beta}_s &= \delta_0\delta_1(v_2^s) \in E_2^2(S) = H^2\,\mathrm{BP}_* && \text{for } v_2^s \in H^0\,\mathrm{BP}_*/I_2,\\
\bar{\gamma}''_{s/j} &= \bar{\delta}_j(v_3^s) \in E_2^1(V(1)) = H^1\,\mathrm{BP}_*/I_2 && \text{for } v_3^s \in H^0\,\mathrm{BP}_*/J_j,
\end{aligned}$$

and $\bar{\gamma}''_s = \bar{\gamma}''_{s/1} \in E_2^1(V(1))$. We notice that $1 \le j \le p^n$ if $p^n \mid s$, so that $v_3^s \in H^0\,\mathrm{BP}_*/J_j$.

Let $\mathbb{Z}$ and $\mathbb{N}$ denote the set of all integers and its subset consisting of all nonnegative integers, respectively. We denote by $\mathbb{Z}^{(p)}$ $(= \mathbb{Z} \setminus p\mathbb{Z})$ and $\mathbb{N}^{(p)}$ $(= \mathbb{N} \setminus p\mathbb{N})$ the sets of the integers prime to p, and decompose $\mathbb{Z}^{(p)}$ into the three summands

$$\mathbb{Z}^{(p)} = \mathbb{Z}_0 \coprod \mathbb{Z}_1 \coprod \mathbb{Z}_2,$$

for

$$(1.7)\qquad \begin{gathered}
\mathbb{Z}_0 = \{s \in \mathbb{Z}^{(p)} : p \nmid (s+1)\}, \quad \mathbb{Z}_1 = \{s \in \mathbb{Z}^{(p)} : p^2 \mid (s+1)\},\\
\mathbb{Z}_2 = \{s \in \mathbb{Z}^{(p)} : p \mid (s+1) \text{ and } p^2 \nmid (s+1)\}.
\end{gathered}$$

We consider subsets of $\mathbb{N}$:

$$\begin{aligned}
2\mathbb{N}_{>0} &= \{s \in \mathbb{N} : s \text{ is even} \ge 2\}, & \mathbb{N}_1 &= \{s \in \mathbb{N}^{(p)} : p^2 \nmid (s+p+1) \text{ or } p^3 \mid (s+p+1)\},\\
\overline{2\mathbb{N}} &= \{s \in \mathbb{N} : s \text{ is odd}\}, & \mathbb{N}_2 &= \{s \in \mathbb{N}^{(p)} : p \nmid (s+2) \text{ or } p^3 \mid (s+2)(s+2+p)\}.
\end{aligned}$$

Furthermore, we put $\mathbb{Z}_i^+ = \mathbb{Z}_i \cap \mathbb{N}$ for $i = 0, 1, 2$. We introduce the subsets U_1, U'_1, U_2 and U'_2 of $\mathbb{N}^{(p)} \times \mathbb{N}$ given by

$$\begin{aligned}
U_1 &= (\mathbb{N}^{(p)} \times 2\mathbb{N}) \cup (\mathbb{Z}_0^+ \times \mathbb{N}),\\
U'_1 &= (\mathbb{N}^{(3)} \times \{0\}) \cup (\mathbb{N}_1 \times 2\mathbb{N}_{>0}) \cup ((\mathbb{Z}_0^+ \cap \mathbb{N}_2) \times \mathbb{N}) \cup (\mathbb{Z}_0^+ \times \{1\}),\\
U_2 &= (\mathbb{N}_1 \times 2\mathbb{N}) \cup (((\mathbb{Z}_0^+ \cap \mathbb{N}_2) \cup \mathbb{Z}_1^+) \times \mathbb{N}) \cup (\mathbb{N}^{(p)} \times \{1\}),\\
U'_2 &= (\mathbb{N}_1 \times \{0\}) \cup (\mathbb{N}^{(3)} \times (\{1\} \cup 2\mathbb{N}_{>0})) \cup ((\mathbb{Z}_0^+ \cup \mathbb{Z}_1^+) \times \mathbb{N}).
\end{aligned}$$

Our main result is the following:

Theorem 1.8 *Let p be an odd prime. In the Adams–Novikov E_2–term for computing $\pi_*(V(1))$, $\bar\beta_1$ and $\bar\beta_2$ act on the gamma elements $\bar\gamma''_{sp^r/j}$ (for $(s,r)\in\mathbb{N}^{(p)}\times\mathbb{N}$ and $1\le j\le p^r$) by*

$$\bar\gamma''_{sp^r/j}\bar\beta_1\neq 0 \quad \text{for } (s,r)\in U_1 \text{ if } p\ge 5, \text{ and for } (s,r)\in U_1' \text{ if } p=3,$$
$$\bar\gamma''_{sp^r/j}\bar\beta_2\neq 0 \quad \text{for } (s,r)\in U_2 \text{ if } p\ge 5, \text{ and for } (s,r)\in U_2' \text{ if } p=3,$$

in $E_2^3(V(1))$.

There is a way to define $\gamma''_{sp^r/j}$ for $j\le a_r$ (a_r is defined in (2.7)) so that $v_2^{j-1}\gamma''_{sp^r/j}=\gamma''_{sp^r}$, and the theorem holds for such extended gamma elements. Also $\bar\beta_s\equiv\binom{s}{2}v_2^{s-2}\bar\beta_2+s(2-s)v_2^{s-1}\bar\beta_1 \mod I_2$ (see Oka and Shimomura [5, Lemma 4.4]), and so

$$\bar\gamma''_{sp^r/j}\bar\beta_t=\binom{t}{2}\bar\gamma''_{sp^r/j-t+2}\bar\beta_2+t(2-t)\bar\gamma''_{sp^r/j-t+1}\bar\beta_1.$$

Thus Theorem 1.8 implies nontriviality of the products of $\bar\gamma''_{sp^r/j}$ and $\bar\beta_t$.

The Adams–Novikov differential d_r is 0 if $q\nmid(r-1)$ by the sparseness of the spectral sequence (1.2). This shows that the products in the theorem are not in the image of any differentials d_r. It is well known that the elements $\bar\beta_1$ and $\bar\beta_2$ converge to the homotopy elements $\beta_1,\beta_2\in\pi_*(S)$, respectively, in the spectral sequence (1.2) for $X=S$.

Corollary 1.9 *Let p be an odd prime. If $\bar\gamma''_{sp^r/j}\in E_2^1(V(1))$ is a permanent cycle detecting $\gamma''_{sp^r/j}\in\pi_*(V(1))$, then $\gamma''_{sp^r/j}\beta_i\neq 0$ for $i=1,2$ in the homotopy groups $\pi_*(V(1))$ for (s,r) given in Theorem 1.8.*

Toda [12, Theorem 1] and Oka [4, Theorem 4.2] showed that γ''_s and $\gamma''_{sp/2}$ are permanent cycles for $p\ge 7$.

Corollary 1.10 *Let $p\ge 7$, and r and s be integers with $(s,r)\in\mathbb{N}^{(p)}\times\mathbb{N}$. Then, in $\pi_*(V(1))$,*

$$\gamma''_{sp^r/j}\beta_1\neq 0 \quad \text{if } r \text{ is even or } p\nmid(s+1),$$
$$\gamma''_{sp^{2r}/j}\beta_2\neq 0 \quad \text{if } p^2\nmid(s+p+1) \text{ or } p^3\mid(s+p+1),$$
$$\gamma''_{sp^{2r+1}/j}\beta_2\neq 0 \quad \text{for } r\ge 1 \text{ if } p\nmid(s+1)(s+2),\ p^2\mid(s+1) \text{ or } p^3\mid(s+2)(s+2+p),$$

and $\gamma''_{sp/j}\beta_2\neq 0$, where $j=1,2$.

Theorem 1.8 follows from Theorem 2.9, which states the structure of the first cohomology of the monochromatic comodule M_2^1. The cohomology $H^1M_2^1$ was determined by the first author [10] based on the computation in [9] at a prime ≥ 5. Here we determine the cohomology based on elementary calculation at an odd prime. The generators are explicitly given so that we can use the result easily in further computation. This result will be a stepping stone for determining the long-desired cohomology $H^*M_0^3$.

This paper is organized as follows: In Section 2, we state the main result, Theorem 2.9, which gives the structure of $H^1M_2^1$. In Section 3, we prove Theorems 2.9 and 1.8 assuming Lemma 3.4, whose proof will be given in the Section 6. Section 4 is devoted to introducing some formulas, cochains and relations for the following sections. We refine the elements $x_{3,i}$ given by Miller, Ravenel and Wilson [2, (5.11)] to define x_i, which induce the cochains $y_{s,i}, y'_{s,i}\in\Omega^1E(3)_*$ in Section 5.

Acknowledgments The authors would like to express their gratitude to the referee for the careful reading of the manuscript and useful suggestions.

2 The structure of $H^1M_2^1$

In this section, we state the structure of $H^1M_2^1$ for an odd prime p. The structure was given in [10] for primes $p \geq 5$.

We begin with defining the monochromatic $\mathrm{BP}_*(\mathrm{BP})$–comodules N_n^s and M_n^s inductively by

$$N_n^0 = \mathrm{BP}_*/I_n, \quad M_n^s = v_{s+n}^{-1}N_n^s,$$

for the ideal I_n in (1.4) and the short exact sequence

$$0 \to N_n^s \xrightarrow{\iota_n^s} M_n^s \xrightarrow{\kappa_n^s} N_n^{s+1} \to 0 \tag{2.1}$$

see [2, Section 3.A]. Since BP_* is a $\mathrm{BP}_*(\mathrm{BP})$–comodule with structure map η_R, the right unit map of the Hopf algebroid $\mathrm{BP}_*(\mathrm{BP})$, these monochromatic comodules have the structure maps induced from η_R.

Let $E(3)$ denote the third Johnson–Wilson spectrum, which yields a Hopf algebroid

$$\big(E(3)_*, E(3)_*(E(3))\big) = (\mathbb{Z}_{(p)}[v_1, v_2, v_3, v_3^{-1}], E(3)_* \otimes_{\mathrm{BP}_*} \mathrm{BP}_*(\mathrm{BP}) \otimes_{\mathrm{BP}_*} E(3)_*).$$

Its structure maps are induced from the Hopf algebroid $(\mathrm{BP}_*, \mathrm{BP}_*(\mathrm{BP}))$ in (1.3). Since we have the Miller–Ravenel change of rings theorem

$$H^*M = \mathrm{Ext}^*_{\mathrm{BP}_*(\mathrm{BP})}(\mathrm{BP}_*, M) \cong \mathrm{Ext}^*_{E(3)_*(E(3))}(E(3)_*, E(3)_* \otimes_{\mathrm{BP}_*} M)$$

for a v_3–local $\mathrm{BP}_*(\mathrm{BP})$–comodule M [1, Theorem 3.10], we denote the cohomology of an $E(3)_*(E(3))$–comodule M also by

$$H^sM = \mathrm{Ext}^s_{E(3)_*(E(3))}(E(3)_*, M).$$

By virtue of the change of rings theorem, we denote simply by M_n^s the $E(3)_*(E(3))$–comodule $E(3)_* \otimes_{\mathrm{BP}_*} M_n^s$. We consider the Ext group as the cohomology group of the cobar complex

$$\Omega^sM = M \otimes_{E(3)_*} E(3)_*(E(3)) \otimes_{E(3)_*} \cdots \otimes_{E(3)_*} E(3)_*(E(3)) \tag{2.2}$$

for s factors of $E(3)_*(E(3))$, with well-known differentials $d_r\colon \Omega^rM \to \Omega^{r+1}M$; see (4.1).

The cohomology $H^tM_n^s$ of the monochromatic comodules with $s+n=3$ are determined in the following cases (see [8, Theorems 6.3.12 and 6.3.14; 2, Theorem 5.10]):

$$\begin{aligned}
H^0M_3^0 &= K(3)_*,\\
H^1M_3^0 &= K(3)_*\{h_0, h_1, h_2, \zeta_3\},\\
H^2M_3^0 &= K(3)_*\{g_i, k_i, b_i, h_i\zeta_3 : i \in \mathbb{Z}/3\},\\
H^0M_2^1 &= K(2)_*/k(2)_* \oplus \bigoplus_{i\geq 0, s\in\mathbb{Z}^{(p)}} k(2)_*/(v_2^{a_i})\{x_i^s/v_2^{a_i}\}.
\end{aligned} \tag{2.3}$$

Indeed, we read off $H^sM_3^0 = K(3)_* \otimes H^sS(3)$ from [8, Proposition 6.2.1], where $S(3)$ is the Hopf algebra defined in [8, Section 6.2]. The cohomology groups $H^*M_3^0$ and $H^0M_1^2$ for $p \geq 5$ are also determined by Ravenel [8, Theorem 6.3.34] and Nakai [3], respectively. Here

$$k(2)_* = \mathbb{Z}/p[v_2], \quad K(2)_* = \mathbb{Z}/p[v_2, v_2^{-1}] \quad \text{and} \quad K(3)_* = \mathbb{Z}/p[v_3, v_3^{-1}],$$

where $K(3)_* = E(3)_*/I_3 = M_3^0$. The elements x_i $(= x_{3,i})$ are introduced in [2, (5.11)] and are such that $x_i \equiv v_3^{p^i} \mod I_3$ (see Lemma 5.1), and the generators h_i, ζ_3, g_i, k_i and b_i are defined by cocycles in the cobar complex $\Omega^*E(3)_*/I_3$ as follows:

$$\textbf{(2.4)} \qquad h_i = [t_1^{p^i}], \quad \zeta_3 = [Z], \quad g_i = [G_i], \quad k_i = [K_i] \quad \text{and} \quad b_i = [b_{1,i}].$$

Hereafter $[x]$ denotes the cohomology class represented by a cocycle x, and the representatives in (2.4) are defined by

$$\textbf{(2.5)} \qquad \begin{aligned} Z &= -v_3^{-1}ct_3 + v_3^{-p}t_3^p + v_3^{-p^2}t_3^{p^2} - v_3^{-p}t_1^pt_2^{p^2}, & G_i &= t_1^{p^i} \otimes t_2^{p^i} + \tfrac{1}{2}t_1^{2p^i} \otimes t_1^{p^{i+1}}, \\ K_i &= t_2^{p^i} \otimes t_1^{p^{i+1}} + \tfrac{1}{2}t_1^{p^i} \otimes t_1^{2p^{i+1}}, & b_{1,i} &= \sum_{k=1}^{p-1} \frac{1}{p}\binom{p}{k} t_1^{kp^i} \otimes t_1^{(p-k)p^i}. \end{aligned}$$

Here ct_3 is the Hopf conjugation of t_3 (see Lemma 4.3). We notice that G_i, K_i and $b_{1,i}$ are also cocycles of $\Omega^*E(3)_*/I_2$, and of $\Omega^*\mathrm{BP}_*/I_2$ in [2, (1.9)].

Remark 2.6 The generators g_i and k_i in (2.3) are given by the Massey products $\langle h_i, h_{i+1}, h_i\rangle$ and $\langle h_{i+1}, h_{i+1}, h_i\rangle$, respectively, in [8, Theorem 6.3.4]. These are represented by cocycles

$$G_i'' = t_2^{p^i} \otimes t_1^{p^{i+1}} + t_1^{p^i} \otimes ct_2^{p^i}$$

and K_i' in (4.20) in the cobar complex $\Omega^*E(3)_*/I_2$, since these Massey products have no indeterminacy. By (4.21), K_i' is homologous to K_i. Also $d_1(t_1^{p^i}t_2^{p^i}) = -2G_i - G_i''$, and G_i'' is homologous to $-2G_i$. Since p is odd, we may replace generators g_i and k_i by $[G_i]$ and $[K_i]$, and set as in (2.4).

We introduce integers $e(n)$, a_n, $j_{s,n}$ and $j'_{s,n}$ for integers $n(\geq 0)$ and s by

$$\textbf{(2.7)} \qquad \begin{aligned} e(n) &= \frac{p^n-1}{p-1} \quad \text{for } n \geq 0, \\ a_n &= \begin{cases} 1 & \text{for } n = 0, \\ p^n + (p^{n-1}-1)/(p+1) & \text{for odd } n \geq 1, \\ p^n + p(p^{n-2}-1)/(p+1) & \text{for even } n \geq 2, \end{cases} \\ j_{s,n} &= \begin{cases} 2 & \text{for } s \in \mathbb{Z}_0 \text{ and } n = 0, \\ 2p^2 - p + 1 & \text{for } s \in \mathbb{Z}_0 \text{ and } n = 2, \\ 2a_n + \bar{1} & \text{for } s \in \mathbb{Z}_0, \text{ even } n \geq 4, \\ a_{n+2} - a_{n+1} & \text{for } s \in \mathbb{Z}_1 \text{ and even } n \geq 0, \\ p+1 & \text{for } s \in \mathbb{Z}^{(p)} \text{ and } n = 1, \\ e(3)p^{n-2} - p + 1 & \text{for } s \in \mathbb{Z}^{(p)} \text{ and odd } n \geq 3, \end{cases} \end{aligned}$$

$$j'_{s,0} = \begin{cases} 2 & \text{for } p \nmid s(s-1), \\ 2p & \text{for } s = tp+1 \text{ and } p \nmid t(t-1), \\ p^2+1 & \text{for } s = tp^2+1 \text{ and } p \nmid t, \\ a_n + p & \text{for } s = tp^n+1 \text{ with } n > 2 \text{ and } p \nmid t, \\ a_n + 1 & \text{for } s = tp^n + e(n) \text{ with even } n \geq 2 \text{ and } p \nmid (t-1), \\ a_n + 2 & \text{for } s = tp^n + e(n) \text{ with odd } n > 2 \text{ and } p \nmid (t-1), \end{cases}$$

$$j'_{s,n} = \begin{cases} 2p & \text{for } s \in \mathbb{Z}_0 \text{ and } n = 1, \\ 2pa_{n-1} + p & \text{for } s \in \mathbb{Z}_0 \text{ and odd } n \geq 3, \\ pa_{n+1} - pa_n & \text{for } s \in \mathbb{Z}_1 \text{ and odd } n \geq 1, \\ p^2 + p & \text{for } s \in \mathbb{Z}^{(p)} \text{ and } n = 2, \\ e(3)p^{n-2} - 1 + \bar{1} & \text{for } s \in \mathbb{Z}^{(p)} \text{ and even } n \geq 4. \end{cases}$$

Here $\bar{1} = 0$ if $p \geq 5$ and 1 if $p = 3$, the $\mathbb{Z}_i$ are the subsets of the integers $\mathbb{Z}$ defined in (1.7), and the integers a_n are $a_{3,n}$ in [2, (5.13)]. We note that

(2.8) $$a_n + a_{n-1} = e(3)p^{n-2} - 1 \quad \text{for } n \geq 2 \qquad \text{and} \qquad p^n + a_{n-2} - p^{n-3} = a_n \quad \text{for } n \geq 3.$$

Theorem 2.9 *Let p be an odd prime. Then $H^1M_2^1$ is the direct sum of the $k(2)_*$–module $B_\infty = K(2)_*/k(2)_*\{h_0, h_1, \tilde{\zeta}_2, \zeta_3\}$ and the $k(2)_*$–cyclic modules generated by*

$$\begin{aligned} &(\zeta_3)_{sp^n/a_n} && \text{for } (s,n) \in \mathbb{Z}^{(p)} \times \mathbb{N}, \\ &(h_0)_{sp^n/j_{s,n}} && \text{for } (s,n) \in ((\mathbb{Z}_0 \cup \mathbb{Z}_1) \times 2\mathbb{N}) \cup (\mathbb{Z}^{(p)} \times \overline{2\mathbb{N}}), \\ &(h_1)_{sp^n/j'_{s,n}} && \text{for } (s,n) \in ((\mathbb{Z}_0 \cup \mathbb{Z}_1) \times \overline{2\mathbb{N}}) \cup ((\mathbb{Z}^{(p)} \times 2\mathbb{N}) \setminus \{(1,0)\}), \\ &(h_2)_{tp-1/p-1} && \text{for } t \in \mathbb{Z}. \end{aligned}$$

There is a little difference between the cases for $p \geq 5$ and $p = 3$. In the theorem, $\tilde{\zeta}_2$ $(= (h_1)_1)$ denotes the homology class of z in (4.18) (see also (3.8)), and the generators $(\xi)_{s/j}$ for $\xi = [X]$ in $H^1M_3^0$ denote

$$(\xi)_{s/j} = [v_3^s X/v_2^j + *]$$

for a cocycle $v_3^s X/v_2^j + *$ of the cobar complex $\Omega^1 M_2^1$ with an element $*$ killed by v_2^{j-1}. The element v_2 acts on $(\xi)_{s/j}$ by

(2.10) $$v_2(\xi)_{s/j} = (\xi)_{s/j-1} \quad \text{and} \quad v_2(\xi)_{s/1} = 0,$$

and so $(\xi)_{s/j}$ generates a cyclic $k(2)_*$–module isomorphic to $k(2)_*/(v_2^j)$:

$$k(2)_*\{(\xi)_{s/j}\} \cong k(2)_*/(v_2^j).$$

3 Proofs of Theorems 2.9 and 1.8

In this section, we assume Lemma 3.4, which will be verified by a routine calculation in Section 6.

3.1 Proof of Theorem 2.9

For the monochromatic comodules defined in Section 2, we have a short exact sequence

(3.1)
$$0 \to M_3^0 \xrightarrow{\eta} M_2^1 \xrightarrow{v_2} M_2^1 \to 0,$$

where $\eta(x) = x/v_2$ (see [2, (3.10)]), which induces the long exact sequence

(3.2)
$$\cdots \to H^0 M_2^1 \xrightarrow{\delta_0} H^1 M_3^0 \xrightarrow{\eta_*} H^1 M_2^1 \xrightarrow{v_2} H^1 M_2^1 \xrightarrow{\delta_1} H^2 M_3^0 \to \cdots.$$

From [2, (5.18)], we read off the following:

Proposition 3.3 *The cokernel of $\delta_0\colon H^0 M_2^1 \to H^1 M_3^0$ is a $\mathbb{Z}/p$–module generated by $(h_0)_0$, $(h_1)_0$,*

$$(h_0)_{sp^{2k}} \text{ for } s \in \mathbb{Z}_0 \cup \mathbb{Z}_1, \quad (h_0)_{tp^{2k+1}} \text{ for } t \in \mathbb{Z}^{(p)}, \quad (h_1)_{tp^{2k}} \text{ for } t \in \mathbb{Z}^{(p)},$$
$$(h_1)_{sp^{2k+1}} \text{ for } s \in \mathbb{Z}_0 \cup \mathbb{Z}_1, \quad (h_2)_{tp-1} \text{ for } t \in \mathbb{Z}, \quad (\zeta_3)_t \text{ for } t \in \mathbb{Z},$$

for $k \geq 0$. Here $\mathbb{Z}_i$ is a subset of $\mathbb{Z}$ given in (1.7), and $(\xi)_s = v_3^s \xi$ for $\xi \in \{h_i, \zeta_3 : i \in \mathbb{Z}/3\}$.

Let $(x)_s \in \Omega^1 E(3)_*$ denote a cochain satisfying

$$(x)_s \equiv v_3^s x \mod I_3.$$

Lemma 3.4 *The following cochains exist in $\Omega^1 E(3)_*/I_2$:*

(1) $(t_1)_{sp^{2k}}$ *and* $(t_1^p)_{sp^{2k+1}}$ *for $s \in \mathbb{Z}_0$ such that*

$$d_1((t_1)_{sp^{2k}}) \equiv \begin{cases} s(s+1)v_2^2 v_3^{s-1-p} G_2 & k=0, \\ s(s+1)v_2^{2p^2-p+1} v_3^{sp^2-2p} G_1 & k=1, \\ -3s(s+1)v_2^{2a_{2k}} v_3^{(sp-2)p^{2k-1}} K_0 & k \geq 2,\ p \geq 5, \\ -2s(s+1)v_2^{2a_{2k}+1} v_3^{3^{2k-1}(3s-2)}(b_{1,0} + t_1^p \otimes Z') & k \geq 2,\ p=3, \end{cases}$$

$$d_1((t_1^p)_{sp^{2k+1}}) \equiv \begin{cases} s(s+1)v_2^{2p} v_3^{sp-2} G_0 & k=0, \\ s(s+1)v_2^{2pa_{2k}+p} v_3^{(sp-2)p^{2k}} b_{1,1} & k \geq 1, \end{cases}$$

(2) $(t_1)_{sp^{2k}}$ *and* $(t_1^p)_{sp^{2k+1}}$ *for $s = tp^2 - 1 \in \mathbb{Z}_1$ such that*

$$d_1((t_1)_{sp^{2k}}) \equiv v_2^{a_{2k+2}-a_{2k+1}} v_3^{(tp-1)p^{2k+1}} b_{1,0}, \quad d_1((t_1^p)_{sp^{2k+1}}) \equiv v_2^{pa_{2k+2}-pa_{2k+1}} v_3^{(tp-1)p^{2k+2}} b_{1,1},$$

for $k \geq 0$,

(3) $(t_1)_{sp^{2k+1}}$ *and* $(t_1^p)_{sp^{2k}}$ *for $s \in \mathbb{Z}^{(p)}$ such that*

$$d_1((t_1^p)_{tp^k+1}) \equiv \begin{cases} t(t-1)v_2^{2p} v_3^{tp-1} G_0 & k=1, \\ -tv_2^{p^2+1} v_3^{(tp-1)p} G_1 & k=2, \\ -2tv_2^{a_k+p} v_3^{(tp-1)p^{k-1}} G_0 & \text{odd } k \geq 3, \\ 2tv_2^{a_k+p} v_3^{(tp-1)p^{k-1}} K_0 & \text{even } k \geq 4, \end{cases}$$

$$d_1((t_1^p)_{tp^k+e(k)}) \equiv \begin{cases} -(t-1)v_2^{a_k+1} v_3^{tp^k+pe(k-2)} G_1 & \text{even } k \geq 2, \\ -(t-1)v_2^{a_k+2} v_3^{tp^k+pe(k-2)} b_{1,1} & \text{odd } k \geq 3, \end{cases}$$

$$d_1((t_1^p)_{sp^{2k}}) \equiv \begin{cases} s(s-1)v_2^2 v_3^{s-2} K_1 & k=0, \\ -sv_2^{p^2+p} v_3^{sp^2-p-1} K_0 & k=1, \\ -3sv_2^{e(3)p^{2k-2}-1} v_3^{(sp^2-p-1)p^{2k-2}} K_0 & p \geq 5, k \geq 2, \\ -sv_2^{3^{2k-2}e(3)} v_3^{(9s-4)3^{2k-2}} (b_{1,0} + Z' \otimes t_1^p) & p=3, k \geq 2, \end{cases}$$

$$d_1((t_1)_{sp^{2k+1}}) \equiv \begin{cases} -sv_2^{p+1} v_3^{(s-2)p} K_2 & k=0, \\ sv_2^{e(3)p^{2k-1}-p+1} v_3^{(sp^2-p-1)p^{2k-1}} b_{1,1} & k \geq 1, \end{cases}$$

(4) $(t_1^{p^2})_{tp-1}$ such that $d_1((t_1^{p^2})_{tp-1}) \equiv v_2^{p-1} v_3^{tp-p} b_{1,2}$.

Here G_i, K_i and $b_{1,i}$ are the cocycles of $\Omega^2 E(3)_*/I_2$ in (2.5), Z' is an element in Lemma 5.1, and $x \equiv v_2^a y$ denotes the congruence modulo J_{a+1}.

Let $d_1((x)_t) \equiv v_2^j y \bmod J_{j+1}$ be a congruence in Lemma 3.4. Then $\delta_1(([x])_{t/j}) = [y]$ for the connecting homomorphism δ_1 in (3.2). Here $([x])_{t/j}$ $(= [(x)_t/v_2^j]) \in H^1 M_2^1$ denotes the cohomology class of the cocycle $(x)_t/v_2^j$ of $\Omega^1 M_2^1$. Thus the cochains in Lemma 3.4 give rise to elements $(h_0)_{sp^r/j_{s,r}}$ and $(h_1)_{sp^r/j'_{s,r}}$ of $H^1 M_2^1$ as well as their δ_1–images. Furthermore, we have elements

$$(\zeta_3)_{tp^n/a_n} = x_n^t \zeta_3 / v_2^{a_n} \in H^1 M_2^1$$

for the elements x_n $(= x_{3,n})$ introduced in [2, (5.11)] (see Lemma 5.1) with

(3.5)
$$\delta_1((\zeta_3)_{tp^n/a_n}) = \begin{cases} (h_2\zeta_3)_{t-1} & n=0, \\ (h_0\zeta_3)_{(tp-1)p^{n-1}} & n \text{ is odd}, \\ (h_1\zeta_3)_{(tp-1)p^{n-1}} & n \text{ is even } \geq 2, \end{cases}$$

by [2, (5.18)] (or Lemma 5.1). As a $k(2)_*$–module, $K(2)_*/k(2)_*\{\xi\} = \mathbb{Z}/p\{(\xi)_{0/j} : j \geq 1\}$ with $v_2(\xi)_{0/j} = (\xi)_{0/j-1}$ and $v_2(\xi)_{0/1} = 0$; see (2.10).

Let B be the $k(2)_*$–module of the theorem. Each direct summand of B is a submodule of $H^1 M_2^1$, which defines a $k(2)_*$–module map $f : B \to H^1 M_2^1$. Furthermore, assigning $(\xi)_{s/1} \in B$ to the generator $(\xi)_s$ of the cokernel of δ_0, we have a homomorphism $\overline{\eta}_* : H^1 M_3^0 \to B$ by Proposition 3.3. These homomorphisms fit in the commutative diagram

$$\begin{array}{ccccccccc} H^0 M_2^1 & \xrightarrow{\delta_0} & H^1 M_3^0 & \xrightarrow{\overline{\eta}_*} & B & \xrightarrow{v_2} & B & \xrightarrow{\delta_1'} & H^2 M_3^0 \\ \| & & \| & & \downarrow f & & \downarrow f & & \| \\ H^0 M_2^1 & \xrightarrow{\delta_0} & H^1 M_3^0 & \xrightarrow{\eta_*} & H^1 M_2^1 & \xrightarrow{v_2} & H^1 M_2^1 & \xrightarrow{\delta_1} & H^2 M_3^0 \end{array}$$

where we define δ_1' by $\delta_1 f$. It suffices to show that the upper sequence is exact by [2, Remark 3.11]. By the definition of B, the subsequence $H^0 M_2^1 \xrightarrow{\delta_0} H^1 M_3^0 \xrightarrow{\overline{\eta}_*} B \xrightarrow{v_2} B$ is exact and the composite $B \xrightarrow{v_2} B \xrightarrow{\delta_1'} H^2 M_3^0$ is zero.

Suppose that the δ_1'–images of the generators are linearly independent, and take $\xi \in \operatorname{Ker}\delta_1'$ to be a homogeneous element. Then

$$\xi = \sum_k c_k\xi_k \quad \text{for generators } \xi_k \text{ of } B \text{ and scalars } c_k \in k(2)_* \qquad \text{and} \qquad 0 = \delta_1'(\xi) = \sum_k \bar{c}_k\delta_1'(\xi_k)$$

for the image $\bar{c}_k$ of c_k under the projection $k(2)_* \to \mathbb{Z}/p$ sending v_2 to zero. Since the $\delta_1'(\xi_k)$ are linearly independent we see $\bar{c}_k = 0$, and so we have $c_k' \in k(2)_*$ such that $c_k = v_2c_k'$. Therefore

$$\xi = \sum_k v_2c_k'\xi_k \in \operatorname{Im} v_2,$$

and we see that the upper sequence of the above diagram is exact if the δ_1'–images of the generators are linearly independent.

The δ_1'–image is a $\mathbb{Z}/p$–submodule of $H^2M_3^0$ in (2.3) with generators of the form $(\rho)_s$ for $\rho \in \{g_i, k_i, b_i, h_i\zeta_3 : i \in \mathbb{Z}/3\}$ by Lemma 3.4 and (3.5). Moreover, Lemma 3.4 and (3.5) show that the δ_1'–image of each generator ξ_k has only one summand of form $(\rho)_s$,

$$(h_0\zeta_3)_{(tp-1)p^{2n}}, \quad (h_1\zeta_3)_{(tp-1)p^{2n-1}}, \quad (h_2\zeta_3)_{t-1}, \quad (g_2)_{s-1-p},$$
$$(k_1)_{s-2}, \quad (k_2)_{(s-2)p}, \quad (b_0)_{(tp-1)p^{2n+1}} \text{ for } p \ge 5, \quad (b_2)_{tp-p},$$

except for

g_0	$(g_0)_{sp-2}$	$(g_0)_{(tp-1)p^{2n}}$			
g_1	$(g_1)_{(sp-2)p}$	$(g_1)_{(tp-1)p}$	$(g_1)_{tp^{2n}+pe(2n-2)}$		
k_0	$(k_0)_{(sp-2)p^{2n-1}}$	$(k_0)_{(tp-1)p^{2n-1}}$	$(k_0)_{(sp^2-p-1)p^{2n}}$		$(p \ge 5)$
k_0	$(k_0)_{3^{2n-1}(3t-1)}$	$(k_0)_{9s-4}$			$(p = 3)$
b_0	$(b_0)_{3^{2n-1}(3s-2)}$	$(b_0)_{3^{2n+1}(3t-1)}$	$(b_0)_{3^{2n-2}(9s-4)}$		$(p = 3)$
b_1	$(b_1)_{(sp-2)p^{2n}}$	$(b_1)_{(tp-1)p^{2n+2}}$	$(b_1)_{tp^{2n+1}+pe(2n-1)}$	$(b_1)_{(sp^2-p-1)p^{2n-1}}$	

These show that the δ_1'–images $\delta_1'(\xi_k)$ for the generators ξ_k of B are different from each other, and so they are linearly independent. □

3.2 Proof of Theorem 1.8

Let $\delta_2^0 \colon H^*N_2^1 \to H^{*+1}N_2^0$ be the connecting homomorphism associated to the short exact sequence (2.1), and consider the diagram

$$\begin{array}{ccccccc} & & H^2M_2^0 & \xrightarrow{(\kappa_2^0)_*} & H^2N_2^1 & \xrightarrow{\delta_2^0} & H^3N_2^0 = E_2^3(V(1)) \\ & & & & \downarrow{\scriptstyle \iota_2^1} & & \\ H^1M_2^1 & \xrightarrow{\delta_1} & H^2M_3^0 & \xrightarrow{\eta_*} & H^2M_2^1 & & \end{array}$$

of exact sequences for δ_1 in (3.2). The connecting homomorphism $\bar{\delta}_j$ associated to (1.6) is factorized into the composite $\bar{\delta}_j \colon H^s\mathrm{BP}_*/J_j \xrightarrow{\hat{\iota}_j} H^sN_2^1 \xrightarrow{\delta_2^0} H^{s+1}N_2^0$ for the homomorphism $\hat{\iota}_j$ given by $\hat{\iota}_j(x) = x/v_2^j$. It follows that

(3.6) $$\bar{\gamma}''_{sp^r/j} = \delta_2^0(v_3^{sp^r}/v_2^j) \in H^1N_2^0 = E_2^1(V(1)) \quad \text{for } v_3^{sp^r}/v_2^j \in H^0N_2^1.$$

Since δ_2^0 is a $k(2)_*$–module map, we have

(3.7) $$v_2^{j-1}\overline{\gamma}''_{sp^r/j} = v_2^{j-1}\delta_2^0(v_3^{sp^r}/v_2^j) = \delta_2^0(v_2^{j-1}v_3^{sp^r}/v_2^j) = \delta_2^0(v_3^{sp^r}/v_2) = \overline{\gamma}''_{sp^r}.$$

It is well known that

$$\bar\beta_1 = -b_0 = [-b_{1,0}] \quad\text{and}\quad \bar\beta_2 = 2k_0 = [2K_0] \in H^2N_3^0$$

for the cocycles $b_{1,0}$ and K_0 in (2.5); see [5, Lemma 4.4]. This defines elements $v_3^{sp^r}\bar\beta_i/v_2 \in H^2N_2^1$ for $i = 1, 2$, and

$$\delta_2^0(v_3^{sp^r}\bar\beta_i/v_2) = \gamma''_{sp^r}\bar\beta_i \in E_2^3(V(1)) \quad \text{(by (3.6))}.$$

We also see that for $v_3^{sp^r}\bar\beta_i \in H^2M_3^0$,

$$\eta_*(v_3^{sp^r}\bar\beta_i) = \iota_2^1(v_3^{sp^r}\bar\beta_i/v_2) \in H^2M_2^1.$$

From Lemma 3.4, the elements $v_3^{sp^r}\bar\beta_1 = -(b_0)_{sp^r}$ and $v_3^{sp^r}\bar\beta_2 = 2(k_0)_{sp^r} \in H^2M_3^0$ may be in the image of δ_1 if

$$p \ge 5 \quad\text{and}\quad (s,r) \in (\mathbb{Z}_1^+ \cup \mathbb{Z}_2^+) \times \overline{2\mathbb{N}}$$

or

$$p = 3 \quad\text{and}\quad (s,r) \in (\overline{\mathbb{N}}_1 \times 2\mathbb{N}_{>0}) \cup ((\mathbb{Z}_1^+ \cup \mathbb{Z}_2^+) \times \overline{2\mathbb{N}}) \cup (\overline{\mathbb{N}}_2 \times \overline{2\mathbb{N}}_{>1}),$$

and if

$$p \ge 5 \quad\text{and}\quad (s,r) \in (\overline{\mathbb{N}}_1 \times 2\mathbb{N}) \cup (\mathbb{Z}_2^+ \times \overline{2\mathbb{N}}_{>1}) \cup (\overline{\mathbb{N}}_2 \times \overline{2\mathbb{N}}_{>1})$$

or

$$p = 3 \quad\text{and}\quad (s,r) \in (\overline{\mathbb{N}}_1 \times \{0\}) \cup (\mathbb{Z}_2^+ \times \overline{2\mathbb{N}}_{>1}),$$

respectively. Here $\overline{\mathbb{N}}_i = \mathbb{N}^{(p)} \setminus \mathbb{N}_i$ for $i = 1, 2$. Therefore, if a pair (s, r) satisfies the condition of the theorem, then the element $v_3^{sp^r}\bar\beta_i$ is not in the image of δ_1, and survives to $\iota_2^1(v_3^{sp^r}\bar\beta_i/v_2)$ under the homomorphism η_*. Thus $v_3^{sp^r}\bar\beta_i/v_2 \neq 0 \in H^2N_2^1$ under the conditions.

Ravenel determined in [8, Theorem 6.3.24; 7, Theorem 3.2] that

(3.8) $$H^2M_2^0 = \begin{cases} K(2)_*\{h_0\tilde\zeta_2, h_1\tilde\zeta_2, b_0, b_1, \xi\} & p = 3, \\ K(2)_*\{h_0\tilde\zeta_2, h_1\tilde\zeta_2, g_0, g_1\} & p \ge 5, \end{cases}$$

where $\tilde\zeta_2 = v_2^{p+1}\zeta_2 = [-z]$ for ζ_2 in [2, Proposition 3.18)] and z in (4.18). This shows that the elements $v_3^{sp^r}\bar\beta_i/v_2$ for $i = 1, 2$ are not in the image of $(\kappa_2^0)_*$, and hence survive to $\gamma''_{sp^r}\bar\beta_i \in E_2^3(V(1))$. Moreover, $\gamma''_{sp^r/j}\bar\beta_i \neq 0 \in E_2^3(V(1))$ if $v_2^{j-1}\gamma''_{sp^r/j}\bar\beta_i = \gamma''_{sp^r}\bar\beta_i$ is not zero, where the equality follows by (3.7). □

4 Some cochains in the cobar complex $\Omega^*E(3)_*$

In the rest of this paper, we consider $E(3)_*(E(3))$–comodules whose structure maps are induced from the right unit map $\eta_R\colon E(3)_* \to E(3)_*(E(3))$. We consider the cobar complex Ω^*M of a comodule M in (2.2), whose differentials are given by

(4.1) $$d_0(v) = \eta_R(v) - v \in \Omega^1E(3)_*, \quad\text{and}\quad d_1(x) = 1 \otimes x - \Delta(x) + x \otimes 1 \in \Omega^2E(3)_*$$

for $v \in \Omega^0 E(3)_* = E(3)_*$ and $x \in \Omega^1 E(3)_* = E(3)_*(E(3))$. For the differentials d_0 and d_1, we have relations (see [11, (2.3.2)])

$$\begin{aligned} d_0(vv') &= vd_0(v') + d_0(v)\eta_R(v'), \\ d_1(vx) &= d_0(v) \otimes x + vd_1(x), \\ d_1(xy) &= -x \otimes y - y \otimes x + d_1(x)\Delta y + (x \otimes 1 + 1 \otimes x)d_1(y), \\ d_1(x\eta_R(v)) &= d_1(x)(1 \otimes \eta_R(v)) - x \otimes d_0(v), \end{aligned} \tag{4.2}$$

for $v, v' \in E(3)_*$ and $x, y \in E(3)_*(E(3))$. A formula for the Hopf conjugation $c\colon \mathrm{BP}_*(\mathrm{BP}) \to \mathrm{BP}_*(\mathrm{BP})$ is given in [6, (3)], and immediately implies the following:

Lemma 4.3 *The Hopf conjugation $c\colon E(3)_*(E(3)) \to E(3)_*(E(3))$ acts as*

$$ct_1 = -t_1, \quad ct_2 = t_1^{p+1} - t_2 \quad \text{and} \quad ct_3 \equiv t_2 t_1^{p^2} - t_1 ct_2^p - t_3 \bmod I_2.$$

For the right unit $\eta_R\colon \mathrm{BP}_* \to \mathrm{BP}_*(\mathrm{BP})$, we have a well-known formula (see [6, (11)])

$$\eta_R(v_n) \equiv v_n + v_{n-1}t_1^{p^{n-1}} - v_{n-1}^p t_1 \mod I_{n-1}. \tag{4.4}$$

A routine calculation using (4.1) and (4.4) shows the following:

Lemma 4.5 *Put $\sigma_n = \sum_{k=0}^{n-1} v_2^{p^{2k}a_{2n-2k-1} - p^{2k+1}} v_3^{p^{2k}} \in E(3)_*$. Then*

$$d_0(\sigma_n) \equiv v_2^{p^{2n-2}} t_1^{p^{2n}} - v_2^{a_{2n-1}} t_1 \mod I_2.$$

In $E(3)_*(E(3))$ we have $\eta_R(v_4) = 0 = \eta_R(v_5)$, which give rise to relations

$$v_3 t_1^{p^3} \equiv t_1\eta_R(v_3)^p - v_2 t_2^{p^2} + v_2^{p^2} t_2 \quad \text{and} \quad v_3 t_2^{p^3} \equiv t_2\eta_R(v_3)^{p^2} - v_2 t_3^{p^2} - v_2 w^p + v_2^{p^3} t_3 \mod I_2 \tag{4.6}$$

(see [6, (12) and (16); 8, Corollary 4.3.21]), where $w \in E(3)_*(E(3))$ $(= w_1(v_3, v_2 t_1^{p^2}, -v_2^p t_1)$ in [8, Corollary 4.3.21]) is an element defined by

$$pw = v_3^p + v_2^p t_1^{p^3} - v_2^{p^2} t_1^p + y^p - \eta_R(v_3)^p \tag{4.7}$$

for $y \in (p, v_1)$ in $\eta_R(v_3) = v_3 + v_2 t_1^{p^2} - v_2^p t_1 + y$; see (4.4).

The diagonal $\Delta\colon E(3)_*(E(3)) \to E(3)_*(E(3)) \otimes_{E(3)_*} E(3)_*(E(3))$ of the Hopf algebroid $E(3)_*(E(3))$ acts on the elements t_i and ct_i by

$$\begin{aligned} \Delta(t_1) &= t_1 \otimes 1 + 1 \otimes t_1, \\ \Delta(t_2) &\equiv t_2 \otimes 1 + t_1 \otimes t_1^p + 1 \otimes t_2 - v_1 b_{1,0} \mod (p, v_1^2), \\ \Delta(t_3) &\equiv t_3 \otimes 1 + t_2 \otimes t_1^{p^2} + t_1 \otimes t_2^p + 1 \otimes t_3 - v_2 b_{1,1} \mod I_2, \\ \Delta(t_4) &\equiv t_4 \otimes 1 + t_3 \otimes t_1^{p^3} + t_2 \otimes t_2^{p^2} + t_1 \otimes t_3^p + 1 \otimes t_4 - v_3 b_{1,2} \mod I_3, \end{aligned} \tag{4.8}$$

(see [6, Theorem 8; 8, Corollary 4.3.15]) and so

(4.9)
$$\begin{aligned} d_1(ct_2) &\equiv -t_1^p \otimes t_1, \\ d_1(ct_3) &\equiv ct_2^p \otimes t_1 + t_1^{p^2} \otimes ct_2 - v_2 b_{1,1} \mod I_2, \\ d_1(ct_4) &\equiv t_1^{p^3} \otimes ct_3 - ct_2^{p^2} \otimes ct_2 + ct_3^p \otimes t_1 - v_3 b_{1,2} \mod I_3, \end{aligned}$$

since $\Delta(cx) = (c \otimes c)T\Delta(x)$ for the switching map T given by $T(x \otimes y) = y \otimes x$, where $b_{1,k}$ is the cocycle in (2.5).

The fact $d_1(t_1^{p^{k+1}}) \equiv -pb_{1,k} \mod (p^2)$ implies not only that the cochain $b_{1,k} \in \Omega^2 E(3)_*/(p)$ is a cocycle, but also the following lemma:

Lemma 4.10 *The cochain w in (4.7) satisfies*

$$w \equiv -v_2 v_3^{p-1} t_1^{p^2} \mod J_2 \quad \text{and} \quad d_1(w) \equiv -v_2^p b_{1,2} + v_2^{p^2} b_{1,0} \mod I_2.$$

Corollary 4.11 *Put $W_n = \sum_{i=0}^{n-1} v_2^{p^{2i} a_{2n-2i} - p^{2i+2}} w^{p^{2i}}$. Then*

$$d_1(W_n) \equiv -v_2^{p^{2n-1}} b_{1,2n} + v_2^{a_{2n}} b_{1,0} \mod I_2.$$

We generalize the relations (4.6) and obtain the following proposition from [8, (4.3.1) and Lemma 4.3.11]; 6, Theorem 1] (see [9, Proposition 2.1]):

Proposition 4.12 *There exist elements T_n for $n \geq 0$ satisfying $T_n \equiv t_n^p \mod I_3$ and*

$$v_2^{p^{k+1}} t_{k+1} + t_k \eta_R(v_3)^{p^k} \equiv v_1 T_{k+2} + v_2 T_{k+1}^p + v_3 T_k^{p^2} \mod (p, v_1^2)$$

for $k \geq 0$. In particular, $T_0 = 1$, $T_1 \equiv t_1^p$, $T_2 \equiv t_2^p$ and $T_3 \equiv t_3^p + w \mod I_2$.

Proof We begin by recalling some notation from [8, Section 4.3]. For a sequence $J = (j_1, j_2, \dots, j_m)$ of positive integers we set $|J| = m$ and $\|J\| = \sum_{i=1}^m j_i$, and an element $v_J \in E(3)_*$ is defined recursively by $v_{(j,J)} = v_j v_J^{p^j}$. Let $w_k(S)$ for a set S be symmetric polynomials of degree p^n such that $w_0(S) = \sum_{x \in S} x$ and $\sum_{x \in S} x^{p^n} = \sum_{k=0}^n p^k w_k(S)^{p^{n-k}}$. We then define sets S_n out of a set $S = \{a_{i,j}\}$ recursively by

$$S_n = \{a_{i,j} : i + j = n\} \cup \bigcup_{|J|>0} \{v_J w_{|J|}(S_{n-\|J\|})^{p^{\|J\|-|J|}}\}.$$

By [8, (4.3.1) and Lemma 4.3.11],

(4.13)
$$w_0(C_n) \equiv \sum_{i+j=n}^F t_i \eta_R(v_j)^{p^i} \equiv \sum_{i+j=n}^F v_i t_j^{p^i} \equiv w_0(D_n) \mod (p)$$

for the sets

$$C = \{t_i \eta_R(v_j)^{p^i}\} \quad \text{and} \quad D = \{v_i t_j^{p^i}\}.$$

In $E(3)_*(E(3))$, put

$$w(S_n) = \sum_J v_J^p w_{|J|+1}(S_{n-\|J\|})^{p^{\|J\|-|J|}} \quad \text{and} \quad T_n = t_n^p - w(C_n) + w(D_n).$$

Then the proposition follows from (4.13) and the congruences

$$w_0(C_n) \equiv v_2^{p^{n-2}} t_{n-2} + t_{n-3}\eta_R(v_3)^{p^{n-3}} + v_1 w(C_{n-1}) + v_2 w(C_{n-2})^p + v_3 w(C_{n-3})^{p^2},$$
$$w_0(D_n) \equiv v_1 t_{n-1}^p + v_2 t_{n-2}^{p^2} + v_3 t_{n-3}^{p^3} + v_1 w(D_{n-1}) + v_2 w(D_{n-2})^p + v_3 w(D_{n-3})^{p^2},$$

seen by the relation

$$v_{(k,J)} w_{|(k,J)|}(S_{n-\|(k,J)\|})^{p^{\|(k,J)\|-|(k,J)|}} = v_k v_J^{p^k} w_{|J|+1}(S_{n-k-\|J\|})^{p^{\|J\|-|J|+k-1}}. \qquad \square$$

Lemma 4.14 *For $n \geq 0$,*

$$\eta_R(v_2^{p-1} v_3^{e(n)}) \equiv \sum_{i=0}^{n} (-1)^{n-i} v_2^{p^{i+1}e(n-i)+p-1} v_3^{e(i)} t_{n-i}^{p^i} - v_2^p w_n^p + v_1 v_2^{p-2} w_{n+1} \mod (p, v_1^2).$$

Here

(4.15) $$w_n = \sum_{i=1}^{n} (-1)^i v_2^{e(i-1)} T_i \eta_R(v_3^{p^{i-1}e(n-i)}).$$

Proof In this proof, every congruence is considered modulo (p, v_1^2). By Proposition 4.12, $t_k \eta_R(v_3^{p^k}) \equiv \widetilde{T}_k - v_2^{p^{k+1}} t_{k+1}$ for $\widetilde{T}_k = v_1 T_{k+2} + v_2 T_{k+1}^p + v_3 T_k^{p^2}$, which implies inductively

$$t_1 \eta_R(v_3^{pe(n)}) \equiv -\sum_{i=1}^{n} (-1)^i v_2^{p^2 e(i-1)} \widetilde{T}_i \eta_R(v_3^{p^{i+1}e(n-i)}) + (-1)^n v_2^{p^2 e(n)} t_{n+1},$$

and hence

(4.16) $$t_1 \eta_R(v_3^{pe(n)}) \equiv -v_1 v_2^{-p-1} w_{n+2} + v_2^{1-p} w_{n+1}^p - v_3 w_n^{p^2} + (-1)^n v_2^{p^2 e(n)} t_{n+1} - v_1 v_2^{-p-1}(t_1^p \eta_R(v_3) - v_2 t_2^p)\eta_R(v_3^{pe(n)}) + v_2^{1-p} t_1^{p^2} \eta_R(v_3^{pe(n)}).$$

Now we prove the lemma by induction. For $n = 0$, it follows from the facts $\eta_R(v_2) \equiv v_2 + v_1 t_1^p$ by (4.4) and $w_1 = -t_1^p$.

Assuming the case for n, we obtain the case for $n+1$ from (4.16) and

$$\eta_R(v_2^{p-1} v_3^{e(n+1)}) \equiv v_2^{-p^2+2p-1} v_3 \eta_R(v_2^{p-1} v_3^{e(n)})^p + v_2^{p-1}(v_2 t_1^{p^2} + v_1 t_2^p)\eta_R(v_3^{pe(n)}) - v_2^{2p-1} t_1 \eta_R(v_3^{pe(n)}) - v_1 v_2^{p-2} t_1^p \eta_R(v_3^{e(n+1)}),$$

given by $\eta_R(v_2^{p-1} v_3) \equiv v_2^{p-1}(v_3 + v_2 t_1^{p^2} - v_2^p t_1 + v_1 t_2^p) - v_1 v_2^{p-2} t_1^p \eta_R(v_3)$. Here $\eta_R(v_3)$ is given in [2, (5.7)]. $\square$

Evaluate the congruence in Lemma 4.14 under d_1, and compare the v_1–multiples. Then we deduce the following corollary; see [9, Proposition 2.3]. Indeed, if $v_1 v_2^{p-2} d_1(w_{n+1}) \equiv A + v_1 B \mod (p, v_1^2)$ for some A and B involving no v_1, then $A \equiv 0 \mod (p, v_1^2)$ and $v_2^{p-2} d_1(w_{n+1}) \equiv B \mod I_2$.

Corollary 4.17 *For the elements w_n in (4.15),*

$$d_1(w_{n+1}) \equiv -\sum_{i=0}^{n-1}(-1)^{n-i} v_2^{p^{i+1}e(n-i)} w_{i+1} \otimes t_{n-i}^{p^i} - (-1)^n v_2^{e(n+1)} \mathfrak{b}_n \mod I_2.$$

Here $\mathfrak{b}_n$ is an element in $d_1(t_n) \equiv \mathfrak{a}_n + v_1 \mathfrak{b}_n \bmod (p, v_1^2)$ for $\mathfrak{a}_n$ and $\mathfrak{b}_n$ involving no v_1. In particular, $\mathfrak{b}_2 = b_{1,0}$ by (4.8).

We have the cocycle z in $\Omega^1 E(3)_*/I_2$ given by

(4.18) $$z = v_3 t_1^p + v_2 c t_2^p - v_2^p t_2 = t_1^p \eta_R(v_3) - v_2 t_2^p + v_2^p c t_2 = -w_2 + v_2^p c t_2,$$

which represents the element $-v_2^{p+1}\zeta_2 \in H^1 M_2^0$; see [2, Proposition 3.18(c)] and (3.8). In particular,

(4.19) $$t_1^p \eta_R(v_3) \equiv z + v_2 t_2^p - v_2^p c t_2 \mod I_2.$$

We further have cocycles G_i' and $K_i' \in \Omega^2 E(3)_*/I_2$ for $i \in \{0,1,2\}$ defined by

(4.20) $$G_i' = c t_2^{p^i} \otimes t_1^{p^i} + \tfrac{1}{2} t_1^{p^{i+1}} \otimes t_1^{2p^i} \quad\text{and}\quad K_i' = t_1^{p^{i+1}} \otimes c t_2^{p^i} + \tfrac{1}{2} t_1^{2p^{i+1}} \otimes t_1^{p^i},$$

which are homologous to G_i and K_i in (2.5), respectively. Indeed,

(4.21) $$d_1(\mathfrak{g}_i) \equiv G_i' - G_i \quad\text{and}\quad d_1(\mathfrak{k}_i) \equiv K_i' - K_i \mod I_2,$$

for $i \in \{0,1,2\}$, and for $\mathfrak{g}_i, \mathfrak{k}_i \in \Omega^1 E(3)_*$ given by

(4.22) $$\mathfrak{g}_i = t_1^{p^i} t_2^{p^i} - \tfrac{1}{2} t_1^{p^{i+1}+2p^i} \quad\text{and}\quad \mathfrak{k}_i = t_1^{p^{i+1}} t_2^{p^i} - \tfrac{1}{2} t_1^{2p^{i+1}+p^i}.$$

We also have a similar relation

(4.23) $$d_1(t_1^p t_2) \equiv -(t_1^p \otimes t_2 + c t_2 \otimes t_1^p) - 2K_0 \mod I_2.$$

Lemma 4.24 *In $\Omega^1 E(3)_*$, put*

$$\omega_1 = \eta_R(v_3) t_2 - v_2 t_3 + v_2^p t_1 t_2, \quad \omega_2 = \tfrac{1}{2}\eta_R(v_3) t_1^{2p} - v_2^p \mathfrak{k}_0 \quad\text{and}\quad \tilde{\omega}_2 = -w_3 - v_2^{pe(2)} t_1^p t_2.$$

Then, modulo I_2,

$$d_1(\omega_1) \equiv -t_1 \otimes z - v_2^2 b_{1,1} - 2v_2^p G_0, \quad d_1(\omega_2) \equiv -t_1^p \otimes z - v_2 G_1 + v_2^p K_0,$$
$$d_1(\tilde{\omega}_2) \equiv v_2^{p^2} z \otimes t_1^p + 2v_2^{p^2+p} K_0 + v_2^{e(3)} b_{1,0}.$$

Proof In this proof we consider congruences modulo I_2. A routine calculation shows the congruence for $d_1(\omega_1)$:

$$d_1(\eta_R(v_3)t_2) \equiv -t_1 \otimes (z + \underbrace{v_2 t_2^p}_{a} + \underset{\sim\sim}{v_2^p t_2} - \underbrace{v_2^p t_1^{p+1}}_{c}) - t_2 \otimes (\underbrace{v_2 t_1^{p^2}}_{b} - \underbrace{v_2^p t_1}_{d}), \qquad \text{(by (4.2) and (4.19))}$$

$$d_1(-v_2 t_3) \equiv v_2(\underbrace{t_1 \otimes t_2^p}_{a} + \underbrace{t_2 \otimes t_1^{p^2}}_{b} - v_2 b_{1,1}), \qquad \text{(by (4.8))}$$

$$d_1(v_2^p t_1 t_2) \equiv -v_2^p(\underset{\sim\sim}{t_1 \otimes t_2} + \underbrace{t_2 \otimes t_1}_{d} + \underset{\sim\sim}{t_1^2 \otimes t_1^p} + \underbrace{t_1 \otimes t_1^{p+1}}_{c}). \qquad \text{(by (4.8) and (4.2))}$$

Here the underlined terms with the same label cancel each other and the wavy underlined terms make $-2v_2^p G_0$.

For $d_1(\omega_2)$, we calculate

$$d_1\big(\tfrac{1}{2}\eta_R(v_3)t_1^{2p}\big) \equiv -t_1^p \otimes (z + \underbrace{v_2 t_2^p}_{G} - \underbrace{v_2^p c t_2}_{K'}) - \underbrace{\tfrac{1}{2} v_2 t_1^{2p} \otimes t_1^{p^2}}_{G} + \underbrace{\tfrac{1}{2} v_2^p t_1^{2p} \otimes t_1}_{K'}. \qquad \text{(by (4.2) and (4.19))}$$

Add $d_1(-v_2^p \text{ŧ}_0)$, and we obtain the desired congruence by (4.21).

We verify $d_1(\widetilde{\omega}_2)$ by

$$\begin{aligned} d_1(w_3) &\equiv -v_2^{pe(2)} w_1 \otimes t_2 + v_2^{p^2} w_2 \otimes t_1^p - v_2^{e(3)} b_{1,0} && \text{(by Corollary 4.17)} \\ &\equiv -\underbrace{v_2^{pe(2)}(-t_1^p) \otimes t_2}_{a} + v_2^{p^2}(-z + \underbrace{v_2^p c t_2}_{b}) \otimes t_1^p - v_2^{e(3)} b_{1,0}, && \text{(by (4.18) and (4.15))} \\ d_1(v_2^{pe(2)} t_1^p t_2) &\equiv -v_2^{pe(2)}((\underbrace{t_1^p \otimes t_2}_{a} + \underbrace{c t_2 \otimes t_1^p}_{b}) + 2K_0). && \text{(by (4.23))} \end{aligned}$$

□

5 The elements x_i and deriving elements y_i and y_i'

In [2, (5.11)], Miller, Ravenel and Wilson introduced elements $x_{3,i} \in v_3^{-1}\mathrm{BP}_*$. We refine them, and define the elements $x_i \in E(3)_*$ by

$$\begin{aligned} x_i &= v_3^{p^i} \quad \text{for } i = 0, 1, 2, \\ x_3 &= x_2^p - v_2^{p^3-1} v_3^{(p-1)p^2+1}, \\ x_4 &= x_3^p - v_2^{e(2)p^3-p-1} v_3^{(p^2-e(2))p^2+p+1}, \\ x_{2k+1} &= x_{2k}^p - v_2^{pa_{2k}-1} x_{2k-1}^{(p-1)p} v_3 - v_2^{e(3)p^{2k-1}-e(3)} v_3^{(p^2-e(2))p^{2k-1}+p+1}, \\ x_{2k+2} &= x_{2k+1}^p - 2v_2^{e(3)p^{2k}-e(3)} v_3^{(p^2-e(2))p^{2k}+p+1}, \end{aligned}$$

for $k \geq 2$.

Lemma 5.1 (see [9, Proposition 3.1]) *In $\Omega^1 E(3)_*$, we have*

$$\begin{aligned} d_0(x_0) &\equiv v_2 t_1^{p^2} - v_2^p t_1 \mod I_2, \\ d_0(x_1) &\equiv v_2^p v_3^{p-1} t_1 - v_2^{p+1} v_3^{-1} t_2^{p^2} \mod J_{2p}, \\ d_0(x_i) &\equiv v_2^{a_i}(x_{i-1}^{p-1} t_1^{p^{\varepsilon_i}} + B_i) \mod J_{e(3)p^{i-2}} \quad \text{for } i \geq 2. \end{aligned}$$

Here $\varepsilon_i = \frac{1}{2}(1 + (-1)^i)$, and the B_i are given by

i	2	3	$2k$	$2k+1$
B_i	$-v_2^p v_3^{c(2)} t_2$	$v_2^{p^2-p} v_3^{c(3)}(z - v_2^p t_1^{p+1})$	$v_2^{a_{2k-1}-p} v_3^{c(2k)}(z - v_2^p t_2)$	$v_2^{a_{2k}-p} v_3^{c(2k+1)}(2z - v_2^p c t_2)$

for $c(k) = (p^2 - p - 1)p^{k-2}$. For $i \geq 4$, add $v_2^{a_{i-1}+1} v_3^{c(i)} Z'$ to B_i if we consider the congruence modulo $J_{e(3)p^{i-2}+1}$. Here Z' is a cocycle homologous to aZ for some $a \in \mathbb{Z}/p$.

Proof This follows from a routine calculation: For $i \leq 2$, it follows from (4.4) and from (4.6).

We obtain $d_0(x_3)$ from (4.19) and

$$d_0(v_3^{(p-1)p^2+1}) \equiv v_3^{(p-1)p^2}(v_2t_1^{p^2}-v_2^pt_1)-v_2^{a_2}v_3^{(p-1)p^2-p}(t_1^p\eta_R(v_3)-v_2^pt_2) \bmod J_{e(3)}$$

by (4.2), (4.4) and the congruence on $d_0(x_2)$. We note that $\eta_R(v_3^{p+1})=v_3^{p+1}+v_2z^p-v_2^{p^2}z$ by [2, (3.20)], and obtain

$$\begin{aligned}&d_0(v_3^{(p^2-e(2))p^2+p+1})\\ &\equiv v_3^{(p^2-e(2))p^2}(v_2z^p-v_2^{p^2}z)-v_2^{a_2}v_3^{(p^2-e(2))p^2-p}t_1^p(v_3^{p+1}+v_2z^p)+v_2^{p^2+p}v_3^{(p^2-e(2))p^2}t_2 \bmod J_{e(3)}.\end{aligned}$$

The congruence on $d_0(x_4)$ follows from this and the congruence on $d_0(x_3)$, together with the definition of the element x_3.

Inductively suppose that

$$d_0(x_{2k}) \equiv v_2^{a_{2k}}x_{2k-1}^{p-1}t_1^p+v_2^{e(3)p^{2k-2}-e(2)}v_3^{(p^2-e(2))p^{2k-2}}(z-v_2^pt_2) \quad \bmod J_{e(3)p^{2k-2}}.$$

Then we calculate

$$\begin{aligned}&d_0(x_{2k}^p) \equiv \underbrace{v_2^{pa_{2k}}x_{2k-1}^{(p-1)p}t_1^{p^2}}_{a}+v_2^{e(3)p^{2k-1}-e(2)p}v_3^{(p^2-e(2))p^{2k-1}}(\underbrace{z^p}_{b}-\underbrace{v_2^{p^2}t_2^p}_{c})\\ &d_0(-v_2^{pa_{2k}-1}x_{2k-1}^{(p-1)p}v_3)\\ &\qquad\equiv -v_2^{pa_{2k}-1}x_{2k-1}^{(p-1)p}(\underbrace{v_2t_1^{p^2}}_{a}-v_2^pt_1)+v_2^{e(3)p^{2k-1}-p-1}x_{2k-1}^{p^2-p-1}(z+\underbrace{v_2t_2^p}_{c}-v_2^pct_2),\end{aligned}$$

where the second congruence follows by (4.2) and (4.19), and

$$d_0(-v_2^{e(3)p^{2k-1}-e(3)}v_3^{(p^2-e(2))p^{2k-1}+p+1}) \equiv -v_2^{e(3)p^{2k-1}-e(3)}v_3^{(p^2-e(2))p^{2k-1}}(\underbrace{v_2z^p}_{b}-v_2^{p^2}z).$$

Therefore

$$d_0(x_{2k+1}) \equiv v_2^{pa_{2k}+p-1}x_{2k-1}^{(p-1)p}t_1+v_2^{e(3)p^{2k-1}-e(2)}v_3^{(p^2-e(2))p^{2k-1}}(2z-v_2^pct_2).$$

Also

$$\begin{aligned}d_0(x_{2k+1}^p)&\equiv v_2^{pa_{2k+1}}x_{2k-1}^{(p-1)p^2}t_1^p+v_2^{e(3)p^{2k}-e(2)p}v_3^{(p^2-e(2))p^{2k}}(2z^p-v_2^{p^2}ct_2^p)\\ &\equiv v_2^{pa_{2k+1}}(x_{2k+1}^{p-1}-v_2^{pa_{2k}-1}x_{2k-1}^{(p^2-p-1)p}v_3)t_1^p+v_2^{e(3)p^{2k}-e(2)p}v_3^{(p^2-e(2))p^{2k}}(2z^p-v_2^{p^2}ct_2^p)\\ &\equiv v_2^{pa_{2k+1}}x_{2k+1}^{p-1}t_1^p+v_2^{e(3)p^{2k}-e(2)p}v_3^{(p^2-e(2))p^{2k}}(2z^p-v_2^{p^2-1}z-v_2^{p^2+p-1}t_2)\end{aligned}$$

and

$$d_0(-2v_2^{e(3)p^{2k}-e(3)}v_3^{(p^2-e(2))p^{2k}+p+1}) \equiv -2v_2^{e(3)p^{2k}-e(3)}v_3^{(p^2-e(2))p^{2k}}(v_2z^p-v_2^{p^2}z).$$

Therefore

$$d_0(x_{2k+2}) \equiv v_2^{pa_{2k+1}}x_{2k+1}^{p-1}t_1^p+v_2^{e(3)p^{2k}-e(2)}v_3^{(p^2-e(2))p^{2k}}(z-v_2^pt_2).$$

These complete the induction.

Put $d_0(x_i) \equiv v_2^{a_i}(x_{i-1}^{p-1}t_1^{p^{\varepsilon_i}}+B_i+v_2^{a_{i-1}+1}C) \bmod J_{e(3)p^{i-1}+1}$ for a cochain C. It is easy to see

$$d_1(v_2^{a_i}(x_{i-1}^{p-1}t_1^{p^{\varepsilon_i}}+B_i)) \equiv 0 \quad \bmod J_{e(3)p^{i-1}+1}.$$

It follows that C is a cocycle of $\Omega^1 M_3^0$, and so C represents a cohomology class $av_3^{c(i)}\zeta_3 \in H^1 M_3^0$ for some $a \in \mathbb{Z}/p$ by (2.3). $\square$

Put
$$d_0(x_i) \equiv v_2^{a_i} A_i + v_2^{a_i} B_i \quad \text{for } A_i = x_{i-1}^{p-1} t_1^{p^{\varepsilon_i}},$$
where $\varepsilon_i = \frac{1}{2}(1+(-1)^i)$. We introduce elements $y_{s,i}$ and $y'_{s,i} \in \Omega^1 E(3)_*$ by
$$y_{s,i} = x_i^s t_1^{p^{\varepsilon_i+1}} - s x_i^{s-p+1} B_{i+1} \quad \text{and} \quad y'_{s,i} = x_i^s t_1^{p^{\varepsilon_i}} + \tfrac{1}{2} s v_2^{a_i} x_i^{s-1} A_i t_1^{p^{\varepsilon_i}}.$$

Lemma 5.2 *For the elements $y_{s,i}$ and $y'_{s,i}$,*
$$d_1(y_{s,0}) \equiv s(s+1) v_2^2 v_3^{s-p-1} G_2, \quad d_1(y_{s,1}) \equiv s(s+1) v_2^{2p} v_3^{sp-2} G_0,$$
$$d_1(y_{s,2}) \equiv -s(s+1) v_2^{2p^2-p} v_3^{sp^2-2p}(t_1^p \otimes z - v_2^p x),$$
$$d_1(y_{s,i}) \equiv \begin{cases} -s(s+1) v_2^{2a_{2k+1}-p} x_{2k}^{sp-2}(t_1 \otimes z - v_2^p G_0) & i = 2k+1, \\ -s(s+1) v_2^{2a_{2k+2}-p} x_{2k+1}^{sp-2}(2t_1^p \otimes z - v_2^p K'_0) & i = 2k+2, \end{cases}$$
$$d_1(y'_{s,1}) \equiv -s v_2^{p+1} v_3^{sp-2p} K_2, \quad d_1(y'_{s,2}) \equiv -s v_2^{p^2+p} v_3^{sp^2-p-1} K_0,$$
$$d_1(y'_{s,3}) \equiv s v_2^{a_3+p^2-p} v_3^{sp^3-p^2-p}(z \otimes t_1 - v_2^p x'),$$
$$d_1(y'_{s,i}) \equiv \begin{cases} s v_2^{e(3)p^{i-2}-p-1} v_3^{(sp^2-p-1)p^{2k-2}}(z \otimes t_1^p - v_2^p K_0) & i = 2k, \\ s v_2^{e(3)p^{i-2}-p-1} v_3^{(sp^2-p-1)p^{2k-1}}(2z \otimes t_1 - v_2^p G'_0) & i = 2k+1. \end{cases}$$
Here $x = (t_2 + t_1^{p+1}) \otimes t_1^p + t_1^p \otimes t_1^{p+1} + \frac{1}{2} t_1^{2p} \otimes t_1$ and $x' = t_1^{p+1} \otimes t_1 + \frac{1}{2} t_1^p \otimes t_1^2$, and these congruences are considered modulo J_{a+1}, where a is the largest power of v_2 in each congruence. Furthermore, replace K'_0 and K_0 in the congruences on $d_1(y_{s,2k+2})$ and $d_1(y'_{s,2k})$ by $K'_0 + v_2 t_1^p \otimes Z'$ and $K_0 + v_2 Z' \otimes t_1^p$, respectively, if we consider the congruences modulo J_{a+2}.

Proof We note that
$$d_1(B_{i+1}) \equiv -d_1(A_{i+1}) \equiv -d_0(x_i^{p-1}) \otimes t_1^{p^{\varepsilon_i+1}} \mod I_2,$$
$$d_0(x_i^s) + s x_i^{s+1-p} d_0(x_i^{p-1}) \equiv \binom{s+1}{2} x_i^{s-2} d_0(x_i)^2 \mod J_{3a_i}.$$
Indeed, $d_0(x_i^s) \equiv s x_i^{s-1} d_0(x_i) + \binom{s}{2} x_i^{s-2} d_0(x_i)^2 \mod J_{3a_i}$. Also,
$$d_1(A_i t_1^{p^{\varepsilon_i}}) \equiv d_0(x_{i-1}^{p-1}) \otimes t_1^{2p^{\varepsilon_i}} - 2x_{i-1}^{p-1} t_1^{p^{\varepsilon_i}} \otimes t_1^{p^{\varepsilon_i}} \equiv d_0(x_{i-1}^{p-1}) \otimes t_1^{2p^{\varepsilon_i}} - 2A_i \otimes t_1^{p^{\varepsilon_i}} \mod J_{a_{i-1}+2}.$$
Then we calculate
$$\begin{aligned} d_1(y_{s,i}) &\equiv d_0(x_i^s) \otimes t_1^{p^{\varepsilon_i+1}} - s d_0(x_i^{s+1-p}) \otimes B_{i+1} + s x_i^{s+1-p} d_0(x_i^{p-1}) \otimes t_1^{p^{\varepsilon_i+1}} && \text{(by (4.2))} \\ &\equiv \binom{s+1}{2} x_i^{s-2} d_0(x_i)^2 \otimes t_1^{p^{\varepsilon_i+1}} - s(s+1) x_i^{s-p} d_0(x_i) \otimes B_{i+1} \mod J_{2a_i+p}, \\ d_1(y'_{s,i}) &\equiv s x_i^{s-1} d_0(x_i) \otimes t_1^{p^{\varepsilon_i}} + \tfrac{1}{2} s v_2^{a_i} x_i^{s-1} d_0(x_{i-1}^{p-1}) \otimes t_1^{2p^{\varepsilon_i}} - s v_2^{a_i} x_i^{s-1} A_i \otimes t_1^{p^{\varepsilon_i}} && \text{(by (4.2))} \\ &\equiv s v_2^{a_i} x_i^{s-1} (B_i \otimes t_1^{p^{\varepsilon_i}} + \tfrac{1}{2} d_0(x_{i-1}^{p-1}) \otimes t_1^{2p^{\varepsilon_i}}) \mod J_{e(3)p^{i-2}+1}. \end{aligned}$$
Now we obtain the lemma from Lemma 5.1. $\square$

6 Proof of Lemma 3.4

In this section, we define the cochains $(t_1^{p^i})_s$ and verify their d_1–differential.

6.1 The cochains $(t_1)_{sp^{2k}}$ and $(t_1^p)_{sp^{2k+1}}$ for $s \in \mathbb{Z}_0$

We define the cochains by

$$(t_1)_s = y_{s,0},\ (t_1^p)_{sp} = y_{s,1}, \quad (t_1)_{sp^2} = y_{s,2} - s(s+1)v_2^{2p^2-p}v_3^{sp^2-2p}\omega_2,$$
$$(t_1^p)_{sp^{2k+1}} = y_{s,2k+1} - s(s+1)v_2^{2a_{2k+1}-p}x_{2k}^{sp-2}\omega_1,$$
$$(t_1)_{sp^{2k+2}} = y_{s,2k+2} - s(s+1)v_2^{2a_{2k+2}-p^2-p}x_{2k+1}^{sp-2}(2\widetilde{\omega}_2 + v_2^{p^2}(2zt_1^p + v_2^p\mathfrak{k}_0)),$$

for $k \ge 1$. Then the lemma for this case follows immediately from Lemmas 5.2, 5.1 and 4.24 together with (4.21). Note also $2a_{2k+1} - p + 2 = 2pa_{2k} + p$. For example, for the case $p = 3$ and $k \ge 2$, we compute modulo $J_{2a_{2k}+2}$

$$\begin{aligned} d_1((t_1)_{3^{2k}s}) &\equiv d_1(y_{s,2k}) - s(s+1)v_2^{2a_{2k}-12}x_{2k-1}^{3s-2}d_1(2\widetilde{\omega}_2 + v_2^9(2zt_1^3 + v_2^3\mathfrak{k}_0)) \\ &\equiv -s(s+1)v_2^{2a_{2k}-3}x_{2k-1}^{3s-2}(\underbrace{2t_1^3\otimes z}_{a} - v_2^3(\underbrace{K_0'}_{b} + v_2t_1^3\otimes Z')) - s(s+1)v_2^{2a_{2k}-12}x_{2k-1}^{3s-2} \\ &\quad \cdot\big(2(\underbrace{v_2^9z\otimes t_1^3}_{c} + \underbrace{2v_2^{12}K_0}_{d} + v_2^{13}b_{1,0}) + v_2^9(-2(\underbrace{z\otimes t_1^3}_{c} + \underbrace{t_1^3\otimes z}_{a}) + v_2^3(\underbrace{K_0'}_{b} - \underbrace{K_0}_{d}))\big), \end{aligned}$$

where the second equivalence follows by Lemmas 5.2 and 4.24, and Equation (4.21)

6.2 The cochains $(t_1)_{sp^{2k}}$ and $(t_1^p)_{sp^{2k+1}}$ for $s \in \mathbb{Z}_1$

We put $s = tp^2 - 1$, and define the cochains $(t_1)_{(tp^2-1)p^{2k}}$ and $(t_1^p)_{(tp^2-1)p^{2k+1}}$ by

$$\begin{aligned} v_2^{a_{2k+1}}(t_1)_{(tp^2-1)p^{2k}} &= -v_3^{(t-1)p^{2k+2}}w^{p^{2k+1}} - d_0(v_2^{p^{2k+1}-p^{2k-2}}v_3^{(tp^2-1)p^{2k}}\sigma_k) \\ &\qquad + v_2^{p^{2k+2}-p^{2k-1}}v_3^{(tp-1)p^{2k+1}}W_k, \\ (t_1^p)_{(tp^2-1)p^{2k+1}} &= (t_1)^p_{(tp^2-1)p^{2k}}, \end{aligned}$$

for the elements σ_k in Lemma 4.5, w in (4.7) and W_k in Corollary 4.11. Then this case follows from Lemmas 4.5 and 4.10, Corollary 4.11 and (2.8). We also use relations $w^{p^{2k+1}} \equiv -v_2^{p^{2k+1}}v_3^{p^{2k+2}-p^{2k}}t_1^{p^{2k}}$ mod $J_{a_{2k+1}+1}$ by Lemma 4.10 and (4.6), and $b_{1,2}^{p^{2k+1}} \equiv b_{1,2k+3} \equiv v_3^{(p-1)p^{2k+1}}b_{1,2k}$ mod I_3 by (4.6). For example,

$$\begin{aligned} v_2^{a_{2k+1}}(t_1)_{(tp^2-1)p^{2k}} &\equiv v_3^{(t-1)p^{2k+2}}(v_2^{p^{2k+1}}v_3^{p^{2k+2}-p^{2k}}t_1^{p^{2k}}) && \text{(by Lemma 4.5)} \\ &\quad - v_2^{p^{2k+1}-p^{2k-2}}v_3^{(tp^2-1)p^{2k}}(v_2^{p^{2k-2}}t_1^{p^{2k}} - v_2^{a_{2k-1}}t_1) \\ &\equiv v_2^{a_{2k+1}}v_3^{(tp^2-1)p^{2k}}t_1 \quad \text{mod } J_{a_{2k+1}+1}, \end{aligned}$$

since $p^{2k+1} - p^{2k-2} + a_{2k-1} = a_{2k+1}$ in (2.8), and

$$v_2^{a_{2k+1}} d_1((t_1)_{(tp^2-1)p^{2k}}) \\ \equiv \underbrace{v_2^{p^{2k+2}} v_3^{(t-1)p^{2k+2}} b_{1,2}^{p^{2k+1}}}_{a} + v_2^{p^{2k+2}-p^{2k-1}} v_3^{(tp-1)p^{2k+1}} (-\underbrace{v_2^{p^{2k-1}} b_{1,2k}}_{a} + v_2^{a_{2k}} b_{1,0}) \mod J_{a_{2k+2}+1}$$

by Lemma 4.10 and Corollary 4.11. Since $p^{2k+2} - p^{2k-1} + a_{2k} = a_{2k+2}$ in (2.8), we obtain the case for $(t_1)_{sp^{2k}}$.

6.3 The cochains $(t_1)_{sp^{2k+1}}$ and $(t_1^p)_{sp^{2k}}$ for $s \in \mathbb{Z}^{(p)}$

We begin by defining

$$(t_1^p)_s = v_3^s t_1^p + s v_2 v_3^{s-1} c t_2^p - s(s-1) v_2^2 v_3^{s-2} \mathfrak{t}_1.$$

Then we calculate by (4.2), (4.4), (4.8) and (4.22), and obtain

$$d_1((t_1^p)_s) \equiv s(s-1) v_2^2 v_3^{s-2} K_1 \mod J_3.$$

Now we consider the cases for $p \mid s(s-1)$.

6.3.1 The cochains $(t_1^p)_{tp^k+1}$ for $k \geq 1$ We define the cochains by

$$\begin{aligned}
(t_1^p)_{tp+1} &= v_3^{tp} z + t v_2^p v_3^{tp} t_2 - t v_2^{p+1} v_3^{tp-p} c t_3^p, \\
(t_1^p)_{tp^2+1} &= x_2^t z + t v_2^{a_2} v_3^{(tp-1)p} \omega_2, \\
(t_1^p)_{tp^{2k+1}+1} &= x_{2k+1}^t z + t v_2^{a_{2k+1}} v_3^{(tp-1)p^{2k}} \omega_1 + t v_2^{a_{2k}+p+1} (t_1^p)_{(tp^2-1)p^{2k-1}}, \\
(t_1^p)_{tp^{2k+2}+1} &= x_{2k+2}^t z + t v_2^{a_{2k+2}-p^2} v_3^{(tp-1)p^{2k+1}} (\widetilde{\omega}_2 + v_2^{p^2} z t_1^p),
\end{aligned}$$

in $\Omega^1 E(3)_*$ for $k \geq 1$, $t \in \mathbb{Z}^{(p)}$, x_n in 5.1, z in (4.18) and ω_i in Lemma 4.24. We verify this case by a routine calculation using (4.2), (4.4), (4.18), (4.8) and (4.9). We see that

$$t_1^{p^3} \otimes z \equiv \eta_R(v_3) t_1^{p^3} \otimes t_1^p + v_2 t_1^{p^3} \otimes c t_2^p - v_2^p v_3^{p-1} t_1 \otimes t_2 \quad \text{and} \quad \eta_R(v_3) t_1^{p^3} \equiv v_3^p t_1 + v_2 c t_2^{p^2} \mod J_{p+1}$$

by (4.18), (4.4) and (4.6). It follows that $t_1^{p^3} \otimes z \equiv -d_1(v_3^p t_2) + v_2 d_1(c t_3^p) \mod J_{p+1}$, and then $d_1(v_3^{tp} z) \equiv t v_2^p v_3^{tp-p}(-d_1(v_3^p t_2) + v_2 d_1(c t_3^p)) + \binom{t}{2} v_2^{2p} v_3^{tp-1} t_1^2 \otimes t_1^p \mod J_{2p+1}$. Thus we obtain $d_1((t_1^p)_{tp+1})$.

The congruences on $d_1((t_1^p)_{tp^k+1})$ for $k \geq 2$ follow directly from Lemmas 5.1 and 4.24 and the results on $d_1((t_1^p)_{(tp^2-1)p^{2k-1}})$ shown in the previous subsection. For example,

$$\begin{aligned}
d_1((t_1^p)_{tp^{2k+1}+1}) &\equiv d_1(x_{2k+1}^t) \otimes z + t v_2^{a_{2k+1}} v_3^{(tp-1)p^{2k}} d_1(\omega_1) + t v_2^{a_{2k}+p+1} d_1((t_1^p)_{(tp^2-1)p^{2k-1}}) \\
&\equiv \underbrace{t v_2^{a_{2k+1}} v_3^{(tp-1)p^{2k}} t_1 \otimes z}_{a} + t v_2^{a_{2k+1}} v_3^{(tp-1)p^{2k}} (-\underbrace{t_1 \otimes z}_{a} - \underbrace{v_2^2 b_{1,1}}_{b} - 2 v_2^p G_0) \\
&\qquad + \underbrace{t v_2^{a_{2k}+p+1+pa_{2k}-pa_{2k-1}} v_3^{(tp-1)p^{2k}} b_{1,1}}_{b} \mod J_{a_{2k+1}+p+1},
\end{aligned}$$

where the second equivalence follows by Lemmas 5.1, 4.24 and 3.4(2).

6.3.2 The cochains $(t_1^p)_{tp^k+e(k)}$ for $k\ge 2$ We put $r=2n-1+\varepsilon$ ($\varepsilon\in\{0,1\}$), and

$$(t_1^p)'_{tp^r+e(r)}=x_r^t(w_{r+1}+v_2^{p^r-p^{r-3}}w_r\eta_R(\sigma_{n-1}^{p^\varepsilon})+v_2^{a_r}w_rt_1^{p^\varepsilon})$$

for w_r in (4.15). Note that $w_r\equiv v_3^{pe(r-2)}w_2\equiv -v_3^{pe(r-2)}z$ mod J_p by (4.15) and (4.18). Then $(t_1^p)'_{tp^r+e(r)}\equiv x_r^tw_{r+1}\equiv -v_3^{tp^r+e(r)}t_1^p$ mod I_3. Furthermore, we calculate

$$\begin{aligned}d_1((t_1^p)'_{tp^r+e(r)})&\equiv \underset{\sim\sim\sim}{tv_2^{a_r}v_3^{(tp-1)p^{r-1}}t_1^{p^\varepsilon}\otimes w_{r+1}} \qquad \text{(by Lemmas 5.1 and 4.5 and Corollary 4.17)}\\ &\quad+x_r^t\big(\underbrace{v_2^{p^r}w_r\otimes t_1^{p^{r-1}}}_{a}-v_2^{p^r-p^{r-3}}w_r\otimes(\underbrace{v_2^{p^{2n-4+\varepsilon}}t_1^{p^{2n-2+\varepsilon}}}_{a}-\underbrace{v_2^{a_{2n-3+\varepsilon}}t_1^{p^\varepsilon}}_{b})\\ &\qquad-v_2^{a_r}(\underbrace{w_r\otimes t_1^{p^\varepsilon}}_{b}+\underset{\sim\sim\sim}{t_1^{p^\varepsilon}\otimes w_r})\big)\\ &\equiv -(t-1)v_2^{a_r}v_3^{tp^r+pe(r-2)}t_1^{p^\varepsilon}\otimes z \quad \text{mod } J_{a_r+p}\end{aligned}$$

together with (4.2) and (2.8). This case now follows from Lemma 4.24 by setting $(t_1^p)_{tp^r+e(r)}=-(t_1^p)'_{tp^r+e(r)}+(t-1)\,v_2^{a_r}v_3^{tp^r+pe(r-2)}\omega_{1+\varepsilon}$.

6.3.3 The cochains $(t_1^p)_{sp^{2k}}$ for $k\ge 1$ and $(t_1)_{sp^{2k+1}}$ for $k\ge 0$ We define $(t_1^{\varepsilon_i})_{sp^i}$ by

$$\begin{aligned}&(t_1)_{sp}=y'_{s,1},\quad (t_1^p)_{sp^2}=y'_{s,2},\quad (t_1)_{sp^3}=y'_{s,3}+sv_2^{e(3)p-p-1}v_3^{(sp^2-p-1)p}(zt_1-\omega_1),\\ &(t_1^p)_{sp^4}=y'_{s,4}-\tfrac12 sv_2^{e(3)p^2-p^2-p-1}v_3^{(sp^2-p-1)p^2}(\tilde\omega_2'-v_2^{p^2}zt_1^p),\\ &(t_1)_{sp^{2k+1}}=y'_{s,2k+1}+2sv_2^{e(3)p^{2k-1}-p-1}v_3^{(sp^2-p-1)p^{2k-1}}(zt_1-\omega_1),\\ &(t_1^p)_{sp^{2k+2}}=y'_{s,2k+2}-sv_2^{e(3)p^{2k}-p^2-p-1}v_3^{(sp^2-p-1)p^{2k}}\tilde\omega_2,\end{aligned}$$

where $\tilde\omega_2'=\tilde\omega_2-v_2^{p^2+p}t_1^pt_2-v_2^{e(3)}v_3^{-p^2}ct_4^p$. Except for $d_1((t_1^p)_{sp^4})$, the lemma for this case follows from Lemmas 5.2 and 4.24 with (4.2).

For $d_1((t_1^p)_{sp^4})$, we make a calculation:

$$\begin{aligned}\tilde\omega_2&\equiv -w_3 &&\text{(by Lemma 4.24)}\\ &\equiv t_1^p\eta_R(v_3^{p+1})-v_2t_2^p\eta_R(v_3^p)+v_2^{p+1}t_3^p &&\text{(by (4.15) and Proposition 4.12)}\\ &\equiv (z+\underbrace{v_2t_2^p}_{a}-v_2^pct_2)\eta_R(v_3^p)-\underbrace{v_2t_2^p\eta_R(v_3^p)}_{a}+v_2^{p+1}t_3^p &&\text{(by (4.19))}\\ &\equiv v_3^p(z+v_2^pt_2)-v_2^{p+1}ct_3^p \quad \text{mod } J_{p+2}. &&\text{(by (4.4))}\end{aligned}$$

Applying the Hopf conjugation c to the congruences of (4.6) shows the relations

$$\textbf{(6.1)}\qquad t_1^{p^3}\eta_R(v_3)\equiv v_3^pt_1+v_2ct_2^{p^2} \quad\text{and}\quad ct_2^{p^3}\eta_R(v_3)\equiv v_3^{p^2}ct_2-v_2ct_3^{p^2} \quad \text{mod } J_{p+1}.$$

Then, modulo J_{p+2},

$$\begin{aligned}
t_1^{p^4}\otimes v_3^p z &\equiv t_1^{p^4}\eta_R(v_3)^p\otimes z \\
&\equiv (v_3^{p^2}t_1^p+v_2^p ct_2^{p^3})\otimes z && \text{(by (6.1))}\\
&\equiv v_3^{p^2}t_1^p\otimes z+v_2^p ct_2^{p^3}\eta_R(v_3)\otimes t_1^p+v_2^{p+1}ct_2^{p^3}\otimes ct_2^p && \text{(by (4.18))}\\
&\equiv v_3^{p^2}t_1^p\otimes z+v_2^p(v_3^{p^2}ct_2-v_2ct_3^{p^2})\otimes t_1^p+v_2^{p+1}ct_2^{p^3}\otimes ct_2^p, && \text{(by (6.1))}\\
t_1^{p^4}\otimes v_2^p v_3^p t_2 &\equiv v_2^p t_1^{p^4}\eta_R(v_3)^p\otimes t_2 && \text{(by (6.1))}\\
&\equiv v_2^p v_3^{p^2}t_1^p\otimes t_2.
\end{aligned}$$

Therefore

$$\begin{aligned}
d_1(v_2^p v_3^{p^2}t_1^p t_2+v_2^{p+1}ct_4^p) &\equiv -v_2^p v_3^{p^2}(\underbrace{t_1^p\otimes t_2}_{a}+\underline{t_2\otimes t_1^p+t_1^{p+1}\otimes t_1^p+t_1\otimes t_1^{2p}})\\
&\qquad +v_2^{p+1}(\underbrace{t_1^{p^4}\otimes ct_3^p}_{b}-\underbrace{ct_2^{p^3}\otimes ct_2^p}_{c}+\underbrace{ct_3^{p^2}\otimes t_1^p}_{d}-v_3^p b_{1,2}^p),\\
t_1^{p^4}\otimes\widetilde{\omega}_2 &\equiv v_3^{p^2}t_1^p\otimes z+v_2^p(\underline{v_3^{p^2}ct_2}-\underbrace{v_2ct_3^{p^2}}_{d})\otimes t_1^p+\underbrace{v_2^{p+1}ct_2^{p^3}\otimes ct_2^p}_{c}\\
&\qquad +\underbrace{v_2^p v_3^{p^2}t_1^p\otimes t_2}_{a}-\underbrace{v_2^{p+1}t_1^{p^4}\otimes ct_3^p}_{b},
\end{aligned}$$

where the first equation follows from (4.2), (4.8) and (4.9). The sum of the wavy underlined terms is $-v_2^p v_3^{p^2}(2t_2\otimes t_1^p+t_1\otimes t_1^{2p})=-v_2^p v_3^{p^2}K_0$, and $b_{1,2}^p\equiv v_3^{p^2-p}b_{1,0}$ mod I_3 by (4.6). Then, modulo J_{p+2},

(6.2) $$t_1^{p^4}\otimes\widetilde{\omega}_2+d_1(v_2^p v_3^{p^2}t_1^p t_2+v_2^{p+1}ct_4^p)\equiv v_3^{p^2}t_1^p\otimes z-2v_2^p v_3^{p^2}K_0-v_2^{p+1}v_3^{p^2}b_{1,0}.$$

Now we calculate $d_1((t_1^p)_{sp^4})$ mod $J_{e(3)p^2+1}$ for odd prime p as

$$d_1(y'_{s,4})\equiv sv_2^{e(3)p^2-p-1}v_3^{(sp^2-p-1)p^2}(\underbrace{z\otimes t_1^p}_{a}-v_2^p(K_0+v_2Z'\otimes t_1^p)) \qquad \text{(by 5.2)}$$

and

$$\begin{aligned}
&d_1\big(-\tfrac12 sv_2^{e(3)p^2-p^2-p-1}v_3^{(sp^2-p-1)p^2}(\widetilde{\omega}'_2-v_2^{p^2}zt_1^p)\big)\\
&\equiv \tfrac12 sv_2^{e(3)p^2-p-1}v_3^{(sp^2-p-2)p^2}\underline{t_1^{p^4}\otimes\widetilde{\omega}_2}-\tfrac12 sv_2^{e(3)p^2-p^2-p-1}v_3^{(sp^2-p-1)p^2} && \text{(by Lemma 4.24)}\\
&\qquad\cdot(v_2^{p^2}z\otimes t_1^p+2v_2^{p^2+p}K_0+v_2^{p^2+p+1}b_{1,0}-\underline{d_1(v_2^{p^2+p}t_1^p t_2+v_2^{e(3)}v_3^{-p^2}ct_4^p)}+v_2^{p^2}(z\otimes t_1^p+t_1^p\otimes z))\\
&\equiv \tfrac12 sv_2^{e(3)p^2-p-1}v_3^{(sp^2-p-2)p^2}(\underbrace{v_3^{p^2}t_1^p\otimes z}_{b}-2v_2^p v_3^{p^2}K_0-v_2^{p+1}v_3^{p^2}b_{1,0}) && \text{(by (6.2))}\\
&\qquad -\tfrac12 sv_2^{e(3)p^2-p^2-p-1}v_3^{(sp^2-p-1)p^2}(\underbrace{v_2^{p^2}z\otimes t_1^p}_{a}+2v_2^{p^2+p}K_0+v_2^{p^2+p+1}b_{1,0}+v_2^{p^2}(\underbrace{z\otimes t_1^p}_{a}+\underbrace{t_1^p\otimes z}_{b})).
\end{aligned}$$

6.4 The cochains $(t_1^{p^2})_{tp-1}$ for $t \in \mathbb{Z}$

Put

$$(t_1^{p^2})_{tp-1} = -v_2^{-1} v_3^{(t-1)p} w.$$

Then the lemma for this case follows from Lemma 4.10.

References

[1] **H R Miller**, **D C Ravenel**, *Morava stabilizer algebras and the localization of Novikov's E_2–term*, Duke Math. J. 44 (1977) 433–447 MR Zbl

[2] **H R Miller**, **D C Ravenel**, **W S Wilson**, *Periodic phenomena in the Adams–Novikov spectral sequence*, Ann. of Math. 106 (1977) 469–516 MR Zbl

[3] **H Nakai**, *The chromatic E_1–term $H^0 M_1^2$ for $p > 3$*, New York J. Math. 6 (2000) 21–54 MR Zbl

[4] **S Oka**, *Multiplicative structure of finite ring spectra and stable homotopy of spheres*, from "Algebraic topology", Lecture Notes in Math. 1051, Springer (1984) 418–441 MR Zbl

[5] **S Oka**, **K Shimomura**, *On products of the β–elements in the stable homotopy of spheres*, Hiroshima Math. J. 12 (1982) 611–626 MR Zbl

[6] **D C Ravenel**, *The structure of* $\mathrm{BP}_*\mathrm{BP}$ *modulo an invariant prime ideal*, Topology 15 (1976) 149–153 MR Zbl

[7] **D C Ravenel**, *The cohomology of the Morava stabilizer algebras*, Math. Z. 152 (1977) 287–297 MR Zbl

[8] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, 2nd edition, Amer. Math. Soc., Providence, RI (2004) Zbl

[9] **K Shimomura**, *Note on the right unit map and some elements of the Brown–Peterson homology*, J. Fac. Educ. Tottori Univ. (Nat. Sci.) 38 (1989) 77–89

[10] **K Shimomura**, *The chromatic E_1–term $H^1 M_2^1$ and its application to the homology of the Toda–Smith spectrum $V(1)$*, J. Fac. Educ. Tottori Univ. (Nat. Sci.) 39 (1990) 63–83 Correction in 41 (1992) 7–11

[11] **K Shimomura**, **H Tamura**, *Nontriviality of some compositions of β–elements in the stable homotopy of the Moore spaces*, Hiroshima Math. J. 16 (1986) 121–133 MR Zbl

[12] **H Toda**, *On spectra realizing exterior parts of the Steenrod algebra*, Topology 10 (1971) 53–65 MR Zbl

Department of Mathematics, Faculty of Science, Kochi University
Kochi, Japan

Department of Mathematics, Faculty of Science and Technology, Kochi University
Kochi, Japan

katsumi@kochi-u.ac.jp, mao.shimomura@outlook.jp

Received: 7 November 2022 Revised: 23 July 2023

Algebraic & Geometric Topology **24**:7 (2024) 3897–3917
DOI: 10.2140/agt.2024.24.3897
Published: 9 December 2024

Phase transition for the existence of van Kampen 2–complexes in random groups

TSUNG-HSUAN TSAI

Gromov (1993) showed that every reduced van Kampen diagram D of a random group at density d satisfies the isoperimetric inequality $|\partial D| \geq (1-2d-\varepsilon)|D|\ell$. Adapting Gruber and Mackay's (2021) method for random triangular groups, we obtain a nonreduced van Kampen 2–complex version of this inequality.

The main result of this article is a phase transition: given a geometric form Y of 2–complexes, we find a critical density $d_c(Y)$ such that, in a random group at density d, if $d < d_c$, then there is no reduced van Kampen 2–complex of the form Y; while if $d > d_c$, then there exists reduced van Kampen 2–complexes of the form Y.

As an application, we exhibit phase transitions for small-cancellation conditions in random groups, giving explicitly the critical densities for the conditions $C'(\lambda)$, $C(p)$, $B(p)$ and $T(q)$.

20F06; 20F05, 20P05

1 Introduction

Random groups The first occurrence of random group presentations is the density model by M Gromov [1993, 9.B]. Formally, a random group is a random variable with values in a given set of groups, often constructed by group presentations with a fixed set of generators and a random set of relators. The goal is to study the asymptotic behaviors of a sequence of random groups (G_ℓ) when the maximal relator lengths ℓ goes to infinity. We say that G_ℓ satisfies some property Q_ℓ *asymptotically almost surely* (a.a.s.) if the probability that G_ℓ satisfies Q_ℓ converges to 1 as ℓ goes to infinity.

Let us consider the *permutation invariant density model* of random groups introduced by Gromov [1993, page 272] and developed in [Tsai 2022]. Fix the set of generators $X_m = \{x_1, \dots, x_m\}$ with $m \geq 2$ for group presentations. Let B_ℓ be the set of cyclically reduced words of $X_m^{\pm}$ of length at most ℓ. We shall construct random groups by *densable* and *permutation invariant* random subsets of B_ℓ.

Definition 1.1 [Gromov 1993, page 272; Tsai 2022, Definitions 1.5 and 2.5] A sequence of random subsets (R_ℓ) of the sequence of sets (B_ℓ) is called *densable with density* $d \in \{-\infty\} \cup [0,1]$ if the sequence of random variables $\mathrm{dens}_{B_\ell}(R_\ell) := \log_{|B_\ell|}(|R_\ell|)$ converges in probability to the constant d.

The sequence (R_ℓ) is called *permutation invariant* if R_ℓ is a permutation measure-invariant random subset of B_ℓ.

Many natural models of random subsets are densable and permutation invariant. For example, the uniform distribution on all subsets of cardinality $\lfloor B_\ell^d \rfloor$ considered in [Ollivier 2004; 2005; 2007], or the Bernoulli sampling of parameter $|B_\ell|^{d-1}$ considered in [Antoniuk et al. 2015] for random triangular groups.

Some other natural models are densable but not permutation invariant. For instance, consider the Bernoulli sampling of parameter $(2m)^{(d-1)\ell}$ on the set of nonreduced words of length ℓ, and reduce these words to form a random subset of B_ℓ. This is also the case for Gromov's expander graph model [2003], in which the random relators are the words read on the simple cycles of a randomly labeled expander graph.

Definition 1.2 [Gromov 1993, page 273; Tsai 2022, Definition 4.1] A sequence of random groups $(G_\ell(m,d))$ with m generators at density d is defined by

$$G_\ell(m,d) = \langle X_m \mid R_\ell \rangle,$$

where (R_ℓ) is a densable sequence of permutation invariant random subsets of (B_ℓ) with density d.

For detailed surveys on random groups, we refer the reader to work by E Ghys [2004], Y Ollivier [2005], I Kapovich and P Schupp [2008], and F Bassino, C Nicaud and P Weil [Bassino et al. 2020].

Van Kampen 2–complexes We consider oriented combinatorial 2–complexes and van Kampen diagrams as in [Lyndon and Schupp 1977, III.2 and III.9], with an additional precision that each face has an orientation given by its boundary path.

A 2*–complex* is hence a triplet $Y = (V, E, F)$ where V is the set of vertices, E is the set of oriented edges and F is the set of oriented faces. Its underlying graph is denote by $Y^{(1)} = (V, E)$. Every edge $e \in E$ has a starting point $\alpha(e) \in V$, an ending point $\omega(e) \in V$ and an inverse edge $e^{-1} \in E$, satisfying $\alpha(e^{-1}) = \omega(e)$, $\omega(e^{-1}) = \alpha(e)$ and $(e^{-1})^{-1} = e$. Every face $f \in F$ has a boundary path ∂f that is a cyclically reduced loop on the underlying graph $Y^{(1)}$, and an inverse face $f^{-1} \in F$ whose boundary path is the inverse. That is to say, it satisfies $(f^{-1})^{-1} = f$ and $\partial(f^{-1}) = (\partial f)^{-1}$. The *starting point* of a face f is the starting point of its boundary path. Note that f^{-1} has the same starting point as f.

A *geometric edge* is a pair of inverse edges $\{e, e^{-1}\}$, denoted by $\bar{e}$. Similarly, a *geometric face* is a pair of inverse faces $\{f, f^{-1}\}$, denoted by $\bar{f}$. Throughout this article, we will carefully distinguish oriented edges (faces) and geometric edges (faces). We denote by $|Y^{(1)}|$ the number of geometric edges and $|Y|$ the number of geometric faces.

Definition 1.3 A *van Kampen* 2*–complex* with respect to a group presentation $G = \langle X \mid R \rangle$ is a 2–complex $Y = (V, E, F)$ with labels on edges by generators $\varphi_1 \colon E \to X^\pm$ and labels on faces by relators $\varphi_2 \colon F \to R^\pm$ such that $\varphi_1(e^{-1}) = \varphi_1(e)^{-1}$, $\varphi_2(f^{-1}) = \varphi_2(f)^{-1}$ and $\varphi_1(\partial f) = \varphi_2(f)$.

We denote briefly $Y = (V, E, F, \varphi_1, \varphi_2)$.

The data of the labels φ_1, φ_2 on Y is equivalently given by a combinatorial map $Y \to K(X, R)$, where $K(X, R)$ is the standard 2–complex with respect to the group presentation $G = \langle X \mid R \rangle$ (with one vertex, an edge for each generator and its inverse, and a face for each relator and its inverse).

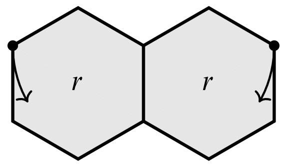


Figure 1: A reducible pair of faces.

A *van Kampen diagram* D is a finite, planar (embedded in a Euclidean plane) and simply connected van Kampen 2–complex. Its boundary length $|\partial D|$ is the length of a boundary path, passing once by every edge adjacent to one face and twice by every edge adjacent to zero faces.

A pair of faces in a van Kampen 2–complex is called *reducible* if they have the same relator label and their boundaries share a common edge at the same respective position (see Figure 1). A van Kampen 2–complex is called *reduced* if there is no reducible pair of faces.

Isoperimetric inequalities In order to prove the hyperbolicity of a random group at density $d < \frac{1}{2}$, Gromov [1993, 9.B] showed that a.a.s. *reduced* local van Kampen diagrams of $G_\ell(m,d)$ satisfy an isoperimetric inequality depending on the density d.

Theorem 1.4 [Gromov 1993, page 274; Ollivier 2004, Chapter 2] *Let $(G_\ell(m,d))$ be a sequence of random groups with $m \geq 2$ generators at density d. For any $\varepsilon > 0$ and $K > 0$, a.a.s. every* ***reduced*** *van Kampen diagram D of $G_\ell(m,d)$ with $|D| \leq K$ satisfies the isoperimetric inequality*

$$|\partial D| \geq (1 - 2d - \varepsilon)|D|\ell.$$

Ollivier's proof [2004] can achieve a slightly stronger[1] inequality,

$$|D^{(1)}| \geq \left(1 - d - \tfrac{1}{2}\varepsilon\right)|D|\ell.$$

One may expect such an inequality to hold for *every* reduced van Kampen 2–complex Y with $|Y| \leq K$. D Gruber and J Mackay [2021, Section 2] showed that in the triangular model of random groups,[2] the above inequality holds for every *nonreduced van Kampen* 2*–complex* Y with $|Y| \leq K$ if the *reduction degree* (Definition 2.1) $\mathrm{Red}(Y)$ is added in the left-hand side of the inequality.

However, the result fails in the regular Gromov density model: the condition $|Y| \leq K$ is not enough (see Remark 2.4). In Section 2 of this paper, we introduce the notion of *complexity* (Definition 2.2) to adapt Gruber and Mackay's inequality in the Gromov density model, establishing a *nonreduced van Kampen* 2*–complex* version of Theorem 1.4. A similar approach was given in the preprint [Odrzygóźdź 2021].

[1] Note that every van Kampen diagram composed of relators of lengths at most ℓ satisfies $2|D^{(1)}| - |\partial D| \leq |D|\ell$, so the given inequality implies the isoperimetric inequality.

[2] A model where the relator length $\ell = 3$ is fixed, and we are interested in asymptotic behaviors when the number of generators m goes to infinity.

Theorem 1.5 *Let $(G_\ell(m,d))$ be a sequence of random groups with $m \geq 2$ generators at density d. Let $\varepsilon > 0$, $K > 0$. For any $d < \frac{1}{2}$, a.a.s. every van Kampen 2–complex Y of complexity K of $G_\ell(m,d)$ satisfies*

$$|Y^{(1)}| + \mathrm{Red}(Y) \geq (1-d-\varepsilon)|Y|\ell.$$

Phase transition for the existence of van Kampen 2–complexes We are now interested in the converse of Theorem 1.5: Given a 2–complex Y satisfying the inequality of Theorem 1.5, is it true that a.a.s. there exists a *reduced* van Kampen 2–complex of $G_\ell(m,d)$ whose underlying 2–complex is Y?

A 2–complex Y is said to be *fillable* by a group presentation $G = \langle X \mid R\rangle$ (or by the set of relators R) if there exists a *reduced* van Kampen 2–complex of G whose underlying 2–complex is Y. An edge of a 2–complex is called *isolated* if it is not adjacent to any face. Since isolated edges do not affect fillability, we will only consider finite 2–complexes without isolated edges in the following.

To better formulate the problem, we consider a sequence of 2–complexes (Y_ℓ) and introduce the notion of *geometric form* of 2–complexes (Y,λ) (Definition 3.1), together with its density $\mathrm{dens}\, Y$ and its *critical density* $\mathrm{dens}_c\, Y$ (Definition 3.2). The main result of this article is the phase transition at density $1-\mathrm{dens}_c(Y)$, for the fillability of the 2–complex Y_ℓ.

Theorem 1.6 *Let $(G_\ell(m,d))$ be a sequence of random groups with $m \geq 2$ generators at density d. Let (Y_ℓ) be a sequence of 2–complexes with some geometric form (Y,λ).*

(i) *If $d < 1-\mathrm{dens}_c\, Y$, then a.a.s. Y_ℓ is not fillable by $G_\ell(m,d)$.*

(ii) *If $d > 1-\mathrm{dens}_c\, Y$ and Y_ℓ is fillable by B_ℓ, then a.a.s. Y_ℓ is fillable by $G_\ell(m,d)$.*

In Section 3, we prove Theorem 1.6 using the multidimensional intersection formula for random subsets (Theorem 3.6, [Tsai 2022, Theorem 3.7]), which generalizes the proof for the $C'(\lambda)$ phase transition in [Tsai 2022, Theorem 1.4]. We will see in Remark 3.3 that the second assertion of the theorem is equivalent to the following corollary.

Corollary 1.7 *Let $(G_\ell(m,d))$ be a sequence of random groups with $m \geq 2$ generators at density d. Let $s > 0$ and $K > 0$. Let (Y_ℓ) be a sequence of 2–complexes of the same geometric form such that Y_ℓ is fillable by B_ℓ. If every sub-2–complex Z_ℓ of Y_ℓ satisfies*

$$|Z_\ell^{(1)}| \geq (1-d+s)|Z_\ell|\ell,$$

then a.a.s. Y_ℓ is fillable by $G_\ell(m,d)$.

Note that we need Y_ℓ to have at least one filling by the set of all possible relators B_ℓ. It is automatically satisfied for planar and simply connected 2–complexes. In addition, if every face boundary length of Y_ℓ is exactly ℓ, then the given inequality is equivalent to an isoperimetric inequality similar the inequality of Theorem 1.4. Hence the following corollary.

Corollary 1.8 *Let $(G_\ell(m,d))$ be a sequence of random groups with $m \geq 2$ generators at density d. Let $s > 0$ and $K > 0$. Let (D_ℓ) be a sequence of finite planar 2–complexes of the same geometric form such that every face boundary length of D_ℓ is exactly ℓ. If every sub-2–complex D'_ℓ of D_ℓ satisfies*

$$|\partial D'_\ell| \geq (1-2d+s)|D'_\ell|\ell,$$

then a.a.s. D_ℓ is fillable by $G_\ell(m,d)$.

It is mentioned in [Ollivier and Wise 2011, Proposition 1.8] that when $d < 1/p$, a.a.s., a random group at density d has the $C(p)$ small cancellation condition. As an application of Theorem 1.6, we show that there is a phase transition: if $d > 1/p$, then a.a.s. a random group at density d *does not* have $C(p)$ (see Proposition 4.2).

Acknowledgements The content of this article was completed during my PhD thesis at the University of Strasbourg. I would like to thank my thesis advisor, Thomas Delzant, for his guidance and interesting discussions on the subject.

2 Isoperimetric inequality for van Kampen 2–complexes

We shall prove Theorem 1.5 in this section.

2.1 Reduction degree and complexity

Given a (nonreduced) van Kampen diagram $Y = (V, E, F, \varphi_1, \varphi_2)$ with respect to a group presentation $\langle X \mid R \rangle$, its *reduction degree* is the total number of geometric edges causing reducible pair of faces, counted with *multiplicity*: for any edge $e \in E$, any relator $r \in R$ and any integer j, we count the number of faces $f \in F$ labeled by r and having e as the j^{th} boundary edge. If this number is k, we add $(k-1)^+$ to the reduction degree where $(\cdot)^+$ is the positive part function. Here is the formal definition given by Gruber and Mackay [2021].

Definition 2.1 (reduction degree [Gruber and Mackay 2021, Definition 2.5]) Let $Y = (V, E, F, \varphi_1, \varphi_2)$ be a van Kampen 2–complex of a group presentation $G = \langle X \mid R \rangle$. Let ℓ be the maximal boundary length of faces of Y. The reduction degree of Y is

$$\operatorname{Red}(Y) = \sum_{e \in E} \sum_{r \in R} \sum_{1 \leq j \leq \ell} \left(\left| \{ f \in F \mid \varphi_2(f) = r,\ e \text{ is the } j^{\text{th}} \text{ edge of } \partial f \} \right| - 1 \right)^+.$$

It is not hard to see that a van Kampen 2–complex Y is reduced if and only if $\operatorname{Red}(Y) = 0$. Since isolated edges (edges that are not attached by any face) do not affect the reduction degree, we will only consider 2–complexes without isolated edges in the following.

A *maximal arc* of a 2–complex is a reduced combinatorial path passing only by vertices of degree 2 whose endpoints are not of degree 2. The *complexity* of a 2–complex encodes the number of maximal arcs with the number of faces.

Definition 2.2 (complexity of a 2–complex) Let Y be a 2–complex. Let $K > 0$. We say that Y is of complexity K if $|Y| \leq K$ and if for any face f of Y, the boundary path ∂f is divided into at most K maximal arcs.

If D is a planar and simply connected 2–complex with $|D| \leq K$, then the complexity of D is $6K$. Indeed, as the rank of its underlying graph is K, the number of its maximal arcs is at most $3K$, and every boundary path is divided into at most $6K$ maximal arcs (an arc may be used twice).

Lemma 2.3 *Let $K > 0$. There exists a number $C(K)$, depending only on K, such that the number of 2–complexes of complexity K with face boundary lengths at most ℓ is bounded by $C(K)\ell^{K^2}$.*

Proof Recall that we only consider 2–complexes without isolated edges, so the number of maximal arcs in a 2–complex of complexity K is at most K^2 (each of the K faces has at most K arcs). Since the face boundary lengths are at most ℓ, these K^2 maximal arcs have lengths at most ℓ. So there are at most ℓ^{K^2} choices for their lengths. Now let $C(K)$ be the number of choices to attach these K^2 maximal arcs to form a 2–complex. The number of ways to construct a 2–complex of complexity K with boundary lengths at most ℓ is hence bounded by $C(K)\ell^{K^2}$. □

Remark 2.4 While the number of 2–complexes with a *bounded complexity* grows polynomially with the maximal face boundary length ℓ, it is not the case for 2–complexes with a *bounded number of faces*, not even for 2–complexes with a *bounded number of maximal arcs*.

For example, consider the set of 2–complexes with one single face of boundary length ℓ whose underlying graph is 8–shaped with one vertex and two edges. There are only two maximal arcs, while the number of such 2–complexes equals to the number of words on two letters and their inverses of length ℓ, which grows exponentially with ℓ. Our polynomial bound will be useful in the proof of Theorem 1.5.

Remark 2.5 Actually, there are van Kampen 2–complexes that contradict the inequality of Theorem 1.5. For instance, D Calegari and A Walker [2015] proved that at any density $d < \frac{1}{2}$, there exists a number K depending only on d such that, in $G_\ell(m,d)$ there is a.a.s. a reduced van Kampen 2–complex Y homeomorphic to a surface of genus $O(\ell)$ (hence with complexity $O(\ell)$) with at most K faces.

Since every edge is adjacent to two faces in a surface, we have $|Y^{(1)}| \leq \frac{1}{2}|Y|\ell$, while according to Theorem 1.5 we expect that

$$|Y^{(1)}| \geq (1-d-\varepsilon)|Y|\ell > \tfrac{1}{2}|Y|\ell.$$

2.2 Abstract van Kampen 2–complexes

Let $(G_\ell(m,d))$ be a sequence of random groups at density d, defined by $G_\ell(m,d) = \langle x_1,\dots,x_m \mid R_\ell\rangle$. Recall that B_ℓ is the set of all cyclically reduced words of length at most ℓ and $|B_\ell| = (2m-1)^{\ell+O(1)}$. Let $0 < \varepsilon < 1-d$. Since $\log_{|B_\ell|}|R_\ell|$ converges in probability to the constant d, the probability event

$$Q_\ell := \left\{(2m-1)^{(d-(\varepsilon/4)\ell)} \leq |R_\ell| \leq (2m-1)^{(d+(\varepsilon/4)\ell)}\right\}$$

is a.a.s. true (see [Tsai 2022, Proposition 1.8]).

If we consider the Bernoulli density model where the events $\{r \in R_\ell\}$ through $r \in B_\ell$ are independent of the same probability $(2m-1)^{(d-1)\ell}$, it is obvious that we have $\mathbf{Pr}(r_1,\dots,r_k \in R_\ell) = (2m-1)^{k(d-1)\ell}$ for distinct $r_1,\dots,r_k$ in B_ℓ. In the permutation invariant density model, we have the following corresponding proposition, which is a variant of [Tsai 2022, Lemma 3.10].

Proposition 2.6 *Let $r_1,\dots,r_k$ be pairwise different relators in B_ℓ. We have*

$$\mathbf{Pr}(r_1,\dots,r_k \in R_\ell \mid Q_\ell) \le (2m-1)^{k(d-1+(\varepsilon/2))\ell}. \qquad \square$$

Abstract van Kampen 2–complexes, as abstract van Kampen diagrams introduced by Ollivier [2004], is a structure between 2–complexes and van Kampen 2–complexes that helps us solve 2–complex problems in random groups. Recall that since isolated edges do not affect fillability, we will only consider finite 2–complexes without isolated edges.

Definition 2.7 (abstract van Kampen 2–complex) An abstract van Kampen 2–complex $\widetilde{Y}$ is a 2–complex (V,E,F) with a labeling function on faces by integer numbers and their inverses

$$\widetilde{\varphi}_2\colon F \to \{1,1^-,2,2^-,\dots,k,k^-\}$$

such that $\widetilde{\varphi}_2(f^{-1}) = \widetilde{\varphi}_2(f)^-$. We denote it simply by $\widetilde{Y} = (V,E,F,\widetilde{\varphi}_2)$.

By convention $(i^-)^- = i$. The integers $\{1,\dots,k\}$ are called abstract relators. Similar to a van Kampen diagram, a pair of faces $f, f' \in F$ is *reducible* if they are labeled by the same abstract relator, and they share an edge at the same position of their boundaries. An abstract diagram is called *reduced* if there is no reducible pair of faces. Let ℓ be the maximal boundary length of faces. The *reduction degree* of the 2–complex $\widetilde{Y}$ can be similarly defined as

$$\operatorname{Red}(\widetilde{Y}) = \sum_{e\in E}\sum_{1\le i\le k}\sum_{1\le j\le\ell}\bigl(\bigl|\{f\in F \mid \widetilde{\varphi}_2(f)=i,\ e \text{ is the } j^{\text{th}} \text{ edge of } \partial f\}\bigr|-1\bigr)^+.$$

We say that an abstract van Kampen 2–complex with k abstract relators $\widetilde{Y} = (V,E,F,\widetilde{\varphi}_2)$ is *fillable* by a group presentation $G = \langle X \mid R\rangle$ (or by a set of relators R) if there exists k *different* relators $r_1,\dots,r_k \in R$ such that the construction $\varphi_2(f) := r_{\widetilde{\varphi}_2(f)}$ gives a van Kampen 2–complex $Y = (V,E,F,\varphi_1,\varphi_2)$[3] of G. The k–tuple of relators $(r_1,\dots,r_k)$, or the van Kampen 2–complex Y, is called a *filling* of $\widetilde{Y}$; see Figure 2, left. As we picked different relators for different abstract relators, if Y is a filling of $\widetilde{Y}$, then $\operatorname{Red}(Y) = \operatorname{Red}(\widetilde{Y})$, and $\widetilde{Y}$ is reduced if and only if Y is reduced.

Denote ℓ_i the length of the abstract relator i for $1 \le i \le k$. Let $\ell = \max\{\ell_1,\dots,\ell_k\}$ be the maximal boundary length of faces. The pairs of integers $(i,1),\dots,(i,\ell_i)$ are called *abstract letters* of i. The set of abstract letters of $\widetilde{Y}$ is then a subset of the product set $\{1,\dots,k\}\times\{1,\dots,\ell\}$. The geometric edges of $\widetilde{Y}$ are decorated by abstract letters and directions: Let $f \in F$ be labeled by i and let $e \in E$ be at the j^{th}

[3] Note that the edge labeling φ_1 is determined by the face labeling φ_2 as there are no isolated edges.

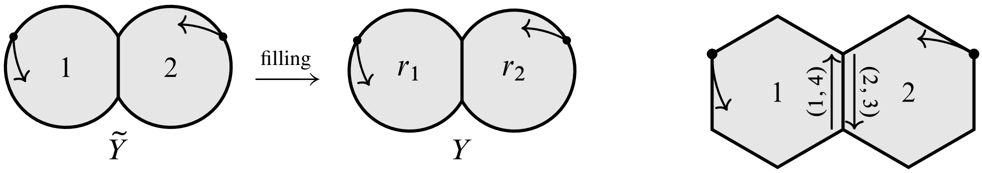


Figure 2: Left: filling an abstract van Kampen 2–complex. Right: a geometric edge decorated by two abstract letters.

position of ∂f. The geometric edge $\bar{e}$ is decorated, on the side of $\bar{f}$, by an arrow indicating the direction of e and the abstract letter (i, j). The number of decorations on a geometric edge is the number of its adjacent faces with multiplicity (an edge may be attached twice by the same face); see Figure 2, right.

Definition 2.8 (free-to-fill) An abstract letter (i, j) of $\widetilde{D}$ is *free-to-fill* if, for any edge $\bar{e}$ decorated by (i, j), it is the minimal decoration on $\bar{e}$.

Denote α_i the number of faces labeled by the abstract relator i and η_i the number of free-to-fill edges of i. We have the following estimation.

Lemma 2.9 *Let $\widetilde{Y} = (V, E, F, \widetilde{\varphi}_2)$ be an abstract van Kampen 2–complex with k abstract relators. Then*

$$\sum_{i=1}^{k} \alpha_i \eta_i \leq |\widetilde{Y}^{(1)}| + \mathrm{Red}(\widetilde{Y}).$$

Proof Denote by $\overline{E}$ the set of geometric edges and $\overline{F}$ the set of geometric faces. For any geometric edge $\bar{e}$, an adjacent face $\bar{f}$ from which the decoration is minimal is called a *preferred face* of $\bar{e}$. For any face $\bar{f}$, let $\overline{E}_{\bar{f}}$ be the set of geometric edges $\bar{e}$ on its boundary such that $\bar{f}$ is a preferred face of $\bar{e}$. Note that an edge will never be counted twice as the decorations given by one face are all different. According to Definition 2.8, for any face f with $\widetilde{\varphi}_2(f) = i$, we have $\eta_i \leq |\overline{E}_{\bar{f}}|$. Hence,

$$\sum_{i=1}^{k} \alpha_i \eta_i \leq \sum_{\bar{f} \in \overline{F}} |\overline{E}_{\bar{f}}|.$$

Denote by $\mathrm{Red}(\bar{e})$ the reduction degree caused by the edge $\bar{e}$. That is,

$$\mathrm{Red}(\bar{e}) := \sum_{1 \leq i \leq k} \sum_{1 \leq j \leq \ell} \bigl(\bigl|\{f \in F \mid \widetilde{\varphi}_2(f) = i,\ e \text{ or } e^{-1} \text{ is the } j^{\text{th}} \text{ edge of } \partial f\}\bigr| - 1\bigr)^+,$$

so that the number of preferred faces of $\bar{e}$ is bounded by $1 + \mathrm{Red}(\bar{e})$. Hence,

$$\sum_{\bar{f} \in \overline{F}} |\overline{E}_{\bar{f}}| \leq \sum_{\bar{e} \in \overline{E}} (1 + \mathrm{Red}(\bar{e})) = |\widetilde{Y}^{(1)}| + \mathrm{Red}(\widetilde{Y}).$$ □

Probability of filling We shall estimate the probability that an abstract van Kampen 2–complex $\widetilde{Y}$ is fillable by a random group $G_\ell(m, d)$. This step is the key to prove Theorem 1.5. Recall that

$$Q_\ell := \bigl\{(2m-1)^{(d-(\varepsilon/4)\ell)} \leq |R_\ell| \leq (2m-1)^{(d+(\varepsilon/4)\ell)}\bigr\}$$

is an a.a.s. true probability event.

Lemma 2.10 *Let $\widetilde{Y}$ be an abstract van Kampen 2–complex with k abstract relators. We have*

$$\mathbf{Pr}(\widetilde{Y} \text{ is fillable by } G_\ell(m,d) \mid Q_\ell) \le \left(\frac{2m}{2m-1}\right)^k (2m-1)^{\sum_{i=1}^k (\eta_i + (d-1+\varepsilon/2)\ell)}.$$

Proof Let us estimate the number of fillings of $\widetilde{Y}$. For every free-to-fill abstract letter (i,j), there are at most $2m$ ways to fill a generator if $j=1$, at most $(2m-1)$ ways to fill if $j \neq 1$ for avoiding reducible word. As there are η_i free-to-fill abstract letters on the i^{th} abstract relator, there are at most $2m(2m-1)^{\eta_i-1}$ ways to fill it. So there are at most $\prod_{i=1}^k (2m(2m-1)^{\eta_i-1})$ ways to fill $\widetilde{Y}$.

Let Y be a van Kampen 2–complex, which is a filling of $\widetilde{Y}$. The 2–complex Y is labeled by k different relators in B_ℓ, denoted $r_1,\dots,r_k$. By Proposition 2.6,

$$\mathbf{Pr}(Y \text{ is a 2–complex of } G_\ell(m,d) \mid Q_\ell) = \mathbf{Pr}(r_1,\dots,r_k \in R_\ell \mid Q_\ell) \le (2m-1)^{k(d-1+\varepsilon/2)\ell}.$$

Hence

$$\begin{aligned}\mathbf{Pr}(\widetilde{Y} \text{ is fillable by } G_\ell(m,d) \mid Q_\ell) &\le \sum_{Y \text{ fills } \widetilde{Y}} \mathbf{Pr}(Y \text{ is a 2–complex of } G_\ell(m,d) \mid Q_\ell)\\ &\le \prod_{i=1}^k (2m(2m-1)^{\eta_i-1})(2m-1)^{k(d-1+\varepsilon/2)\ell}\\ &\le \left(\frac{2m}{2m-1}\right)^k (2m-1)^{\sum_{i=1}^k (\eta_i+(d-1+\varepsilon/2)\ell)}. \qquad \square\end{aligned}$$

Lemma 2.11 *Let $\widetilde{Y}$ be an abstract van Kampen 2–complex with k abstract relators. Suppose that $\widetilde{Y}$ does not satisfy the inequality given in Theorem 1.5, ie*

$$|\widetilde{Y}^{(1)}| + \mathrm{Red}(\widetilde{Y}) < (1-d-\varepsilon)|\widetilde{Y}|\ell,$$

then

$$\mathbf{Pr}(\widetilde{Y} \text{ is fillable by } G_\ell(m,d) \mid Q_\ell) \le \left(\frac{2m}{2m-1}\right)(2m-1)^{-(\varepsilon/2)\ell}.$$

Proof Let $\widetilde{Y}_i$ be the sub-2–complex of $\widetilde{Y}$ consisting of faces labeled by the i first abstract relators. Let $P_i = \mathbf{Pr}(\widetilde{Y}_i \text{ is fillable by } G_\ell(m,d) \mid Q_\ell)$. Apply Lemma 2.10 on $\widetilde{Y}_i$; we have

$$P_i \le \left(\frac{2m}{2m-1}\right)^i (2m-1)^{\sum_{j=1}^i (\eta_j+(d-1+\varepsilon/2)\ell)}.$$

Note that if $\widetilde{Y}$ is fillable by $G_\ell(m,d)$ then its sub-2–complex $\widetilde{Y}_i$ is fillable by the same group. So for any $1 \le i \le k$,

$$\log_{2m-1}(P_k) \le \log_{2m-1}(P_i) \le \sum_{j=1}^i \left(\eta_j + \left(d-1+\tfrac{1}{2}\varepsilon\right)\ell + \log_{2m-1}\left(\frac{2m}{2m-1}\right)\right).$$

Without loss of generality, suppose that $\alpha_1 \ge \alpha_2 \ge \dots \ge \alpha_k$. Note that $\log_{2m-1}(P_k)$ is negative and

$\alpha_1 \le |\widetilde{Y}|$, so $|\widetilde{Y}|\log_{2m-1}(P_k) \le \alpha_1 \log_{2m-1}(P_k)$. By Abel's summation formula, with convention $\alpha_{k+1}=0$,

$$\begin{aligned}|\widetilde{Y}|\log_{2m-1}(P_k) \le \alpha_1\log_{2m-1}(P_k) &= \sum_{i=1}^{k}(\alpha_i-\alpha_{i+1})\log_{2m-1}(P_k)\\ &\le \sum_{i=1}^{k}(\alpha_i-\alpha_{i+1})\sum_{j=1}^{i}\Big[\eta_i+\big(d-1+\tfrac{1}{2}\varepsilon\big)\ell+\log_{2m-1}\Big(\frac{2m}{2m-1}\Big)\Big]\\ &= \sum_{i=1}^{k}\alpha_i\Big[\eta_i+\big(d-1+\tfrac{1}{2}\varepsilon\big)\ell+\log_{2m-1}\Big(\frac{2m}{2m-1}\Big)\Big]\\ &= \sum_{i=1}^{k}\alpha_i\eta_i+\Big(\sum_{i=1}^{k}\alpha_i\Big)\Big[\big(d-1+\tfrac{1}{2}\varepsilon\big)\ell+\log_{2m-1}\Big(\frac{2m}{2m-1}\Big)\Big].\end{aligned}$$

Note that $\sum_{i=1}^{k}\alpha_i=|\widetilde{Y}|$. By Lemma 2.9 and the hypothesis of the current lemma,

$$\sum_{i=1}^{k}\alpha_i\eta_i \le |\widetilde{Y}^{(1)}|+\mathrm{Red}(\widetilde{Y}) < (1-d-\varepsilon)|\widetilde{Y}|\ell.$$

Hence,

$$\begin{aligned}|\widetilde{Y}|\log_{2m-1}(P_k) &\le (1-d-\varepsilon)|\widetilde{Y}|\ell+|\widetilde{Y}|\Big[\big(d-1+\tfrac{1}{2}\varepsilon\big)\ell+\log_{2m-1}\Big(\frac{2m}{2m-1}\Big)\Big]\\ &\le |\widetilde{Y}|\Big[-\tfrac{1}{2}\varepsilon\ell+\log_{2m-1}\Big(\frac{2m}{2m-1}\Big)\Big].\end{aligned}$$

□

2.3 Proof of Theorem 1.5

Under the condition $Q_\ell := \{(2m-1)^{(d-(\varepsilon/4)\ell)} \le |R_\ell| \le (2m-1)^{(d+(\varepsilon/4)\ell)}\}$, the probability that there exists a van Kampen 2–complex of complexity K of $G_\ell(m,d)$ satisfying the inverse inequality

$$(*)\qquad |Y^{(1)}|+\mathrm{Red}(Y) < (1-d-\varepsilon)|Y|\ell$$

is bounded by

$$\sum_{\widetilde{Y}\text{ of complexity }K,\text{ satisfying }(*)} \mathbf{Pr}(\widetilde{Y}\text{ is fillable by }G_\ell(m,d)\mid Q_\ell).$$

By Lemma 2.3 and the face that there at most K^{2K} ways to label a 2–complex with K faces by abstract relators $\{1^\pm,\dots,K^\pm\}$, there are at most $\ell^{3K}\times K^{2K}$ terms in the sum. By Lemma 2.11, every term is bounded by

$$\Big(\frac{2m}{2m-1}\Big)(2m-1)^{-(\varepsilon/2)\ell}.$$

So the sum is smaller than

$$\ell^{3K}K^{2K}\Big(\frac{2m}{2m-1}\Big)(2m-1)^{-(\varepsilon/2)\ell},$$

which converges to 0 as $\ell\to\infty$.

By definition $\mathbf{Pr}(Q_\ell) \xrightarrow{\ell\to\infty} 1$, so the probability that there exists a van Kampen 2–complex of $G_\ell(m,d)$ of complexity K satisfying $(*)$ converges to 0 as ℓ goes to infinity. That is to say, a.a.s. every van Kampen diagram of $G_\ell(m,d)$ of complexity K satisfies the inequality

$$|Y^{(1)}| + \mathrm{Red}(Y) \geq (1-d-\varepsilon)|Y|\ell. \qquad \square$$

Collapsible 2–complexes and closed surfaces Recall that an *elementary collapse* of a 2–complex, in the sense of Whitehead [1939], is the removal of a face together with one of its edges that is not adjacent to other faces. A 2–complex is called *collapsible*[4] to a graph if it can be collapsed to a graph by a sequence of elementary collapses.

Let Y be a 2–complex of complexity K. If Y is *not* collapsible, then after all possible elementary collapses, we obtain a sub-2–complex Y' having only edges that are adjacent to at least 2 faces, which gives $|Y'^{(1)}| \leq \frac{1}{2}|Y'|\ell$, where ℓ is the maximal boundary length of faces. Since it contradicts the inequality of Theorem 1.5 for *any* density $d < \frac{1}{2}$, the 2–complex Y cannot be fillable by any random group. Hence the following proposition.

Proposition 2.12 *Let $(G_\ell(m,d))$ be a sequence of random groups with $m \geq 2$ generators at density d. For any $d < \frac{1}{2}$ and $K > 0$, a.a.s. every reduced van Kampen 2–complex of complexity K of $G_\ell(m,d)$ is collapsible to a graph.*

Consequently, a 2–complex with K faces that is homeomorphic to a *closed surface* of a fixed genus[5] g is not fillable by any random group, since a surface is not collapsible and the complexity is bounded by a number depending only on K and g.

3 Phase transition for the existence of van Kampen 2–complexes

In this section, we work on the proof of Theorem 1.6.

Motivation and a counterexample Let $(G_\ell(m,d))$ be a sequence of random groups at density d. We are interested in the converse of Theorem 1.5 without the reduction part: if a 2–complex Y_ℓ with bounded complexity satisfies the inequality

$$|Y_\ell^{(1)}| \geq (1-d+s)|Y_\ell|\ell$$

with some $s > 0$, does there exist a face labeling by relators and an edge labeling by generators, so that Y_ℓ becomes a reduced van Kampen 2–complex of $G_\ell(m,d)$?

The motivation for this question comes from the well-known phase transition at density $d = \frac{1}{2}\lambda$, mentioned in [Gromov 1993, page 274]: if $d < \frac{1}{2}\lambda$ then a.a.s. $G_\ell(m,d)$ has the $C'(\lambda)$ small cancellation condition;

[4] In the original context [Whitehead 1939], the removal of an isolated edge is also an elementary collapse, and a 2–complex is *collapsible* if it can be collapsed to a point.

[5] Note that the genus g need to be fixed, otherwise by Calegari and Walker's result [2015] there exists a closed surface (see Remark 2.5).

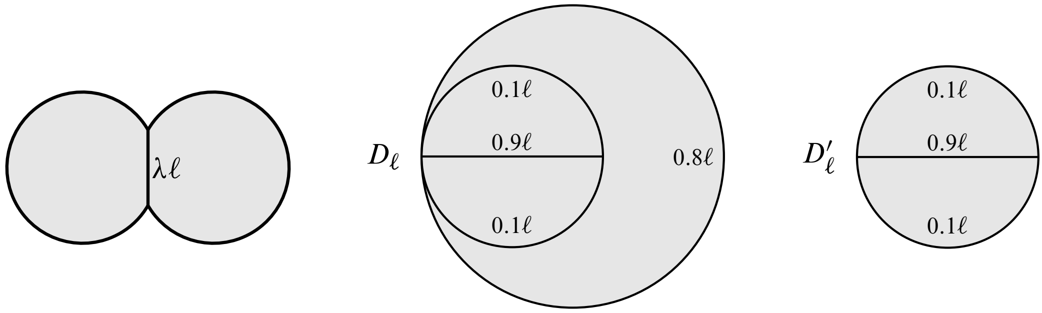


Figure 3: Left: a van Kampen diagram denying the $C'(\lambda)$ condition. Middle and right: A 2–complex that satisfies the isoperimetric inequality with a sub-2–complex that does not.

while if $d > \frac{1}{2}\lambda$ then a.a.s. $G_\ell(m,d)$ does not have $C'(\lambda)$. The first assertion is a simple application of Theorem 1.4. For the second assertion, we need to show that a.a.s. there exists a van Kampen 2–complex D of $G_\ell(m,d)$ with exactly 2 faces of boundary length ℓ, sharing a common path of length at least $\lambda\ell$ (Figure 3, left).

The first detailed proof of such an existence is given in [Bassino et al. 2020, Theorem 2.1], using an analog of the probabilistic pigeonhole principle. Another proof is given in [Tsai 2022, Theorem 1.4]. An intuitive explanation using the "dimension reasoning" is given in [Ollivier 2005, page 30]: The dimension of the set of couples $R_\ell \times R_\ell$ is $2d\ell$. Sharing a common subword of length L imposes L equations, so the "dimension" of the set of couples of relators sharing a common subword of length $\lambda\ell$ is $2d\ell - \lambda\ell$. If $d > \lambda/2$, then there will exist such a couple because the dimension will be positive. However, this argument is not true for *any* 2–complex in general. Here is a counterexample:

At density $d = 0.4$, let (D_ℓ) be a sequence of 2–complexes where D_ℓ is given in Figure 3, middle. The given inequality is satisfied because $|D_\ell^{(1)}| = 1.9\ell > 1.8\ell = (1-d)|D_\ell|\ell$. However, the subdiagram D'_ℓ (Figure 3, right) gives $|D_\ell'^{(1)}| = 1.1\ell < 1.2\ell = (1-d)|D'_\ell|\ell)$, which contradicts the isoperimetric inequality of Theorem 1.5 and cannot be a van Kampen diagram of $G_\ell(m,d)$.

3.1 Geometric form and critical density

Let us define the *geometric form* of 2–complexes and the *critical density* of a geometric form. To simplify the notation, for a 2–complex $Y = (V, E, F)$, we denote by $\mathrm{Edge}(Y)$ the set of geometric edges of Y and e instead of $\bar{e}$ for geometric edges.

Definition 3.1 A *geometric form* of 2–complexes is a couple (Y, λ) where $Y = (V, E, F)$ is a finite connected 2–complex without isolated edges, and λ is a length labeled on edges defined by

$$\lambda\colon \mathrm{Edge}(Y) \to \,]0,1], \quad e \mapsto \lambda_e,$$

such that for every face f of Y, the boundary length $|\partial f|$ is bounded by 1.

A sequence of 2–complexes (Y_ℓ) is said to be *of the geometric form* (Y,λ) if Y_ℓ is obtained from Y by dividing every edge e of Y into $\lfloor \lambda_e \ell \rfloor$ edges[6] of length 1.

A sequence of 2–complexes (Y_ℓ) is briefly said to be *of the same geometric form* if the geometric form (Y,λ) is not specified. Note that the boundary length of every face f of Y_ℓ is at most ℓ. If Z is a sub-2–complex of Y, we denote $Z \le Y$. By convention, if (Z_ℓ) is a sequence of 2–complexes of the geometric form $(Z,\lambda|_Z)$, we have $Z_\ell \le Y_\ell$ for any integer ℓ.

Definition 3.2 Let (Y,λ) be a geometric form of 2–complexes. The *density* of Y is

$$\operatorname{dens}(Y) := \frac{\sum_{e\in \operatorname{Edge}(Y)} \lambda_e}{|Y|}.$$

The *critical density* of Y is

$$\operatorname{dens}_c(Y) := \min_{Z\le Y}\{\operatorname{dens}(Z)\}.$$

The intuition of this definition can be found in Lemma 3.8: the density of Y is actually the density of all possible van Kampen 2–complexes that fill Y_ℓ.

Remark 3.3 Taking Definitions 3.2 and 3.1 together, we have

$$\operatorname{dens}(Y) = \frac{\sum_{e\in \operatorname{Edge}(Y)} \lambda_e}{|Y|} = \lim_{\ell\to\infty} \frac{\sum_{e\in \operatorname{Edge}(Y)} \lfloor \lambda_e \ell \rfloor}{|Y_\ell|\ell} = \lim_{\ell\to\infty} \frac{|Y_\ell^{(1)}|}{|Y_\ell|\ell}.$$

Hence, the condition "$\operatorname{dens}_c(Y) + d > 1$" is equivalent to the following statement: Given $s > 0$, for ℓ large enough, every sub-2–complex Z_ℓ of Y_ℓ satisfies

$$|Z_\ell^{(1)}| \ge (1-d+s)|Z_\ell|\ell.$$

This argument shows that the second assertion of Theorem 1.6 is equivalent to Corollary 1.7.

Proof of Theorem 1.6(i) We will use Theorem 1.5 without the reduction part. Let $(G_\ell(m,d))$ be a sequence of random groups with m generators at density d. Recall that a 2–complex Y_ℓ is said to be *fillable* by $G_\ell(m,d)$ if there exists a *reduced* van Kampen 2–complex of $G_\ell(m,d)$ whose underlying 2–complex is Y_ℓ.

Let (Y,λ) be a geometric form of 2–complexes with $\operatorname{dens}_c Y + d < 1$. Let (Y_ℓ) be a sequence of 2–complexes of the geometric form (Y,λ). We shall prove that a.a.s. the 2–complex Y_ℓ is *not* fillable by the random group $G_\ell(m,d)$. By the definition of critical density, there exists a sub-2–complex $Z \le Y$ satisfying $\operatorname{dens} Z + d < 1$. Let (Z_ℓ) be the sequence of 2–complexes of the geometric form $(Z,\lambda|_Z)$. We shall prove that a.a.s. Z_ℓ is not fillable by $G_\ell(m,d)$.

[6] We can replace $\lfloor \lambda_e \ell \rfloor$ by any function with $\lambda\ell + o(\ell)$ and slightly smaller than $\lambda\ell$. Note that the sum of edge lengths on every face boundary of Y_ℓ is at most ℓ

Let $\varepsilon > 0$ such that $\operatorname{dens} Z = 1 - d - 3\varepsilon$. By definition,

$$\lim_{\ell\to\infty} \frac{|Z_\ell^{(1)}|}{|Z_\ell|\ell} = 1 - d - 3\varepsilon,$$

so for ℓ large enough,

$$|Z_\ell^{(1)}| \le (1-d-2\varepsilon)|Z_\ell|\ell < (1-d-\varepsilon)|Z_\ell|\ell.$$

The complexity of Z_ℓ is

$$K = \max\Big\{|Z|, |Z^{(1)}|, \max\Big\{\frac{1}{\lambda_e} \;\Big|\; e \in \operatorname{Edge}(Z)\Big\}\Big\},$$

independent of ℓ. By Theorem 1.5 with ε and K given above, a.a.s. every van Kampen 2–complex Z_ℓ of $G_\ell(m,d)$ of complexity K should satisfy

$$|Z_\ell^{(1)}| \ge (1-d-\varepsilon)|Z_\ell|\ell.$$

Hence, a.a.s. the given 2–complex Z_ℓ is not fillable by $G_\ell(m,d)$, which implies that a.a.s. Y_ℓ is not fillable by $G_\ell(m,d)$. □

3.2 The multidimensional intersection formula for random subsets

To prove the second assertion of Theorem 1.6, we need the *multidimensional intersection formula* for random subsets with density, introduced in [Tsai 2022, Section 3].

Recall that B_ℓ is the set of cyclically reduced words of $X_m^{\pm} = \{x_1^{\pm}, \dots, x_m^{\pm}\}$ of length at most ℓ, and that $|B_\ell| = (2m-1)^{\ell+o(\ell)}$. Let $k \ge 1$ be an integer. Denote by $B_\ell^{(k)}$ the set of k–tuples of pairwise distinct relators $(r_1, \dots, r_k)$ in B_ℓ. Such notation can be used for any set or any random set.

Note that $|B_\ell^{(k)}| = (2m-1)^{k\ell+o(\ell)}$. Recall that a sequence of fixed subsets $(\mathcal{Y}_\ell)$ of the sequence $(B_\ell^{(k)})$ is called *densable with density* $\alpha \in \{-\infty\} \cup [0,1]$ if the sequence of real numbers $(\log_{|B_\ell^{(k)}|} |\mathcal{Y}_\ell|)$ converges to α (see [Gromov 1993, page 272; Tsai 2022, Definition 1.5]). That is to say, $|\mathcal{Y}_\ell| = (2m-1)^{\alpha k\ell+o(\ell)}$.

Definition 3.4 (self-intersection partition [Tsai 2022, Definition 3.4]) Let $(\mathcal{Y}_\ell)$ be a sequence of fixed subsets of the sequence $(B_\ell^{(k)})$. Let $0 \le i \le k$ be an integer. The i^{th} self-intersection of $\mathcal{Y}_\ell$ is

$$S_{i,\ell} := \{(x,y) \in \mathcal{Y}_\ell^2 \mid |x \cap y| = i\},$$

where $|x \cap y|$ is the number of common elements between the sets $x = (r_1, \dots, r_k)$ and $y = (r_1', \dots, r_k')$.

The family of subsets $\{S_{i,\ell} \mid 0 \le i \le k\}$ is a partition of $\mathcal{Y}_\ell^2$, called the *self-intersection partition* of $\mathcal{Y}_\ell$. Note that $(S_{i,\ell})_{\ell\in\mathbb{N}}$ is a sequence of subsets of the sequence $((B_\ell^{(k)})^2)_{\ell\in\mathbb{N}}$, with density smaller than $\operatorname{dens}_{((B_\ell^{(k)})^2)}(\mathcal{Y}_\ell^2) = \operatorname{dens}_{(B_\ell^{(k)})}(\mathcal{Y}_\ell)$.

Definition 3.5 (d–small self-intersection condition [Tsai 2022, Definition 3.5]) Let $(\mathcal{Y}_\ell)$ be a sequence of fixed subsets of $(B_\ell^{(k)})$ with density α. Let $S_{i,\ell}$ with $0 \le i \le k$ be its self-intersection partition. Let $d > 1-\alpha$. We say that $(\mathcal{Y}_\ell)$ satisfies the d–small self-intersection condition if, for every $1 \le i \le k-1$,

$$\operatorname{dens}_{((B_\ell^{(k)})^2)}(S_{i,\ell}) < \alpha - (1-d) \times \frac{i}{2k}.$$

Theorem 3.6 (multidimensional intersection formula [Tsai 2022, Theorem 3.6]) *Let (R_ℓ) be a sequence of permutation invariant random subsets of (B_ℓ) of density d. Let $(\mathcal{Y}_\ell)$ be a sequence of fixed subsets of $(B_\ell^{(k)})$ of density $\alpha > 1-d$. If $(\mathcal{Y}_\ell)$ satisfies the d–small self-intersection condition, then the sequence of random subsets $(\mathcal{Y}_\ell \cap R_\ell^{(k)})$ is densable with density $\alpha + d - 1$.*

In particular, a.a.s. the random subset $\mathcal{Y}_\ell \cap R_\ell^{(k)}$ of $B_\ell^{(k)}$ is not empty.

3.3 Proof of Theorem 1.6(ii)

Let (Y_ℓ) be a sequence of 2–complexes of the same geometric form (Y, λ) with k faces. In the following, we denote by $\mathcal{Y}_\ell$ the set of pairwise distinct relators in B_ℓ that fill Y_ℓ, which is a subset of $B_\ell^{(k)}$.

Let $(G_\ell(m,d))$ be a sequence of random groups at density d, defined by $G_\ell(m,d) = \langle X_m \mid R_\ell \rangle$, where (R_ℓ) is a sequence of random subsets with density d. The intersection $\mathcal{Y}_\ell \cap R_\ell^{(k)}$ is hence the set of k–tuples of pairwise distinct relators in R_ℓ that fill Y_ℓ. We want to prove that this intersection is not empty, so that Y_ℓ is fillable by $G_\ell(m,d)$. According to Theorem 3.6, it remains to prove that if $\operatorname{dens}_c Y > 1-d$, then the sequence $(\mathcal{Y}_\ell)$ is densable and satisfies the d–small self-intersection condition.

We will prove in Lemma 3.8 that $(\mathcal{Y}_\ell)$ is densable with density exactly $\operatorname{dens}(Y)$, and in Lemma 3.9 that it satisfies the d–small self-intersection condition.

Lemma 3.7 *Let $\overline{\mathcal{Y}_\ell}$ be the set of k–tuples of relators in B_ℓ that fill Y_ℓ, not necessarily pairwise distinct. If Y_ℓ is fillable by B_ℓ, then*

$$\operatorname{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}) = \operatorname{dens} Y.$$

Proof We shall estimate the number $|\overline{\mathcal{Y}_\ell}|$ by counting the number of labelings on edges of Y_ℓ that produce van Kampen 2–complexes with respect to all possible relators B_ℓ.

We start by filling edges in the neighborhoods of vertices that are originally vertices of the geometric form Y (before dividing). Consider the set of oriented edges of Y_ℓ starting at some vertex that is originally a vertex of Y before dividing. A *vertex labeling* is a labeling on these edges by $X_m^\pm$ that does not produce any reducible pair of edges on face boundaries: for every pair of different edges e_1, e_2 starting at the same vertex, if they are labeled by the same generator $x \in X_m^\pm$, then the path $e_1^{-1}e_2$ is not cyclically part of any face boundary loop. Since the 2–complex Y_ℓ is fillable, the set of vertex labelings is not empty. Denote by $C \ge 1$ the number of vertex labelings of Y_ℓ.

As $m \ge 2$ and $\lfloor \lambda_e \ell \rfloor \ge 3$ for ℓ large enough, if there exists a vertex labeling, then the other edges of Y_ℓ can be completed as a van Kampen 2–complex of B_ℓ, and the number C depends only on the geometric form Y.

To label the remaining $\lfloor \lambda \ell \rfloor - 2$ edges on the arc divided from the edge $e \in \mathrm{Edge}(Y)$, there are $2m-1$ choices for the first $\lfloor \lambda \ell \rfloor - 3$ edges, and $2m-2$ or $2m-1$ choices for the last edge. So

$$C \prod_{e \in \mathrm{Edge}(Y)} (2m-1)^{\lfloor \lambda_e \ell \rfloor - 3}(2m-2) \le |\overline{\mathcal{Y}_\ell}| \le C \prod_{e \in \mathrm{Edge}(Y)} (2m-1)^{\lfloor \lambda_e \ell \rfloor - 2}.$$

Recall that $k = |Y| = |Y_\ell|$ and that $|B_\ell^k| = (2m-1)^{k\ell + o(\ell)}$. We have

$$\mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}) = \frac{\sum_{e \in \mathrm{Edge}(Y)} \lambda_e}{|Y|} = \mathrm{dens}\, Y. \qquad \square$$

Lemma 3.8 *If* $\mathrm{dens}_c\, Y > \frac{1}{2}$ *and* Y_ℓ *is fillable by* B_ℓ*, then* $(\mathcal{Y}_\ell)$ *is densable in* $(B_\ell^{(k)})$ *and*

$$\mathrm{dens}_{(B_\ell^{(k)})}(\mathcal{Y}_\ell) = \mathrm{dens}\, Y.$$

Proof Suppose that $|Y| \ge 2$. The case $|Y| = 1$ is trivial. Let Z be a sub-2–complex of Y with exactly two faces f_1, f_2. As $\mathrm{dens}\, Z \ge \mathrm{dens}_c\, Y > \frac{1}{2}$, by Definition 3.2, we have

$$\sum_{e \in \mathrm{Edge}(Z)} \lambda_e > \tfrac{1}{2}|Z| = 1 \ge |\partial f_1|.$$

Let $\overline{\mathcal{Y}_\ell^Z}$ be the set of fillings of Y_ℓ by B_ℓ such that the two faces of Z are filled by the same relator. By the same arguments of the previous lemma,

$$|\overline{\mathcal{Y}_\ell^Z}| \le C(2m-1)^{|\partial f_1|} \prod_{e \in \mathrm{Edge}(Y) \setminus \mathrm{Edge}(Z)} (2m-1)^{\lfloor \lambda_e \ell \rfloor - 2},$$

so

$$\begin{aligned} \mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell^Z}) &\le \frac{1}{|Y|}\Bigg[\sum_{e \in \mathrm{Edge}(Y)} \lambda_e + \Bigg(|\partial f_1| - \sum_{e \in \mathrm{Edge}(Z)} \lambda_e\Bigg)\Bigg] \\ &< \frac{\sum_{e \in \mathrm{Edge}(Y)} \lambda_e}{|Y|} = \mathrm{dens}\, Y = \mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}). \end{aligned}$$

Knowing that

$$\mathcal{Y}_\ell = \overline{\mathcal{Y}_\ell} \Big\backslash \bigcup_{Z < Y,\ |Z| = 2} \overline{\mathcal{Y}_\ell^Z},$$

we have

$$|\overline{\mathcal{Y}_\ell}| - \sum_{Z < Y,\ |Z| = 2} |\overline{\mathcal{Y}_\ell^Z}| \le |\mathcal{Y}_\ell| \le |\overline{\mathcal{Y}_\ell}|.$$

There are $\binom{|Y|}{2}$ terms in the sum, in every term we have $\mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell^Z}) < \mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell})$, so (see [Tsai 2022, Propositions 2.7 and 2.8])

$$\mathrm{dens}_{(B_\ell^k)}(\mathcal{Y}_\ell) = \mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}).$$

Together with Lemma 3.7, we have $\mathrm{dens}_{(B_\ell^k)}(\mathcal{Y}_\ell) = \mathrm{dens}\, Y$. As $\mathrm{dens}_{(B_\ell^k)}(B_\ell^{(k)}) = 1$, we get

$$\mathrm{dens}_{(B_\ell^{(k)})}(\mathcal{Y}_\ell) = \mathrm{dens}\, Y. \qquad \square$$

Lemma 3.9 *Suppose that* $\operatorname{dens}_c Y > 1-d$. *Let* $S_{i,\ell}$ *be the* i^{th} *self-intersection of the set* $\mathcal{Y}_\ell$. *We have*

$$\operatorname{dens}_{((B_\ell^{(k)})^2)}(S_{i,\ell}) < \operatorname{dens} Y - (1-d) \times \frac{i}{2k}.$$

Proof Let Z, W be two sub-2–complexes of Y with $|Z| = |W| = i < k = |Y|$. Let $(Z_\ell), (W_\ell)$ be the corresponding sequences of 2–complexes of the geometric forms Z and W, respectively. Denote by $S_\ell(Z, W)$ the set of pairs of pairwise distinct fillings $((r_1, \dots, r_k), (r'_1, \dots, r'_k))$ of Y_ℓ by all possible relators B_ℓ such that, the i relators in the first filling $(r_1, \dots, r_k)$ corresponding to Z_ℓ are identical to the i relators in the second filling $(r'_1, \dots, r'_k)$ corresponding to W_ℓ, and that the remaining $2k-2i$ relators are pairwise different, not repeating the relators in Z_ℓ and W_ℓ.

Let us estimate the cardinality $|S_\ell(Z, W)|$. First, fill the k–tuple $(r_1, \dots, r_k)$ so the i relators in the next k–tuple $(r'_1, \dots, r'_k)$ corresponding to the sub-2–complex W_ℓ is determined. There are at most $i!$ choices for ordering these i relators. To fill the remaining $k-i$ relators in $(r'_1, \dots, r'_k)$, by the same arguments of Lemma 3.7, we get

$$|S_\ell(Z, W)| \le |\mathcal{Y}_\ell| \times i! \times C \prod_{e \in \operatorname{Edge}(Y) \setminus \operatorname{Edge}(W)} (2m-1)^{\lfloor \lambda_e \ell \rfloor - 2}.$$

Recall that the density of Y is defined by $(1/|Y|)\big(\sum_{e \in \operatorname{Edge}(Y)} \lambda_e\big)$, and that $\operatorname{dens} W \ge \operatorname{dens}_c Y > 1-d$ by Definition 3.2. Together with the hypothesis $\operatorname{dens}_c Y > 1-d$, we have

$$\begin{aligned}
\operatorname{dens}_{((B_\ell^{(k)})^2)}(S_\ell(Z, W)) &\le \frac{1}{2k}\bigg(\sum_{e \in \operatorname{Edge}(Y)} \lambda_e + \sum_{e \in \operatorname{Edge}(Y) \setminus \operatorname{Edge}(W)} \lambda_e\bigg) \\
&= \frac{1}{2k}\bigg(2 \sum_{e \in \operatorname{Edge}(Y)} \lambda_e - \sum_{e \in \operatorname{Edge}(W)} \lambda_e\bigg) \\
&= \operatorname{dens} Y - \frac{i}{2k} \operatorname{dens} W \\
&< \operatorname{dens} Y - \frac{i}{2k}(1-d).
\end{aligned}$$

Note that

$$S_{i,\ell} = \bigcup_{\substack{Z < Y,\ W < Y \\ |Z| = |W| = i}} S_\ell(Z, W).$$

It is a union of $\binom{k}{i}^2$ subsets of densities strictly smaller than $\operatorname{dens} Y - \frac{i}{2k}(1-d)$. According to [Tsai 2022, Proposition 2.7], we have

$$\operatorname{dens}_{((B_\ell^{(k)})^2)}(S_{i,\ell}) < \operatorname{dens} Y - \frac{i}{2k}(1-d). \qquad \square$$

This completes the proof of Theorem 1.6.

4 Phase transitions for small cancellation conditions

Let us recall small cancellation notions in [Lyndon and Schupp 1977, page 240]. A *piece* with respect to a set of relators is a cyclic subword that appears at least twice. A group presentation satisfies the $C'(\lambda)$ small cancellation condition for some $0 < \lambda < 1$ if the length of a piece is at most λ times the length of any relator in which it appears. It satisfies the $C(p)$ small cancellation condition for some integer $p \geq 2$ if no relator is a product of fewer than p pieces.

The $C'(\lambda)$ condition Let $(G_\ell(m,d))$ be a sequence of random groups at density d. It is known that there is a phase transition at density $d = \lambda/2$ for the $C'(\lambda)$ condition (see [Gromov 1993, page 274; Bassino et al. 2020, Theorem 2.1; Tsai 2022, Theorem 1.4]). We give here a much simpler proof using Theorem 1.6.

Proposition 4.1 *Let $0 < \lambda < 1$. Let $(G_\ell(m,d))$ be a sequence of random groups at density d. There is a phase transition at density $d = \lambda/2$:*

(i) *If $d < \lambda/2$, then a.a.s. $G_\ell(m,d)$ satisfies $C'(\lambda)$.*

(ii) *If $d > \lambda/2$, then a.a.s. $G_\ell(m,d)$ does not satisfy $C'(\lambda)$.*

Proof (i) Let us prove by contradiction. Suppose that a.a.s. $G_\ell(m,d)$ does not satisfy $C'(\lambda)$. That is to say, a.a.s. there exists a piece w that appears in relators r_1, r_2 with $|w| > \lambda|r_1|$. It is possible that $r_1 = r_2$, but the piece should be at different positions.

Construct a van Kampen diagram D by gluing two combinatorial disks with one face, labeled respectively by r_1 and r_2, along with the paths where the piece w appears (Figure 4, left). As $r_1 \neq r_2$ or $r_1 = r_2$ but the piece appears at different positions, we obtain a reduced van Kampen diagram. The diagram satisfies $|D^{(1)}| = |r_1| + |r_2| + |w| < \ell + \ell + \lambda\ell < (1 - \lambda/2)|D|\ell$, which contradicts Theorem 1.5.

(ii) Consider a geometric form Y with two faces sharing a common edge of length λ, the other two edges are of length $1-\lambda$ (Figure 4, right). We have dens $Y = \frac{1}{2}(2(1-\lambda)+\lambda) > 1-d$, and every sub-2–complex with one face is with density $1 > 1-d$. So $\mathrm{dens}_c\, Y > 1-d$.

Let (Y_ℓ) be a sequence of 2–complexes of the geometric form Y. By Theorem 1.6, a.a.s. Y_ℓ is fillable by $G_\ell(m,d)$, hence a.a.s. $G_\ell(m,d)$ does not satisfy $C'(\lambda)$. □

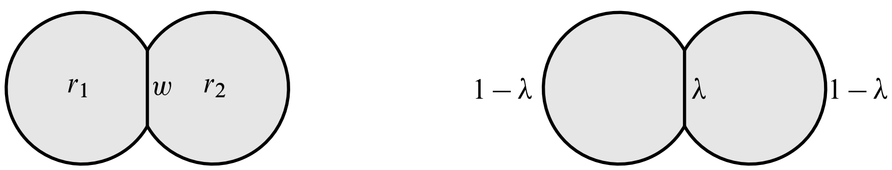


Figure 4: Left: a van Kampen 2–complex constructed from a $C'(\lambda)$ group. Right: the geometric form for the $C'(\lambda)$ condition.

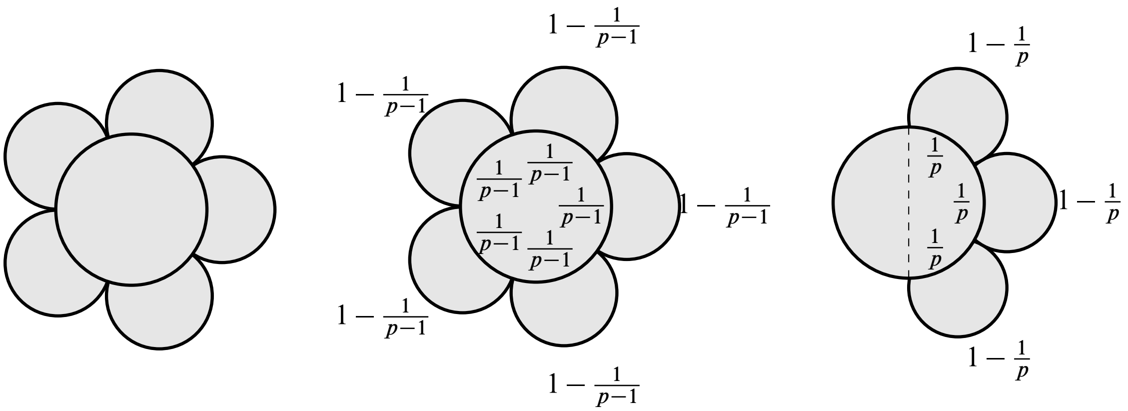


Figure 5: Left: a van Kampen 2–complex constructed from a non-$C(p)$ group. Middle: the geometric form for the $C(p)$ condition. Right: the geometric form for the $B(2p)$ condition.

The $C(p)$ condition We shall prove by Theorem 1.6 that for random groups with density, there is a phase transition at density $1/p$ for the $C(p)$ condition.

Proposition 4.2 *Let $p \geq 2$ be an integer. Let $(G_\ell(m,d))$ be a sequence of random groups at density d. There is a phase transition at density $1/p$:*

(i) *If $d < 1/p$, then a.a.s. $G_\ell(m,d)$ satisfies $C(p)$.*

(ii) *If $d > 1/p$, then a.a.s. $G_\ell(m,d)$ does not satisfy $C(p)$.*

Proof (i) Let us prove by contradiction. Suppose that a.a.s. $G_\ell(m,d)$ does not satisfy $C(p)$. That is to say, a.a.s. there is a relator that is a product of q pieces with $q \leq p-1$. By this relator we construct a reduced van Kampen diagram D with $q+1$ faces, one face is placed in the center, attached by the other q faces on the whole boundary, and there is no other attachments (Figure 5, left).

Observer that $|D| = q+1$ and $|D^{(1)}| \leq q\ell$ (sum of the boundary lengths of the outer q faces). Let $\varepsilon = (1/(q+1) - d)/2$, which is positive since $d < 1/p \leq 1/(q+1)$. We have

$$1 - d - \varepsilon = \frac{q}{q+1} + \varepsilon > \frac{q}{q+1}.$$

Hence $|D^{(1)}| < (1-d-\varepsilon)|D|\ell$, which contradicts Theorem 1.5.

(ii) Consider a geometric form Y with p faces, one of the faces is placed in the center, having $p-1$ edges of length $1/(p-1)$, such that every edge is attached by another face with two edges of lengths $1/(p-1)$ and $1-1/(p-1)$. There are no other attachments (Figure 5, middle).

The density of Y is $(p-1)/p > 1-d$. If Z is a sub-2–complex of Y not containing the center face, then $\operatorname{dens} Z = 1 > 1-d$. If Z contains the center face and $i \leq p$ other faces, then

$$\operatorname{dens} Z = \frac{1 + i(1 - 1/(p-1))}{i+1} > 1 - d.$$

So $\operatorname{dens}_c Y > 1-d$.

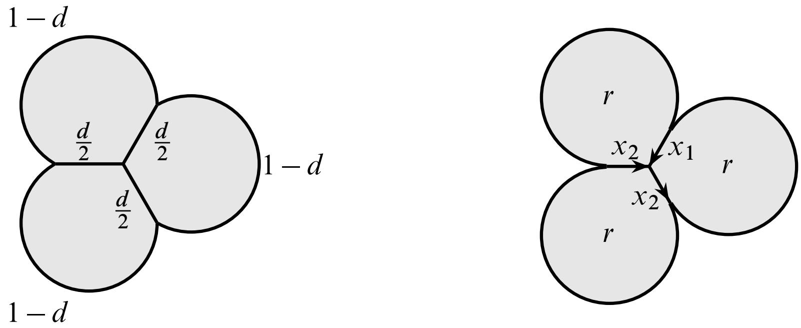


Figure 6: Left: the geometric form for the $T(q)$ condition. Right: a random relator r denying the $T(q)$ condition.

Let (Y_ℓ) be a sequence of 2–complexes of the geometric form Y. By Theorem 1.6, a.a.s. Y_ℓ is fillable by $G_\ell(m,d)$, hence a.a.s. $G_\ell(m,d)$ does not satisfy $C(p)$. □

The $B(2p)$ condition The same argument holds for the $B(2p)$ condition, introduced in [Ollivier and Wise 2011, Definition 1.7]: half of a relator cannot be the product of fewer than p pieces. One can construct a geometric form with $p+1$ faces, one of the faces is in the center, with half of its boundary attached by the other p faces, each with length $1/p$ (Figure 5, right). Its critical density is $\left(p+\frac{1}{2}\right)/(p+1)$, so a phase transition occurs at density $d=1/(2p+2)$.

Proposition 4.3 *Let $p\geq 1$ be an integer. Let $(G_\ell(m,d))$ be a sequence of random groups at density d. There is a phase transition at density $d=1/(2p+2)$:*

(i) *If $d<1/(2p+2)$, then a.a.s. $G_\ell(m,d)$ satisfies $B(2p)$.*

(ii) *If $d>1/(2p+2)$, then a.a.s. $G_\ell(m,d)$ does not satisfy $B(2p)$.* □

The $T(q)$ condition Recall that [Lyndon and Schupp 1977, page 241] a group presentation satisfies the $T(q)$ small cancellation condition for some $q\geq 4$ if, in every of its reduced van Kampen diagram, every vertex of valency at least 3 is actually of valency at least q.

Proposition 4.4 *For any density $0<d\leq 1$, a.a.s. $G_\ell(m,d)$ does not satisfy $T(4)$.*

Proof We shall construct a reduced van Kampen diagram with a vertex of valency exactly 3. Consider the geometric form Y with 3 faces sharing one common vertex, attaching to each other with common segments of length $d/2$ (Figure 6, left). The critical density of Y is $1-d/2>1-d$, so by Theorem 1.6, a.a.s. the random group $G_\ell(m,d)$ has a van Kampen diagram of the form Y. □

Remark 4.5 Proposition 4.4 holds for the few relator model. For example, for a one relator random group $\langle x_1,\dots,x_m \mid r\rangle$ with $m\geq 2$, a.a.s. (when $|r|\to\infty$) the three subwords x_1x_2, x_2^{-2} and $x_2x_1^{-1}$ appear in the random relator r at different places. By these subwords, we can construct a reduced van Kampen diagram with 3 faces that has a vertex of valency exactly 3 (Figure 6, right), denying the $T(4)$ small cancellation condition.

References

[Antoniuk et al. 2015] **S Antoniuk**, **T Łuczak**, **J Świątkowski**, *Random triangular groups at density $\frac{1}{3}$*, Compos. Math. 151 (2015) 167–178 MR Zbl

[Bassino et al. 2020] **F Bassino**, **C Nicaud**, **P Weil**, *Random presentations and random subgroups: a survey*, from "Complexity and randomness in group theory", de Gruyter, Berlin (2020) 45–76

[Calegari and Walker 2015] **D Calegari**, **A Walker**, *Random groups contain surface subgroups*, J. Amer. Math. Soc. 28 (2015) 383–419 MR Zbl

[Ghys 2004] **É Ghys**, *Groupes aléatoires (d'après Misha Gromov,...)*, from "Séminaire Bourbaki, 2002/2003", Astérisque 294, Soc. Math. France, Paris (2004) exposé 916, 173–204 MR Zbl

[Gromov 1993] **M Gromov**, *Finitely presented groups*, from "Geometric group theory, II: Asymptotic invariants of infinite groups", Lond. Math. Soc. Lect. Note Ser. 182, Cambridge Univ. Press (1993) 269–282 Zbl

[Gromov 2003] **M Gromov**, *Random walk in random groups*, Geom. Funct. Anal. 13 (2003) 73–146 MR Zbl

[Gruber and Mackay 2021] **D Gruber**, **J M Mackay**, *Random triangular Burnside groups*, Israel J. Math. 244 (2021) 75–94 MR Zbl

[Kapovich and Schupp 2008] **I Kapovich**, **P Schupp**, *On group-theoretic models of randomness and genericity*, Groups Geom. Dyn. 2 (2008) 383–404 MR Zbl

[Lyndon and Schupp 1977] **R C Lyndon**, **P E Schupp**, *Combinatorial group theory*, Ergebnisse der Math. 89, Springer (1977) MR Zbl

[Odrzygóźdź 2021] **T Odrzygóźdź**, *Nonplanar isoperimetric inequality for random groups*, preprint (2021) arXiv 2104.13903

[Ollivier 2004] **Y Ollivier**, *Sharp phase transition theorems for hyperbolicity of random groups*, Geom. Funct. Anal. 14 (2004) 595–679 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 2007] **Y Ollivier**, *Some small cancellation properties of random groups*, Int. J. Algebra Comput. 17 (2007) 37–51 MR Zbl

[Ollivier and Wise 2011] **Y Ollivier**, **D T Wise**, *Cubulating random groups at density less than $1/6$*, Trans. Amer. Math. Soc. 363 (2011) 4701–4733 MR Zbl

[Tsai 2022] **T-H Tsai**, *Density of random subsets and applications to group theory*, J. Comb. Algebra 6 (2022) 223–263 MR Zbl

[Whitehead 1939] **J H C Whitehead**, *Simplicial spaces, nuclei and m–groups*, Proc. Lond. Math. Soc. 45 (1939) 243–327 MR Zbl

Institut de Recherche Mathématique Avancée, Université de Strasbourg
Strasbourg, France

Current address: *Institut Camille Jordan, Université Claude Bernard Lyon 1*
Villeurbanne, France

tsai71517@gmail.com

Received: 23 November 2022 Revised: 4 September 2023

Algebraic & Geometric Topology **24**:7 (2024) 3919–3984
DOI: 10.2140/agt.2024.24.3919
Published: 9 December 2024

A qualitative description of the horoboundary of the Teichmüller metric

AITOR AZEMAR

Two commonly studied compactifications of Teichmüller spaces of finite type surfaces with respect to the Teichmüller metric are the horofunction and visual compactifications. We show that these two compactifications are related, by proving that the horofunction compactification is finer than the visual compactification. This allows us to use the straightforwardness of the visual compactification to obtain topological properties of the horofunction compactification. Among other things, we show that Busemann points of the Teichmüller metric are not dense within the horoboundary, answering a question of Liu and Su. We also show that the horoboundary of Teichmüller space is path connected, determine for which surfaces the horofunction compactification is isomorphic to the visual one and show that some horocycles diverge in the visual compactification based at some point. As an ingredient in one of the proofs we show that extremal length is not C^2 along some paths that are smooth with respect to the piecewise linear structure on measured foliations.

30F60, 32G15; 51F30

1 Introduction

The horofunction compactification of a metric space is defined in terms of the metric, so its properties are well aligned for studying the metric properties of the space. For example, all geodesic rays converge to points and isometries of the space can be extended to homeomorphisms of the compactification. This compactification was first introduced by Gromov [16] as a natural, general compactification, based on previous ideas of Busemann. The horofunction compactification has since found many applications; it was used to obtain asymptotic properties of random walks on weakly hyperbolic spaces by Maher and Tiozzo [29], to determine the isometry group of some Hilbert geometries by Lemmens and Walsh [25] and to obtain properties of quantum metric spaces by Rieffel [39]. The compactification is obtained by embedding the metric space X into the space $C(X)$ of continuous functions on X via the map $h: X \hookrightarrow C(X)$ defined by

$$h(p)(\cdot) = d(p,\cdot) - d(p,b), \tag{1}$$

where $b \in X$ is an arbitrarily chosen basepoint. As explained, for example, by Walsh [42, Section 2], if the space X is proper then h is an embedding, the closure of $h(X)$ is compact and the *horofunction compactification* of X is defined as the pair $(h, \overline{h(X)})$. By considering two functions equivalent if they differ by a constant one can show that the compactification does not depend on the basepoint b. While this compactification has been rather useful, it is sometimes hard to visualize, and there are not that many examples where the horofunction boundary is explicitly known. Some cases where the horofunction compactification is understood include Hadamard manifolds and some of their quotients, by Dal'bo, Peigné, and Sambusetti [8], as well as the Heisenberg group with the Carnot–Carathéodory metric, by Klein and Nicas [24], and Hilbert geometries, by Walsh [43].

On the other hand, for a proper, uniquely geodesic, straight metric space X (see Section 2 for definitions) the visual compactification based at some point $b \in X$ is defined by pasting the set of geodesic rays exiting b, denoted D_b, to the space X in such a way that a sequence $(x_n) \subset X$ converges to some ray $\gamma \in D_b$ if the distance $d(b, x_n)$ goes to infinity as $n \to \infty$, and the geodesic ray between b and x_n converges uniformly on compacts to γ. This compactification was introduced by Eberlein and O'Neil [10] as a generalization of the Poincaré disk model, and we give a brief introduction in Section 2. This compactification may depend on the basepoint b, which restricts its usefulness. It can even happen that isometries of X that move the basepoint cannot be extended continuously to the compactification, as Kerckhoff showed for Teichmüller spaces [23]. However, the visual compactification usually has a simple geometric interpretation. For example, for a Hadamard manifold, as well as for a Teichmüller space with the Teichmüller metric, this compactification is homeomorphic to a closed ball of the same dimension as the space, where the boundary of that ball is the space of geodesic rays based at b. In the context of Teichmüller spaces with the Teichmüller metric, the visual compactification is often called the Teichmüller compactification.

1.1 Horoboundary of proper, uniquely geodesic, straight metric spaces

To make this work as general as possible, we begin our analysis by using the aforementioned metric properties of the Teichmüller metric. The relationship between the horofunction compactification and the visual compactification is established by observing that, for such a metric space, a sequence converging to a point in the horofunction compactification also converges in the visual compactification. This allows us to build a continuous map Π_b from the horofunction compactification $\overline{h(X)}$ to the visual compactification

$X \cup D_b$, showing that the former is finer than the latter. In the context of Teichmüller spaces without boundary, the map Π_b coincides with the one defined by Liu and Shi [27, Definition 3.3]. We may denote this map as simply Π when the basepoint is not relevant to the discussion.

Given a geodesic γ, the path $\gamma(t)$ converges, as $t \to \infty$, to the *Busemann point* associated to γ in the horofunction compactification, which we denote by B_γ. As the map Π is defined in terms of sequences it follows that $\Pi(B_\gamma) = \gamma$. The existence of the map Π shows a strong relation between the horofunction and the visual compactification, which we state in the following result.

Theorem 1.1 *Let (X, d) be a proper, uniquely geodesic, straight metric space. For any basepoint $b \in X$, there is a continuous surjection Π from the horofunction compactification to the visual compactification based at b such that $\Pi(B_\gamma) = \gamma$ for every ray γ starting at b and $\Pi(h(p)) = p$ for every $p \in X$.*

In particular, the horofunction compactification of X is finer than the visual compactification of X based at any point.

Most of the subsequent results in the paper follow as applications of this theorem.

It is not the first time that a map such as Π appears in the literature. Similar maps have been found for δ–hyperbolic spaces by Webster and Winchester [45]. Walsh [43] defined such a map for Hilbert geometries, which satisfy the hypothesis of the theorem whenever there are no coplanar noncollinear segments in the boundary of the convex set, as shown by de la Harpe [17, Proposition 2].

The map Π does not induce a fiber bundle, as its fibers $\Pi^{-1}(\gamma)$ vary from points to higher dimensional sets (see Theorem 6.10). Still, Theorem 1.1 characterizes the horoboundary as the disjoint union of all the fibers $\Pi^{-1}(\gamma)$. Furthermore, our analysis of the topology of these fibers shows that they are path connected (see Proposition 3.11), which gives the following characterization of the connectivity of the horoboundary.

Proposition 1.2 *The horoboundary of a proper, uniquely geodesic straight metric space is connected if and only if its visual boundary based at some point (and hence, any) is connected.*

The *Busemann map* B from the visual compactification $X \cup D_b$ to the horofunction compactification is defined by setting $B(\gamma) = B_\gamma$ for each geodesic ray $\gamma \in D_b$ and $B(p) = h(p)$ for each $p \in X$. With this definition, the map satisfies $\Pi \circ B = \mathrm{id}$. As the next result shows, the continuity of this map is related with the topology of the horofunction compactification.

Proposition 1.3 *The visual compactification of a proper, uniquely geodesic, straight metric space based at some point is isomorphic to its horofunction compactification if and only if the Busemann map is continuous.*

The Busemann map is essentially the identity inside X, so the only possible points of discontinuity are at the boundary. It is therefore of interest to find a criterion for the continuity of B at the boundary, which turns out to give a criterion for when the fibers $\Pi^{-1}(\gamma)$ are singletons.

Proposition 1.4 *Let X be a proper, uniquely geodesic, straight metric space, $b \in X$ a basepoint and B the corresponding Busemann map. Furthermore, let γ be a geodesic ray based at b. Then the following three statements are equivalent*:

(1) *The Busemann map B restricted to the boundary is continuous at γ.*

(2) *The fiber $\Pi^{-1}(\gamma)$ is a singleton.*

(3) *The Busemann map B is continuous at γ.*

In other words, we have reduced the continuity of B to the continuity restricted to the boundary. This result can then be applied to different settings to obtain a more precise characterization. In the case of Teichmüller spaces, Proposition 1.4 can be used to get an explicit criterion for the continuity of the Busemann map in terms of the quadratic differentials associated to the geodesic rays, giving us a characterization of the fibers that are singletons.

1.2 Horoboundary of the Teichmüller metric

Many compactifications have been defined for Teichmüller space, such as Thurston's compactification, the visual compactification (also known as the Teichmüller compactification) and the Gardiner–Masur compactification. These compactifications play an important role in the study of mapping class groups and asymptotic aspects of Teichmüller space. See for example the articles by Thurston [41], Kerckhoff [23] or Ohshika [36]. The main reason multiple compactifications have been introduced is that each one has been designed with a certain application in mind.

Thurston's compactification takes the rather simple shape of a ball, upon which the mapping class groups acts as homeomorphisms. This facts make this compactification well suited for studying properties of the mapping class group. Indeed, Thurston's classification of the elements of the mapping class group relied on this compactification [41]. However, the Teichmüller metric is not directly related to the compactification, which results in some quirks when trying to use it to study the asymptotic geometry. For example, Lenzhen, Modami, and Rafi [26] prove that there exist geodesic rays with high-dimensional limit sets.

The visual compactification is defined directly using the metric, and takes the shape of a sphere where each point in the boundary has a clear geometric interpretation. This makes the compactification a good tool to interpret asymptotic geometric results. For example, Walsh [44, Theorem 7] has proven that all geodesic rays converge to points in the visual boundary. However, as proved by Kerckhoff [23], the action of the mapping class group does not extend continuously to this compactification, which implies that the compactification depends on the choice of basepoint. This fact limits the usability of the visual compactification.

The Gardiner–Masur compactification was initially defined to study the asymptotic properties of extremal lengths, following an analogous construction to that of the Thurston's compactification. It was later proved by Liu and Su [28] that this compactification is isomorphic to the horofunction compactification with respect to the Teichmüller metric, giving it a geometric meaning. Furthermore, the mapping class group extends continuously to the compactification. These two properties make the Gardiner–Masur compactification a good candidate to study asymptotic properties of the Teichmüller metric. However, as noted by Miyachi [32] and Liu and Su [28], there is a lack of information on the shape of this compactification. In this paper, we start working towards an understanding of the shape of this boundary.

Let S be a compact surface with (possibly empty) boundary and finitely many marked points, where we allow marked points to be on the boundary. Denote by $\mathcal{T}(S)$ its Teichmüller space equipped with the Teichmüller metric. Furthermore, for any quadratic differential q based at some basepoint $b \in \mathcal{T}(S)$, denote by $R(q;\cdot)$ the geodesic ray in $\mathcal{T}(S)$ starting at b in the direction q, and $V(q)$ the vertical foliation associated to q, see Section 4 for a quick introduction or the book by Farb and Margalit [11] for a more in-depth explanation of these concepts. Recall that a measured foliation is *indecomposable* if it is either a thickened curve, or a component with a transverse measure that cannot be expressed as the sum of two projectively distinct nonzero transverse measures. Furthermore, each measured foliation can be decomposed uniquely into finitely many indecomposable components (see Section 4.1 for detailed definitions). Walsh has shown the following characterization of the convergence of Busemann points in terms of the convergence of the associated quadratic differentials.

Theorem 1.5 (Walsh [44, Theorem 10]) *Let (q_n) be a sequence of unit area quadratic differentials based at $b \in \mathcal{T}(S)$. Then, $B_{R(q_n;\cdot)}$ converges to $B_{R(q;\cdot)}$ if and only if both of the following hold*:

(1) *(q_n) converges to q with respect to the L^1 norm on $T_b^*\mathcal{T}(S)$.*

(2) *For every subsequence $(G^n)_n$ of indecomposable measured foliations such that*, *for each $n \in \mathbb{N}$*, *G^n is a component of $V(q_n)$, we have that every limit point of G^n is indecomposable.*

While Walsh's proof is done in the context of surfaces without boundary, it can be easily extended to our setting. In view of this theorem, we say that a sequence of quadratic differentials (q_n) *converges strongly* to q if it satisfies the two conditions of Theorem 1.5. Furthermore, we say that q is *infusible* if every sequence of quadratic differentials converging to q converges strongly. By Proposition 1.4, a quadratic differential q is infusible if and only if the Busemann map is continuous at $R(q;\cdot)$. In Theorem 5.4, we derive a topological characterization of the vertical foliations of infusible quadratic differentials. This allows us to determine precisely which surfaces only admit infusible quadratic differentials, yielding the following result.

Theorem 1.6 *Let S be a compact surface of genus g with b_m and b_u boundary components with and without marked points respectively and p interior marked points. Then the horofunction compactification of $\mathcal{T}(S)$ is isomorphic to the visual compactification if and only if $3g + 2b_m + b_u + p \leq 4$.*

This result had been previously proven by Miyachi [32] for surfaces without boundary, that is, when $b_m = b_u = 0$. For the cases where we do not have an isomorphism Miyachi found non-Busemann points in the boundary. These points are in the closure of Busemann points, which prompted Liu and Su to ask the following question:

Question 1.7 (Liu and Su [28, Question 1.4.2]) *Is the set of Busemann points dense in the horofunction boundary?*

We give a negative answer to this question, summed up in the following result.

Theorem 1.8 *Let S be a closed surface of genus g with p marked points. Then the Busemann points are not dense in the horofunction boundary of $\mathcal{T}(S)$ whenever $3g + p \geq 5$.*

To achieve this result we use Liu and Su's [28] and Walsh's [44] characterization of the horofunction compactification as the Gardiner–Masur compactification. We use an equivalent but slightly different definition than usual for the Gardiner–Masur compactification, as the definition we use is more well suited for our computations, and more easily extendable to surfaces with boundary (see Section 4.4 for the precise definition). For each point in the horofunction compactification there is an associated real-valued function on the set of measured foliations. We show that the functions associated to elements in the closure of Busemann points are polynomials of degree 2 with respect to some variables (see Proposition 6.2 for the precise statement). We then show that the elements of the Gardiner–Masur boundary found by Fortier Bourque [13] do not satisfy that condition. The main ingredient for this last part of the reasoning is the following result, which shows that extremal length is not C^2 along certain smooth paths in $\mathcal{MF}$.

Theorem 1.9 *Let S be a closed surface of genus g with p marked points and empty boundary satisfying $3g + p \geq 5$. Then there is a point $X \in \mathcal{T}(S)$ and a path G_t, $t \in [0, t_0]$, in the space of measured foliations on X, smooth with respect to the canonical piecewise linear structure of the space of measured foliations, such that $\operatorname{Ext}(G_t)$ is not C^2.*

The canonical piecewise linear structure of the space of measured foliations was developed by Bonahon [3; 4; 5]. The first derivative of the extremal length along such a path was determined by Miyachi [33], so our proof is based on finding cases where Miyachi's expression is not C^1. This follows from an explicit computation, whose complication is greatly reduced by using previous estimates established by Markovic [30].

The relation between the Thurston compactification and the horofunction compactification was studied by Miyachi [34]. He proves that, while neither Thurston's nor the horofunction compactification is finer than the other, there is a bicontinuous map from the union of $\mathcal{T}(S)$ and uniquely ergodic foliations in Thurston's boundary to a subset of the horofunction compactification. Masur showed [31] that this result can be interpreted to say that these two compactifications are the same almost everywhere according to the Lebesgue measure on Thurston's boundary. The image of uniquely ergodic foliations by the bicontinuous map is the set of Busemann points associated to uniquely ergodic foliations. As we show in Theorem 7.5, this set is nowhere dense within the horoboundary. Hence the map defined by Miyachi does not show that these two are the same almost everywhere according to any strictly positive measure on the horoboundary. In fact, any attempt to extend the identity map from the interior of the Thurston compactification to the interior of the horoboundary compactification to a set of full measure within the Thurston compactification results in the same problem.

Corollary 1.10 *Let ν be any finite strictly positive measure on the horoboundary and let μ be the Lebesgue measure on the Thurston boundary. Furthermore, let ϕ be a map from the Thurston compactification to the horofunction compactification satisfying $\phi|_{\mathcal{T}(S)} = h$, where h is as in (1). Then there is no subset U of the Thurston boundary with full μ–measure such that ϕ is continuous at every point in U and $\phi(U)$ has full ν–measure.*

Under some smoothness assumptions satisfied by Teichmüller metric, we are able to use the maps Π_b to give an alternative definition of the horofunction compactification based on geometric notions. This definition characterizes the horofunction compactification as the reachable subset of the product of all visual compactifications obtained by choosing different basepoints (see Section 3.3 for details). Hence, the horofunction compactification can be interpreted as a collection of the asymptotic information provided by all visual compactifications. As a straightforward result of this alternative definition we get the following characterization of converging sequences in the horofunction compactification.

Corollary 1.11 *A sequence $(x_n) \subset \mathcal{T}(S)$ converges in the horofunction compactification if and only if the sequence converges in all the visual compactifications.*

Considering the horocycles diverging in the horofunction compactification found by Fortier Bourque [13] we get that there is some visual compactification in which these horocycles do not converge.

Corollary 1.12 *Let S be a closed surface of genus g with p marked points, such that $3g + p \geq 5$. There is a basepoint such that a horocycle diverges in the visual compactification based at that point.*

This contrasts with the behavior of Teichmüller rays, which converge in all visual compactifications (see [44, Theorem 7] by Walsh).

The structure of the horoboundary provided by Theorem 1.1, as well as the path-connectivity of the fibers, allows us to prove the following connectivity result.

Theorem 1.13 *The horoboundary of any Teichmüller space of real dimension at least 2 is path connected.*

Furthermore, we also prove that whenever the surface has empty boundary the map Π restricted to the horoboundary admits a section, while it only admits a section for surfaces of low complexity if the boundary is nonempty (see Theorem 8.1 for details).

Figure 1 shows a sketch of what we think the horoboundary looks like based on the results of this paper. The outer circle represents the section given by Theorem 8.1. Each line perpendicular to the sphere represents one of the fibers induced by the map Π, so it is associated with a unique Teichmüller ray starting at b. Note that while by Proposition 3.11 the fibers are path connected, by Theorem 6.10 they are bigger than segments in some cases. Furthermore, a priori they might not be contractible.

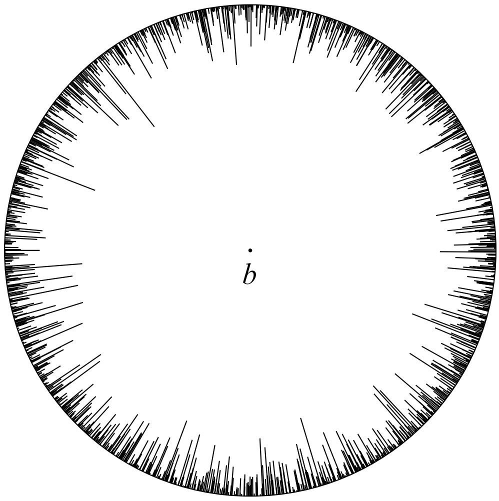


Figure 1: The shape of the horoboundary of the Teichmüller metric for surfaces without boundary.

The nearest point to the basepoint b of each fiber represents the Busemann point associated to the geodesic joining b to the fiber. This point could indeed be considered the nearest point to b from the fiber, as one can access it in a straight way, through a geodesic exiting b. On the other hand, the points in the outer circle represents the points associated to the section alluded to earlier. These can be accessed through a sequence of Busemann points whose associated fiber is a point, which can be considered as the most tangentially possible way to reach points in the boundary. Following a result by Masur [31], with respect to the measure on the fibers induced by the Lebesgue measure on the set of Teichmüller rays exiting b, almost all the fibers are actually points. As we shall see in Theorem 7.5 these points are nowhere dense in the boundary.

Note that there exist paths within the horoboundary connecting the fibers without passing through the section, and a priori there may be paths not represented in the sketch along which the fibers vary continuously. For surfaces for which the map Π does not admit a global section, a similar sketch could be drawn, although there would be no continuous global section in some cases. Hence, the outer circle would be broken at some places.

Finally, Liu and Su's and Walsh's characterization of the horofunction compactification as the Gardiner–Masur compactification can be used to translate some of these findings to results regarding the asymptotic value of extremal length functions. For example, we get the following estimate.

Theorem 1.14 *Let (q_n) be a sequence of unit quadratic differentials converging strongly to a unit quadratic differential q. Denote by G_j the components of the vertical foliation associated to q, and $H(q)$ the horizontal foliation. Then, for any $F \in \mathcal{MF}$ and sequence (t_n) of real values converging to positive infinity we have*

$$\lim_{n\to\infty} e^{-2t_n} \operatorname{Ext}_{R(q_n;t_n)}(F) = \sum_j \frac{i(G_j, F)^2}{i(G_j, H(q))}.$$

This generalizes a previous result by Walsh [44, Theorem 1], where the same is shown for q_n constant.

1.3 Outline of the paper and a note for the reader interested in surfaces without boundary

The paper is structured as follows. In Section 2 we introduce the necessary metric notions used in the paper. We follow in Section 3 by proving the results related to the more general metric setting, such as showing that the horofunction compactification is finer than the visual one. In Section 4 we give a short review of the necessary background on Teichmüller spaces. In Section 5 we determine which quadratic differentials are infusible, and find which surfaces admit infusible quadratic differentials, getting a proof of Theorem 1.6. In Section 6 we characterize the points in the closure of Busemann points, and get some bounds on the dimension of the fibers of the map Π. In Section 7 we show that Busemann points are not dense. In Section 8 we determine which surfaces result in the map Π having a section, and prove that the horoboundary is path connected. Finally, in Section 9 we use the previous results to obtain estimates regarding asymptotic values of extremal lengths.

Some of the most dense parts of this paper are due to the added complexity of considering surfaces with boundary. As such, the reader focused on surfaces with empty boundary might want to omit the corresponding sections on a first reading. One of the largest related parts starts after the remark following Theorem 5.4 and ends before the start of Section 5.2. The other sizable part starts with Proposition 8.3 and ends at the start of the proof of Theorem 1.13, where we note that the proof is significantly simpler in the case of surfaces without boundary.

Acknowledgments

The author would like to thank Maxime Fortier Bourque and Vaibhav Gadre for many helpful discussions and corrections.

2 Metric definitions

2.1 Compactifications

A compactification of a space serves, among other things, as a way of characterizing convergence to infinity. Formally, a *compactification* of a topological space X is a pair $(f, \overline{X})$, where $\overline{X}$ is a compact topological space and $f: X \to \overline{X}$ is an embedding such that $f(X)$ is dense in $\overline{X}$. The boundary of a compactification $\partial\overline{X} = \overline{X} - X$ describes the different ways of converging to infinity provided by that compactification. We shall usually identify the points in X with the ones in $\overline{X}$ via the map f, and say that a sequence $(x_n) \subset X$ converges in $\overline{X}$ if $f(x_n)$ converges.

A compactification (f_1, X_1) is *finer* than another one (f_2, X_2) if there exists a continuous map $\bar{f_2}: X_1 \to X_2$ such that $\bar{f_2} \circ f_1 = f_2$. Since $f_2(X)$ is dense in X_2, the continuous extension $\bar{f_2}$ is surjective. Furthermore, we can restrict the map $\bar{f_2}$ to the boundary to get a surjective map $\bar{f_2}|_{\partial X_1}: \partial X_1 \to \partial X_2$, which can be seen as a projection. Having a compactification finer than another ones means, from an intuitive point of view, that the finer compactification catalogs more ways of converging to infinity than the other one. Namely, any sequence in X converging in the finer compactification converges also in the coarser one, while the opposite may not be true.

We say that two compactifications are isomorphic if each one is finer than the other one. The following lemma found in [44, Lemma 17] coincides with the intuitive notion of finer compactifications.

Lemma 2.1 *Let (f_1, X_1) and (f_2, X_2) be two compactifications of a Hausdorff topological space X such that f_2 extends continuously to an injective map $\bar{f_2}\colon X_1 \to X_2$. Then the two compactifications are isomorphic.*

We will usually refer to the space $\overline{X}$ as the compactification when the embedding is clear from the context. Since the images of X by the embedding are dense, the extensions we get to compare the compactifications are unique. That is, we have the following result:

Lemma 2.2 *Let (f_1, X_1) and (f_2, X_2) be two compactifications of a Hausdorff topological space X such that X_1 is finer than X_2. Then the extension $\bar{f_2}\colon X_1 \to X_2$ is unique.*

Proof For any $x \in X$ we have $\bar{f_2}(f_1(x)) = f_2(x)$. Hence, the image of $\bar{f_2}$ is determined on a dense subset of X_1, so by continuity it is determined on X_1. □

2.2 Visual compactification of proper, uniquely geodesic, straight spaces

Let (X, d) be a metric space. We shall say that a map γ from an interval $I \subset \mathbb{R}$ to X is a *geodesic* if it is an isometric embedding, that is, if $d(\gamma(t), \gamma(s)) = |t-s|$. We shall consider two geodesics to be equal if their image is equal and have the same orientation. A space is *uniquely geodesic* if for any two distinct points $a, b \in X$ there is a unique geodesic starting at a and ending at b.

Furthermore, we say that the space is *proper* if the closed balls $D(x,r) = \{p \in X \mid d(p,x) \leq r\}$ are compact.

If geodesic segments can be extended uniquely, that is, if for any geodesic segment γ_1 there is a unique biinfinite geodesic γ_2 such that $\gamma_1 \cap \gamma_2 = \gamma_1$, we say that the space is *straight*.

Let then X be a proper, uniquely geodesic, straight space and let D_b be the set of infinite geodesic rays starting at b, with the topology given by uniform convergence on compact sets. Furthermore, denote by $S_b^1 = \{x \in X \mid d(x,b) = 1\}$ the sphere of radius 1 around b.

Lemma 2.3 *The map from D_b to S_b^1 defined by sending $\gamma \in D_b$ to $\gamma(1)$ is a homeomorphism.*

Proof Since the topology on D_b is given by uniform convergence on compact sets, the point $\gamma(1)$ varies continuously with respect to γ.

On the other hand, since the space is straight and has unique geodesics, given any point $a \in S_b^1$ there is a unique geodesic ray starting at b and passing through a. This is the inverse to the map obtained by evaluating the geodesics. To see that the relation is continuous we consider a sequence $(a_n) \subset S_b^1$ converging to some a, and denote by (γ_n) and γ the associated geodesics. Assume γ_n does not converge to γ. Then we have a subsequence without γ as an accumulation point. For any $t > 0$, the geodesic segments $\gamma|_{[0,t]}$ are contained in the ball of radius t, which is compact, as X is proper. As these are geodesics we have equicontinuity, so by Arzelà–Ascoli we can take a subsequence converging uniformly to some path γ'. Since the distance function is continuous, γ' is a geodesic. Furthermore, $\gamma'(1) = \lim_{n\to\infty} \gamma_n(1) = \lim_{n\to\infty} a_n = a$. By uniqueness of geodesics, γ' and γ are equal when restricted to $[0,1]$, which by straightness implies they are equal. Hence, γ_n converges to γ uniformly on the compact $[0,t]$. □

Following a similar reasoning it is possible to show the following, still under the same hypotheses on X.

Lemma 2.4 *The space X is homeomorphic to $D_b \times [0,\infty)/D_b \times \{0\}$.*

Proof We define the map $C\colon D_b \times [0,\infty)/D_b \times \{0\} \to X$ given by $C(\theta, r) = \theta(r)$. This is well defined, as $C(\theta, 0) = b$ for any $\theta \in D_b$. Furthermore, this is a bijection, since for every $x \in X - \{b\}$ there is a unique geodesic ray from b to x. The map is continuous, as the topology on D_b is given by uniform convergence on compact sets. To see that the inverse is continuous consider a sequence $a_n \in X$ converging to some $a \in X$. If $a = b$, then $d(a_n, b) \to 0$, so we have continuity. Otherwise we let $r_n = d(a_n, b)$ and $r = d(a, b)$. We have $r_n \to r$, so denoting (γ_n) and γ the unique geodesic in D_b such that $\gamma_n(r_n) = a_n$ and $\gamma(r) = a$ and applying Arzelà–Ascoli's theorem in the same way as in Lemma 2.3, we have that γ_n converges to γ. □

The space $D_b \times [0,\infty)/D_b \times \{0\}$ can be included into the compact space $D_b \times [0,\infty]/D_b \times \{0\}$, which can be written as $(D_b \times [0,\infty)/D_b \times \{0\}) \cup D_b \times \{\infty\}$. Using the homeomorphism from Lemma 2.4, we can use this inclusion to give a compact topology on the space $X \cup D_b$. The *visual compactification* is defined as the pair $(i, X \cup D_b)$, where i is the inclusion $i\colon X \to X \cup D_b$ and the topology on the space $X \cup D_b$ is the one we just defined. We shall denote $X \cup D_b$ as $\overline{X}{}^v_b$, or $\overline{X}{}^v$ when the basepoint is not relevant to the discussion.

2.3 Horofunction compactification

The second compactification that will play a part in this paper is slightly more involved and difficult to visualize.

Let X be a proper, uniquely geodesic, straight metric space. Given a basepoint $b \in X$, one can embed X into the space of continuous functions from X to $\mathbb{R}$ via the map $h\colon X \to C(X)$ defined by

$$h(x)(\cdot) := d(x,\cdot) - d(x,b).$$

The topology given to $C(X)$ is that of uniform convergence on compact sets. The map h is indeed continuous, as the distance function is continuous. Furthermore, h is injective, as $h(x)$ has a strict global minimum at x. It can also be proven that since X is proper, h is an embedding. For more details about this construction see [42, Section 2]. Furthermore, the properness of X implies it is second countable, so the closure of $h(X)$ is compact, Hausdorff and second countable. We shall denote the closure of $h(X)$ on $C(X)$ as $\overline{X}{}^h$. The *horofunction compactification* is defined as the pair $(h, \overline{X}{}^h)$. We call the set $\partial\overline{X}{}^h = \overline{X}{}^h - X$ the *horofunction boundary* or *horoboundary*, and we call its members *horofunctions*. If we want to specify the chosen basepoint we write $\overline{X}^h_b$. However, it is possible to see that quotienting the compactification by letting $f \sim g$ whenever the difference is constant we get an isomorphic compactification, showing that the horofunction compactification does not depend on the basepoint.

Usually the easier points to identify in the horoboundary are the Busemann points. These are the ones that can be reached as a limit along almost geodesics, which is a slight weakening of the notion of geodesic by allowing an additive constant approaching 0. That is, a path $\gamma\colon [0,\infty) \to X$ is an *almost geodesic* if for each $\varepsilon > 0$,

$$\left| d(\gamma(0),\gamma(s)) + d(\gamma(s),\gamma(t)) - t \right| < \varepsilon$$

for all s and t large enough, with $s \le t$. Rieffel [39] proved that every almost geodesic converges to a limit in $\partial\overline{X}^h$. A horofunction is called a *Busemann point* if there exists an almost geodesic converging to it. We shall denote the Busemann point associated in this way to the almost geodesic γ by B_γ.

3 Horofunction compactification of proper, uniquely geodesic, straight metric spaces

3.1 The relation between the horofunction compactification and the visual compactification

Fix a uniquely geodesic, proper and straight metric space (X,d) and a basepoint $b \in X$. We will assume X satisfies these hypotheses through this section. For each geodesic ray $\gamma \in \partial\overline{X}^v$ starting at b there is an associated Busemann point $B_\gamma \in \partial\overline{X}^h$. We can extend this map to all the visual compactification by setting it as the identification with the map h on X given by the horofunction compactification. That is, we define the *Busemann map* $B\colon \overline{X}^v \to \overline{X}^h$ by setting $B(\gamma) = B_\gamma$ for $\gamma \in \partial\overline{X}^v$ and $B(x) = h(x)$ for $x \in X$. The relevance of this map can be seen with the following result.

Lemma 3.1 *The visual compactification $(i, \overline{X}^v)$ is finer than the horofunction compactification $(h, \overline{X}^h)$ if and only if the map B is continuous.*

Proof We have that $B(i(x)) = h(x)$, so B is an extension of h to $\overline{X}^v$. Hence, if B is continuous, then the visual compactification is finer than the horofunction compactification.

On the other hand, if the visual compactification is finer than the horofunction compactification, then we have a continuous map $f\colon \overline{X}^v \to \overline{X}^h$. For every $x \in X$, we have $f(i(x)) = h(x) = B(i(x))$. Furthermore, for any ray γ starting at the basepoint we have $f(\gamma) = \lim_{t\to\infty} f(i(\gamma(t))) = \lim h(\gamma(t)) = B(\gamma)$. Hence, $B = f$, and B is continuous. □

In general, the Busemann map may not be surjective nor continuous. However, we have the following.

Proposition 3.2 *For a proper, uniquely geodesic, straight metric space (X,d) the Busemann map is injective.*

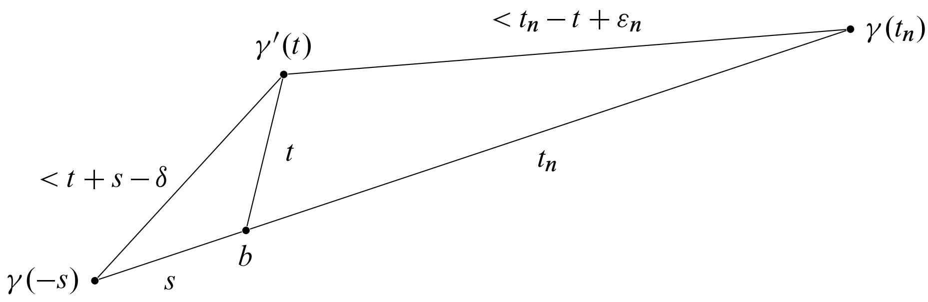


Figure 2: The triangles involved in the proof of Proposition 3.2.

Proof For each $x \in X$, the associated function $h(x)$ has a global minimum at x, while B_γ is unbounded below for every $\gamma \in \partial \overline{X}^v$. Hence, in the interior of $\overline{X}^v$ the map is injective and $B(X) \cap B(\partial \overline{X}^v) = \varnothing$. Assume we have $\gamma, \gamma' \in \partial \overline{X}^v$ such that $\gamma \neq \gamma'$ and $B(\gamma) = B(\gamma') = \xi$. Then, for a given sequence $t_n \to \infty$ we have $\lim_{n\to\infty} h(\gamma(t_n)) = \lim_{n\to\infty} h(\gamma'(t_n)) = \xi$. For any $t \in \mathbb{R}$ and any n such that $t_n > t$ we have

$$h(\gamma(t_n))(\gamma(t)) = d(\gamma(t_n), \gamma(t)) - d(\gamma(t_n), \gamma(0)) = t_n - t - t_n = -t,$$

and similarly for γ'. Hence $\xi(\gamma(t)) = \xi(\gamma'(t)) = -t$ for all t.

Fix now a $t > 0$. We have

$$-t = \xi(\gamma'(t)) = \lim_{n\to\infty} \bigl(d(\gamma'(t), \gamma(t_n)) - d(b, \gamma(t_n))\bigr) = \lim_{n\to\infty} \bigl(d(\gamma'(t), \gamma(t_n)) - t_n\bigr).$$

That is, there is a sequence ε_n with $\varepsilon_n \to 0$ such that

$$t_n - t + \varepsilon_n \geq d(\gamma'(t), \gamma(t_n)) \geq t_n - t - \varepsilon_n.$$

for every n.

By straightness we can extend γ in the negative direction towards $\gamma(-s)$ for some $s > 0$. We shall now show that the geodesic γ does not minimize the distance between $\gamma(-s)$ and $\gamma(t_n)$ for n big enough. Since the space is straight, the geodesic segment between $\gamma(-s)$ and b can be extended uniquely, so concatenating it with the segment between b and $\gamma'(t)$ does not result in a geodesic. Hence, the distance between $\gamma'(t)$ and $\gamma(-s)$ is strictly smaller than $s + t$. That is, there is some $\delta > 0$ such that $d(\gamma(-s), \gamma'(t)) < t + s - \delta$. As shown in Figure 2 we get a path going from $\gamma(-s)$, to $\gamma(t_n)$, passing through $\gamma'(t)$ that has length less than $t + s - \delta + t_n - t + \varepsilon_n = t_n + s - \delta + \varepsilon_n$. Hence, taking n big enough so that $\varepsilon_n < \delta$ we get that the geodesic segment between $\gamma(-s)$ and $\gamma(t_n)$ is not minimizing. This is a contradiction, from which we conclude that $\gamma = \gamma'$. Therefore, B is injective. □

Hence, given a Busemann point ξ in $B(\partial \overline{X}^v)$ we have a unique associated geodesic ray $\gamma \in \partial \overline{X}^v$ such that $\xi(\gamma(t)) = -t$ for all t. Our next aim is to build a similar relation for all other horofunctions. Our approach is similar to the one used by Walsh in [44, Section 7].

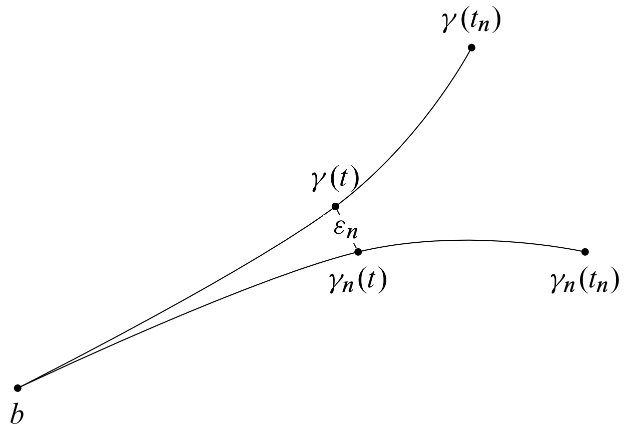


Figure 3: In the proof of Lemma 3.3, γ_n converges to γ, so $\gamma_n(t)$ converges to $\gamma(t)$, and hence the distance between $\gamma_n(t_n)$ and $\gamma_n(t)$ gets arbitrarily close to the distance between $\gamma_n(t_n)$ and $\gamma(t)$.

We say that a geodesic γ is an *optimal geodesic* for a certain horofunction $\xi \in \overline{X}^h$ if $\xi(\gamma(t)) - \xi(\gamma(0)) = -t$ for all $t \in \mathbb{R}$. We shall now see that each function in the horoboundary has at least one optimal geodesic.

Lemma 3.3 *Let X be a proper, uniquely geodesic, straight metric space and let $\xi \in \partial\overline{X}^h$ be a horofunction. Suppose that $(x_n) \subset X$ converges to ξ, with $x_n = \gamma_n(t_n)$, $\gamma_n \in \partial\overline{X}^v$ and (γ_n) converging to γ as $n \to \infty$. Then $\xi(\gamma(t)) = -t$ for every $t \in \mathbb{R}$. That is, $\gamma(t)$ is an optimal geodesic for ξ.*

Proof Fix t. We have that

$$\xi(\gamma(t)) = \lim_{n\to\infty} \big(d(\gamma(t), \gamma_n(t_n)) - d(b, \gamma_n(t_n))\big) = \lim_{n\to\infty} \big(d(\gamma(t), \gamma_n(t_n)) - t_n\big).$$

As n goes to infinity, γ_n converges to γ. Hence by the given topology on the visual boundary, the maps $\gamma_n(\cdot)$ converge uniformly on compact sets to the geodesic $\gamma(\cdot)$. In particular, denoting $d(\gamma(t), \gamma_n(t)) = \varepsilon_n$ we have $\varepsilon_n \to 0$. We get then Figure 3, so by the triangle inequality,

$$|d(\gamma(t), \gamma_n(t_n)) - (t_n - t)| = \big|d(\gamma(t), \gamma_n(t_n)) - d(\gamma_n(t), \gamma_n(t_n))\big| \le \varepsilon_n,$$

and so $\xi(\gamma(t)) = -t$. □

Since $\partial\overline{X}^v$ is compact, for any horofunction $\xi \in \partial\overline{X}^h$ and sequence $(x_n) \subset X$ converging to ξ we can take a subsequence such that the hypotheses of Lemma 3.3 are satisfied, so each $\xi \in \partial\overline{X}^v$ does have at least one optimal geodesic.

If ξ has another optimal geodesic γ' with $\gamma'(0) = \gamma(0)$ we have at least two geodesics along which $\xi(\gamma(t)) = \xi(\gamma'(t)) = -t$ for all t. Following a reasoning similar to the one in the proof of Proposition 3.2, we get a contradiction. This time, however, we have to be a bit more careful about the distances, as instead of two fixed rays we have a fixed ray and a sequence converging to a distinct fixed ray.

Proposition 3.4 *Let $\xi \in \partial \overline{X}^h$ and $b \in X$. Then there is a unique optimal geodesic for ξ passing through b.*

Proof Let $(x_n) = (\gamma_n(t_n))$ be a sequence converging to ξ, with $(\gamma_n) \subset \partial \overline{X}^v$, and take a subsequence such that γ_n converges to some geodesic γ. By Lemma 3.3, γ is an optimal geodesic. Assume that we have a different optimal geodesic γ' passing through b.

Using that $h(\gamma_n(t_n))$ converges pointwise to ξ, we have

$$-t = \xi(\gamma'(t)) = \lim_{n\to\infty} \bigl(d(\gamma'(t), \gamma_n(t_n)) - d(b, \gamma_n(t_n))\bigr) = \lim_{n\to\infty} \bigl(d(\gamma'(t), \gamma_n(t_n)) - t_n\bigr).$$

Hence, there is a sequence ε_n with $\varepsilon_n \to 0$ such that

$$t_n - t + \varepsilon_n \geq d(\gamma'(t), \gamma_n(t_n)) \geq t_n - t - \varepsilon_n.$$

We proceed by showing that for n big enough there is some $s > 0$ such that the geodesic γ_n does not minimize the distance between $\gamma_n(-s)$ and $\gamma_n(t_n)$. As in the proof of Proposition 3.2, by applying the triangle inequality between $\gamma'(t), \gamma(-s)$ and b we have $d(\gamma'(t), \gamma(-s)) < s + t$. Fix $s > 0$ and pick $\delta > 0$ such that $d(\gamma'(t), \gamma(-s)) < t + s - \delta$. Since γ_n converges to γ uniformly on compact sets, $\gamma_n(-s)$ converges to $\gamma(-s)$. Hence, $d(\gamma'(t), \gamma_n(-s))$ converges to $d(\gamma'(t), \gamma(-s))$. Then for n big enough we have $d(\gamma'(t), \gamma_n(-s)) < t + s - \delta$. Consider then n big enough so that $\varepsilon_n \leq \delta/2$ as well. The triangle between $\gamma'(t), \gamma_n(-s)$ and $\gamma_n(t_n)$ gives

$$d(\gamma_n(-s), \gamma_n(t_n)) \leq d(\gamma_n(-s), \gamma'(t)) + d(\gamma'(t), \gamma_n(t_n)) < (t + s - \delta) + (t_n - t + \varepsilon_n) < t_n + s.$$

This is a contradiction, which proves the uniqueness of γ. □

Given a basepoint $b \in X$ we can now define a map $\Pi_b \colon \overline{X}^h \to \overline{X}^v_b$ by sending any $\xi \in \partial \overline{X}^h$ to the unique optimal geodesic γ of ξ with $\gamma(0) = b$, and by sending $h(x)$ to x for any $x \in X$. This map is indeed an extension of the relation we had established for Busemann points in $\mathcal{B}(\partial \overline{X}^v)$, since if $\xi = B(\gamma)$ for $\gamma \in D_b$ then γ is an optimal geodesic of ξ, giving us $\Pi_b(B(\gamma)) = \gamma$.

We will often write Π instead of Π_b whenever the basepoint is not relevant to the discussion. To prove that Π is continuous, we first have to see the following result.

Proposition 3.5 *Let $(x_n) \subset X$ be a sequence converging to $\xi \in \partial \overline{X}^h$. Then, (x_n) has a unique accumulation point in the visual compactification. Further, this accumulation point depends only on ξ.*

Proof Since $\partial \overline{X}^v$ is compact, (x_n) has accumulation points in the visual compactification. If (x_n) has two accumulation points we can take two subsequences converging to two different geodesics, which by Lemma 3.3 are optimal geodesics, contradicting Proposition 3.4.

If there is another sequence (y_n) converging to ξ with a different accumulation point the result follows by merging both sequences and repeating the reasoning. □

Hence, Π can be alternatively defined by sending any $\xi \in \partial \overline{X}^h$ to the unique accumulation point in $\overline{X}^v$ of the sequences converging to ξ in $\overline{X}^h$, and by sending $h(x)$ to x for any $x \in X$. By Proposition 3.5, this definition is equivalent to the previous one.

By this second definition of the map Π, we see how it is mostly related to the convergence of sequences, so using a diagonal sequence argument we can prove its continuity.

Proposition 3.6 *The map Π is continuous.*

Proof Take a sequence $(\xi_n) \subset \overline{X}^h$ converging to ξ. If $\xi \in h(X)$ we have that, as $h(X)$ is open, $\xi_n \in h(X)$ for n big enough. Hence, $\Pi(\xi_n) = h^{-1}(\xi_n)$, which converges to $h^{-1}(\xi)$, as h is a homeomorphism with its image.

If $\xi \in \partial \overline{X}^h$ we split the sequence into two subsequences, one contained in $h(X)$ and one contained in $\partial \overline{X}^h$. The one contained in $h(X)$ converges to ξ, so by definition of Π and we have $\Pi(\xi) = \lim_{n\to\infty} h^{-1}(\xi_n)$.

Assume then that $(\xi_n) \subset \partial \overline{X}^h$ converges to ξ. We want to see that $\gamma_n = \Pi(\xi_n)$ converges to $\gamma = \Pi(\xi)$. For each ξ_n we can take a sequence $\big(h(\gamma_n^m(t_n^m))\big)_m$ converging, as $m \to \infty$ to ξ_n. By Proposition 3.5 the sequence $\gamma_n^m(t_n^m)$ converges to γ_n. Let γ' be an accumulation point of γ_n. Take a convergent subsequence of γ_n converging to γ', and relabel it as γ_n. Let (V_n) be a nested sequence of open neighborhoods of ξ in $\overline{X}^h$ such that $\xi_n \in V_n$ and $\bigcap_n V_n = \{\xi\}$ and let (W_n) be a nested sequence of open neighborhoods of γ' in $\overline{X}^v$ such that $\gamma_n \in W_n$ and $\bigcap_n W_n = \{\gamma'\}$. We can take such sequences of sets, as both spaces are metrizable.

For each n, there exists $m(n)$ big enough so that $\gamma_n^{m(n)} \in W_n$ and $h(\gamma_n^{m(n)}(t_n^{m(n)})) \in V_n$. By the first condition on $m(n)$, we have that $\gamma_n^{m(n)}$ converges to γ'. By the second condition, $h(\gamma_n^{m(n)}(t_n^{m(n)}))$ converges to ξ, so by the definition of Π and Proposition 3.5 the sequence $\gamma_n^{m(n)}$ converges to $\Pi(\xi) = \gamma$. Hence, $\gamma = \gamma'$, so the only accumulation point of (γ_n) is γ and by compactness of $\partial \overline{X}^v$ the sequence (γ_n) converges to γ. □

By combining Propositions 3.5 and 3.6 we get that Π is the map announced at the introduction, giving us a proof of Theorem 1.1. As mentioned in the introduction, this map shows that the horofunction compactification is finer than the visual compactification. By using the Busemann map to insert the visual boundary inside the horoboundary, we can consider the map Π as a projection.

One straightforward consequence of the continuity of Π_b is as follows.

Corollary 3.7 *Let γ be a geodesic ray, not necessarily starting at the basepoint $b \in X$. Then, γ converges in the visual compactification of X based at b.*

Proof The ray γ converges in the horofunction compactification to B_γ. Since Π_b is continuous, the ray also converges in the visual compactification based at b to $\Pi_b(B_\gamma)$. □

For Teichmüller spaces with the Teichmüller metric this result was first proved by Walsh [44, Theorem 7].

By Lemma 3.1, the visual compactification is finer than the horofunction compactification if and only if the Busemann map is continuous. Hence, since the horofunction compactification is always finer than the visual compactification, we obtain an isomorphism whenever this is the case, resulting in Proposition 1.3.

3.2 The fiber structure

To get a better picture of the shape of the horoboundary we shall study the shape of the preimages of the projection Π restricted to the boundary. That is, for a given point γ in the visual boundary we are interested in finding out information about the fiber $\Pi^{-1}(\gamma)$. We first prove the following lemma, which we will use to get bounds on the values of $\Pi^{-1}(\gamma)$.

Lemma 3.8 *Fix a geodesic ray $\gamma \in \partial \overline{X}^v$ and $p \in X$ not in the biinfinite extension of the geodesic ray γ. Then, the function $h(\gamma(\cdot))(p)$, with domain $[0,\infty)$, is strictly decreasing.*

Proof Take $t, s \geq 0$ with $s < t$. By the triangle inequality we have

$$d(\gamma(t), p) \leq d(\gamma(s), p) + d(\gamma(t), \gamma(s)) = d(\gamma(s), p) + t - s.$$

Further, we have strict inequality, as equality would give us two different paths with the same length between $\gamma(t)$ and p, with one of them being geodesic. Hence,

$$h(\gamma(t))(p) = d(\gamma(t), p) - d(\gamma(t), b) < d(\gamma(s), p) + t - s - t = h(\gamma(s))(p). \qquad \square$$

The set $C(X)$ can be partially ordered by saying that $f \geq g$ whenever $f(x) \geq g(x)$ for all $x \in X$. If $f \geq g$ and $f \neq g$ then we write $f > g$. If $p = \gamma(r)$ for some r and $s < t$ we have $h(\gamma(s))(p) = h(\gamma(t))(p) = -r$ for $r \leq s$ and $-s = h(\gamma(s))(p) > h(\gamma(t))(p) = -\min(r, t)$ otherwise. Hence, adding the previous lemma we have $h(\gamma(s)) > h(\gamma(t))$ whenever $s < t$. By attempting to extend this relation to the horofunction boundary we get that Busemann points are maximal in their fibers.

Proposition 3.9 *Let $\gamma \in \partial \overline{X}^v$ and $\xi \in \Pi^{-1}(\gamma)$. Then, $\xi \leq B(\gamma)$.*

Proof Choose any sequence $(x_n) \subset X$ such that $h(x_n)$ converges to ξ. Since $\xi \in \Pi^{-1}(\gamma)$ the sequence (x_n) converges to γ in $\overline{X}^v$, so we can write $x_n = \gamma_n(t_n)$ with t_n converging to infinity and γ_n converging to γ.

Fix $p \in X$ and let $\varepsilon > 0$. Denote $s_n = \sup\{t : d(\gamma(t), \gamma_n(t)) < \varepsilon \text{ and } t < t_n\}$. The geodesics γ_n converge to γ uniformly on compact sets, so $s_n \to \infty$ as $n \to \infty$. Hence, by definition of the Busemann point and since $d(\gamma_n(s_n), \gamma(s_n)) < \varepsilon$,

$$B_\gamma(p) = \lim_{n\to\infty} h(\gamma(s_n))(p) \geq \limsup_{n\to\infty} h(\gamma_n(s_n))(p) - 2\varepsilon.$$

Furthermore, $s_n \leq t_n$, so by Lemma 3.8,

$$\xi(p) = \lim_{n\to\infty} h(\gamma_n(t_n))(p) \leq \limsup_{n\to\infty} h(\gamma_n(s_n))(p) \leq B_\gamma(p) + 2\varepsilon.$$

Since ε can be arbitrarily small we get the proposition. $\square$

While it might not be possible to get a similar unique minimum in each fiber, we can get the following result.

Proposition 3.10 *Let $\gamma \in \partial \overline{X}^v$ and $\xi \in \Pi^{-1}(\gamma)$. Furthermore, let $(x_n) \subset X$ be a sequence converging to ξ with $x_n = \gamma_n(t_n)$. For any p, define $\eta(p) = \liminf_{n\to\infty} B(\gamma_n)(p)$. Then $\xi \geq \eta$.*

Proof The proof follows a similar reasoning as the last one.

Fix $p \in X$, choose a subsequence so $B(\gamma_n)(p)$ converges to $\eta(p)$ and let (ε_m) be a sequence of positive numbers converging to 0. For each ε_m, take $n(m)$ big enough so that $B(\gamma_{n(m)})(p) \geq \eta(p) - \varepsilon_m$. Further, take s_m bigger than $t_{n(m)}$, and big enough so that

$$h(\gamma_{n(m)}(s_m))(p) \geq B(\gamma_{n(m)})(p) - \varepsilon_m.$$

Such an s_m always exists by the definition of $B(\gamma_{n(m)})$. In particular, we have that

$$\liminf_{m\to\infty} h(\gamma_{n(m)}(s_m))(p) \geq \eta(p).$$

By Lemma 3.8 we have

$$\xi(p) = \lim_{m\to\infty} h(\gamma_{n(m)}(t_{n(m)}))(p) \geq \liminf_{m\to\infty} h(\gamma_{n(m)}(s_m))(p) \geq \eta(p). \qquad \square$$

The intuition one might get from these propositions is that approaching γ "through the boundary", that is, through the furthest way possible from the interior of X, gives a lower bound on the possible values of approaching through other angles, and approaching γ in a straight way, that is, through the geodesic, gives an upper bound. Hence, when these two ways of approaching γ are the same, every other possible angle of approach should also yield the same limit. Following this reasoning we get our next result, announced in the introduction.

Proposition 1.4 *Let X be a proper, uniquely geodesic, straight metric space, $b \in X$ a basepoint and B the corresponding Busemann map. Furthermore, let γ be a geodesic ray based at b. Then the following three statements are equivalent*:

(1) *The Busemann map B restricted to the boundary is continuous at γ.*
(2) *The fiber $\Pi^{-1}(\gamma)$ is a singleton.*
(3) *The Busemann map B is continuous at γ.*

Proof (1) $\Longrightarrow$ (2) Take $\xi \in \Pi^{-1}(\gamma)$. By Proposition 3.9 we have $\xi \leq B(\gamma)$. Since B is continuous at γ when restricted to the boundary we have that for any $\gamma_n \to \gamma$ the horofunctions $B(\gamma_n)$ converge to $B(\gamma)$. Hence, by Proposition 3.10, $\xi \geq B(\gamma)$, so $\xi = B(\gamma)$ and we have (2).

(2) $\Longrightarrow$ (3) Take then any $(x_n) \subset \overline{X}^v$ converging to γ, consider the sequence $(B(x_n)) \subset \overline{X}^h$ and let η be an accumulation point. By the definition of Π we have $\eta \in \Pi^{-1}(\gamma)$, so $\eta = B(\gamma)$ since we assumed that $\Pi^{-1}(\gamma)$ is a singleton. This shows that B is continuous at γ.

Finally, it is clear that (3) $\Longrightarrow$ (1). $\square$

The relation obtained in Lemma 3.8 can be exploited further. Indeed, trying to carry it to the boundary in a more delicate manner we can see that the fibers are path connected.

Proposition 3.11 *Let $\gamma \in \partial \overline{X}^v$. For any $\xi \in \Pi^{-1}(\gamma)$ there exists a continuous path from $B(\gamma)$ to ξ contained in $\Pi^{-1}(\gamma)$.*

Proof Take a sequence $(x_n) \subset X$ converging to ξ in the horofunction compactification, and write $x_n = \gamma_n(u_n)$. As we have seen in the proof of Proposition 3.10, we can take a sequence $(l_n) \subset \mathbb{R}$ with $\gamma_n(l_n)$ converging to B_γ such that $l_n < u_n$ for all n. For each n we have a path $\tilde{\alpha}^n(t)$ connecting $\gamma_n(l_n)$ and $\gamma_n(u_n)$ by setting $\tilde{\alpha}^n(t) = \gamma_n(tu_n + (1-t)l_n)$ for $t \in [0,1]$. We would like to carry this path to the limit, getting a path between ξ and $B(q)$. However, directly taking such a limit might result in some discontinuities, so we have to choose a parametrization carefully.

To find a good parametrization we shall use a certain functional as a control. We want the functional to carry discontinuities and strict increases in the path of functions to discontinuities and strict increases in the value of the functional. Since X is proper, it is separable, so let $(p_i)_{i\in\mathbb{N}}$ be a countable dense set in X. We define the functional $I\colon \overline{X}^h \to \mathbb{R}$ given by

$$I(f) = \sum_{i\in\mathbb{N}} \frac{f(p_i)}{2^i d(b,p_i)}.$$

Since $|f(x)| \le d(b,x)$ for all $f \in \overline{X}^h$, the summation in the definition of $I(f)$ is absolutely convergent, so $I(f)$ is defined, finite, continuous with respect to f, and for any two $f, g \in \overline{X}^h$ we have $I(f+g) = I(f) + I(g)$. Furthermore, since (p_n) is dense and we are taking continuous functions, we have that the functional translates strict inequalities. That is, $f > g$ implies $I(f) > I(g)$. Hence, if $I(f) = 0$ and $f \ge 0$ we have $f = 0$.

We define then the function $F_n(t) = I(h(\gamma_n(t)))$. By continuity of I this function is continuous, and by Lemma 3.8 it is strictly decreasing with respect to t. That is, we have continuous strictly decreasing functions $F_n\colon [l_n, u_n] \to [F_n(u_n), F_n(l_n)]$. Hence, we can define implicitly the continuous parametrizations $s_n\colon [0,1] \to [l_n, u_n]$ by taking the unique value $s_n(t)$ such that

$$F_n(s_n(t)) = (1-t)F_n(l_n) + tF_n(u_n).$$

Denote the $F_n(s_n(t))$ as $E_n(t)$. By the continuity of I we have that $E_n(t)$ converges to $(1-t)I(B_\gamma) + tI(\xi)$ as $n \to \infty$, which we denote by $E(t)$.

Take now a countable dense set $(t^k)_{k\in\mathbb{N}} \subset [0,1]$ containing 0 and 1. We are now ready to start defining the path $\alpha\colon [0,1] \to \Pi^{-1}(\gamma)$, and we begin defining it for the dense set (t^k). For $k = 1$ we define $\alpha(t^1)$ as an accumulation point of $h\big(\gamma_n(s_n(t^1))\big)$. Denote $(\gamma_{m^1(n)})$ the subsequence of γ_n such that $h\big(\gamma_{m^1(n)}(s_{m^1(n)}(t^1))\big)$ converges to $\alpha(t^1)$. Define inductively $\alpha(t^k)$ and $(\gamma_{m^k(n)})$ by taking an accumulation point and a corresponding converging subsequence of $h\big(\gamma_{m^{k-1}(n)}(s_{m^{k-1}(n)}(t^k))\big)$. By the continuity of I we have

$$I(\alpha(t^k)) = \lim_{n\to\infty} \big(F_{m^k(n)}(s_{m^k(n)}(t^k))\big) = E(t^k).$$

For each pair $i > j$ we have that $m^i(n)$ is a subsequence of $m^j(n)$, so $h\big(\gamma_{m^i(n)}(s_{m^i(n)}(t^j))\big)$ converges to $\alpha(t^j)$. Assume $t^i > t^j$. By Lemma 3.8 we have that $h\big(\gamma_{m^i(n)}(s_{m^i(n)}(t^i))\big) < h\big(\gamma_{m^i(n)}(s_{m^i(n)}(t^j))\big)$, so $\alpha(t^i) \le \alpha(t^j)$.

We now have to prove that the definition we have given for α on (t^k) can be extended continuously to $[0,1]$. Fix any $t \notin (t^k)$ and take a subsequence of t^k, labeled t^{k_n}, such that $t^{k_n} \to t$. We shall now see that $\alpha(t^{k_n})$ converges to a function which does not depend on the chosen subsequence, and define $\alpha(t)$ as that limit. We can split and reorder the sequence (t^{k_n}) into (t_n^+) and (t_n^-) satisfying $t_n^+ > t_{n+1}^+ > t > t_{n+1}^- > t_n^-$. The associated $\alpha(t_n^\pm)$ are ordered, so for any $p \in X$ the sequence $\alpha(t_n^\pm)(p)$ is an increasing (or decreasing) sequence of values in $\mathbb{R}$, bounded above (or below) by $\alpha(0)(p)$ (or $\alpha(1)(p)$). Hence, both sequences converge pointwise, which implies uniform convergence on compact sets, as these functions are 1–Lipschitz. Furthermore, these limits do not depend on the chosen sequence, since if we had any other we could intercalate them and the sequences would still converge. Denote then α^+ the limit associated to t_n^+, and α^- the limit associated to t_n^-. Since $\alpha(t_n^+) < \alpha(t_m^-)$ for all n, m we have $\alpha^+ \le \alpha^-$. For each $\alpha(t^k)$ we have $I(\alpha(t^k)) = E(t^k)$. Hence by the continuity of I we have that

$$I(\alpha^+) = E(t) = I(\alpha^-).$$

That is, we have

$$I(\alpha^- - \alpha^+) = 0.$$

Since α^- and α^+ are continuous and $\alpha^- - \alpha^+ \ge 0$ we have $\alpha^- = \alpha^+$. We thus define $\alpha(t)$ to be either one. The same reasoning shows that α is continuous. □

We would like to remark that several choices where made in the proof of the previous lemma, and the obtained path may not be unique.

We can use the previous result to observe that the horoboundary is connected if and only if the visual boundary is connected.

Proof of Proposition 1.2 Assume that the visual boundary is not connected. Then we have $U, V \subset \partial \overline{X}^v$ nonempty and open such that $U \cap V = \varnothing$ and $U \cup V = \partial \overline{X}^v$. As Π is continuous, the sets $\Pi^{-1}(U)$ and $\Pi^{-1}(V)$ are open, so the horoboundary is not connected.

For the other implication, assume that the visual boundary is connected while the horoboundary is not connected. Then we have $U, V \subset \partial \overline{X}^h$ nonempty and open such that $U \cap V = \varnothing$ and $U \cup V = \partial \overline{X}^h$. Since fibers are path connected, each of them is contained in only one of U or V, so $\Pi(U)$ and $\Pi(V)$ are disjoint. Since $U \cup V = \partial \overline{X}^h$ we have $\Pi(U) \cup \Pi(V) = \partial \overline{X}^v$, and since both U and V are nonempty, so are the images. Hence, both images cannot be open at the same time, as $\partial \overline{X}^v$ is connected. Therefore, these sets cannot be both closed. Assume $\Pi(U)$ is not closed. We then have a sequence $(\gamma_n) \subset \Pi(U)$ converging to a point in $\Pi(V)$. Again, since $U \cup V = \partial \overline{X}^h$, we have that $U = \Pi^{-1}\Pi(U)$ and $V = \Pi^{-1}\Pi(V)$. Hence, any lift of the sequence (γ_n) to $\Pi^{-1}\Pi(U)$ is contained in U and, since $\partial \overline{X}^h$ is compact, has accumulation points which, by the continuity of the projection map, are be contained in $\Pi^{-1}\Pi(V) = V$. Hence, U is not closed and we get a contradiction. □

3.3 An alternative definition of the horofunction compactification

Under what a priori seem to be more restrictive hypotheses on the space X it is possible to characterize the horofunction compactification as a subset of the product of all of its visual compactifications. We detail the construction in this section.

The new extra hypotheses are both related to the differentiability of the distance function. We say a that a uniquely geodesic metric space X is C^1 *along geodesics* if given a point $p \in X$ and a geodesic segment γ that does not intersect p, the distance function $d(\gamma(t), p)$ is first differentiable and the value of the derivative depends continuously on both t and p. Furthermore, the space X has *constant distance variation* if for any two distinct geodesics γ, η with $\gamma(0) = \eta(0)$ we have either

$$\frac{d}{dt} d(\gamma(t), \eta(s))\Big|_{t=0} = \frac{d}{dt} d(\gamma(t), \eta(1))\Big|_{t=0} \tag{2}$$

for all $s > 0$, or $\frac{d}{dt} d(\gamma(t), \eta(s))\big|_{t=0}$ does not exist for any $s > 0$.

Many commonly studied metric spaces have constant distance variation. For example, spaces with bounded curvature, either above of below, have constant distance variation, as explained in the book by Burago, Burago, and Ivanov [6, Section 4]. Importantly to our case, Teichmüller spaces with the Teichmüller distance satisfy both hypotheses. Earle [9] proved that the distance function is C^1 by providing a formula for its derivative. Applying the formula to (2) we get that the derivative depends only on the tangential vector to γ at 0 and the unit area quadratic differential associated to η at 0, so we also have constant distance variation. Furthermore, Teichmüller spaces with the Teichmüller distance are also straight and proper, so the results from this section can be applied to them.

Consider the product of all the possible visual compactifications obtained by changing the basepoint,

$$E = \prod_{b \in X} \overline{X}^v_b,$$

with the usual product topology. See the book by Munkres [35, Chapters 2.19 and 5.37] for some background on infinite products of topological spaces. Denote by π_b the projection from E to $\overline{X}^v_b$. By definition of the product topology, the diagonal inclusion $i \colon X \hookrightarrow E$ such that by $\pi_b(i(x)) = x$ for every $x, b \in X$ is continuous, and has continuous inverse restricted to $i(X)$ given by π_b. Hence, $i(X)$ is homeomorphic to X. That is, i is an embedding. Furthermore, by Tychonoff's theorem the product is compact, as each factor of the product is compact. Hence the closure $\overline{i(X)}$, which we shall denote by $\overline{X}^V$, is compact. The pair $(i, \overline{X}^V)$ is then a compactification of X, which tracks the information given by the visual boundary at each point. That is, a sequence in X converges in the topology of $\overline{X}^V$ if and only if it converges for every possible visual compactification $\overline{X}^v_b$. The main interest of this compactification comes from the following result.

Theorem 3.12 *Let X be a proper, uniquely geodesic, straight metric space which is C^1 along geodesics and has constant distance variation. Then $(i, \overline{X}^V)$ is isomorphic to $(h, \overline{X}^h)$.*

Denote by Π_b the continuous map from $\overline{X}^h$ to $\overline{X}^v_b$ given by Theorem 1.1. The isomorphism between $\overline{X}^h$ and $\overline{X}^V$ is defined by recording the value of each possible Π_b within $\overline{X}^V$. That is, we define $\widetilde{\Pi}\colon \overline{X}^h \to \overline{X}^V$ in such a way that $\pi_b \circ \widetilde{\Pi} := \Pi_b$ for each $b \in X$. The only property required to prove that $\widetilde{\Pi}$ is an isomorphism not following directly from previous results is the injectivity. By Proposition 3.4 we know that if $f \in \Pi_b^{-1}(\gamma)$ then γ is an optimal geodesic of f. That is, $f(\gamma(t)) - f(\gamma(s)) = -(t-s)$. Hence, if $f, g \in \Pi_b^{-1}(\gamma)$, then they differ by a constant along the geodesic γ. If f and g are horofunctions in the preimage of a point by $\widetilde{\Pi}$, then they differ by a constant along infinitely many geodesics, which cover X. However, the constant might depend on the geodesic, so we need a way to connect these constants. We proceed by strengthening Proposition 3.4 to show that any two functions in $\Pi_b^{-1}(\gamma)$ also have the same directional derivatives at points in γ, which allows us to connect the geodesics. Precisely, we prove the following.

Proposition 3.13 *Let X be a proper, uniquely geodesic, straight metric space which is C^1 along geodesics and has constant distance variation. Furthermore, let γ be a geodesic ray starting at b, and let α be a geodesic starting at some point on γ. Then, $\frac{d}{dt} f \circ \alpha(t)\big|_{t=0}$ exists and its value is the same for all $f \in \Pi_b^{-1}(\gamma)$.*

Proof For any $b' \in \gamma$ we have that γ is an optimal geodesic of f passing through b'. Denoting $\gamma_{b'}$ the geodesic ray starting at b' with the same biinfinite extension as γ we have that $f \in \Pi_{b'}^{-1}(\gamma_{b'})$, by Proposition 3.4. Hence, we can assume that $\alpha(0) = b$ by changing the basepoint if necessary. Let x_n be a sequence converging to f. Furthermore, let η_t^n be the geodesic from $\alpha(t)$ to x_n and $g_n(t)$ be the value of $\frac{d}{ds} h(x_n) \circ \alpha(s)\big|_{s=t}$. By the definition of the map h we have $g_n(t) = \frac{d}{ds} d(\alpha(s), x_n)\big|_{s=t}$. By the constant distant variation we have $g_n(t) = \frac{d}{ds} d(\alpha(s), \eta_t^n(1))\big|_{s=t}$, which since X is C^1 along geodesics depends continuously on $\eta_t^n(1)$ and t.

By Proposition 3.5 the geodesics η_t^n converge as $n \to \infty$ to some geodesics η_t, so $\eta_t^n(1)$ converges to $\eta_t(1)$. Since the space is C^1 along geodesics, the value of $\frac{d}{ds} d(\alpha(s), \eta_t^n(1))\big|_{s=t}$ depends continuously on $\eta_t^n(1)$, and so g_n converges pointwise to $g(t) = \frac{d}{ds} d(\alpha(s), \eta_t(1))\big|_{s=t}$.

Take some $\delta > 0$ and assume the convergence is not uniform on $[-\delta, \delta]$. Then there is some $\varepsilon > 0$ such that for each n there is at least one $t_n \in [-\delta, \delta]$ such that $|g_n(t_n) - g(t_n)| > \varepsilon$. Since $[-\delta, \delta]$ is compact we can take a converging subsequence such that t_n converges to some $T \in [-\delta, \delta]$. Hence, the point $\eta_{t_n}^n(1)$ does not converge to $\eta_T(1)$, so by properness of X we can take a subsequence such that $\eta_{t_n}^n(1)$ converges to some $p \in X$ different from $\eta_T(1)$. Let β be the geodesic starting at $\alpha(T)$ passing through p. The geodesics $\eta_{t_n}^n$ converge uniformly to β, and $\beta \neq \eta_T$. For any fixed $t > 0$ we have, following the same reasoning than in the proof of Proposition 3.5,

$$f(\beta(t)) - f(\beta(0)) = \lim_{n\to\infty} d(x_n, \beta(t)) - d(x_n, \beta(0)) = -t.$$

Hence, β is an optimal geodesic of f passing through $\alpha(T)$. However, $f \in \Pi_{\alpha(T)}^{-1}(\eta_T)$, so η_T is also an optimal geodesic passing through $\alpha(T)$, contradicting Proposition 3.4.

Hence, the convergence of $(h(x_n)\circ\alpha)' = g_n$ to g is uniform on $[-\delta,\delta]$. Therefore, f is differentiable and $f'(0) = g(0) = \frac{d}{ds} d(\alpha(s),\gamma(1))\big|_{s=0}$, which is the same for all $f \in \Pi^{-1}(\gamma)$. □

Proof of Theorem 3.12 Each Π_b is continuous, so by the definition of the product topology the map $\widetilde{\Pi}$ is continuous. Hence, by Lemma 2.1 to see that $\widetilde{\Pi}$ is an isomorphism it is enough to show that $\widetilde{\Pi}$ is injective. Let $f, g \in \overline{X}^h$ be such that $\widetilde{\Pi}(f) = \widetilde{\Pi}(g)$. If there is some $b \in X$ such that $\pi_b \circ \widetilde{\Pi}(f) \in X$ then $f = h(\pi_b \circ \widetilde{\Pi}(f)) = g$. Assume then $\pi_b \circ \widetilde{\Pi}(f) \in \partial \overline{X}^v_b$ for all $b \in X$. By Proposition 3.13 they have the same directional derivatives at every point. Let α be a geodesic from a fixed basepoint b to any other point. We have $(f\circ\alpha)' = (g\circ\alpha)'$, so $f - g$ is constant along α, and hence everywhere, since any point can be connected to b by a geodesic. Hence, f and g are the same horofunctions. □

By the definition of the convergence in the product topology, this characterization gives us the following equivalence for the convergence to points in the horoboundary.

Corollary 3.14 *Let X be a proper, uniquely geodesic, straight metric space, C^1 along geodesics and with constant distance variation. A sequence $(x_n) \subset X$ converges in the horofunction compactification if and only if the sequence converges in all the visual compactifications.*

Restricting the result to the Teichmüller metric we get Corollary 1.11 announced in the introduction.

4 Background on Teichmüller spaces

A *surface with marked points* S is a pair (Σ, P), where Σ is a compact, orientable surface with possibly empty boundary, and $P \subset \Sigma$ is a finite, possibly empty, set of points, where we allow points to be on the boundary. The *Teichmüller space* $\mathcal{T}(S)$ is the set of equivalence classes of pairs (X, f), where X is a Riemann surface and $f\colon \Sigma \to X$ is an orientation-preserving homeomorphism. Two pairs (X, f) and (Y, g) are equivalent if there is a conformal diffeomorphism $h\colon X \to Y$ such that $g^{-1}\circ h\circ f$ is isotopic to identity rel P.

The *Teichmüller distance* between two points $[(X, f)], [(Y, g)] \in \mathcal{T}(S)$ is defined as the value $\frac{1}{2}\log\inf K$, where the infimum is taken over all $K \ge 1$ such that there exists a K–quasiconformal homeomorphism $h\colon X \to Y$ with $g^{-1}\circ h\circ f$ isotopic to identity rel P. Together with the smooth structure provided by the Fenchel–Nielsen coordinates $\mathcal{T}(S)$ satisfies all the metric properties discussed in the previous section. That is, $\mathcal{T}(S)$ with the Teichmüller distance is a proper, uniquely geodesic and straight metric space which is C^1 along geodesics and has constant distance variation. See [11, Part 2] for some background on the Teichmüller metric and the Fenchel–Nielsen coordinates.

A *quadratic differential* on a Riemann surface X is a map $q\colon TX \to \mathbb{C}$ such that $q(\lambda v) = \lambda^2 q(v)$ for every $\lambda \in \mathbb{C}$ and $v \in TX$. Considering only holomorphic quadratic differentials with finite area $\int_X |q|$ we get a characterization of the cotangent space to the Teichmüller space based at $[(X, f)]$. Given a point $p \in \mathcal{T}(S)$ and a quadratic differential $q \in T_p^* \mathcal{T}(S)$ there is a unique geodesic γ such that $\gamma(0) = p$ and $\gamma'(0) = |q|/q$. We shall denote such a geodesic as $R(q; \cdot)$ and denote the associated Busemann points as $B(q)$ or B_q.

4.1 Measured foliations

A *multicurve* on S is an embedded 1–dimensional submanifold of $\Sigma \setminus P$ with boundary in $\partial\Sigma \setminus P$ such that

- no circle component bounds a disk with at most 1 marked point;
- no arc component bounds a disk with no interior marked points and at most 1 marked point on $\partial\Sigma$; and
- no two components are isotopic to each other in Σ rel P.

Each of the components is called *curve.* A *weighted multicurve* is a multicurve together with a positive weight associated to each curve. We shall consider (weighted) multicurves up to isotopy rel P. If a simple curve is a circle we shall call it a *closed curve*, and a *proper arc* otherwise.

A *measured foliation* on S is a foliation with isolated prong singularities, where we allow 1–prong singularities at marked points, equipped with an invariant transverse measure μ_F [12, exposé 5]. Denoting α_i and w_i the components and the weights of α respectively, the intersection number $i(\alpha, F)$ is defined as $\inf \sum_i w_i \int_{\alpha_i} |\mu_F| \, d\alpha_i$, where the infimum is taken over all representatives of α. Two measured foliations F and G are *equivalent* if $i(\alpha, F) = i(\alpha, G)$ for every multicurve α. We shall always consider measured foliations up to this equivalence relation. The set of measured foliations is usually denoted as $\mathcal{MF}$, and its topology is defined in such a way that a sequence $(F_n) \subset \mathcal{MF}$ converges to F if and only if $i(\alpha, F_n)$ converges to $i(\alpha, F)$ for every multicurve α.

Given a quadratic differential one can define the *vertical foliation* as the union of *vertical trajectories*, that is, maximal smooth paths γ such that $q(\gamma'(t)) < 0$ for every t in the interior of the domain. This foliation can be equipped with the transverse measure given by $|\operatorname{Re}\sqrt{q}|$. This measured foliation is called the *vertical measured foliation* of q, and shall be denoted as $V(q)$. This map is actually a homeomorphism. As such, given a measured foliation F and a complex structure X, there is a unique quadratic differential $q_{F,X}$ on X such that $V(q_{F,X}) = F$. We call this quadratic differential the *Hubbard–Masur* differential associated to F on X [19]. Furthermore, for each $\lambda > 0$ we have $q_{\lambda F,X} = \lambda q_{F,X}$. Similarly, the *horizontal foliation* $H(q)$ can be defined as the union of maximal smooth paths γ such that $q(\gamma'(t)) > 0$, with the transverse measure $|\operatorname{Im}\sqrt{q}|$.

It is possible to associate a measured foliation to each weighted multicurve by thickening each proper arc and closed curve to a rectangle or cylinder respectively with width equal to the weight of the curve, and then collapsing the rest of the surface. The intersection numbers are maintained by this construction. This association is injective, and hence we shall consider the set of weighted multicurves as a subset of the measured foliations, and use both expressions of weighted multicurve indistinctly.

By removing the critical graph, a measured foliation is decomposed into a finite number of connected components, each of which is either a thickened curve, or a minimal component which does not intersect the boundary, in which every leaf is dense [40, Chapter 24.3]. Each transverse measure within the minimal components can be further decomposed into a sum of finitely many projectively distinct ergodic measures. A foliation F' is an indecomposable component of F if it is either a thickened curve or a minimal component with a transverse measure that cannot be decomposed as a sum of more than one projectively distinct ergodic measure. Every foliation can be decomposed uniquely into a union of indecomposable foliations. For a surface of genus g with no boundaries nor marked points Papadopoulos shows [37] that the maximum number of indecomposable components for any foliation is $3g-3$. It is possible to get an upper bound for foliations on surfaces with boundary and marked points by swapping the marked points for boundaries and using the doubling trick we will explain in Section 4.3.

It was shown by Thurston that for surfaces without boundary it is possible to achieve a dense subset by restricting to simple closed curves, see Fathi, Laudenbach, and Poénaru [12] for a reference. When there are boundaries the picture gets slightly more complicated, but it has been shown by Kahn, Pilgrim and Thurston [21, Proposition 2.12] that multicurves can be seen as a dense subset. More precisely, they show the following.

Proposition 4.1 (Kahn–Pilgrim–Thurston) *Let F be a measured foliation in S not containing proper arcs. Then there exists a sequence of multicurves composed solely of closed curves approaching F.*

The result can be extended to any foliation by cutting along the proper arcs and approaching the foliation in the resulting surfaces by multicurves. Then, joining the multicurves from the proposition with the proper arcs and the adequate weights we get a sequence of multicurves converging to our original foliation.

4.2 Extremal length

Given a marked conformal structure on S, that is, a point $X \in \mathcal{T}$, the *extremal length* of F on X is defined as

$$\mathrm{Ext}_X(F) := \int_X |q_{F,X}|.$$

The map $\mathrm{Ext}\colon \mathcal{MF}(S) \times \mathcal{T}(S) \to \mathbb{R}$ is continuous and homogeneous of degree 2 in the first variable.

Given two points $x, y \in \mathcal{T}(S)$ we can define the function

$$K_{x,y} := \sup_{F \in P_b} \frac{\mathrm{Ext}_x(F)}{\mathrm{Ext}_y(F)},$$

where P_b is the set of measured foliations F satisfying $\mathrm{Ext}_b(F) = 1$. As revealed by Kerckhoff's formula [23], the value $\frac{1}{2}\log K_{x,y}$ coincides with the usual definition of the Teichmüller distance $d(x,y)$.

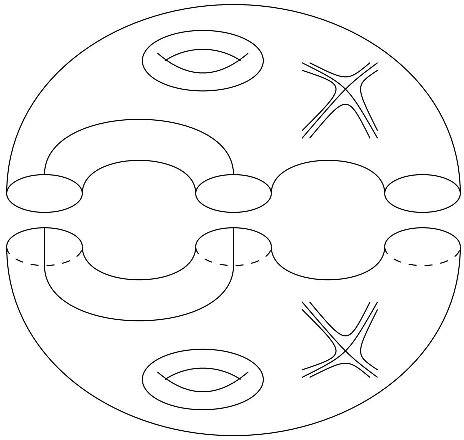


Figure 4: Visual representation of the doubling trick.

4.3 The doubling trick

Let X be a Riemann surface with nonempty boundary. Denote by $\overline{X}$ the mirror surface, obtained by composing each atlas of X with the complex conjugation. Gluing X to $\overline{X}$ along the corresponding boundary components we obtain the *conformal double* $X^d = X \cup \overline{X}/\sim$ of X. Note that X^d has empty boundary. See Figure 4. Given a foliation F or a quadratic differential q on X, we can repeat the same process, obtaining the corresponding conformal doubles F^d and q^d on X^d. For a more detailed treatment of this argument see [1, Section II.1.5].

The main interest of the conformal doubles is that these are surfaces without boundary, so most of the results relating to Teichmüller theory of surfaces without boundary can be translated to surfaces with boundary. We have the following.

Proposition 4.2 *Let X be a Riemann surface with boundary, and F be a foliation on X. Then,*

$$\operatorname{Ext}_{X^d}(F^d) = 2\operatorname{Ext}_X(F).$$

Proof We have $q_{F^d,X^d} = q^d_{F,X}$, so the result follows, as $\int_{X^d} |q^d_{F,X}| = 2\int_X |q_{F^d,X^d}|$. □

4.4 The Gardiner–Masur compactification

For a surface S with marked points and empty boundary we can embed $\mathcal{T}(S)$ into the space of continuous functions from the set $\mathcal{S}$ of simple closed curves on S to $\mathbb{R}$ via the map $\phi\colon \mathcal{T}(S) \to P(\mathbb{R}^{\mathcal{S}})$ defined by

$$\phi(X) = [\operatorname{Ext}_X(\alpha)^{1/2}]_{\alpha\in\mathcal{S}},$$

where the square brackets indicate a projective vector. Gardiner and Masur show [14] that this map is indeed an embedding, and that $\phi(\mathcal{T}(S))$ is precompact. The Gardiner–Masur compactification of a surface without boundary is then defined as the pair $\big(\phi, \overline{\phi(\mathcal{T}(S))}\big)$.

Alternatively, after choosing a basepoint $b \in \mathcal{T}(S)$, it is also possible to consider the map

$$\mathcal{E}\colon \mathcal{T}(S) \to C(\mathcal{MF}) \quad \text{defined by} \quad \mathcal{E}(X)(\cdot) := \left(\frac{\operatorname{Ext}_X(\cdot)}{K_{b,X}}\right)^{1/2},$$

This map is quite similar to the original map ϕ, the differences being that $\mathcal{E}$ considers all measured foliations instead of just the closed curves, and normalizes instead of projectivizing. Walsh proves [44] that, for surfaces without boundary, the map $\mathcal{E}$ defines a compactification in the same way that ϕ does, and in fact this compactification is isomorphic to the one defined by ϕ.

The compactification defined by $\mathcal{E}$ fits better our goal, so we shall define the Gardiner–Masur compactification of Teichmüller spaces of surfaces with boundary as the one obtained by using $\mathcal{E}$. With this in mind, we first need the following result.

Proposition 4.3 *Let S be a compact surface with possibly boundary and marked points. Then the map $\mathcal{E}\colon \mathcal{T}(S) \to C(\mathcal{MF})$ is injective.*

Proof Assume we have $x, y \in \mathcal{T}(S)$ with $\mathcal{E}(x)(F) = \mathcal{E}(y)(F)$ for all $F \in \mathcal{MF}$. Then,

$$K_{x,y} = \sup_{F \in P_b} \frac{\operatorname{Ext}_x(F)}{\operatorname{Ext}_y(F)} = \frac{K_{b,x}}{K_{b,y}} \quad \text{and} \quad K_{y,x} = \sup_{F \in P_b} \frac{\operatorname{Ext}_y(F)}{\operatorname{Ext}_x(F)} = \frac{K_{b,y}}{K_{b,x}} = K_{x,y}^{-1}.$$

However, $K_{y,x} = K_{x,y}$, since the Teichmüller distance is symmetric. Hence, $K_{x,y} = 1$ and, by Kerckhoff's formula, $d(x, y) = \frac{1}{2}\log K_{x,y} = 0$. □

Miyachi shows [32] that the set $E(S) := \{\mathcal{E}(X) \mid X \in \mathcal{T}(S)\}$ is precompact when S is a surface without boundary. Given a surface with boundary S, denote by $\mathcal{MF}^d(S)$ the set of measured foliations on S^d obtained by doubling the foliations $\mathcal{MF}(S)$. The set $E(S^d)|_{\mathcal{MF}^d(S)} = \{\mathcal{E}(X)|_{\mathcal{MF}^d(S)} \mid X \in \mathcal{T}(S^d)\}$, obtained by restricting the functions in $E(S^d)$ to $\mathcal{MF}^d$, is precompact. Furthermore, we can embed $E(S)$ into $E(S^d)|_{\mathcal{MF}^d(S)}$ by sending $f \in E(S)$ to $f^d \in E(S^d)|_{\mathcal{MF}^d(S)}$ defined by $f^d(F^d) = f(F)$. Hence, $E(S)$ is precompact.

We define the *Gardiner–Masur compactification* for a surface with boundary as the closure $\overline{E}$ of $E(S)$, together with the map $\mathcal{E}$. We shall be using the same characterization for surfaces without boundary.

One of the relevant features of the Gardiner–Masur compactification is that it coincides with the horofunction compactification. Indeed, Liu and Su [28] and Walsh [44] prove that for surfaces without boundary these two compactifications are isomorphic. In the following, we shall extend the relevant results to surfaces with boundary. We begin with the driving theorem from Walsh's paper.

Theorem 4.4 (extension of [44, Theorem 1] to surfaces with boundary) *Let $R(q;\cdot)\colon \mathbb{R}_+ \to \mathcal{T}(S)$ be the Teichmüller ray with initial unit-area quadratic differential q, and let F be a measured foliation. Then,*

$$\lim_{t\to\infty} e^{-2t} \operatorname{Ext}_{R(q;t)}(F) = \sum_j \frac{i(G_j, F)^2}{i(G_j, H(q))},$$

where the $\{G_j\}$ are the indecomposable components of the vertical foliation $V(q)$, and $H(q)$ is the horizontal foliation.

Proof If S does not have boundary the result follows from Walsh's paper. Assume then that S has boundary. Let p be the number of proper arcs of $V(q)$, and reorder the components so G_j is a proper arc for $j \leq p$. The conformal double G_j^d is indecomposable whenever G_j is a proper arc, and decomposes into two components otherwise, as it is not incident to the boundary of S. Denote G_j^1 and G_j^2 the two components of G_j for $j > p$. We have

$$2 \lim_{t\to\infty} e^{-2t} \operatorname{Ext}_{R(q;t)}(F) = \lim_{t\to\infty} e^{-2t} \operatorname{Ext}_{R(q^d;t)}(F^d) = \sum_{j\leq p} \frac{i(G_j^d, F^d)^2}{i(G_j^d, H(q)^d)} + \sum_{i\in\{1,2\}} \sum_{j>p} \frac{i(G_j^i, F^d)^2}{i(G_j^i, H(q)^d)}.$$

For foliations $G, F \in \mathcal{MF}(S)$ we have $i(G^d, F^d) = 2i(G, F)$. Hence, $i(G_j^d, F^d) = 2i(G_j, F)$. Using the symmetry, $i(G_j^1, F^d) = i(G_j^2, F^d)$, so for $j > p$ we have $i(G_j^1, F^d) = i(G_j, F)$. Using these identities we get the result. □

Following the same reasoning we can extend as well the next result.

Lemma 4.5 (extension of [44, Lemma 3] to surfaces with boundary) *Let q be a unit area quadratic differential. Then,*

$$e^{-2t} \operatorname{Ext}_{R(q;t)}(F) \geq \sum_j \frac{i(G_j, F)^2}{i(G_j, H(q))},$$

where $t \in \mathbb{R}_+$ and $\{G_j\}$ are the indecomposable components of the vertical foliation $V(q)$.

Most of the results in Walsh's paper use the previous theorem. In particular, we have the following.

Corollary 4.6 (extension of [44, Corollary 1] to surfaces with boundary) *Let q be a quadratic differential and denote by G_j the indecomposable components of its vertical foliation. Then, the Teichmüller ray $R(q;\cdot)$ converges in the Gardiner–Masur compactification to*

$$\left(\sum_j \frac{i(G_j, \cdot)^2}{i(G_j, H(q))}\right)^{1/2}.$$

The relation between the Gardiner–Masur compactification and the horoboundary compactification is given by the map $\Xi : \overline{E} \to \overline{\mathcal{T}(S)}^h$ defined by

$$\Xi(f)(x) := \tfrac{1}{2} \log \sup_{F\in\mathcal{P}} \frac{f(F)^2}{\operatorname{Ext}_x(F)}.$$

The following result can be extended to surfaces with boundary by repeating the proof found in Walsh's paper in this context.

Theorem 4.7 (extension of [44, Lemma 21] to surfaces with boundary) *The map Ξ is an isomorphism between the compactifications $(\mathcal{E}, \overline{E})$ and $(h, \overline{\mathcal{T}(S)}^h)$.*

Directly from the definition of Ξ we have the following.

Corollary 4.8 *Let $f, g \in \overline{E}$. If $f \geq g$, then $\Xi(f) \geq \Xi(g)$.*

We shall denote the representation of the Busemann point $B(q)$ in the Gardiner–Masur compactification as $\mathcal{E}(q)$. By Corollary 4.6 we have an explicit representation of $\mathcal{E}(q)$. As we have seen in Propositions 1.3 and 1.4, the continuity of the Busemann map has some interesting implications, and it is enough to look for continuity of the map restricted to the boundary. Related to this question we have the following result, which can also be derived by the same proof found in Walsh's paper, applied to this context.

Theorem 4.9 (extension of [44, Theorem 10] to surfaces with boundary) *Let (q_n) be a sequence of quadratic differentials based at $b \in \mathcal{T}(S)$. Then $B(q_n)$ converges to $B(q)$ if and only if both of the following hold*:

1. *(q_n) converges to q.*
2. *For every subsequence $(G^n)_n$ of indecomposable elements of $\mathcal{MF}$ such that, for each $n \in \mathbb{N}$, G^n is a component of $V(q_n)$, we have that every limit point of G^n is indecomposable.*

In view of this theorem, we say that a sequence of quadratic differentials (q_n) converges *strongly* to q if it does so in the sense described by the theorem.

Finally, while the following result may be extendable to surfaces with boundary, we only use it in the context of surfaces without boundary, so we shall not be working on finding an extension.

Theorem 4.10 [44, Theorem 3] *For the Teichmüller space of a surface without boundary with the Teichmüller metric, for any basepoint $X \in \mathcal{T}(S)$, all Busemann points can be expressed as $B(q)$ for some quadratic differential q based at X.*

5 Horoboundary convergence for Teichmüller spaces

5.1 Continuity of the Busemann map

We begin by using Proposition 1.4 to determine when the Busemann map is continuous. Recall that a sequence (q_n) converges to q strongly if and only if the sequence satisfies the conditions of Theorem 4.9. That is, a sequence (q_n) converges to q strongly if and only if the associated Busemann points $B(q_n)$ converge to $B(q)$. With this in mind we introduce the following notion.

Definition 5.1 Let q be a quadratic differential. We say that q is *infusible* if any sequence of quadratic differentials converging to q converges strongly. We say that q is *fusible* if it is not infusible.

In other words, we say that q is fusible when it can be approached by a sequence of quadratic differentials (q_n) such that there is some sequence (G^n) of measured foliations with each G^n being an indecomposable component of $V(q_n)$, with (G^n) having at least one decomposable accumulation point. The following statement follows directly from this definition, Proposition 1.4 and Walsh's result.

Proposition 5.2 *Let q be a unit area quadratic differential. The Busemann map B is continuous at q if and only if q is infusible.*

Proof If q is fusible then we have a sequence converging to q but not strongly. Hence, by Theorem 4.9 the sequence $(B(q_n))$ does not converge to $B(q)$, and so the Busemann map is not continuous at q.

If q is infusible we have that any sequence (q_n) converging to q does so strongly, and so we have that $B(q_n)$ converges to $B(q)$, so B is continuous at q when restricted to the boundary. By Proposition 1.4 this implies that B is continuous at q. □

We shall now find a criterion on the vertical foliation to determine when a unit area quadratic differential is infusible.

Definition 5.3 Let F be a measured foliation on a surface S and let G be one of its indecomposable components. We say that G is a *boundary annulus* if it is an annulus parallel to a boundary with no marked points, and a *boundary component* if it is a boundary annulus or a proper arc. If G is not a boundary component, we shall call it an *interior component*. Each of the connected components of the surface obtained after removing the proper arcs shall be called *interior part*. If each of these interior parts has at most one interior component, then we say that F is *internally indecomposable*. If F is not internally indecomposable we say that it is *internally decomposable*.

For surfaces without boundary, a foliation F is internally indecomposable if and only if it is indecomposable, as we do not have boundary components. Given these definitions we can state our main result of this section.

Theorem 5.4 *Let q be a quadratic differential. Then q is infusible if and only if its vertical foliation $V(q)$ is internally indecomposable.*

This result is somewhat straightforward whenever S does not have boundary, as in order to have a sequence (q_n) that converges to q but not strongly we need a sequence of components of $V(q_n)$ converging to a decomposable component of $V(q)$, but if S is closed and $V(q)$ is internally indecomposable, then $V(q)$ only has one indecomposable component. Conversely, if $V(q)$ has more than one indecomposable component, as S does not have boundary $V(q)$ can be approached by a sequence of simple closed curves, so the associated sequence of quadratic differentials converges to q but not strongly.

For surfaces with boundary the proof is more involved, as simple closed curves are no longer dense. However, the density of multicurves from Proposition 4.1 allows us to follow a slightly similar reasoning. We begin by proving some results regarding the shape that foliations have to take when approaching a foliation with boundary components, namely, boundary components have to be eventually included in the approaching foliations.

Proposition 5.5 *Let (F_n) be a sequence of measured foliations converging to a measured foliation F, let G be the union of the boundary components of F and let H be such that $F = H + G$. Then, for n big enough, $F_n = H_n + a_n G$, with a_n converging to 1 and H_n converging to H.*

In particular, the proper arcs of the limiting foliation have to be included in the approaching foliations. Hence, we will be able to separate the surface along these proper arcs into the interior parts of the limiting foliation, and study the convergence in each of these parts.

We say that a subset of a boundary component is a *boundary arc* if it is homeomorphic to an open interval or a circle, does not contain marked points and, if it is homeomorphic to an open interval, it is delimited by marked points.

Repeating the argument by Chen, Chernov, Flores, Fortier Bourque, Lee, and Yang [7] to a more general setting we get the following characterization of foliations on simple surfaces, which we shall use to solve the simpler cases.

Lemma 5.6 *Let S be a sphere with one boundary component possibly containing boundary marked points and one interior marked point. Then every indecomposable foliation on S is a proper arc and there are finitely many distinct proper arcs.*

Proof Assuming that there is some foliation F with a recurrent leaf to some part of S we get a contradiction, as explained in the proof of [7, Lemma 4.1]. Hence, each indecomposable foliation is a curve. Any closed curve in S is contractible to the marked point. Hence, a each indecomposable foliation is a proper arc.

A proper arc in S must have two endpoints, which must be contained in the boundary arcs in the boundary component of S. Denote b_1 and b_2 these two boundary arcs, which might be the same. We aim to show that there are at most two classes of arcs with endpoints in b_1 and b_2. Fix three proper arcs with endpoints on b_1 and b_2. Any intersection between these arcs can be removed by doing isotopies moving the endpoints along the arcs b_1 or b_2. Hence, these arcs can be isotoped to not intersect each other. Since there is only one interior marked point, two of these arcs delimit a rectangle with no marked interior marked points, so are isotopic. Hence, there are at most two different proper arcs between b_1 and b_2. There are finitely marked points in the boundary component, so there are finitely many boundary arcs. Therefore, there are finitely many pairs of boundary arcs, and since we have at most two proper arcs per pair, there are also finitely many different proper arcs. □

We shall first see the proposition for the case where G contains a proper arc and we are approaching with a sequence of indecomposable foliations.

Lemma 5.7 *Let S be a surface and let (F_n) be a sequence of indecomposable foliations on S converging to a measured foliation G. Then G is either a multiple of a proper arc γ, in which case F_n is also a multiple of γ for n big enough, or G does not contain a proper arc.*

Proof Assume G contains a proper arc γ with weight $w > 0$ and denote by b one of the boundary arcs where γ is incident.

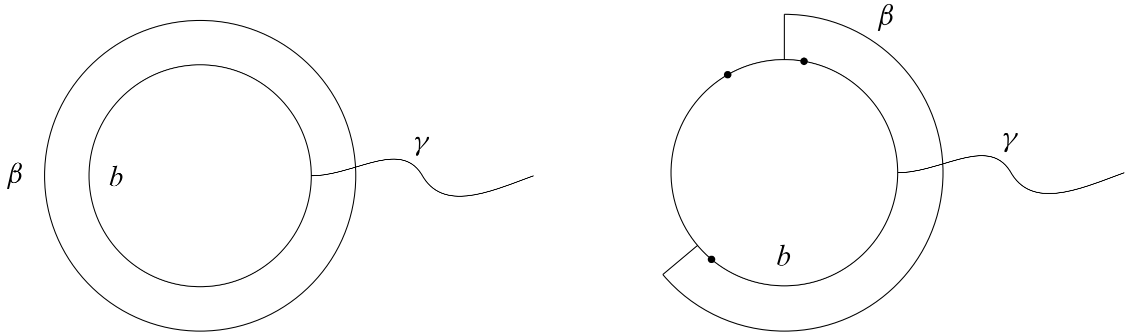


Figure 5: Sample curves used in the proof of Lemma 5.7.

Our first step is seeing that, for n big enough, F_n intersects b. We shall do this by finding different test curves β depending on the shape of b. If the boundary component containing b has at most one marked point, we consider β to be a curve parallel to that boundary component as in Figure 5, left. Otherwise we consider β to be the curve defined by taking a small arc starting at the boundary arc next to b, concatenating with a curve parallel to b, and concatenating another segment with endpoint in the boundary arc after b, as shown in Figure 5, right.

If the curve β is contractible then S is a sphere with one boundary component and at most one interior marked point, so by Lemma 5.6 the result follows. Assume then that β is not contractible. We have $i(\gamma, \beta) > 0$, so $i(G, \beta) > 0$ and hence $i(F_n, \beta) > 0$ for n big enough, which implies that F_n intersects b. Hence, since F_n is indecomposable, it is a weighted proper arc, which we denote by $w_n \gamma_n$, where $w_n > 0$ is the weight at γ_n is a proper arc.

Denote b_1 and b_2 the boundary arcs where γ has its endpoints, and denote by β_1 and β_2 the associated test curves shown in Figure 5. If both endpoints are in the same boundary arc we set b_2 and β_2 as null curves. We shall now find a multicurve A surrounding γ, b_1 and b_2 such that any leaf of G intersecting A but not γ has an endpoint in either b_1 or b_2. The multicurve A is chosen so that, together with the boundaries where γ has its endpoints, delimits the smallest surface containing γ. The precise shape of A depends on whether the endpoints of γ are in the same boundary component or not, and the distribution of marked points in these boundaries.

If both endpoints of γ are in different boundary components we proceed differently according to the distribution of marked points at these boundaries. If each of the boundaries contains at most one marked point then we define A as the curve shown in Figure 6, left. If one of the boundary components has two or more marked points, but the other has at most one marked point we define A as the arc shown in Figure 6, middle. Finally, if each of the boundaries contains at least two marked points we define A as the multicurve formed by the curves A_1 and A_2 as shown in Figure 6, right.

If both endpoints γ are in the same boundary we also proceed differently according to the distribution of marked points. In all cases A is defined as a multicurve formed by two curves. If each possible segment

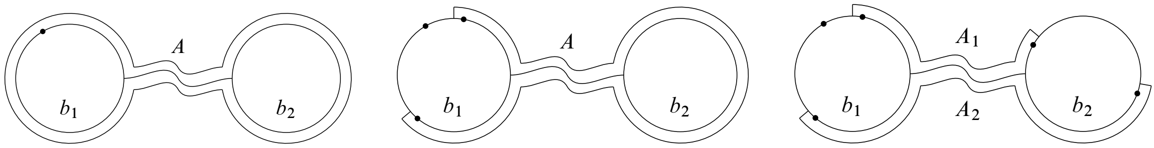


Figure 6: Construction of the curves A_1 and A_2 whenever γ has endpoints in different boundary components in the proof of Lemma 5.7.

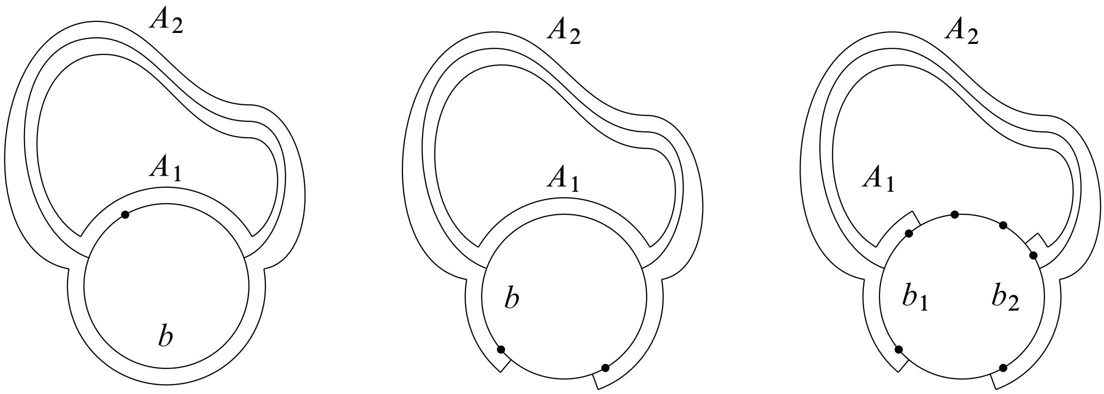


Figure 7: Construction of the curves A_1 and A_2 whenever γ has endpoints in the same boundary component in the proof of Lemma 5.7.

within the boundary component joining the two endpoints has at most one marked points we proceed as in Figure 7, left. If one of these segments has two or more marked points, while the other has at most one we proceed as in Figure 7, middle. Finally, if both of these segments have two or more marked points we proceed as in Figure 7, right.

In any of the cases above if a component of A is nonessential we remove it from A. The following argument also applies whenever A is a null curve. Put A and G in minimal position and denote by P the surface containing γ, delimited by A and the boundary components where γ has its endpoints. Let α be a connected component of a noncritical leaf of G restricted to P intersecting A. Since G contains γ the proper arc α cannot intersect γ. Furthermore, by observing the possible configurations, if α has one endpoint in A_1, the other one cannot be in A_2, as whenever we have both A_1 and A_2, these are separated within P by the proper arc γ. Furthermore, if both endpoints are in A_1 then α can be isotoped to not intersect A. Therefore, the other endpoint of α is in either b_1 or b_2. Hence, $i(G,\beta_1)+i(G,\beta_2)\geq i(G,A)+w\,i(\gamma,\beta_1)+w\,i(\gamma,\beta_2)>i(G,A)$. Since $w_n\gamma_n$ converges to G, this last inequality implies that for n big enough,

$$i(\gamma_n,\beta_1)+i(\gamma_n,\beta_2)>i(\gamma_n,A).$$

Fix n such that γ_n satisfies the previous inequality. Assume γ_n has just one endpoint inside P. Then, $i(\gamma_n,\beta_1)+i(\gamma_n,\beta_2)=1$, so $i(\gamma_n,A)=0$ and γ_n cannot leave P. If γ_n has both endpoints in P then

$i(\gamma_n, \beta_1) + i(\gamma_n, \beta_2) = 2$. Furthermore, if γ_n leaves P, then it has to reenter at some point, resulting in $i(\gamma_n, A_1 + A_2) = 2$. Hence, γ_n stays inside P.

The weights w_n do not converge to 0, as $w_n i(\gamma_n, \beta)$ converges to $i(G, \beta)$, but $i(\gamma_n, \beta) \leq 2$. Since γ is contained in G we have $i(G, \gamma) = 0$. Therefore, for any $\varepsilon > 0$ and n big enough we have $w_n i(\gamma_n, \gamma) < \varepsilon$, so for n big enough $i(\gamma_n, \gamma) = 0$. Since γ_n does not intersect γ and stays inside P, γ_n can be isotoped to stay inside one of the components obtained after removing γ from P. Denote such a component by C. The component C has either one or two boundary components and no interior marked points or one boundary component and one interior marked point. By Lemma 5.6 the only case where we do not have finitely many different proper arcs is when C has two boundary components. However, in that case one of the boundary components is associated to a curve in A, so γ_n does not intersect it and that boundary can be treated as a marked point. Hence, in all cases there are finitely many possible proper arcs, and so γ_n is a multiple of γ for n big enough. □

When the boundary component is an annulus we have to be a bit more careful, so we start by proving it for approaching curves.

Lemma 5.8 *Let S be a surface and let $(w_n\gamma_n)$ be a sequence of weighted curves on S converging to a foliation G, where (w_n) are the weights and (γ_n) are the curves. Then G is either a multiple of a boundary annulus γ, in which case γ_n is γ for n big enough, or G does not contain a boundary annulus.*

Proof If S is a polygon with at most one interior marked point, then G cannot contain a boundary annulus. If S is a cylinder then, since we have a boundary annulus, at least one of the boundaries must not contain marked points. Hence, the number of curves is finite, as there is only one possible closed curve, and for counting the proper arcs we can consider the boundary without marked points as a marked point and apply Lemma 5.6. In that case, the conclusion follows.

Assume then that S is neither a disk with at most one interior marked point nor a cylinder with no interior marked points. Then there is a pair of pants P in S containing γ where each boundary component of P is either noncontractible or contractible to a marked point. Denote by B_1 the boundary component parallel to γ and B_2 and B_3 the other two boundary components of P. Furthermore, assume that G contains γ with weight w.

Begin by assuming that B_2 and B_3 are not contractible to marked points. Let C be the proper arc contained in P with both endpoints in B_1. Put B_2, B_3 and C in a minimal position with respect to G, and consider a connected component of a noncritical leaf of G intersecting C restricted to P. This noncritical leaf either is isotopic to γ, or to the curves F, E and D shown in Figure 8. Since the leaves of G do not intersect, there cannot be leaves isotopic to E and leaves isotopic to D at the same time. Breaking symmetry, assume there are no leaves isotopic to D. Then, $i(C, G) = i(C, \gamma) + i(B_3, G) > i(B_3, G) \geq i(B_2, G)$. Doing the same reasoning assuming that there are no leaves isotopic to E we get $i(C, G) > \max(i(B_2, G), i(B_3, G))$.

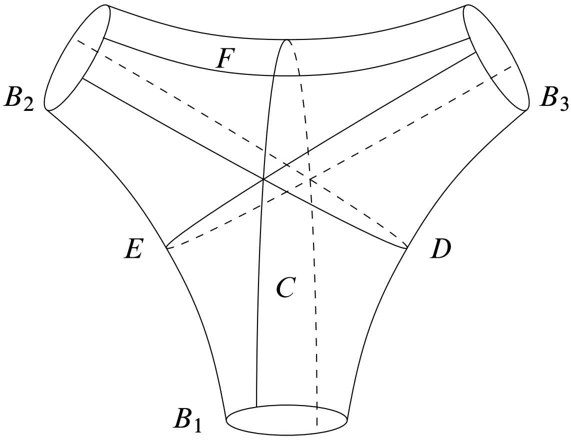


Figure 8: Curve labeling for the proof of Lemma 5.8.

Hence, since $w_n\gamma_n$ converges to G, γ_n has to satisfy

$$i(C,\gamma_n) > \max(i(B_2,\gamma_n), i(B_3,\gamma_n))$$

for n big enough.

For each n put B_3, B_2 and C in a minimal position with respect to γ_n, and consider the restriction of γ_n to P. Assume γ_n is not γ. Then, the curves on the restriction of γ_n to P intersecting C are isotopic to either E, F and D, but not γ. As before, this restriction cannot contain curves isotopic to E and curves isotopic to D for the same n, so assuming there are no curves isotopic to D we have $i(C,\gamma_n) = i(B_3,\gamma_n)$ which is a contradiction. Doing the same reasoning assuming that there are no curves isotopic to E also gives a contradiction. Hence, γ_n is γ for n big enough.

If B_2 or B_3 are contractible to marked points we have $i(G, B_2)$ or $i(G, B_3)$ is 0, and a similar reasoning yields the same result. □

Proof of Proposition 5.5 Let (F_n) be a sequence of measured foliations converging to F. As pointed out before, Proposition 4.1 can be extended to get sequences of weighted multicurves $(\gamma_n^m)_m$ converging to each F_n. Denote by $\gamma_{n,1}^m, \gamma_{n,2}^m, \dots, \gamma_{n,k(n,m)}^m$ the weighted curves of γ_n^m. For each n we take a subsequence such that $k(n,m)$ is constant with respect to m, and $\gamma_{n,i}^m$ converges for each i as $m \to \infty$. Denoting $F_{n,i}$ the limit of $\gamma_{n,i}^m$ as $m \to \infty$, we can write $F_n = \sum F_{n,i}$.

Denote by β_j the boundary components of F. That is, $\sum \beta_j = G$. Furthermore, denote by $b_{n,j}$ and $b_{n,j}^m$ the weights of β_j on F_n and γ_n^m, where we set the weight to be 0 if β_j is not contained in the foliation. It is clear that if $b_{n,j} = 0$ then $b_{n,j}^m \to 0$, as we must have $b_{n,j} \geq \liminf_{m\to\infty} b_{n,j}^m$. If $b_{n,j} > 0$ for some n, then $F_{n,i}$ contains β_j for some i. Hence, by Lemmas 5.7 and 5.8 we have $F_{n,i}$ and $\gamma_{n,i}^m$ are both multiples of β_j for m big enough. Then, since each of the multicurves in γ_n^m has to be different, β_j is not contained in any other foliation $F_{n,i}$ for that given n, so $F_{n_i} = b_{n,j}\beta_j$ and $\gamma_{n,i}^m$ can be written as $b_{n,i}^m\beta_j$ for m big enough, with $b_{n,i}^m$ converging to $b_{n,j}$ as $m \to \infty$.

Assume for some j we have $b_{n,j}$ not converging to 1. We can then take a subsequence such that $b_{n,j}$ converges to some $\lambda \neq 1$. Denote $\delta = |1-\lambda|/2$. For each n, there exists some $m_0(n)$ big enough so that $|1-b^m_{n,j}| > \delta$ for all $m \geq m_0(n)$. We can then take a diagonal sequence $\gamma_n^{m(n)}$ converging to F with $m(n) \geq m_0(n)$. However, following the previous reasoning we get that $\gamma_n^{m(n)}$ should contain β_j for n big enough, and the weight should converge to the weight in G, that is, to 1. However, $|1-b^{m(n)}_{n,j}| > \delta$, giving us a contradiction. Hence, $b_{n,j}$ converges to 1 for all j. Let then $a_n = \min_j(b_{n,j})$. Since $b_{n,j} \geq a_n$ we can define $H_n = F_n - a_n G$ and we have $F_n = H_n + a_n G$. Finally, $a_n \to 1$ as $n \to \infty$, so the proposition is proved. □

Proposition 5.9 *Let q be a unit area quadratic differential such that $V(q)$ is internally indecomposable. Then q is infusible.*

Proof Assume q is fusible, that is, we have a sequence of quadratic differentials (q_n) converging to q but not strongly. Let F_i^n be the indecomposable components of $V(q_n)$. To have nonstrong convergence we must have at least one sequence of indecomposable components converging to a decomposable component G, which we assume is $(F_1^n)_n$. Let β be a boundary component of $V(q)$. By Proposition 5.5 for n big enough a multiple of β must be contained in $V(q_n)$. Furthermore, β cannot be contained in G. Since G cannot contain boundary components, it must contain at least two interior components. On the other hand, since $V(q)$ is internally indecomposable, each interior part obtained by removing the proper arcs contains at most one interior component. Hence, for n big enough F_1^n must intersect at least two interior parts, that is, F_1^n must cross at least one proper arc. However, for each proper arc γ there is some n big enough such that γ is contained in the foliation $V(q_n)$, so F_1^n, a component of $V(q_n)$, intersects the foliation $V(q_n)$, giving us a contradiction. □

To prove the other direction we shall first see the following lemma.

Lemma 5.10 *Let S be a compact surface with possibly nonempty boundary and finitely many marked points, let $k \geq 2$ and let $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ be a collection of nonintersecting closed curves on S. Furthermore, let p be the number of curves in α parallel to a boundary. Then there exists a collection of* $\max(\lceil (p/2) \rceil), 1)$ *nonintersecting curves intersecting each α_i.*

Our main interest in the lemma is that the amount of curves needed is strictly smaller than the amount of closed curves in α. This will allow us, by doing Dehn twists along the closed curves in α, to create a sequence of foliations converging to a foliation with strictly more components, which can be translated to a sequence of quadratic differentials that converge but not strongly. The proof of this lemma is based on a reasoning found in [11, Proposition 3.5].

Proof We start by replacing all boundaries of S without parallel curves in α by marked points. Let then α' be a completion of α to a pair of pants decomposition. Glue the remaining boundaries pairwise until we have at most one left. After cutting the surface along the closed curves that were not parallel to boundaries we get a collection of $\lceil p/2 \rceil$ tori with one boundary component and some spheres with b

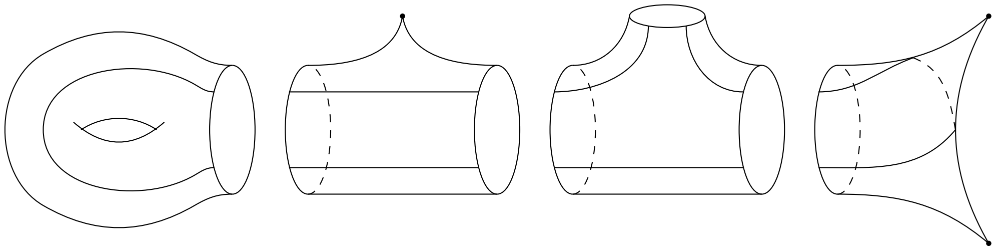


Figure 9: Laying out of curve segments for the proof of Lemma 5.10.

boundary components and n marked points, with $b+n=3$ and $b \geq 1$. If p is odd, one of these spheres has a boundary of S as a boundary. We join the boundaries of each of these surfaces with nonintersecting arcs, as shown in Figure 9, that is, in such a way that each boundary component has two arcs incident to it. We can then paste these surfaces back together in order to obtain a collection $\beta_1, \beta_2, \dots, \beta_l$ of pairwise disjoint curves in S. If p is odd this collection contains precisely one proper arc, as we only have two endpoints coming from the boundary we did not paste. If p is even the collection does not contain any proper arc. By the bigon criterion each β_j is in minimal position with respect to each α_i, and each α_i intersects either one or two of the β_j. Furthermore, since we did not cut along the original boundaries we pasted from S, each α_i parallel to a boundary of S intersects precisely one of the β_j. Suppose we have β_j and $\beta_{j'}$ intersecting a curve $\kappa \in \alpha'$ and that β_j and $\beta_{j'}$ are distinct. Since we have at most one proper arc, at least one of β_j and $\beta_{j'}$ is a closed curve. Hence, doing a half twist about κ, β_j and $\beta_{j'}$ become a single curve. Since this process does not create any bigons, the resulting collection is still in minimal position with α. Continuing this way we obtain a single curve γ intersecting each curve in κ. Furthermore, γ intersects each pasted boundary once. Cutting along the pasted boundaries, we get the curves from the lemma. If p is odd, β is a proper arc, so each cut along a pasted boundary increases the curve count by one, totaling $(p+1)/2$ curves. If p is even, β is a closed curve, so the first cut transforms it into a proper arc, and the following ones increase the curve count by one, giving a total of $\max(p/2, 1)$ curves. □

Proposition 5.11 *Let F be an internally decomposable measured foliation. Then, F can be approached by a sequence of weighted multicurves with fewer components than F.*

Proof By the extension to Proposition 4.1, we have a sequence of weighted multicurves γ^n converging to F, with the only proper arcs being the ones contained in F. Cutting the surface along the proper arcs of γ^n and quotienting these proper arcs to points we get k many surfaces $Z_1, Z_2, \dots, Z_k$ with boundary. Let γ_i^n be the restriction of γ^n to Z_i, and let F_i be the limit of γ_i^n. The foliation F is the union of the foliations F_i and the proper arcs.

Fix some i such that F_i is nonempty, and let $\alpha_1, \dots, \alpha_b$ be the closed curves parallel to the boundaries of Z_i. Let $a_1^n, \dots, a_b^n$ be the weights of $\alpha_1, \dots, \alpha_b$ in γ_i^n. We can take a subsequence such that a_j^n converges for each j to some a_j. If $a_j > 0$, the closed curve α_j is contained in F_i. If $a_j = 0$, then the weights a_j^n can be set to 0 on the multicurves γ_i^n while leaving the limit intact. Hence, we can assume that $a_j^n = 0$ for all j such that $a_j = 0$. Let p and u be the number of closed curves with $a_j > 0$ parallel to boundaries with or without marked points respectively. Since we have removed all the closed curves with $a_j = 0$, the multicurve γ_i^n contains precisely p and u closed curves parallel to boundaries with or without marked points for n big enough. Denote by B the set of closed curves parallel to boundary components without marked points. Applying Lemma 5.10 to the multicurve γ_i^n minus B we get $\max(\lceil (p/2) \rceil), 1)$ curves β_i^n intersecting all closed curves in γ_i^n except the ones parallel to boundaries without marked points. Doing the appropriate Dehn twists along the closed curves of γ_i^n and rescaling to the curves β_i^n, and adding with the corresponding weights the curves in B, we get a sequence converging to γ_i^n with $\max(\lceil (p/2) \rceil), 1) + u$ many components. As such, taking a diagonal sequence we can get a sequence of multicurves converging to F_i with each multicurve containing $\max(\lceil (p/2) \rceil), 1) + u$ components.

Finally, since F is internally decomposable, there is at least one F_i with at least 2 interior components, so one of these multicurves has strictly less components than the limiting foliations, and we have nonstrong convergence. □

Theorem 5.4 follows by combining Propositions 5.9 and 5.11.

We do not need S to have a lot of topology to find internally decomposable foliations. In fact, determining which surfaces do not support internally decomposable foliations we get the following result.

Proposition 5.12 *Let $S_{g,b_m,b_u,p}$ be a surface of genus g with b_m and b_u boundaries with and without marked points respectively and p interior marked points. Then the Busemann map is continuous if and only if $3g + 2b_m + b_u + p \leq 4$.*

We shall split the proof in the following two lemmas

Lemma 5.13 *Let $S_{g,b_m,b_u,p}$ be a surface with $3g + 2b_m + b_u + p > 4$. Then it admits an internally decomposable foliation.*

Proof A multicurve consisting of two interior closed curves generates an internally decomposable foliation, so we just have to find such a pair for each possible surface satisfying the hypothesis. If S has genus at least 2 we can take a multicurve consisting of 2 nonseparating closed curves. If S is a torus with at least 2 boundaries or marked points, or a boundary with marked points, we can take a nonseparating closed curve and a separating closed curve around 2 boundaries or marked points, or around a boundary with marked points. If S is a sphere with at least 5 marked points or boundaries, we can take a closed curve around two interior points or boundaries, and a closed curve around two different interior points or boundaries. If S is a sphere with 1 boundary with marked points and at least 3 other boundaries or interior points we can take a closed curve around the boundary with marked points, and a closed curve around two other interior points or boundaries. Lastly, if S is a sphere with 2 boundaries with marked points and another interior marked point or boundary we take a closed curve around each boundary with marked points. □

Lemma 5.14 *Let $S_{g,b_m,b_u,p}$ be a surface with $3g + 2b_m + b_u + p \leq 4$. Then every foliation on S is internally indecomposable.*

Proof Assume we have an internally decomposable foliation on $S_{g,b_m,b_u,p}$. Then we can get an internally decomposable foliation on $S_{g,0,0,b_u+p+2b_m}$ by removing the boundary components, replacing the boundaries without marked points with marked points and each boundary with marked points for 2 marked points. Furthermore, if we have at least one marked point, we can get an internally decomposable foliation in $S_{g,0,0,b_u+p+2b_m+k}$, $k \in \mathbb{N}$, by replacing a marked point with a $k+1$ marked points.

Hence, we only need to prove that a torus with one marked point and a sphere with 4 marked points do not admit internally decomposable foliations. However, since these do not have boundaries, a foliation being internally decomposable translates to a foliation having at least two indecomposable components.

Assume the torus with one marked point admits a foliation with two indecomposable components. We can replace the marked point with a boundary, and add to the foliation a boundary component parallel to that boundary. Considering the doubled surface explained in Section 4.3 we get a closed surface of genus 2 without boundaries nor marked points, with at least 5 indecomposable components. Recall that the maximum number of indecomposable components for a foliation on a surface of genus g is $3g-3$, so for genus 2 the maximum is 3, giving us a contradiction. A similar process applies for the sphere with 4 marked points. □

Proof of Proposition 5.12 The Busemann map is continuous at every point in the interior of Teichmüller space, as it is the identity when restricted in there and $\partial \overline{X}^v$ is closed. Hence, we only need to prove continuity or discontinuity at the points on the boundary. By Lemma 5.13 if $3g+2b_m+b_u+p>4$ then S admits an internally decomposable foliation F, so by Theorem 5.4 the Hubbard–Masur quadratic differential associated to F at the basepoint X is fusible and hence the Busemann map is not continuous at that point. On the other hand, if $3g+2b_m+b_u+p \leq 4$ then by Lemma 5.14 for any quadratic differential q, the vertical foliation $V(q)$ is internally indecomposable, so again by Theorem 5.4 every quadratic differential is infusible an B is continuous at every boundary point. □

By combining Proposition 5.12 with Proposition 1.3, we get the precise classification of surfaces with horofunction compactification isomorphic to visual compactification announced in Theorem 1.6 from the introduction.

Proof of Theorem 1.6 As shown in Proposition 1.3, the visual compactification and the horofunction compactification are isomorphic if and only if the Busemann map is continuous, so the theorem follows by applying Proposition 5.12. □

5.2 Criteria for convergence

One straightforward consequence of the horofunction compactification being finer than the visual compactification is the following criterion regarding the convergence of sequences in the horofunction compactification.

Corollary 5.15 *Let $(x_n) \subset \mathcal{T}(S)$ be a sequence. If (x_n) converges to a quadratic differential q in the visual compactification, then all accumulation points of (x_n) in the horofunction compactification are contained in $\Pi^{-1}(q)$. In particular, if $V(q)$ is internally indecomposable, then (x_n) converges in the horofunction compactification.*

Furthermore, if (x_n) does not converge in the visual compactification, then it does not converge in the horofunction compactification.

Proof If x_n converges in the visual compactification to a quadratic differential q then by the continuity of Π all its accumulation points are in $\Pi^{-1}(q)$. If $V(q)$ is internally indecomposable, then by Theorem 5.4 the quadratic differential q is infusible, so the Busemann map is continuous at q and by Proposition 1.4 the fiber $\Pi^{-1}(q)$ is a singleton. Therefore x_n converges to $\Pi^{-1}(q)$, as that is the only accumulation point of x_n and the horofunction compactification is compact.

On the other hand, if x_n converges to ξ in the horofunction compactification, by continuity of Π, x_n converges to $\Pi(\xi)$ in the visual compactification. □

A frequent topic in the study of compactifications of Teichmüller spaces is the convergence of certain measure-preserving paths. We shall see now how the previous results can be applied in that study.

Let $X \in \mathcal{T}(S)$ be a point in Teichmüller space and q be a unit quadratic differential based at X. It is a well known fact that there exists a unique orientation-preserving isometric embedding $\iota\colon \mathbb{H} \to \mathcal{T}(S)$ from the hyperbolic plane $\mathbb{H}$ to the Teichmüller space such that $\iota(i) = X$ and $\iota^*(q) = i$, see the work of Herrlich and Schmithüsen [18] for a detailed explanation. The path $\iota(i+t)$ for $t \in \mathbb{R}_+$ is called the *horocycle* generated by q. Since ι is an isometric embedding, $h(X)(p) = d(\iota^{-1}X, \iota^{-1}p) - d(\iota^{-1}X, \iota^{-1}b)$ for $X, b, p \in \iota(\mathbb{H})$. That is, if we restrict the evaluations of horofunctions to the image of the Teichmüller disc, the value coincides with the values in the hyperbolic plane. Hence, since the path $i+t$ is a horocycle of the Busemann point obtained by moving along the geodesic $e^t i$ along the hyperbolic plane, the path $\iota(i+t)$ is also a horocycle of the corresponding Busemann point $B(q)$, obtained by moving along the geodesic $\iota(e^t i)$.

Since ι is an isometric embedding, the geodesic between X and $\iota(i+t)$ is contained in $\iota(\mathbb{H})$. Furthermore, the pushforward and pullback maps are continuous, so denoting q_t the unit quadratic differential spawning the geodesic between X and $\iota(i+t)$, we have $\lim_{t\to\infty} \iota^*(q_t) = i$, and $\iota_*(i) = q$, so $\lim_{t\to\infty} q_t = q$. The distance between $\iota(i+t)$ and X grows to infinity, so any horocycle path generated by some q based at X converges to q in the visual compactification based at X. Hence, horocycles generated by infusible quadratic differentials converge in the horofunction compactification, which had been previously shown by Jiang and Su [20] and Alberge [2] in the context of surfaces without boundary.

Corollary 5.16 *Let S be a compact surface with possibly nonempty boundary and finitely many marked points and let q be an infusible quadratic differential based at any $X \in \mathcal{T}(S)$. Then the horocycle generated by q converges in the horofunction compactification.*

Proof The horocycle path converges to q in the visual compactification based at X, so by Corollary 5.15 all accumulation points in the horofunction compactification are contained in $\Pi_X^{-1}(q)$. Furthermore, since q is infusible, $\Pi_X^{-1}(q)$ is a singleton, so the horocycle path has a unique accumulation point in the horofunction compactification, and hence it converges. □

On the other hand, Fortier Bourque found some diverging horocycles in the horofunction compactification.

Theorem 5.17 (Fortier Bourque [13, Theorem 1.1]) *Let S be a closed surface of genus g with p marked points, such that $3g + p \geq 5$. Then there is some fusible quadratic differential q based at some basepoint $X \in \mathcal{T}(S)$ such that the associated horocycle path does not converge in the horofunction compactification.*

Corollary 5.15 gives an upper limit on the set of accumulation points, as it has to be contained in $\Pi_X^{-1}(q)$. Furthermore, by Corollary 3.14 we have that a path converges in the horofunction compactification if and only if it converges in each visual compactification. Hence, such a divergent horocycle also diverges in some visual compactification. That is, we get Corollary 1.12. This contrasts with the behavior of Teichmüller rays, which by Corollary 3.7 or [44, Theorem 7] converge in all visual compactifications.

6 Dimension of the fibers

Our first approach in determining the shape of the fibers is looking at the limits of Busemann points, which by Proposition 3.10 give us bounds on the elements of $\Pi^{-1}(q)$. For a given quadratic differential q and a foliation G we define $\mathcal{W}^q(G)$ as the map from measured foliations to $\mathbb{R}$ given by

$$\mathcal{W}^q(G) = \frac{i(G, \cdot)^2}{i(G, H(q))}$$

if $i(G, H(q)) > 0$, and $\mathcal{W}^q(G) = 0$ otherwise. By the extension of Walsh's Corollary 4.6 describing Busemann points in the Gardiner–Masur compactification, we see that the element $\mathcal{E}_q = \Xi^{-1} B_q$ has the form $\sqrt{\sum_i \mathcal{W}^q(V_i)}$, where V_i are the indecomposable components of $V(q)$. Hence, a reasonable path to follow for understanding the limits of Busemann points is understanding the limits of $\mathcal{W}^q$ as q varies.

Lemma 6.1 *Let q_n be a sequence of quadratic differentials on X converging to q, and let V_j^n, $1 \leq j \leq c(n)$ be the indecomposable components of $V(q_n)$. Let G^n be a sequence of nonzero measured foliations of the form $\sum \alpha_j^n V_j^n$, converging to a measured foliation G. Then*

$$\lim_{n\to\infty} \mathcal{W}^{q_n}(G^n) = \mathcal{W}^q(G)$$

if G is nonzero and $\lim_{n\to\infty} \mathcal{W}^{q_n}(G^n) = 0$ if G is zero, where the convergence is pointwise in both cases.

Proof For any measured foliation F we have

$$\mathcal{W}^{q_n}(G^n)(F) = \frac{i(G^n, F)^2}{i(G^n, H(q_n))},$$

so if G is nonzero the lemma follows by continuity of the intersection number.

If G is zero the result follows from applying the same proof than in [44, Lemma 27]. □

Denote $\mathcal{B}$ the set of Busemann points, $\bar{\mathcal{B}}$ its closure and $\bar{\mathcal{B}}(q)$ the intersection $\bar{\mathcal{B}} \cap \Pi^{-1}(q)$. We can use the previous lemma to show that the elements of $\bar{\mathcal{B}}(q)$ satisfy certain properties.

Proposition 6.2 *Let S be a closed surface with possibly marked points, $\xi \in \bar{\mathcal{B}}(q)$ and V_i, $i \in \{1, \dots, k\}$ be the indecomposable components of $V(q)$. Let $x_i = i(V_i, \cdot)/i(V_i, H(q))$. Then, the square of the representation of ξ in the Gardiner–Masur compactification, $(\Xi^{-1}\xi)^2$, is a homogeneous polynomial of degree 2 in the variables x_i, whose coefficients sum to 1.*

Recall that we are using a normalized version of the Gardiner–Masur compactification. Under the projectivized version the sum of the coefficients cannot have any fixed value.

Proof Since the surface does not have boundary, all Busemann points are of the form $B(q')$ for some quadratic differential of unit area q'. Consider a sequence (q_n) such that $B(q_n)$ converges to ξ and q_n converges to q. Let $c(n)$ be the number of indecomposable vertical components of $V(q_n)$, and let V_j^n, $0 < j \le c(n)$ be those components. We know that $c(n)$ is bounded by some number depending on the topology of the surface. Take a subsequence such that $c(n)$ is equal to some constant c and V_j^n converges for each j. The sum $\sum_{j=1}^c V_j^n$ converges as $n \to \infty$ to $\sum_{i=1}^k V_i$, so the limit of each V_j^n has to be of the form $\sum_{i=1}^k \alpha_j^i V_i$. Furthermore, $\sum_{j=1}^c \alpha_j^i = 1$, since

$$\sum_{i=1}^k V_i = V(q) = \lim_{n\to\infty} V(q_n) = \lim_{n\to\infty} \sum_{j=1}^c V_j^n = \sum_{j=1}^c \sum_{i=1}^k \alpha_j^i V_i = \sum_{i=1}^k \left(\sum_{j=1}^c \alpha_j^i \right) V_i.$$

The element associated to the Busemann point $B(q_n)$ in the Gardiner–Masur compactification satisfies

$$\mathcal{E}_{q_n}^2 = \sum_{j=1}^c \mathcal{W}^{q_n}(V_j^n).$$

Hence, applying Lemma 6.1 we get the following expressions for the square of the limit of Busemann points:

$$(\Xi^{-1}\xi)^2 = \sum_{j=1}^c \mathcal{W}^q \left(\sum_{i=1}^k \alpha_j^i V_i \right) = \sum_{j=1}^c \frac{\left(\sum_{i=1}^k \alpha_j^i i(V_i, H(q)) x_i\right)^2}{\sum_{i=1}^k \alpha_j^i i(V_i, H(q))}.$$

That is, we get a homogeneous polynomial of degree 2 in the variables x_i. Since q has unit area, the sum of the coefficients is

$$\sum_{j=1}^c \sum_{i=1}^k \alpha_j^i i(V_i, H(q)) = \sum_{i=1}^k i(V_i, H(q)) = 1,$$

which completes our claim. □

By Proposition 3.9, the Busemann point $B(q)$ gives an upper bound on all functions in $\Pi^{-1}(q)$. While Proposition 3.10 does not give us a lower bound directly, we can use Lemma 2.1 to get one. For a unit area quadratic differential q, let Z_j be the interior parts of $V(q)$, and denote by G_j the union of interior indecomposable components within Z_j. Further, let P_i be the boundary components of $V(q)$. We define the *minimal point* at q as

$$M(q) = \Xi\biggl(\sum_i \mathcal{W}^q(P_i) + \sum_j \mathcal{W}^q(G_j)\biggr)^{1/2}.$$

Proposition 6.3 *Let q be a quadratic differential. Then, for any $\xi \in \Pi^{-1}(q)$, we have*

$$\Xi^{-1}\xi \geq \Xi^{-1}M(q)$$

in the Gardiner–Masur compactification. Furthermore, $M(q) \in \Pi^{-1}(q)$ whenever each G_j has at most two annuli parallel to the boundaries of Z_j with marked points.

In the context of surfaces without boundary the previous result has been also proven by Liu and Shi [27, Lemma 3.10]. In such context we have $M(q) = \Xi i(V(q), \cdot)^2$, which by the proposition is always contained in $\Pi^{-1}(q)$.

The minimality is essentially derived from the following well-known inequality.

Lemma 6.4 (Titu's lemma) *For any positive reals $a_1, \dots, a_n$ and $b_1, \dots, b_n$ we have*

$$\sum_j \frac{a_j^2}{b_j} \geq \frac{\bigl(\sum_j a_j\bigr)^2}{\sum_j b_j}.$$

Proof The inequality can be written as

$$\sum_i b_i \sum_j \frac{a_j^2}{b_j} \geq \biggl(\sum_j a_j\biggr)^2,$$

so the result follows after applying the Cauchy–Schwartz inequality. □

The implication this lemma has for our discussion is that $\mathcal{W}^q(\cdot)$ is convex, in the sense that for any $G = \sum_i G_i$ and any measured foliation F we have

$$\sum_i \mathcal{W}^q(G_i)(F) \geq \mathcal{W}^q(G)(F).$$

Proof of Proposition 6.3 If q is infusible then each G_j is indecomposable, so $M(q) = B(q)$, the fiber $\Pi^{-1}(q)$ has one point and the proposition is satisfied.

Consider then q fusible and $\xi \in \Pi^{-1}(q)$. Let $(x_n) = (R(q_n; t_n)) \subset \mathcal{T}$ converging to ξ. By Lemma 4.5 we have $\Xi^{-1}(h(x_n)) \geq \Xi^{-1}B(q_n)$. Hence, $\Xi^{-1}\xi \geq \liminf_{n\to\infty} \Xi^{-1}B(q_n)$.

Given a measured foliation F, take a subsequence so that

$$\liminf_{n\to\infty} \Xi^{-1} B(q_n)(F) = \lim_{n\to\infty} \Xi^{-1} B(q_n)(F).$$

The foliations $V(q_n)$ converge to $V(q)$, so by Proposition 5.5 for n big enough all boundary components P_i are contained within $V(q_n)$. Hence, for n big enough the foliations $V(q_n)$ can be split to the interior parts Z_j by cutting along the proper arcs. Denote by G_j^n the interior components of the foliation $V(q_n)$ restricted to Z_j. Let $G_{j,k}^n$ be the indecomposable components of G_j^n. The sequence G_j^n converges to G_j, so we can take a subsequence such that each $G_{j,k}^n$ converges to some foliation $G_{j,k}$ with $\sum_k G_{j,k} = G_j$. Applying Lemma 6.1 we have

$$\lim_{n\to\infty} \Xi^{-1} B(q_n)(F) = \lim_{n\to\infty} \sum_i^n \mathcal{W}^{q_n}(P_i) + \sum_j \sum_k \mathcal{W}^{q_n}(G_{j,k}^n) = \sum_i \mathcal{W}^q(P_i) + \sum_j \sum_k \mathcal{W}^q(G_{j,k}).$$

Hence, applying Lemma 6.4 to the second sum we get the first part of the proposition.

To observe that the limit is actually reached we can repeat the proof of Proposition 5.11 and observe that a proper arc for each interior part is enough to approach the foliation whenever each interior part of the foliation has at most two annuli parallel to boundaries with marked points. □

By Corollary 4.8 this lower bound is carried to the horofunction representation and by Proposition 3.9 we have an upper bound. Hence, we have the chain of inequalities

$$M(q) \le \xi \le B(q),$$

for any $\xi \in \Pi^{-1}(q)$. As we see in the next proposition, this chain can be translated as well to the Gardiner–Masur compactification.

Proposition 6.5 *Let $\xi \in \Pi^{-1}(q)$. Then,*

$$\Xi^{-1}\xi \le \Xi^{-1} B(q).$$

Proof We have a sequence of points $R(q_n; t_n)$ converging to ξ, with q_n converging to q. By Lemma 3.3 we have $\xi(R(q;t)) = -t$. Further, $R(q_n;t_n)$ converges in the Gardiner–Masur compactification to the function $f(G)^2 = \lim_{n\to\infty} e^{-2t_n} \operatorname{Ext}_{R(q_n;t_n)}(G)$, and we have $\Xi f(x) = \xi(x)$. Hence,

$$\tfrac{1}{2}\log \frac{f(F)}{\operatorname{Ext}_{R(q;t)}(F)} \le \tfrac{1}{2}\log \sup_{G\in P} \frac{f(G)}{\operatorname{Ext}_{R(q;t)}(G)} = -t.$$

Upon exponentiating and reordering the terms, we get

$$\lim_{n\to\infty} e^{-2t_n} \operatorname{Ext}_{R(q_n;t_n)}(F) = f^2(F) \le e^{-2t} \operatorname{Ext}_{R(q;t)}(F)$$

for all t. Letting $t\to\infty$, the right hand side converges to $(\Xi^{-1}B(q)(F))^2$, so we get the proposition. □

Using these bounds we can further refine the characterization of points in $\Xi^{-1}\Pi^{-1}(q)$.

Proposition 6.6 *Let q be a quadratic differential, let V_i, $i \in \{1,\dots,k\}$ be the indecomposable components of $V(q)$ and let $x_i(F) = i(V_i, F)/i(V_i, H(q))$. Given $f \in \Xi^{-1}\Pi^{-1}(q)$ and $c > 0$ we have, for all $F \in \mathcal{MF}$,*

$$f^2(F) = c^2 + 2c\sum_i i(V_i, H(q))(x_i(F)-c) + \sum_{i,j} O\big((x_i(F)-c)(x_j(F)-c)\big).$$

In particular, as a function of the values $x_i(F)$ at the point $x_i = c$ for all i, $f^2(x_1,\dots,x_k)$ takes value c^2, is differentiable and satisfies

$$\frac{\partial}{\partial x_i} f^2(x_1,\dots,x_k) = 2c\, i(V_i, H(q)).$$

Proof We have that $(\Xi^{-1}M(q))^2 \le f^2 \le (\Xi^{-1}B(q))^2$. Letting $a_i = i(V_i, H(q))$ and $x_i = x_i(\cdot)$, we have by Lemmas 6.4 and 6.3 that $\left(\sum a_i x_i\right)^2 \le (\Xi^{-1}M(q))^2$. Writing the bounds on f^2 in terms of the variables x_i, we obtain

$$\left(\sum a_i x_i\right)^2 \le f^2 \le \sum a_i x_i^2.$$

Adding that $\sum a_i = 1$, we have that f^2 is bounded below by the arithmetic mean, and above by the quadratic mean. Rewriting both sides as a polynomial in $x_i - c$, we get

$$c^2 + 2c\sum a_i(x_i-c) + \left(\sum a_i(x_i-c)\right)^2 \le f^2 \le c^2 + 2c\sum a_i(x_i-c) + \sum a_i(x_i-c)^2,$$

so the first part of the proposition is satisfied. Subbing in the value $x_i(F) = c$ we get the second part. □

By Propositions 3.4 and 3.13 all members of $\Pi^{-1}(q)$ share their values along $R(q;\cdot)$, as well as the directional derivatives at the points of the geodesic. For a given q we have $x_i(\lambda H(q)) = \lambda$ for all i and all $\lambda > 0$. Hence, Proposition 6.6 shows a similar relation for the representations of the elements of $\Pi^{-1}(q)$ in the Gardiner–Masur compactification, as they share their value, as well as some derivatives, at all foliations of the form $\lambda H(q)$.

As shown by Fortier Bourque [13], the Gardiner–Masur boundary contains extremal length functions, so we can use Proposition 6.6 to get some information on the differentials of these functions. Namely, we recover in a more restricted setting the following result, proven in [33, Theorem 1.1].

Theorem 6.7 (Miyachi) *Let G_t, $t \in [0, t_0]$ be a path in the space of measured foliations on X which admits a tangent vector $\dot G_0$ at $t = 0$ with respect to the canonical piecewise linear structure. Then, the extremal length $\mathrm{Ext}(G, X)$ is right-differentiable at $t = 0$ and satisfies*

$$\frac{d}{dt^+}\mathrm{Ext}(G_t, X)\bigg|_{t=0} = 2i(\dot G_0, F_{G_0,X}),$$

where $F_{G_0,X}$ is the horizontal foliation of the Hubbard–Masur differential associated to G_0 on X.

The concrete extremal length functions in the Gardiner–Masur boundary we are going to use are given by the following theorem.

Theorem 6.8 (Fortier Bourque) *Let $\{w_1,\dots,w_k\}$ be weights with $w_i>0$, let $\phi_n=\tau_1^{\lfloor nw_1\rfloor}\circ\dots\circ\tau_k^{\lfloor nw_k\rfloor}$ be a sequence of Dehn multitwist around a multicurve $\{\alpha_1,\dots,\alpha_k\}$ in a surface S and let $X\in\mathcal{T}(S)$. Then the sequence $\phi_n(X)$ converges to*

$$\left[\operatorname{Ext}^{1/2}\left(\sum_{i=1}^k w_i i(F,\alpha_i)\alpha_i, X\right)\right]_{F\in\mathcal{MF}(S)}$$

in the projective Gardiner–Masur compactification as $n\to\infty$.

The precise statement of this result is slightly weaker [13, Corollary 3.4], but the same proof yields this extension.

Fix a multicurve $\{\alpha_1,\dots,\alpha_k\}$, weights $\{w_1,\dots,w_k\}$ and let $\alpha=\sum w_i\alpha_i$. Furthermore, normalize the weights $\{w_1,\dots,w_k\}$ so that there is a unit area quadratic differential q such that $V(q)=\alpha$. Denote by V_i the vertical components of $V(q)$. That is, $V_i=w_i\alpha_i$. We are able to recover Miyachi's formula when $i(V_i,H(q))=w_i$ for all i. The sequence $\phi_n(X)$ converges in the visual compactification based at X to $q\in T_X\mathcal{T}(S)$. By Theorem 6.8 the function $f(F)=\lambda^{1/2}\operatorname{Ext}^{1/2}\big(\sum_{i=1}^k w_i i(F,\alpha_i)\alpha_i,X\big)$ is in $\Xi^{-1}\Pi^{-1}(q)$ for some $\lambda>0$. We have $i(F,\alpha_i)=x_i(F)i(V_i,H(q))/w_i$. So, assuming $i(V_i,H(q))=w_i$ we can write

$$f^2(F)=\lambda\operatorname{Ext}\left(\sum_{i=1}^k x_i(F)V_i,X\right).$$

We have $x_i(H(q))=1$ for all i, so by Proposition 6.6 the value of λ satisfies

$$f^2(H(q))=\lambda\operatorname{Ext}(V(q),X)=1.$$

Since q has unit area, $\operatorname{Ext}(V(q),X)=1$, so $\lambda=1$. Let I be any foliation such that $H(q)+I$ is well defined, and let $F_t=H(q)+tI$. We have

$$f^2(F_t)=\operatorname{Ext}\left(\sum_i V_i+t\sum_i x_i(I)V_i,X\right).$$

Hence, letting $J=\sum x_i(I)V_i$ and $G_t=V(q)+tJ$, we can apply Proposition 6.6 to get

$$\frac{d}{dt^+}\operatorname{Ext}(G_t,X)\bigg|_{t=0}=\sum_i\frac{dx_i}{dt}\bigg|_{t=0}\frac{\partial f^2}{\partial x_i}\bigg|_{x_i=1}=\sum_i\frac{i(V_i,I)}{i(V_i,H(q))}\cdot 2i(V_i,H(q))=2i(V(q),I).$$

On the other hand, applying Miyachi's Theorem 6.7 directly, we get

$$\begin{aligned}\frac{d}{dt^+}\operatorname{Ext}(G_t,X)\bigg|_{t=0}&=2i(H(q),J)=2\sum_i i(H(q),V_i)x_i(I)\\&=2\sum_i i(H(q),V_i)\frac{i(V_i,I)}{i(H(q),V_i)}=2i(V(q),I),\end{aligned}$$

so both expressions coincide, and we have recovered Theorem 6.7 in this rather restricted setting. We would like to note that Proposition 6.6 also gives some information for finding the second derivatives around the point $H(q)$. Namely, the second derivatives cannot diverge to infinity as we approach $H(q)$.

Combining Proposition 6.6 with Proposition 6.2 we get fairly restrictive necessary conditions for the points in $\overline{\mathcal{B}}(q)$ for surfaces without boundary. We shall be using these conditions in Section 7 to prove that Busemann points are not dense in the horoboundary. Now we prove a more straightforward consequence. For a topological space U, denote by $\dim(U)$ its Lebesgue dimension. See the book by Munkres [35, Chapter 5.80] for some background on basic dimension theory. Given an embedding $U \hookrightarrow V$ we have $\dim(U) \le \dim(V)$, so the conditions for the points on $\overline{\mathcal{B}}(q)$ gives us the following result.

Corollary 6.9 *Let S be a surface without boundary. Let q be a quadratic differential such that $V(q)$ has n indecomposable components. Then,*

$$\dim(\overline{\mathcal{B}}(q)) \le \tfrac{1}{2}(n(n-1)).$$

Proof By Proposition 6.2 we have an embedding of $\overline{\mathcal{B}}(q)$ into the space of homogeneous polynomials of degree 2. For a given $\xi \in \overline{\mathcal{B}}(q)$, let $b^{\xi}_{i,j}$ be the coefficient of $x_i x_j$. Adding the restriction $b_{i,j} = b_{j,i}$ we have a coefficient for each possible pair, so the dimension of homogeneous polynomials of degree 2 is equal to the number of possible pairs, that is, $n(n+1)/2$. Furthermore, by Proposition 6.6 we know the value of the first derivatives at $x_i = c$ for all i. For each i this gives us the linear equation

$$\sum_{j \ne i} b^{\xi}_{i,j} + 2b^{\xi}_{i,i} = 2i(V_i, H(q)).$$

These n equations are linearly independent, as $b^{\xi}_{i,i}$ is only contained on the equation related to x_i. As such, the dimension of the coefficients is at most $n(n+1)/2 - n = n(n-1)/2$.

We note that the sum of the coefficients being 1 is the equation we get when summing the n equations given by the derivatives, so we cannot use that to restrict further the dimension. □

Recall that the number of indecomposable components n is bounded in terms of the topology of the surface. Hence, the previous corollary gives us a uniform upper bound on the dimension of $\overline{\mathcal{B}}(q)$. More interestingly, we can also get a lower bound for the dimension of $\overline{\mathcal{B}}(q)$. This allows us to get a lower bound on the dimension of $\Pi^{-1}(q)$. Furthermore, as this is a lower bound, we do not need to restrict ourselves to surfaces without boundary, as the set of Busemann points always contains the set of Busemann points of the form $B(q)$. The bound is obtained by finding a dimensionally big set of different ways to approach a certain q along the boundary and showing that each of these different approaches results in different limits for the associated Busemann points.

Theorem 6.10 *Let S be a surface of genus g with b_m and b_u boundaries with and without marked points respectively and p interior marked points. Then there is some unit quadratic differential q such that*

$$\dim(\overline{\mathcal{B}}(q)) \ge 2\left\lfloor \tfrac{1}{2}(g+b_m) + \tfrac{1}{4}(b_u+p) - \sigma(g, b_u+p)\right\rfloor,$$

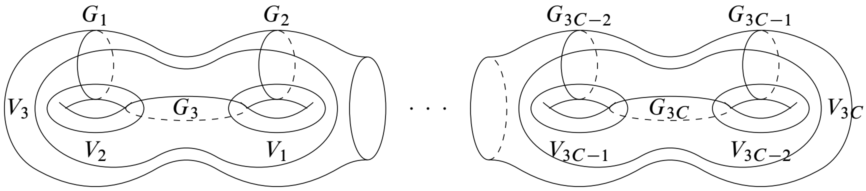


Figure 10: Labeling of the curves when the surface has no boundaries nor marked points. If g is odd then there is an unused handle.

where

$$\sigma = \begin{cases} 0 & \text{if } g \geq 2, \\ \frac{1}{4} & \text{if } g = 1 \text{ and } b_u + p \geq 1, \\ \frac{1}{2} & \text{if } g = 1 \text{ and } b_u + p = 0 \text{ or } g = 0 \text{ and } b_u + p \geq 2, \\ \frac{3}{4} & \text{if } g = 0 \text{ and } b_u + p = 1, \\ 1 & \text{if } g = 0 \text{ and } b_u + p = 0. \end{cases}$$

Proof For simplicity we shall first do the proof in the case where $b_m = b_u = p = 0$, and $g \geq 2$. Let q be the quadratic differential such that $V(q)$ is the union of the closed curves $V_1, \ldots, V_{3C}$ shown in Figure 10, where $C = \lfloor g/2 \rfloor$. Let $U \subset \mathbb{R}^{3C}$ be the space of vectors $(\alpha_1, \alpha_2, \ldots, \alpha_{3C})$ with positive coefficients and such that

$$\alpha_{3k+1} + \alpha_{3k+2} + \alpha_{3k+3} = \frac{1}{C}. \tag{3}$$

Each independent linear restriction reduces the dimension of the set U by 1, so $\dim U = 2C$. Hence, to prove the simplest case of the theorem it suffices to build an injective continuous map from U to $\bar{\mathcal{B}}(q)$.

Choose $\alpha \in U$ and consider the multicurve $\gamma^\alpha = \sum \alpha_i G_i$, where G_i are as in Figure 10. We will shortly show that by applying Dehn twists about the closed curves V_i to γ^α we can get a sequence of multicurves approaching $V(q)$. We can then take the sequences of associated Busemann points, which as we will see converge to distinct points in $\Pi^{-1}(q)$. We will define the injective continuous map from U to $\Pi^{-1}(q)$ by setting it as the limit of the associated sequence of Busemann points, giving us the theorem.

Let τ_i be the Dehn twist around V_i, and let w_i^α be such that

$$w_{3k+1}^\alpha(\alpha_{3k+2} + \alpha_{3k+3}) = w_{3k+2}^\alpha(\alpha_{3k+3} + \alpha_{3k+1}) = w_{3k+3}(\alpha_{3k+1} + \alpha_{3k+2}) = \frac{1}{3C}. \tag{4}$$

Define

$$\phi_n^\alpha = \tau_1^{\lfloor w_1^\alpha n \rfloor} \circ \tau_2^{\lfloor w_2^\alpha n \rfloor} \circ \cdots \circ \tau_{3C}^{\lfloor w_{3C}^\alpha n \rfloor}.$$

For $1 \leq k \leq C$ and $j \in \{1, 2, 3\}$, let

$$F_{k,j}^\alpha = \sum_{i \in \{1,2,3\} - j} w_{3k+i}^\alpha V_{3k+i}.$$

By counting the intersections between the curves V_i and G_i we have that there is some sequence λ_n such that $\lambda_n \phi_n^\alpha G_{3k+j}$ converges to $F_{k,j}^\alpha$ for all k, j as $n \to \infty$. By the conditions on the weights, $\lambda_n \phi_n^\alpha \gamma^\alpha$

converges to $V(q)$. Let q_n^α be the quadratic differential associated to $\lambda_n\phi_n^\alpha\gamma^\alpha$. Since $\lambda_n\phi_n^\alpha\gamma^\alpha$ converges to $V(q)$, we have that q_n converges to q, so all accumulation points of $(B(q_n))$ are in $\Pi^{-1}(q)$. We know that $(\Xi^{-1}B(q_n^\alpha))^2=\sum_i \mathcal{W}^q(\alpha_i\lambda_n\phi_n^\alpha G_i)$, so by Lemma 6.1 we have

$$(\xi^\alpha)^2=\lim_{n\to\infty}(\Xi^{-1}B(q_n^\alpha))^2=\sum_{k=0}^{C-1}\sum_{j\in\{1,2,3\}}\alpha_{3k+j}\mathcal{W}^q(F_{k,j}^\alpha).$$

Define then the map from U to $\Pi^{-1}(q)$ sending $\alpha\in U$ to $\Xi\xi^\alpha\in\Pi^{-1}(q)$. As before, we shall let $x_i:=i(V_i,\cdot)/i(V_i,H(q))=3Ci(V_i,\cdot)$. With this notation we have

$$\mathcal{W}^q(F_{k,j}^\alpha)=\frac{i(F_{k,j}^\alpha,\cdot)^2}{i(F_{k,j}^\alpha,H(q))}=\frac{\left(\sum_{i\notin\{1,2,3\}-j}w_{3k+i}^\alpha x_{3k+i}\right)^2}{3C\sum_{i\notin\{1,2,3\}-j}w_{3k+i}^\alpha}.$$

That is, given α we know precisely the shape of the polynomial ξ^α. Since α has positive coefficients, each of the w_i^α depends continuously on α, so ξ^α depends continuously on α.

It remains to show injectivity. Let $\beta\in U$ be such that $\xi^\alpha=\xi^\beta$. While we have equated two polynomials, we cannot conclude directly that the coefficients are equal, as these cannot be evaluated for arbitrary values. However, we can evaluate at elements of the form $b_1G_{3k+1}+b_2G_{3k+2}+b_3G_{3k+3}$ for $b_1,b_2,b_3\geq 0$, which is enough to prove that ξ^α and ξ^β have the same coefficients.

Equating then the coefficients for $x_{3k+1}x_{3k+2}$, $x_{3k+2}x_{3k+3}$ and $x_{3k+1}x_{3k+3}$ we get

$$\frac{\alpha_{3k+1}w_{3k+2}^\alpha w_{3k+3}^\alpha}{w_{3k+2}^\alpha+w_{3k+3}^\alpha}=\frac{\beta_{3k+1}w_{3k+2}^\beta w_{3k+3}^\beta}{w_{3k+2}^\beta+w_{3k+3}^\beta},$$
$$\frac{\alpha_{3k+2}w_{3k+1}^\alpha w_{3k+3}^\alpha}{w_{3k+1}^\alpha+w_{3k+3}^\alpha}=\frac{\beta_{3k+2}w_{3k+1}^\beta w_{3k+3}^\beta}{w_{3k+1}^\beta+w_{3k+3}^\beta}\quad\text{and}$$
$$\frac{\alpha_{3k+3}w_{3k+1}^\alpha w_{3k+2}^\alpha}{w_{3k+1}^\alpha+w_{3k+2}^\alpha}=\frac{\beta_{3k+3}w_{3k+1}^\beta w_{3k+2}^\beta}{w_{3k+1}^\beta+w_{3k+2}^\beta}.$$

Dividing these equalities and using equations (3) and (4) we get

$$\frac{\alpha_{3k+1}}{\alpha_{3k+2}}\frac{(1/C+\alpha_{3k+2})}{(1/C+\alpha_{3k+1})}=\frac{\beta_{3k+1}}{\beta_{3k+2}}\frac{(1/C+\beta_{3k+2})}{(1/C+\beta_{3k+1})},$$
$$\frac{\alpha_{3k+2}}{\alpha_{3k+3}}\frac{(1/C+\alpha_{3k+3})}{(1/C+\alpha_{3k+2})}=\frac{\beta_{3k+2}}{\beta_{3k+3}}\frac{(1/C+\beta_{3k+3})}{(1/C+\beta_{3k+2})}\quad\text{and}$$
$$\frac{\alpha_{3k+3}}{\alpha_{3k+1}}\frac{(1/C+\alpha_{3k+1})}{(1/C+\alpha_{3k+3})}=\frac{\beta_{3k+3}}{\beta_{3k+1}}\frac{(1/C+\beta_{3k+1})}{(1/C+\beta_{3k+3})}.$$

Rearranging the first equality we have

$$\frac{\alpha_{3k+1}}{\beta_{3k+1}}\frac{\beta_{3k+2}}{\alpha_{3k+2}}=\frac{(1/C+\alpha_{3k+1})}{(1/C+\beta_{3k+1})}\frac{(1/C+\beta_{3k+2})}{(1/C+\alpha_{3k+1})}.\tag{5}$$

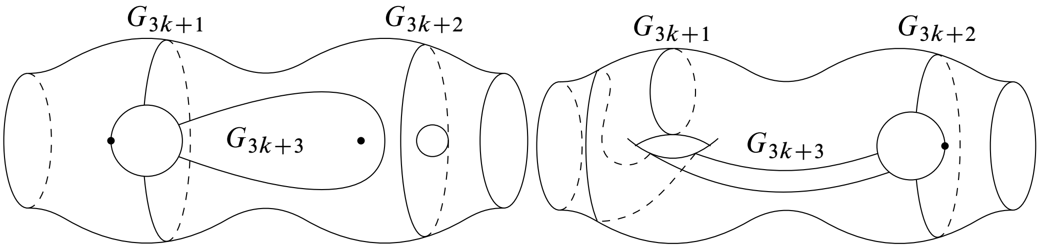


Figure 11: Each pair of marked points and boundary components without marked points can replace a genus, as well as each boundary with marked points.

If

$$\frac{\alpha_{3k+1}}{\beta_{3k+1}} < 1 \quad \text{then} \quad \frac{(1/C + \alpha_{3k+1})}{(1/C + \beta_{3k+1})} > \frac{\alpha_{3k+1}}{\beta_{3k+1}},$$

and if

$$\frac{\alpha_{3k+2}}{\beta_{3k+2}} > 1 \quad \text{then} \quad \frac{(1/C + \alpha_{3k+1})}{(1/C + \beta_{3k+1})} < \frac{\alpha_{3k+1}}{\beta_{3k+1}}.$$

Assume then that $\alpha_{3k+1} < \beta_{3k+1}$. One of the factors of the left hand side of the product in (5) is replaced in the right hand side by a larger value. Hence, the other factor has to be replaced by a smaller value. That is, the inequality $\alpha_{3k+2} < \beta_{3k+2}$ has to be satisfied. Similarly, if $\alpha_{3k+2} < \beta_{3k+2}$ we have $\alpha_{3k+3} < \beta_{3k+3}$. Equation (3) leads to

$$\frac{1}{C} = \alpha_{3k+1} + \alpha_{3k+2} + \alpha_{3k+3} < \beta_{3k+1} + \beta_{3k+2} + \beta_{3k+3} = \frac{1}{C},$$

which is a contradiction. Similarly, $\alpha_{3k+1} > \beta_{3k+1}$ leads to another contradiction, so $\alpha_{3k+1} = \beta_{3k+1}$, which leads to $\alpha = \beta$. Therefore, $\dim(\overline{\mathcal{B}}(q)) \geq \dim(U) = 2\lfloor g/2 \rfloor$.

Assume now that $g \geq 2$ and there are some marked points or boundaries. For each pair of marked points or unmarked boundaries, or for each marked boundary we can repeat the proof with an extra genus, by replacing the curves G_i by the curves shown in Figure 11, and halving the associated weights for w_i, as the curves intersect now twice the vertical components instead of once.

If $g = 1$ we need to place at least one feature at one of the ends to prevent the curve G_1 from being contractible or parallel to a unmarked boundary, so if we have marked points or boundaries without marked points we place these, as boundaries with marked points are more effective at increasing the dimension. In this way we get that if $b_u + p \geq 1$ then

$$\dim(\overline{\mathcal{B}}(q)) \geq 2\left\lfloor \tfrac{1}{2}(g + b_m) + \tfrac{1}{4}(b_u + p - 1) \right\rfloor$$

and if $b_u + p = 0$ then

$$\dim(\overline{\mathcal{B}}(q)) \geq 2\left\lfloor \tfrac{1}{2}(g + b_m - 1) \right\rfloor.$$

Lastly, if $g=0$ we need to place two elements, one at each end. Using the same choice as we took for $g=1$ we get

$$\begin{aligned} \dim(\bar{\mathcal{B}}(q)) &\geq 2\left\lfloor \tfrac{1}{2}(b_m)+\tfrac{1}{4}(b_u+p-2)\right\rfloor && \text{for } b_u+p\geq 2,\\ \dim(\bar{\mathcal{B}}(q)) &\geq 2\left\lfloor \tfrac{1}{2}(b_m-1)\right\rfloor && \text{for } b_u+p=1,\\ \dim(\bar{\mathcal{B}}(q)) &\geq 2\left\lfloor \tfrac{1}{2}(b_m-2)\right\rfloor && \text{for } b_u+p=0. \end{aligned}$$ □

We would like to note that this lower bound is does not look optimal to us. Furthermore, the method used is restricted to getting to the dimension of the closure of Busemann points, so the dimension of the whole fiber may be significantly larger than what could be achieved by refining the strategy from the proof.

7 Nondensity of the Busemann points

7.1 Busemann points are not dense in the horoboundary

By Proposition 6.2 we know that points in the closure of Busemann points are smooth in the Gardiner–Masur representation with respect to certain variables. By showing that at least one point in the horoboundary is not smooth with respect to the corresponding variables we will prove that Busemann points are not dense. The points we use for this analysis are once again the ones found by Fortier Bourque in Theorem 6.8.

Following Fortier Bourque's reasoning, we shall first prove the nondensity for the sphere with five marked points, and then lift to general closed surfaces by using the branched coverings given by the following lemma, found in [15, Lemma 7.1].

Lemma 7.1 (Gekhtman–Markovic) *Let S be a closed surface of genus g with p marked points, such that $3g+p\geq 5$. Then there is a branched cover $\overline{S_{g,p}}\to\overline{S_{0,5}}$ that branches at all preimages of marked points that are not marked and induces an isometric embedding $\mathcal{T}(S_{0,5})\hookrightarrow\mathcal{T}(S_{g,p})$.*

The particular conformal structure given to $S_{0,5}$ is obtained as follows. Let $S^1=\mathbb{R}/\mathbb{Z}$ and let $C=S^1\times[-1,1]$. We obtain a sphere Σ by sealing the top and bottom of C via the relation $(x,y)\sim(-x,y)$ for all $(x,y)\in S^1\times\{-1,1\}$. Let P be set consisting of the five points $(0,\pm 1)$, $\left(\frac{1}{2},\pm 1\right)$ and $(0,0)$. The pair $S=(\Sigma,P)$, where we view Σ as a topological space, is the sphere with five marked points. We get a point X in $\mathcal{T}(S)$ by considering the complex structure on Σ obtained by the construction, using the identity map as our marking.

Let $\alpha(t)=\left(t,\frac{1}{2}\right)$ and $\beta(t)=\left(t,-\frac{1}{2}\right)$ for $t\in S^1$. Denote by τ_α and τ_β the Dehn twists along α and β. By Fortier Bourque's theorem, the sequence $(X_n)=((\tau_\alpha\circ\tau_\beta)^n X)$ converges to a multiple of $\operatorname{Ext}^{1/2}(i(\alpha,\cdot)\alpha+i(\beta,\cdot)\beta,X))$ in the Gardiner–Masur compactification. Furthermore, the sequence (X_n) converges in the visual compactification based at X to the geodesic spawned by the quadratic differential

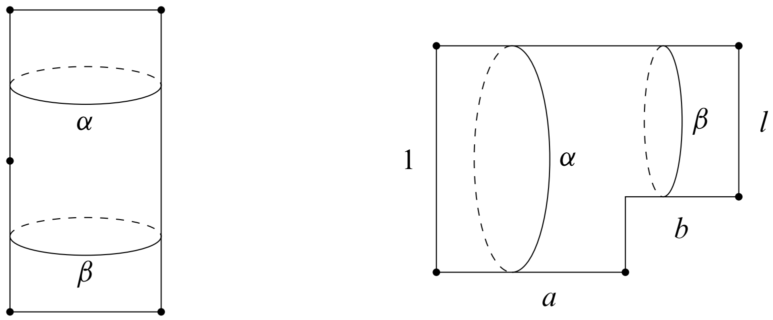


Figure 12: (Lemma 7.2) Left: sphere with five marked points, with curves α and β. We show that the extremal length is not C^2 along the path $\alpha + t\beta,\ t \in [0, t_0]$. Right: doubling of the L–shaped polygon together with the curves α and β.

$q_{\alpha+\beta,X}$. Indeed, as detailed in [13, Section 4], the elements (X_n) diverge to infinity along the horocycle defined by the quadratic differential $q_{\alpha+\beta,X}$. Hence, inside embedded hyperbolic plane associated to $q_{\alpha+\beta,X}$, the sequence (X_n) converges in the visual boundary to the geodesic spawned by $q_{\alpha+\beta,X}$, and so the same occurs in the ambient space. That is, $\Xi\,\mathrm{Ext}^{1/2}(i(\alpha,\cdot)\alpha + i(\beta,\cdot)\beta, X) \in \Pi^{-1}(q_{\alpha+\beta,X})$, so by Proposition 6.2 if we show that $\mathrm{Ext}(i(\alpha,\cdot)\alpha + i(\beta,\cdot)\beta, X)$ is not smooth with respect to the values of $i(\alpha,\cdot)$ and $i(\beta,\cdot)$, then $\Xi\,\mathrm{Ext}^{1/2}(i(\alpha,\cdot)\alpha + i(\beta,\cdot)\beta, X) \notin \bar{\mathcal{B}}(q_{\alpha+\beta,X})$, and hence it is also not in $\bar{\mathcal{B}}$.

Lemma 7.2 *Let $X \in \mathcal{T}(S_{0,5})$ and G_t, $t \in [0, t_0]$ be the foliation $\alpha + t\beta$ on $S_{0,5}$. The map $f(t) := \mathrm{Ext}(G_t, X)$ is not C^2.*

Proof By Miyachi's Theorem 6.7 we have

$$\frac{d}{dt}\mathrm{Ext}(G_t, X) = 2i(\beta, F_{G_t,X}),$$

where we remind that $F_{G_t,X}$ is the horizontal foliation of the unique Hubbard–Masur differential associated to G_t on X. Hence, the Lemma is equivalent to proving that $g(t) = i(\beta, F_{G_t,X})$ is not C^1.

For a general surface finding a precise expression of $F_{G,X}$ is a complicated problem, as the relation established by Hubbard and Masur is not explicit. However, in our case the surface is topologically simple, and one can use Schwartz–Christoffel maps to get a map from G to $F_{G,X}$. In particular, it is possible to show that the sphere with 5 marked points is conformally equivalent to the Riemannian surface obtained by doubling an L–shaped polygon, marking the inner angles as shown in Figure 12, right, and setting certain values for a, b and l. Furthermore, the quadratic differential obtained by dz^2 has α and β as vertical foliations, with weights a and b. Hence $q_{G_t,X}$ is dz^2 on the L–shaped pillowcase where $a = 1$ and $b = t$, so $i(\beta, F_{G_t,X}) = 2l$. Markovic estimated in [30, Section 9] the values of a, b and l around $b = 0$ depending on a common parameter r. Up to rescaling, these values are given by

$$a(r) = a(0) + D_1 r + O(r^2), \quad b(r) = D_2 r + O(r^2) \quad \text{and} \quad l(r) = l(0) + D_3 r \log\frac{1}{r} + o\Big(r \log\frac{1}{r}\Big),$$

where $A(r) = B(r) + O(f(r))$ means $|A(r) - B(r)|/f(r)$ is bounded around $r = 0$, and $A(r) = B(r) + o(f(r))$ means $|A(r) - B(r)|/f(r)$ converges to 0 as r converges to 0.

Rescaling the pillowcase by $1/a(r)$ we see that the parameter t can be expressed as $t(r) = b(r)/a(r)$, and $g(t(r)) = i(\beta, F_{G_t,X}) = 2l(r)/a(r)$. Observing that $t(0) = 0$, we can evaluate the first derivative of $g(t)$ at 0 by evaluating the limit

$$\begin{aligned}\lim_{h\to 0}\frac{g(h)-g(0)}{h} = \lim_{r\to 0}\frac{g(t(r))-g(0)}{t(r)} &= \lim_{r\to 0}\frac{2l(r)/a(r)-2l(0)/a(0)}{b(r)/a(r)}\\ &= 2\lim_{r\to 0}\frac{l(r)-l(0)a(r)/a(0)}{b(r)}\\ &= 2\lim_{r\to 0}\frac{D_3 r\log(1/r)+o(r\log(1/r))-(l(0)D_1/a(0))r}{D_2 r+O(r^2)}\\ &= \infty.\end{aligned}$$

And so, $g(t)$ is not differentiable at $t = 0$, and hence $f(t)$ is not C^2. □

Repeating Fortier Bourque's reasoning we can lift this example to any surface of genus g with p marked points as long as $3g + p \geq 5$. Besides the Gekhtman–Markovic lemma (Lemma 7.1), the other key ingredient for the lifting is the following result.

Lemma 7.3 (Fortier Bourque) *Let $\pi : S_{g,p} \to S_{0,5}$ be a branched cover of degree d and let*

$$\iota : \mathcal{T}(S_{0,5}) \hookrightarrow \mathcal{T}(S_{g,p})$$

be the induced isometric embedding. For any measured foliation F on $S_{0,5}$ and any $X \in \mathcal{T}(S_{0,5})$, we have the identity

$$\operatorname{Ext}(\pi^{-1}(F), \iota(X)) = d\operatorname{Ext}(F, X).$$

Proof Recall that $q_{F,X}$ is the Hubbard–Masur differential associated to γ. We have that $\pi^* q_{F,X} = q_{\pi^{-1}(F),\iota(X)}$, so

$$\operatorname{Ext}(\pi^{-1}(F), \iota(X)) = \int_{\iota(X)} |q_{\pi^{-1}(F),\iota(X)}| = d\int_X |q_{F,X}| = d\operatorname{Ext}(F, X). \qquad \square$$

Lifting the foliation G_t from Lemma 7.2 we get an upper bound for the smoothness of the extremal length.

Theorem 7.4 *Let S be a closed surface of genus g with p marked points, such that $3g + p \geq 5$. Then there exist two nonintersecting multicurves $\hat\alpha$, $\hat\beta$ and some $X \in \mathcal{T}(S)$ such that the map $f(t) := \operatorname{Ext}(\hat\alpha + t\hat\beta, X)$, $t \in [0, t_0]$ is not C^2.*

Proof Since $3g + p \geq 5$ we have a map $\pi\colon S_{g,p} \to S_{0,5}$, with an induced isometric embedding $\iota\colon \mathcal{T}(S_{0,5}) \hookrightarrow \mathcal{T}(S_{g,p})$. By Lemma 7.2 we have two curves $\alpha, \beta \in S_{0,5}$ such that, for any $X \in \mathcal{T}(S_{0,5})$ the map $t \to \operatorname{Ext}(\alpha + t\beta, X)$ is not C^2. Let $\hat{\alpha} = \pi^{-1}(\alpha)$ and $\hat{\beta} = \pi^{-1}(\beta)$. We have $\hat{\alpha} + t\hat{\beta} = \pi^{-1}(\alpha + t\beta)$, so applying Lemma 7.3 we get $\operatorname{Ext}(\hat{\alpha} + t\hat{\beta}, i(X)) = d \operatorname{Ext}(\alpha + t\beta, X)$. By Lemma 7.2 the function $\operatorname{Ext}(\alpha + t\beta, X)$ is not C^2, so we get the theorem. □

Theorem 1.9 is essentially a rephrasing of the previous theorem. Finally, we are able to prove that Busemann points are not dense.

Proof of Theorem 1.8 Let α and β be as in Lemma 7.2. Furthermore, let $\pi\colon S_{g,p} \to S_{0,5}$ and $\iota\colon \mathcal{T}(S_{0,5}) \hookrightarrow \mathcal{T}(S_{g,p})$ be as in Lemma 7.1. For the $X \in \mathcal{T}(S_{0,5})$ described before Lemma 7.2 the sequence $(X_n) = (\tau_\beta \circ \tau_\alpha)^n X$ is contained in the horocycle generated by $q_{\alpha+\beta,X}$ and the distance $d(X_n, X)$ goes to infinity. Therefore (X_n) converges in $\overline{\mathcal{T}(S_{0,5})}^{v}_{X}$ to the geodesic spawned by $q_{\alpha+\beta,X}$. Following Fortier Bourque's reasoning in the proof of [13, Theorem 1.1], using half translation structures, applying the Dehn twist $\tau_\alpha \circ \tau_\beta$ to X is equivalent to applying the shearing transformation

$$h_m = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$

to the half translation structure defined by $q_{\alpha+\beta,X}$. This action commutes with the pull-back coming from the branched cover, so the elements (X_n) are associated with the half translation structure defined by $h_n \pi^*(q_{\alpha+\beta,X})$. These points diverge to infinity along the horocycle defined by $\pi^*(q_{\alpha+\beta,X})$, and so converge in $\overline{\mathcal{T}(S_{g,p})}^{v}_{\iota(X)}$ to the geodesic spawned by $q_{\pi^{-1}(\alpha)+\pi^{-1}(\beta),\iota(X)}$.

Let c_i, $1 \leq i \leq k$, be the components of the half translation structure associated to $\pi^{-1}(\alpha + \beta, X)$. Each c_i covers either α or β with some degree $d_i \in \mathbb{N}$. Hence, each component c_i corresponds to a curve and is a cylindrical with height 1 and circumference d_i. Therefore, if m is the common multiple between all d_i, and γ_i is the curve associated to the component c_i, shifting the flat metric via the matrix h_m is equivalent to performing m/d_i Dehn twists around each curve γ_i. Letting ϕ be the composition of such Dehn twists, we have $\iota(X_{mn}) = \phi^n \iota(X)$. Hence, by Fortier Bourque's Theorem 6.8, in the Gardiner–Masur compactification the sequence $(\iota(X_{mn}))_n$ converges, as $n \to \infty$, to

$$\xi = \left[\operatorname{Ext}^{1/2}\left(\sum_{i=1}^{k} \frac{1}{d_i} i(F, \gamma_i)\gamma_i, \iota(X) \right) \right]_{F \in \mathcal{MF}(S_{g,n})}.$$

Therefore, $\Xi\xi \in \Pi^{-1}(q_{\pi^{-1}(\alpha)+\pi^{-1}(\beta),\iota(X)})$. To see that $\Xi\xi$ is not in $\overline{\mathcal{B}}$ it remains to see that it is not in $\overline{\mathcal{B}}(q_{\pi^{-1}(\alpha)+\pi^{-1}(\beta),\iota(X)})$. We have, $i(c_i, H(q_{\pi^{-1}(\alpha)+\pi^{-1}(\beta),\iota(X)}) = d_i$, so by Proposition 6.2 it remains to prove that there is some path of foliations G_t such that the functions $x_i = i(\gamma_i, G_t)/d_i$ vary smoothly, while the function $f(x_1, \dots, x_k) = \operatorname{Ext}\bigl(\sum_{i=1}^{k} (1/d_i) x_i \gamma_i, \iota(X)\bigr)$ does not. Reorder the curves so there is some $p \geq 1$ such that $\pi^{-1}\alpha = \gamma_1 + \cdots + \gamma_p$ and $\pi^{-1}\beta = \gamma_{p+1} + \cdots + \gamma_k$. It follows from Dehn–Thurston coordinates that for any natural numbers n_j, $1 \leq j \leq k$ there is a multicurve $G_{(n_j)}$ such that $i(G_{(n_j)}, \gamma_i) = n_i$. See, for example, the book by Penner and Harer [38, Theorem 1.2.1]. Allowing renormalizations of the multicurves we get that n_j can be any nonnegative rationals. Finally, doing a limit argument in the space of projective measured foliations we can take n_j to be any nonnegative real numbers. That is, for any $t \geq 0$ there exists a measured foliation G_t such that $i(G_t, \gamma_i) = d_i$ for $i \leq p$, and $i(G_t, \gamma_i) = t d_i$ otherwise. Hence, along such foliations we have $x_i = 1$ for $i \leq p$ and $x_i = t$ otherwise. Therefore, along this path,

$$f(1, \ldots, 1, t, \ldots, t) = \operatorname{Ext}\bigl(\pi^{-1}(\alpha) + t\pi^{-1}(\beta), \iota(X)\bigr),$$

which by Theorem 7.4 is not smooth, as $\pi^{-1}(\alpha)$ and $\pi^{-1}(\beta)$ are the curves used in the proof of the theorem. □

7.2 Busemann points with one indecomposable component are nowhere dense

The Thurston compactification can be build in a similar way as the Gardiner–Masur compactification, by using the hyperbolic length of the curves instead of the extremal length. Let ϕ be the map between $\mathcal{T}(S)$ and $P\mathbb{R}_+^{\mathcal{S}}$ defined by sending $X \in \mathcal{T}(S)$ to the projective vector $[\ell(\alpha, X)]_{\alpha \in \mathcal{S}}$. The pair $\bigl(\phi, \overline{\phi(\mathcal{T}(S))}\bigr)$ defines a compactification, and the boundary is given by the space of projective measured foliations, denoted $\mathcal{PMF}$.

As explained by Miyachi [34], neither the Thurston nor the horofunction compactification is finer than the other one. However, it is possible to get some relation. Let $\mathcal{PMF}^{\mathrm{UE}} \subset \mathcal{PMF}$ be the set of uniquely ergodic foliations. Following the work of Masur [31], $\mathcal{PMF}^{\mathrm{UE}}$ has full Lebesgue measure within $\mathcal{PMF}$. Miyachi [34, Corollary 1] shows that the mapping ϕ on $\mathcal{T}(S)$ can be extended to an homeomorphism f between $\phi(\mathcal{T}(S)) \cup \mathcal{PMF}^{\mathrm{UE}}$ and $h(\mathcal{T}(S)) \cup B_{\mathrm{UE}}$ such that for $x \in \mathcal{T}(S)$ we have $f(\phi(x)) = h$, where B_{UE} are the Busemann points associated to quadratic differentials whose vertical foliation is uniquely ergodic. One might understand this result as stating that the two compactifications are the same almost everywhere with respect to the Lebesgue measure on $\mathcal{PMF}$. As we shall see, the same does not follow with respect to any strictly positive measure on the horoboundary.

The homeomorphism f described by Miyachi is obtained by first defining a map between the boundaries. For a given $x \in \mathcal{T}(S)$, the map on the boundary is denoted $\mathcal{G}_x$, and by its definition we have $\mathcal{G}_x(F) = B(q_{F,x})$, where we recall that $q_{F,x}$ is the quadratic differential on x with $V(q_{F,x}) = F$. Denote by $\mathcal{B}_1$ the set of Busemann points associated to foliations with one indecomposable component. We have $\mathcal{G}_x(\mathcal{PMF}^{\mathrm{UE}}) = \mathcal{B}_{\mathrm{UE}} \subset \mathcal{B}_1$. However, the following is also satisfied.

Theorem 7.5 *Let S be a closed surface of genus g with p marked points, such that $3g + p \geq 5$. Then the set $\mathcal{B}_1$ is nowhere dense in the horoboundary.*

Proof The action of $\mathrm{MCG}(S)$ on $\mathcal{T}(S)$ is extended to the projectivized version of the Gardiner–Masur compactification by $\psi[f(\alpha)]_{\alpha \in \mathcal{S}} = [f(\psi\alpha)]_{\alpha \in \mathcal{S}}$. For any q such that $V(q)$ is an indecomposable measured foliation, $\mathcal{E}_q = \Xi^{-1} B(q) = [i(V(q), \alpha)]_{\alpha \in \mathcal{S}}$, so $\psi\mathcal{E}_q = [i(V(q), \psi(\alpha))]_{\alpha \in \mathcal{S}} = [i(\psi^{-1}(V(q)), \alpha)]_{\alpha \in \mathcal{S}}$.

Hence, $\psi\mathcal{E}_q$ is equal to the representation of the Busemann point in the Gardiner–Masur compactification associated to the quadratic differential with vertical foliation $\psi^{-1}V(q)$, which also is an indecomposable measured foliation. Therefore, $\mathcal{B}_1$ is invariant under the action of $\mathrm{MCG}(S)$, and since $\mathrm{MCG}(S)$ acts by homeomorphisms, the complement of the closure is also invariant.

Let q_0 be a quadratic differential such that there is some $f \in \Xi^{-1}\Pi^{-1}(q_0)$ not in $\Xi^{-1}\overline{\mathcal{B}}$. Such a quadratic differential exists, by Theorem 1.8. By the proof of the theorem, we can assume that $V(q_0)$ is a multicurve. Furthermore, let q be a quadratic differential such that $V(q)$ and $H(q)$ are the stable and unstable foliations respectively of some pseudo-Anosov element $\phi \in \mathrm{MCG}(S)$. It is well known [12, exposé 12] that for any closed curve α we have that $\lambda^{-n}\phi^n(\alpha)$ converges to $\big(i(\alpha, V(q))/i(H(q), V(q))\big)H(q)$, where λ is the stretch factor of ϕ. For any foliation F we have that $\Xi^{-1}M(q_0)(F) = 0$ if and only if $i(V(q_0), F) = 0$, where $M(q_0)$ is the minimal point defined in Section 6. Hence, since $H(q)$ is the unstable foliation of a pseudo-Anosov element and $V(q_0)$ is a multicurve, we have $i(V(q_0), H(q)) \neq 0$, and so $f(H(q)) \geq \Xi^{-1}M(q_0)(H(q)) > 0$. We have $\phi^n[f(\alpha)]_{\alpha\in\mathcal{S}} = [f(\phi^n(\alpha))]_{\alpha\in\mathcal{S}}$. Taking limits and using that the functions in the Gardiner–Masur compactification are homogeneous of degree 1, we get that

$$\lim_{n\to\infty}[\phi^n f(\alpha)]_{\alpha\in\mathcal{S}} = \left[i(\alpha, V(q))\, f\left(\frac{H(q)}{i(V(q), H(q))}\right)\right]_{\alpha\in\mathcal{S}} = [i(\alpha, V(q))]_{\alpha\in\mathcal{S}},$$

Hence, in the normalized version, $\phi^n f$ converges to $i(\cdot, V(q)) = \Xi^{-1}B(q)$, as $V(q)$ is uniquely ergodic and therefore indecomposable. That is, $B(q)$ can be approached through a sequence of elements contained in the complement of the closure of $\mathcal{B}_1$.

Let $B(q')$ be any element in $\mathcal{B}_1$, where q' is any quadratic differential such that $V(q')$ has one indecomposable component. The set of pseudo-Anosov foliations is dense in $\mathcal{MF}(S)$, so we have a sequence of quadratic differentials (q_n) converging to q' with $V(q_n)$ being a pseudo-Anosov foliation. Since q' has one indecomposable component, the convergence is strong, and so $B(q_n)$ converges to $B(q')$. Each $B(q_n)$ can be approached through a sequence of elements contained in the complement of the closure of $\mathcal{B}_1$, so taking a diagonal sequence the same can be said for $B(q')$. □

Corollary 7.6 *Let S be a closed surface of genus g with p marked points, such that $3g + p \geq 5$. Then, for any finite strictly positive measure ν on the horoboundary, the set $\overline{\mathcal{B}}_1$ does not have full ν–measure.*

Proof By Theorem 7.5, the complement of $\overline{\mathcal{B}}_1$ is open and nonempty, so it must have positive ν–measure. □

This last result tells us that the image of Miyachi's homeomorphism does not have full ν–measure within the horoboundary for any strictly positive measure ν. However, as announced in the introduction, any attempt to extend the identity from the Thurston compactification to the horoboundary compactification to a set of full measure within the Thurston compactification results in the same problem. We restate here the result as we shall use the notation for the proof.

Corollary 1.10 *Let ν be any finite strictly positive measure on the horoboundary and let μ be the Lebesgue measure on the Thurston boundary. Furthermore, let ϕ be a map from the Thurston compactification to the horofunction compactification satisfying $\phi|_{\mathcal{T}(S)} = h$, where h is as in (1). Then there is no subset U of the Thurston boundary with full μ–measure such that ϕ is continuous at every point in U and $\phi(U)$ has full ν–measure.*

Proof Assume such a U exists. Choose then a basepoint $x \in \mathcal{T}(S)$ and let $U' = U \cap \mathcal{PMF}^{\mathrm{UE}}$. For each element of $F \in U'$ the associated Hubbard–Masur quadratic differential $q_{F,x}$ satisfies $R(q_{F,x}; t) \to F$ as $t \to \infty$. Hence, since ϕ is continuous at F we have $\phi(F) = B(q_{F,x})$. That is, $\phi(U') \subset \mathcal{B}_1$.

Let $G \in U$. The set $\mathcal{PMF}^{\mathrm{UE}}$ has full μ measure, so $U' = \mathcal{PMF}^{\mathrm{UE}} \cap U$ also has full measure. Hence, since the Lebesgue measure is strictly positive, U' is dense within $\mathcal{PMF}$. Therefore G can be accessed through a sequence $(F_n) \subset U'$. Hence, since ϕ is continuous in G we have $\phi(G) = \lim \phi(F_n)$, so $\phi(U) \subset \bar{\mathcal{B}}_1$ and $\phi(U)$ cannot have full ν–measure. □

Another natural family of measures on the boundary is obtained by considering harmonic measures. Given a nonelementary measure μ on $\mathrm{MCG}(S)$ it is possible to define a random walk (w_n) as the sequence of random variables defined by

$$w_n = g_0 g_1 g_2 \cdots g_n,$$

where g_i are independent, identically distributed random variables on $\mathrm{MCG}(S)$ sampled according to the distribution μ. As proven by Kaimanovich and Masur [22, Theorem 2.2.4], random walks generated by a nonelementary probability measure converges almost surely in Thurston's compactification, so we can define the hitting measure ν in $\mathcal{PMF}$. Furthermore, the walk converges almost surely to uniquely ergodic projective foliations, so we can translate this result to the horofunction compactification in the following way.

Corollary 7.7 *Let μ be a nonelementary measure on* $\mathrm{MCG}(S)$*. Then the associated harmonic measure on the horoboundary is supported in a nowhere dense set.*

Proof For any $x \in \mathcal{T}(S)$ the sequence $(w_n x)$ converges almost surely in Thurston compactification to some $F \in \mathcal{PMF}^{\mathrm{UE}}$. Hence, by [34, Corollary 1], the sequence $(w_n x)$ converges almost surely to the Busemann point generated by a quadratic differential q with $V(q)$ being a multiple of F. Hence, the support of the harmonic measure is contained in $\mathcal{B}_1$, which is nowhere dense by Theorem 7.5. □

8 Topology of the horoboundary

In this section we make some progress towards determining the global topology of the horoboundary. We begin by showing that the minimal point $M(q)$ introduced in Proposition 6.3 serves as a section for the map Π whenever S does not have a boundary. Our main goal for this section is proving the following theorem.

Theorem 8.1 *Let S be a surface of genus g with b_m and b_u boundaries with and without marked points respectively and p interior marked points. Then, the map Π restricted to the boundary has a global continuous section $\partial\overline{\mathcal{T}}^v \to \partial\overline{\mathcal{T}}^h$ if and only if at least one of the two following conditions is satisfied:*

- $b_m = b_u = 0$, *or*
- $2g + 2b_m + b_u + p - \max(1 - b_u, 0) \leq 4$.

The section is given by sending the ray in the direction of q to the point $M(q)$ defined before Proposition 6.3.

Furthermore, if the map does not admit a global section, then it does not admit any local section around some points.

We begin by proving the theorem for surfaces without boundary, as it is significantly easier to prove.

Proposition 8.2 *Let S be a surface without boundary. Then the projection map Π restricted to the boundary admits a global section, given by the map $M : \partial\overline{\mathcal{T}}^v \to \partial\overline{\mathcal{T}}^h$.*

Proof By Proposition 6.3 every preimage $\Pi^{-1}(q)$ contains $M(q)$. We have $M(q) = \Xi\big(i(V(q), \cdot)\big)$, which is continuous, as the map Ξ is continuous. □

The rest of the cases of Theorem 8.1 require a more careful analysis.

Proposition 8.3 *Let S be either*

- *a torus with up at most two unmarked boundaries or interior marked points,*
- *a torus with one marked boundary and one interior marked point,*
- *a sphere with one marked boundary and up to three interior marked points, or*
- *a sphere with two marked boundaries and interior marked point.*

Then the projection map Π restricted to the boundary admits a global section, given by the map $M : \partial\overline{\mathcal{T}}^v \to \partial\overline{\mathcal{T}}^h$.

Proof We shall build the section in the same way we built it in Proposition 8.2, that is, sending q to $M(q)$.

Our first step in the proof is seeing that if $V(q)$ contains a separating proper arc then only one of the two parts separated by the proper arc admit interior components. We shall do this by inspecting each possible case. Assume then that $V(q)$ has a separating proper arc.

If S is a torus with up to two unmarked boundaries or marked points or a torus with one marked boundary and one marked point, then the separating proper arc splits the surface into a torus with a marked boundary and a sphere with a marked boundary and a marked point or unmarked boundary. The latter does not admit an interior component.

If S is a sphere with one marked boundary and up to three boundaries then the separating proper arc splits the surface into two spheres, both with one marked boundary, one of them with two marked points and the other one with one marked point. Again, the latter does not admit an interior component.

Finally, if S is a sphere with two marked boundaries and one marked point or unmarked boundary, the proper arc splits the surface into one sphere with two marked boundaries and a sphere with one marked boundary and one marked point, which again does not admit an interior component.

Take then a sequence of unit quadratic differentials (q_n) converging to q. Let P_i, $i \in \{1, \dots, c\}$ be the boundary components of $V(q)$. Furthermore, denote by G the union of the interior components. By the first part of the proof, all the interior components are contained in the same interior part. We thus have

$$\Xi^{-1} M(q) = \Bigl(\sum_i \mathcal{W}^q(P_i) + \mathcal{W}^q(G)\Bigr)^{1/2}.$$

By Proposition 5.5 all boundary components of $V(q)$ are contained in $V(q_n)$ for n big enough, and all other boundary components of $V(q_n)$, denoted P^n, vanish in the limit. Denote by G^n the union of the interior components of $V(q_n)$. As before, each indecomposable component of G^n is contained in the same interior part, so we have

$$\Xi^{-1} M(q_n) = \Bigl(\sum_i \mathcal{W}^{q_n}(\alpha_i^n P_i) + \mathcal{W}^{q_n}(P^n) + \mathcal{W}^{q_n}(G^n)\Bigr)^{1/2},$$

which converges to $\Xi^{-1} M(q)$. □

Proposition 8.4 *Let S be either*

- *a surface of genus at least two and at least one boundary*,
- *a torus with at least one boundary and two more boundaries or interior marked points*,
- *a torus with at least two boundaries*, *one being marked*, *and possibly interior marked points*,
- *a sphere with at least one boundary*, *and four more boundaries or interior marked points*,
- *a sphere with at least two boundaries*, *one being marked*, *and two interior marked points*, *or*
- *a sphere with at least three boundaries*, *two being marked*, *and possibly interior marked points.*

Then the projection map Π restricted to the boundary does not admit a local section around some points.

Proof We shall prove this by finding a quadratic differential q and sequences (q_n^1) and (q_n^2) converging to q such that their preimages by Π are singletons, but such that $\Pi^{-1}(q_n^1)$ and $\Pi^{-1}(q_n^2)$ converge to different points in $\Pi^{-1}(q)$. If we had a section around q, then its value at q_n^1 and q_n^2 would be $\Pi^{-1}(q_n^1)$ and $\Pi^{-1}(q_n^2)$ respectively, giving us a contradiction.

In all cases the construction will be similar. For q_n^1 we build a foliation with a separating proper arc P such that each of the parts has precisely one interior component consisting of a closed curve, which we denote by G_1 and G_2. Letting the weight of the proper arc diminish to 0 we can get a sequence of quadratic differentials (q_n^1) converging to a quadratic differential q such that $V(q) = G_1 + G_2$. Let $F_n^1 = P + nG_1 + nG_2$, A_n^1 and A the area of the Hubbard–Masur differentials $q_{F_n^1,X}$ and $q_{G_1+G_2,X}$, respectively. Denote $(1/\sqrt{A_n^1})q_{F_n^1,X}$ as q_n^1. These quadratic differentials have unit area, and converge to $(1/\sqrt{A})q_{G_1+G_2,X}$, which we denote by q. By construction, $V(q_n^1)$ is internally indecomposable, so $\Pi^{-1}(q_n^1)$ is a singleton, and

$$\Xi^{-1}\Pi^{-1}(q_n^1) = \left\{\left(\frac{\mathcal{W}^{q_n^1}(P) + n\mathcal{W}^{q_n^1}(G_1) + n\mathcal{W}^{q_n^1}(G_2)}{\sqrt{A_n^1}}\right)^{1/2}\right\}.$$

The sequences $P/\sqrt{A_n^1}$, $nG_1/\sqrt{A_n^1}$ and $nG_2/\sqrt{A_n^1}$ converge, respectively, to 0, $G_1/\sqrt{A}$ and $G_2/\sqrt{A}$. Hence, by Lemma 6.1

$$\text{the sequence } \Pi^{-1}(q_n^1) \quad \text{converges to } \left\{\left(\frac{\mathcal{W}^{q}(G_1) + \mathcal{W}^{q}(G_2)}{\sqrt{A}}\right)^{1/2}\right\}.$$

For building q_n^2 we take a curve γ intersecting G_1 and G_2 at b_1 and b_2 times, where $b_1, b_2 \in \{1,2\}$. Denote by τ_1 and τ_2 the Dehn twists around G_1 and G_2. Let $F_n^2 = \tau_1^{2n/b_1}\tau_2^{2n/b_2}\gamma$ and A_n^2 the area of the Hubbard–Masur differential $q_{F_n^2,X}$. As before, denote by $(1/\sqrt{A_n^2})q_n^2$ the quadratic differentials $(1/\sqrt{A_n^2})q_{F_n^2,X}$. These quadratic differentials have unit area, and converge to q. Furthermore, each $V(q_n^2)$ is a singleton and

$$\Xi^{-1}\Pi^{-1}(q_n^2) = \left\{\left(\frac{\mathcal{W}^{q_n^2}((\tau_1\tau_2)^n\gamma)}{\sqrt{A_n^2}}\right)^{1/2}\right\}.$$

The sequence $(\tau_1\tau_2)^n\gamma/\sqrt{A_n^2}$ converges to $(G_1+G_2)/A$, so by Lemma 6.1

$$\text{the sequence } \Xi^{-1}\Pi^{-1}(q_n^2) \quad \text{converges to } \left\{\left(\frac{\mathcal{W}^{q}(G_1 + G_2)}{\sqrt{A}}\right)^{1/2}\right\},$$

which is different than the limit of $\Xi^{-1}\Pi^{-1}(q_n^1)$.

It remains then to find such a multicurve. For genus at least two we take P to be a separating proper arc such that each of the parts is of genus at least one, and G_1 and G_2 to be noncontractible curves, not parallel to unmarked boundaries on each part, as shown in Figure 13, left.

For the torus we take P to be a separating proper arc with both endpoints in the unmarked boundary, or a marked boundary if there are no unmarked boundaries. Further, we choose the proper arc such that, after cutting along the arc, one part is a torus with one boundary. That is, every other feature of the surface lies in the other part. Then we let G_1 and G_2 be noncontractible curves on each part, as shown in Figure 13, middle.

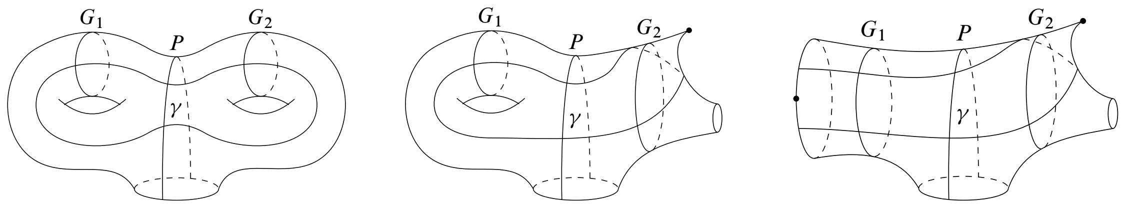


Figure 13: Curves chosen in the proof of Proposition 8.4.

Finally, for the sphere we let P be a separating proper arc with both endpoints on an unmarked boundary, or a marked boundary if there are no boundaries without marked points. Further, we choose the arc such that each interior part has at least either a combination of two marked points or boundaries without marked points, or a boundary with marked points. Hence, each interior part supports an interior component formed by a curve, as shown in Figure 13, right. □

Proof of Theorem 8.1 This is a combination of the results from Propositions 8.2, 8.3 and 8.4. □

By Proposition 1.2 we know that the horoboundary is connected whenever the real dimension of Teichmüller space is at least 2. In the following result we go a bit further, by showing that it is actually path connected.

Proof of Theorem 1.13 Let $x, y \in \partial\overline{\mathcal{T}(S)}^h$. If S does not have boundary then Π has a global section, so we can lift any path between $\Pi(x)$ and $\Pi(y)$ to a path between $M(\Pi(x))$ and $M(\Pi(y))$. Then, since $\Pi^{-1}(x)$ and $\Pi^{-1}(y)$ are path connected, we can connect x to $M(\Pi(x))$ and y to $M(\Pi(y))$ via paths.

If S has boundary we might have to be a bit more careful, as we might not have a global section. However, as we shall see, we can take a path q_t between $\Pi(x)$ and $\Pi(y)$ such that $B(q_t)$ has finitely many discontinuities. Then, since each of the preimages is path connected these discontinuities can be fixed by using paths in the fibers, so we will have a path between x and y.

Choose a boundary component of S, denote by b a curve parallel to that boundary and let $F_x = V(\Pi(x))$. If F_x contains b then all the expressions of the form $(1-t)F_x + tb$ with $t \in [0, 1]$ correspond to foliations on S, which we denote by F_t. Denote by q_t the unit area quadratic differential such that $V(q_t)$ is a multiple of F_t. This defines a continuous path joining $\Pi(x)$ and the unit area quadratic differential associated to a multiple of b. Let V_i be the vertical components of F_x that are not b, and let w_0 be the weight of b in F_x. Then,

$$B(q_t)^2 = \frac{1}{\sqrt{\operatorname{Area}(q_{F_t,X})}}\Big((1-t)\sum \mathcal{W}^{q_t}(V_i) + (t + (1-t)w_0)\mathcal{W}^{q_t}(b)\Big),$$

which gives a continuous path from $B(q_0) \in \Pi^{-1}\Pi(x)$ to $B(q_1) \in \Pi^{-1}(q_1)$. If F_x does not contain b, but b can be added to the foliation then we proceed just as before. Hence, if both x and y result in foliations where b can be added, we create a path by concatenating the paths between x, the Busemann point in $\Pi^{-1}\Pi(x)$, the Busemann point associated to b, the Busemann point in $\Pi^{-1}\Pi(y)$ and y.

If b cannot be added to the foliation F_x then there must be some set P of proper arcs in F_x incident to the boundary component associated to b. Let F'_x be the foliation F_x without the proper arcs P and assume F'_x is nonempty. Denote by F_t the foliations $(1-t)P + (1+t)F'_x$, $t \in [0,1]$, and q_t the unit area quadratic differentials such that $V(q_t)$ is a multiple of F_t. Denoting V_i the vertical components of F'_x, and P_j the proper arcs incident to the boundary component associated to b, we have

$$B(q_t)^2 = \frac{1}{\sqrt{\mathrm{Area}(q_{F_t,X})}}\left((1-t)\sum_j \mathcal{W}^{q_t}(P_j) + (1+t)\sum \mathcal{W}^{q_t}(V_i)\right)$$

for $t < 1$, which is continuous. Furthermore, $\lim_{t\to 1} B(q_t) \in \Pi^{-1}(q_1)$. Hence, we can concatenate a paths between x, the Busemann point in $\Pi^{-1}\Pi(x)$, the limit $\lim_{t\to 1} B(q_t)$, the Busemann point $B(q_1)$ and Busemann point associated to b.

If F'_x is empty we want to add some other components to F_x. If it admits some other component k then we repeat the previous reasoning with $F_t = \left(1-\frac{t}{2}\right)F_x + \frac{t}{2}k$, which does not result in any discontinuity. If F_x does not admit any other component then there must be at least 2 proper arcs incident to the boundary component associated to b, so we choose one of them, denoted p, and repeat the previous reasoning with $F_t = (1-t)F_x + tp$, which does not result in any discontinuity. Finally, we concatenate this last path with the previous paths. □

9 Formulas for limits of extremal lengths

We finish by reframing the bounds we got for the elements of $\Xi^{-1}\Pi^{-1}(q)$ as results regarding limits of extremal lengths, getting in this way some extensions of [44, Theorem 1].

Proposition 9.1 *Let F be a measured foliation, (q_n) be a sequence of unit area quadratic differentials converging to a quadratic differential q and (t_n) be a sequence of real numbers converging to infinity. Then,*

$$(\Xi^{-1}M(q))^2 \le \liminf_{n\to\infty} e^{-2t_n}\,\mathrm{Ext}_{R(q_n;t_n)}(F) \le \limsup_{n\to\infty} e^{-2t_n}\,\mathrm{Ext}_{R(q_n;t_n)}(F) \le (\Xi^{-1}B(q))^2.$$

Proof Take a subsequence such that $e^{-2t_n}\,\mathrm{Ext}_{R(q_n;t_n)}(F)$ converges to the liminf. Furthermore, take a subsequence such that $R(q_n;t_n)$ converge to a point $\xi \in \Pi^{-1}(q)$. By Proposition 6.3 we have $(\Xi^{-1}M(q))^2 \le \xi^2$. Since $e^{-2t_n}\,\mathrm{Ext}_{R(q_n;t_n)}(F)$ converges to $\xi^2(F)$ we have the lower bound. For the upper bound we repeat the process taking the limsup and using Proposition 6.5. □

By noting that $\Xi^{-1}M(q)(F)$ and $\Xi^{-1}B(q)(F)$ evaluate to 0 if and only if $i(V(q),F) = 0$, we get the following corollary, which has also been proven for surfaces without boundary by Liu and Shi [27, Corollary 3.11].

Corollary 9.2 *Let (q_n) be a sequence of unit area quadratic differentials converging to a quadratic differential q, and (t_n) be a sequence of real numbers converging to infinity. Then,*

$$\liminf_{n\to\infty} e^{-2t_n} \operatorname{Ext}_{R(q_n;t_n)}(F) = 0 \iff i(V(q), F) = 0.$$

Proposition 9.1 can be strengthened slightly in the following manner.

Proposition 9.3 *Let (q_n) be a sequence of unit area quadratic differentials converging to a quadratic differential q. Furthermore, denote by V_i^n the indecomposable components of q_n. If the vertical components can be reordered so that for each i we have that V_i^n converges to a foliation V_i, then*

$$\liminf_{n\to\infty} e^{-2t_n} \operatorname{Ext}_{R(q_n;t_n)}(F) \geq \sum_i \mathcal{W}^q(V_i).$$

Proof Take a sequence such that the limit is equal to the liminf, and such that we have convergence in the Gardiner–Masur compactification. Let ξ be the limit in the horofunction compactification. By Lemma 4.5 we have $e^{-2t_n} \operatorname{Ext}_{R(q_n;t_n)}(F) \geq (\Xi^{-1} B(q_n))^2$, and by Corollary 4.6 we have $(\Xi^{-1} B(q_n))^2 = \sum_i \mathcal{W}^{q_n}(V_i^n)$. Hence, by Lemma 6.1, taking limits on both sides we get the proposition. □

If we have strong convergence the upper bound from Proposition 9.1 and the lower bound from Proposition 9.3 coincide, so adding Walsh's formula for the Busemann points [44, Theorem 1] we have a proof of Theorem 1.14.

Finally, the path connectedness of the fibers can be translated to the following result.

Proposition 9.4 *Let (q_n) be a sequence of unit quadratic differentials converging to q, and (t_n) be a sequence of times converging to infinity. Further, for any $F \in \mathcal{MF}$ let $L(F) := \liminf_{n\to\infty} \operatorname{Ext}_{R(q_n;t_n)}(F)$. Then, for any $s \in [L(F), \mathcal{E}_q^2(F)]$ there is a subsequence of $q_{n_k^s}$ and a sequence (t_k^s) of times such that, for any $G \in \mathcal{MF}$ the limit*

$$\lim_{k\to\infty} e^{-2t_k^s} \operatorname{Ext}_{R(q_{n_k^s};t_k^s)}(G)$$

is defined, and if $G = F$ it has value s.

Proof We can take a subsequence such that $\lim_{n\to\infty} \operatorname{Ext}_{R(q_n;t_n)}(F)$ converges to the liminf, and a further subsequence such that we have convergence in the Gardiner–Masur compactification to a point $\Xi^{-1}\xi \in \Xi^{-1}\Pi^{-1}(q)$. By Proposition 3.11 we have a path between ξ and $B(q)$ contained in $\Pi^{-1}(q)$, and hence a path γ between $\Xi^{-1}\xi$ and $\Xi^{-1}B(q)$ contained in $\Xi^{-1}\Pi^{-1}(q)$. By continuity there is a point in that path such that $\gamma_t(F) = \sqrt{s}$, and by the way we constructed γ_t, it is reached by taking a subsequence of $(q_{n_k^s})$ and a sequence (t_k^s) of times converging to infinity. Finally, since γ_t is a point in the Gardiner–Masur compactification approached by $R(q_{n_k^s}^s; t_k^s)$, the value of $\gamma_t(G)^2$ is equal to the limit from the proposition. □

References

[1] **W Abikoff**, *The real analytic theory of Teichmüller space*, Lecture Notes in Math. 820, Springer (1980) MR Zbl

[2] **V Alberge**, *Convergence of some horocyclic deformations to the Gardiner–Masur boundary*, Ann. Acad. Sci. Fenn. Math. 41 (2016) 439–455 MR Zbl

[3] **F Bonahon**, *Shearing hyperbolic surfaces, bending pleated surfaces and Thurston's symplectic form*, Ann. Fac. Sci. Toulouse Math. 5 (1996) 233–297 MR Zbl

[4] **F Bonahon**, *Geodesic laminations with transverse Hölder distributions*, Ann. Sci. École Norm. Sup. 30 (1997) 205–240 MR Zbl

[5] **F Bonahon**, *Transverse Hölder distributions for geodesic laminations*, Topology 36 (1997) 103–122 MR Zbl

[6] **D Burago**, **Y Burago**, **S Ivanov**, *A course in metric geometry*, Grad. Stud. Math. 33, Amer. Math. Soc., Providence, RI (2001) MR Zbl

[7] **Y Chen**, **R Chernov**, **M Flores**, **M F Bourque**, **S Lee**, **B Yang**, *Toy Teichmüller spaces of real dimension 2: the pentagon and the punctured triangle*, Geom. Dedicata 197 (2018) 193–227 MR Zbl

[8] **F Dal'bo**, **M Peigné**, **A Sambusetti**, *On the horoboundary and the geometry of rays of negatively curved manifolds*, Pacific J. Math. 259 (2012) 55–100 MR Zbl

[9] **C J Earle**, *The Teichmüller distance is differentiable*, Duke Math. J. 44 (1977) 389–397 MR Zbl

[10] **P Eberlein**, **B O'Neill**, *Visibility manifolds*, Pacific J. Math. 46 (1973) 45–109 MR

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

[12] **A Fathi**, **F Laudenbach**, **V Poénaru**, *Thurston's work on surfaces*, Math. Notes 48, Princeton Univ. Press (2012) MR Zbl

[13] **M Fortier Bourque**, *A divergent horocycle in the horofunction compactification of the Teichmüller metric*, Ann. Inst. Fourier (Grenoble) 73 (2023) 1885–1902 MR Zbl

[14] **F P Gardiner**, **H Masur**, *Extremal length geometry of Teichmüller space*, Complex Variables Theory Appl. 16 (1991) 209–237 MR Zbl

[15] **D Gekhtman**, **V Markovic**, *Classifying complex geodesics for the Carathéodory metric on low-dimensional Teichmüller spaces*, J. Anal. Math. 140 (2020) 669–694 MR Zbl

[16] **M Gromov**, *Hyperbolic manifolds, groups and actions*, from "Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference", Ann. of Math. Stud. 97, Princeton Univ. Press (1981) 183–213 MR Zbl

[17] **P de la Harpe**, *On Hilbert's metric for simplices*, from "Geometric group theory, I", Lond. Math. Soc. Lect. Note Ser. 181, Cambridge Univ. Press (1993) 97–119 MR Zbl

[18] **F Herrlich**, **G Schmithüsen**, *On the boundary of Teichmüller disks in Teichmüller and in Schottky space*, from "Handbook of Teichmüller theory, I", IRMA Lect. Math. Theor. Phys. 11, Eur. Math. Soc., Zürich (2007) 293–349 MR Zbl

[19] **J Hubbard**, **H Masur**, *Quadratic differentials and foliations*, Acta Math. 142 (1979) 221–274 MR Zbl

[20] **M Jiang**, **W Su**, *Convergence of earthquake and horocycle paths to the boundary of Teichmüller space*, Sci. China Math. 59 (2016) 1937–1948 MR Zbl

[21] **J Kahn**, **K M Pilgrim**, **D P Thurston**, *Conformal surface embeddings and extremal length*, Groups Geom. Dyn. 16 (2022) 403–435 MR Zbl

[22] **V A Kaimanovich**, **H Masur**, *The Poisson boundary of the mapping class group*, Invent. Math. 125 (1996) 221–264 MR Zbl

[23] **S P Kerckhoff**, *The asymptotic geometry of Teichmüller space*, Topology 19 (1980) 23–41 MR Zbl

[24] **T Klein**, **A Nicas**, *The horofunction boundary of the Heisenberg group: the Carnot–Carathéodory metric*, Conform. Geom. Dyn. 14 (2010) 269–295 MR Zbl

[25] **B Lemmens**, **C Walsh**, *Isometries of polyhedral Hilbert geometries*, J. Topol. Anal. 3 (2011) 213–241 MR Zbl

[26] **A Lenzhen**, **B Modami**, **K Rafi**, *Teichmüller geodesics with d–dimensional limit sets*, J. Mod. Dyn. 12 (2018) 261–283 MR Zbl

[27] **L Liu**, **Y Shi**, *On the properties of various compactifications of Teichmüller space*, Monatsh. Math. 198 (2022) 371–391 MR Zbl

[28] **L Liu**, **W Su**, *The horofunction compactification of the Teichmüller metric*, from "Handbook of Teichmüller theory, IV", IRMA Lect. Math. Theor. Phys. 19, Eur. Math. Soc., Zürich (2014) 355–374 MR Zbl

[29] **J Maher**, **G Tiozzo**, *Random walks on weakly hyperbolic groups*, J. Reine Angew. Math. 742 (2018) 187–239 MR Zbl

[30] **V Markovic**, *Carathéodory's metrics on Teichmüller spaces and L–shaped pillowcases*, Duke Math. J. 167 (2018) 497–535 MR Zbl

[31] **H Masur**, *Interval exchange transformations and measured foliations*, Ann. of Math. 115 (1982) 169–200 MR Zbl

[32] **H Miyachi**, *Teichmüller rays and the Gardiner–Masur boundary of Teichmüller space*, Geom. Dedicata 137 (2008) 113–141 MR Zbl

[33] **H Miyachi**, *A differential formula for extremal length*, from "In the tradition of Ahlfors–Bers, VI", Contemp. Math. 590, Amer. Math. Soc., Providence, RI (2013) 137–152 MR Zbl

[34] **H Miyachi**, *Teichmüller rays and the Gardiner–Masur boundary of Teichmüller space, II*, Geom. Dedicata 162 (2013) 283–304 MR Zbl

[35] **J R Munkres**, *Topology: a first course*, Prentice-Hall, Englewood Cliffs, NJ (1975) MR Zbl

[36] **K Ohshika**, *Compactifications of Teichmüller spaces*, from "Handbook of Teichmüller theory, IV", IRMA Lect. Math. Theor. Phys. 19, Eur. Math. Soc., Zürich (2014) 235–254 MR Zbl

[37] **A Papadopoulos**, *Deux remarques sur la géométrie symplectique de l'espace des feuilletages mesurés sur une surface*, Ann. Inst. Fourier (Grenoble) 36 (1986) 127–141 MR Zbl

[38] **R C Penner**, **J L Harer**, *Combinatorics of train tracks*, Ann. of Math. Stud. 125, Princeton Univ. Press (1992) MR Zbl

[39] **M A Rieffel**, *Group C^*–algebras as compact quantum metric spaces*, Doc. Math. 7 (2002) 605–651 MR Zbl

[40] **K Strebel**, *Quadratic differentials: a survey*, from "On the work of Leonhard Euler", Birkhäuser, Basel (1984) 219–238 MR Zbl

[41] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. 19 (1988) 417–431 MR Zbl

[42] **C Walsh**, *The horoboundary and isometry group of Thurston's Lipschitz metric*, from "Handbook of Teichmüller theory, IV", IRMA Lect. Math. Theor. Phys. 19, Eur. Math. Soc., Zürich (2014) 327–353 MR Zbl

[43] **C Walsh**, *The horofunction boundary and isometry group of the Hilbert geometry*, from "Handbook of Hilbert geometry", IRMA Lect. Math. Theor. Phys. 22, Eur. Math. Soc., Zürich (2014) 127–146 MR Zbl

[44] **C Walsh**, *The asymptotic geometry of the Teichmüller metric*, Geom. Dedicata 200 (2019) 115–152 MR Zbl

[45] **C Webster**, **A Winchester**, *Boundaries of hyperbolic metric spaces*, Pacific J. Math. 221 (2005) 147–158 MR Zbl

School of Mathematics and Statistics, University of Glasgow
Glasgow, United Kingdom

aitor@azemar.xyz

https://aitor.azemar.xyz

Received: 28 November 2022 Revised: 3 November 2023

Algebraic & Geometric Topology **24**:7 (2024) 3985–3996
DOI: 10.2140/agt.2024.24.3985
Published: 9 December 2024

Vector fields on noncompact manifolds

TSUYOSHI KATO
DAISUKE KISHIMOTO
MITSUNOBU TSUTAYA

Let M be a noncompact connected manifold with a cocompact and properly discontinuous action of a discrete group G. We establish a Poincaré–Hopf theorem for a bounded vector field on M satisfying a mild condition on zeros. As an application, we show that such a vector field must have infinitely many zeros whenever G is amenable and the Euler characteristic of M/G is nonzero.

57R25, 58K45

1 Introduction

Let M be a noncompact connected manifold. Then M admits a nonvanishing vector field as M admits a vector field with isolated zeros which can be swept out to infinity. However, the resulting nonvanishing vector field is not satisfactory because it may not be bounded in the following sense. A vector field v on a Riemannian manifold is *bounded* if both $|v|$ and $|dv|$ are bounded, where dv denotes the derivative of v. Note that our boundedness condition is different from the one in [Cima and Llibre 1990] and its related works. Bounded vector fields appear in the study of manifolds of bounded geometry. Now we ask whether or not a nonvanishing bounded vector field exists on M. Weinberger [2009, Theorem 1] proved that a manifold M of bounded geometry has a nonvanishing vector field v with $|v|$ constant and $|dv|$ bounded if and only if the Euler class of M in the bounded de Rham cohomology $\widehat{H}^*(M)$ is trivial. As one may think of the Poincaré–Hopf theorem as a refinement of the Euler class criterion for the existence of a nonvanishing vector field on a compact orientable manifold, it is natural to ask whether or not one can establish the Poincaré–Hopf theorem for bounded vector fields on noncompact manifolds of bounded geometry.

A typical manifold of bounded geometry is a covering space of a compact manifold, which we equip with a lift of a metric on the base compact manifold. We say that an action of a group G on a space X is properly discontinuous if every point $x \in X$ has a neighborhood U such that $gU \cap U \neq \varnothing$ implies $g = 1$. Equivalently, the quotient map $X \to X/G$ is a covering. So we consider a connected noncompact manifold M on which a cocompact and properly discontinuous action of a group G is given, and will establish the Poincaré–Hopf theorem for a bounded vector field on M. To state it, we set notation. Let $\ell^\infty(G)$ denote the vector space of bounded functions $G \to \mathbb{R}$, and let G act on $\ell^\infty(G)$ from the left by

$$(g\phi)(h) = \phi(hg)$$

for $g, h \in G$ and $\phi \in \ell^\infty(G)$. We define the module of coinvariants of $\ell^\infty(G)$ by

$$\ell^\infty(G)_G = \ell^\infty(G)/\langle \phi - g\phi \mid \phi \in \ell^\infty(G),\, g \in G\rangle,$$

where $\langle S\rangle$ denotes the vector subspace of $\ell^\infty(G)$ generated by a subset $S \subset \ell^\infty(G)$. Let $\mathbb{1} \in \ell^\infty(G)$ denote the constant function with value 1. Let $D \subset M$ be a fundamental domain (its definition is given in Section 3). We will define the index $\mathrm{ind}(v)$ of a bounded vector field v on M as an element of $\ell^\infty(G)_G$, and will prove

$$\mathrm{ind}(v)(g) = \sum_{x \in \mathrm{Zero}(v) \cap gD} \mathrm{ind}_x(v)$$

whenever v is strongly tame, which is a mild condition on the zeros of v defined in Definition 5.5, where $\mathrm{ind}_x(v)$ denotes the local index of a vector field v at the zero $x \in \mathrm{Zero}(v)$. Here, the above equality means that there is a representative $\phi \in \ell^\infty(G)$ of $\mathrm{ind}(v) \in \ell^\infty(G)_G$ such that $\phi(g)$ is the right-hand side of the equality. Now we are ready to state the Poincaré–Hopf theorem for a bounded vector field on M.

Theorem 1.1 *Let M be a noncompact connected manifold equipped with a cocompact and properly discontinuous action of a group G such that M/G is orientable. If a vector field v on M is strongly tame and bounded, then*

$$\mathrm{ind}(v) = \chi(M/G)\mathbb{1} \quad \text{in } \ell^\infty(G)_G.$$

As an application of Theorem 1.1, we will prove:

Theorem 1.2 *Let M and G be as in Theorem 1.1. If G is amenable and $\chi(M/G) \neq 0$, then every tame bounded vector field on M must have infinitely many zeros.*

Let M and G be as in Theorem 1.1. Then by the abovementioned result of Weinberger [2009, Theorem 1] together with [Attie et al. 1992], one can deduce that a vector bundle v on M with $|v|$ constant and $|dv|$ bounded must have a zero. But one cannot deduce further information on zeros, such as their numbers, from these results. As an application of Theorem 1.2 we will get the following result, where tameness of a diffeomorphism is defined in Definition 5.6 in an analogous way to tameness of a vector field:

Corollary 1.3 (cf. [Weinberger 2009, Corollary to Theorem 1]) *Let M and G be as in Theorem 1.1. If G is amenable and $\chi(M/G) \neq 0$, then every tame diffeomorphism of M which is C^1 close to the identity map must have infinitely many fixed points.*

Example 1.4 Let L be the noncompact surface called Jacob's ladder, a surface with infinite genus and two ends, which admits an infinite cyclic covering map onto the closed oriented surface of genus 2. Then we can apply Corollary 1.3 to L, and conclude that any tame diffeomorphism of L which is C^1 close to the identity map must have infinitely many zeros. This can be easily generalized to the infinite connected sum $M = \#^\infty N$ of a closed connected oriented even-dimensional manifold N with $\chi(N) \neq 2$.

We briefly describe the strategy of our proof, as well as some of the tools we exploit. Let M and G be as in Theorem 1.1. Recall that the Poincaré–Hopf theorem for a compact manifold can be proved by using a suitable integral in top-dimensional de Rham cohomology. Motivated by the compact case, we will define the integral

$$\int_M : \hat{H}^n(M) \to \ell^\infty(G)_G, \tag{1}$$

where $\dim M = n$, and will prove Theorem 1.1 by using it similarly to the compact case. So our approach is an extension of the classical case by means of $\ell^\infty(G)_G$. However, unlike in the compact case, the target module $\ell^\infty(G)_G$ of the integral has some interesting algebraic properties we will use to deduce Theorem 1.2.

Let us observe possible connections of our results to other contexts. Our results could be connected to the index theory on open manifolds by Roe [1988]. More specifically, our index could be related to the index of the Dirac operator

$$d + d^* : \hat{\Omega}^{\mathrm{even}}(M) \to \hat{\Omega}^{\mathrm{odd}}(M)$$

on the bounded de Rham complex $\hat{\Omega}^*(M)$, which lives in the operator K–theory $K_*(C_u^*(|G|))$ of the uniform Roe algebra $C_u^*(|G|)$. There is another possible connection. In [Kato et al. 2024], the pushforward of a vector bundle on M to M/G is defined, and its structure group is the group of unitary operators with finite propagation on the Hilbert space of square integrable functions $G \to \mathbb{C}$. On the other hand, as in [Kato et al. 2023a; 2023b], the module of coinvariants of bounded functions $\mathbb{Z} \to \mathbb{Z}$ appears in the homotopy groups of such a group of unitary operators of finite propagation for $G = \mathbb{Z}$. Then our results could be connected to the obstruction theory for the pushforward of TM onto M/G.

As mentioned in [Block and Weinberger 1992], there is an isomorphism $\hat{H}^n(M) \cong H_0^{\mathrm{uf}}(M)$ (see [Attie and Block 1998] for the proof), where $\dim M = n$ and $H_*^{\mathrm{uf}}(M)$ denotes the uniformly finite homology of M as in [Block and Weinberger 1992]. Since uniformly finite homology is a quasi-isometry invariant, there is an isomorphism $H_0^{\mathrm{uf}}(M) \cong H_0^{\mathrm{uf}}(G)$. On the other hand, as in [Brodzki et al. 2012], there is an isomorphism $H_0^{\mathrm{uf}}(G) \cong \ell^\infty(G)_G$. Then we get an isomorphism

$$\hat{H}^n(M) \cong \ell^\infty(G)_G.$$

However, this isomorphism is not explicit as it is given by a zigzag of several isomorphisms. We believe that the integral (1) gives a direct and explicit description of this isomorphism. Our intuition relies on the case $M = \mathbb{R}$ and $G = \mathbb{Z}$, which we treat in Proposition 4.5, and we propose the following:

Conjecture 1.5 The integral (1) is an isomorphism.

Throughout this paper manifolds will be smooth and without boundary, unless otherwise specified, and group actions on manifolds will be smooth too.

Acknowledgments The authors were partially supported by JSPS KAKENHI grants JP22H01123 (Kato), JP22K03284 and JP19K03473 (Kishimoto), and JP22K03317 (Tsutaya). The authors deeply appreciate the referees' useful advice and comments, which especially improved Section 2.

2 Module of coinvariants

In this section, we collect properties of the module of coinvariants $\ell^\infty(G)_G$ that we are going to use. Block and Weinberger [1992] introduced the uniformly finite homology $H_*^{\mathrm{uf}}(X)$ of a metric space X, and showed its basic properties. Later, Brodzki, Niblo, and Wright [Brodzki et al. 2012] studied amenability of discrete groups by using the uniformly finite homology, where every discrete group will be equipped with a word metric. They observed that if G is finitely generated, then the uniformly finite chain complex $C_*^{\mathrm{uf}}(G)$ is naturally isomorphic to the chain complex $C_*(G;\ell^\infty(G))$. Then since $H_0(G,\ell^\infty(G)) = \ell^\infty(G)_G$, there is a natural isomorphism

$$H_0^{\mathrm{uf}}(G) \cong \ell^\infty(G)_G \tag{2}$$

whenever G is finitely generated.

Proposition 2.1 *Let G and H be finitely generated groups. Then a quasi-isometric homomorphism $G \to H$ induces an isomorphism*

$$\ell^\infty(H)_H \xrightarrow{\cong} \ell^\infty(G)_G.$$

Proof By [Block and Weinberger 1992, Corollary 2.2], a quasi-isometric homomorphism $G \to H$ induces an isomorphism $H_*^{\mathrm{uf}}(G) \xrightarrow{\cong} H_*^{\mathrm{uf}}(H)$. Then the statement follows from (2). □

Corollary 2.2 *If G is a finite group, then*

$$\ell^\infty(G)_G \cong \mathbb{R}.$$

Proof Let 1 denote the trivial group. Since G is finite, the inclusion $1 \to G$ is a quasi-isometry. Then since $\ell^\infty(1)_1 \cong \mathbb{R}$, the statement is proved by Proposition 2.1. □

Proposition 2.3 *Let G be a finitely generated infinite group, and let $\phi \in \ell^\infty(G)$. If $\phi(g) = 0$ for all but finitely many $g \in G$, then ϕ is zero in $\ell^\infty(G)_G$.*

Proof Whyte [1999, Theorem 7.6] gave a necessary and sufficient condition for an element of $C_0^{\mathrm{uf}}(G)$ to be trivial in $H_0^{\mathrm{uf}}(G)$. Through the natural isomorphism $C_0^{\mathrm{uf}}(G) \cong C_0(G;\ell^\infty(G)) = \ell^\infty(G)$, this condition is stated as follows: an element $\phi \in \ell^\infty(G)$ is zero in $\ell^\infty(G)_G$ if and only if there are $C > 0$ and $r > 0$ such that for any finite subset $S \subset G$,

$$\left|\sum_{g\in S} \phi(g)\right| \le C \cdot \#\{g \in G \mid 0 < d(g,S) \le r\},$$

where d denotes a word metric of G. If G is infinite, then for any nonempty finite subset $S \subset G$, we have $\#\{g \in G \mid 0 < d(g,S) \le 1\} \ge 1$. Suppose $\phi \in \ell^\infty(G)$ satisfies $\phi(g) = 0$ for all but finitely many $g \in G$. Then if we set $C = \sum_{g\in G} |\phi(g)|$ and $r = 1$, the above inequality holds for any finite subset $S \subset G$, and so ϕ is zero in $\ell^\infty(G)_G$. □

Recall that a mean on a group G is a linear map

$$\mu\colon \ell^\infty(G) \to \mathbb{R}$$

such that $\mu(\mathbb{1}) = 1$ and $\mu(\phi) \geq 0$ whenever $\phi(g) \geq 0$ for all $g \in G$, where $\mathbb{1} \in \ell^\infty(G)$ denotes the constant function with value 1 as in Section 1. A group G is *amenable* if it admits a G–invariant mean. The proof of [Block and Weinberger 1992, Theorem 3.1] together with (2) implies the following:

Proposition 2.4 *For a finitely generated group G, the following statements are equivalent*:

(i) *G is amenable.*

(ii) $\ell^\infty(G)_G \neq 0$.

(iii) *$\mathbb{1} \in \ell^\infty(G)$ is nonzero in $\ell^\infty(G)_G$.*

3 Basic properties of fundamental domains

In this section, we define a fundamental domain of a manifold with a free group action, and show its basic properties. Throughout this section, let M be a connected manifold of dimension n, possibly with boundary, on which a cocompact and properly discontinuous action of a group G is given. Since G is a quotient of the fundamental group of a compact manifold M/G, which is finitely generated, G is finitely generated.

We define a *fundamental domain* D of M as the closure of an open set of M such that

$$M = \bigcup_{g \in G} gD \quad \text{and} \quad \mathrm{Int}(D) \cap \mathrm{Int}(gD) = \varnothing$$

for all $1 \neq g \in G$. Remark that D need not be connected. A manifold M admits a fundamental domain. Indeed, given a triangulation of M/G, we can lift it to get a triangulation of M such that the G–action is free and simplicial. We choose one lift of the interior of each maximal simplex of M/G to M, so the closure of the union of these open simplices of M is a fundamental domain of M. We choose such a fundamental domain, so that D is a simplicial complex such that each $D \cap gD$ is a subcomplex of D and

$$\partial D = \Big(\bigcup_{1 \neq g \in G} D \cap gD\Big) \cup (D \cap \partial M). \tag{3}$$

If $gD \cap hD$ is $(n-1)$–dimensional, then we call it a *facet* of gD (and hD). We also call $gD \cap \partial M$ a facet of gD when $\partial M \neq \varnothing$. Then the boundary of D is the union of its facets. Clearly the G–action on M restricts to ∂M, and $D \cap \partial M$ is a fundamental domain of ∂M.

We construct a generating set of G by using a fundamental domain D. Let S be a subset of G consisting of elements $g \in G$ such that $D \cap gD$ is a facet of D.

Proposition 3.1 *The set S is a symmetric generating set of G.*

Proof Let $g \in G$ and $x \in \mathrm{Int}(D)$. Then gx belongs to $\mathrm{Int}(gD)$, and so since M is connected there is a path ℓ from x to gx which passes $g_0 D, g_1 D, \dots, g_k D$ in order for $1 = g_0, g_1, \dots, g_{k-1}, g_k = g \in G$ such that $g_i D \cap g_{i+1} D$ is a facet and $\ell \cap g_i D \cap g_{i+1} D$ is a single point sitting in the interior of a facet $g_i D \cap g_{i+1} D$ of $g_i D$ for $i = 0, 1, \dots, k-1$. Since $g_i D \cap g_{i+1} D = g_i(D \cap g_i^{-1} g_{i+1} D)$ is a facet of $g_i D$, $D \cap g_i^{-1} g_{i+1} D$ is a facet of D, implying $g_i^{-1} g_{i+1} \in S$. Thus since

$$g = g_k = (g_0^{-1} g_1)(g_1^{-1} g_2) \cdots (g_{k-1}^{-1} g_k),$$

we obtain that S is a generating set of G. If $g \in S$, then $g(D \cap g^{-1} D) = gD \cap D$ is a facet of D, and so $D \cap g^{-1} D$ is a facet of D too. Hence $g^{-1} \in S$, that is, S is symmetric, completing the proof. □

Corollary 3.2 *There is a partition $S = S^+ \sqcup S^- \sqcup S^0$ such that $(S^+)^{-1} = S^-$ and $(S^0)^2 = \{1\}$.*

Proof Let S^0 be the subset of S consisting of elements of order 2. Then the statement follows because S is symmetric. □

Let $S^+ = \{s_1, \dots, s_k\}$ and $S^0 = \{t_1, \dots, t_l\}$, where S^+ and S^0 are finite because G is finitely generated as mentioned above. We put

$$E = D \cap \partial M, \quad F_i^+ = D \cap s_i D, \quad F_i^- = D \cap s_i^{-1} D, \quad \text{and} \quad F_j^0 = D \cap t_j D$$

for $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, l$.

Lemma 3.3 *The facets of D are $E, F_1^+, \dots, F_k^+, F_1^-, \dots, F_k^-, F_1^0, \dots, F_l^0$.*

Proof The statement follows from Corollary 3.2. □

We consider an orientation of a facet of gD:

Lemma 3.4 *Suppose that M is oriented. If $F = gD \cap hD$ is a facet for $g, h \in G$, then the orientations of F induced from gD and hD are opposite.*

Proof An outward vector of gD rooted at F is an inward vector of hD. Then the statement follows. □

4 The integral in bounded cohomology

In this section, we define the integral in bounded cohomology. Let M be a connected Riemannian manifold of dimension n, possibly with boundary. As in [Roe 1988], we say that a differential form ω on M is *bounded* if both $|\omega|$ and $|d\omega|$ are bounded. Let $\widehat{\Omega}^p(M)$ denote the set of bounded p–forms on M. Then by definition, $\widehat{\Omega}^*(M)$ is closed under differential, and so it is a differential graded algebra. We define the *bounded de Rham cohomology* of M as the cohomology of $\widehat{\Omega}^*(M)$, which we denote by $\widehat{H}^*(M)$. We record the following obvious fact:

Lemma 4.1 *If a map $f : M \to N$ between manifolds has bounded derivative, then it induces a map $f^* : \widehat{\Omega}^*(N) \to \widehat{\Omega}^*(M)$.*

Now we consider a cocompact and properly discontinuous action of a discrete group G on a manifold M, and choose a fundamental domain $D \subset M$. A Riemannian metric of M will be chosen to be the lift of a Riemannian metric of M/G. We assume that M/G is oriented. Then in particular, the fundamental domain D is oriented. We define the *integral* of a bounded differential form on M by

$$\int_M : \widehat{\Omega}^n(M) \to \ell^\infty(G), \quad \Big(\int_M \omega\Big)(g) = \int_{gD} \omega. \tag{4}$$

We may think of the above integral as the external transfer of the covering $M \to M/G$. Note that we can similarly define the integral for ∂M by using a fundamental domain $D \cap \partial M$ of ∂M. We prove Stokes' theorem:

Proposition 4.2 *For $\omega \in \widehat{\Omega}^{n-1}(M)$, we have*

$$\int_M d\omega = \int_{\partial M} \omega \quad \text{in } \ell^\infty(G)_G.$$

Proof We consider the facets of D described in Lemma 3.3. Define $\phi_i^\pm, \phi_j^0 \in \ell^\infty(G)$ by

$$\phi_i^\pm(g) = \int_{gF_i^\pm} \omega \quad \text{and} \quad \phi_j^0(g) = \int_{gF_i^0} \omega$$

for $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, l$, where the orientations of $gF_i^\pm$ and gF_i^0 are induced from gD. Then by Lemma 3.3 and the usual Stokes' theorem, we have

$$\int_{gD} d\omega = \int_{gE} \omega + \sum_{i=1}^{k} (\phi_i^+(g) + \phi_i^-(g)) + \sum_{j=1}^{l} \phi_j^0(g),$$

where the orientation of gE is induced from gD. Since $gF_i^- = gs_i^{-1}F_i^+$, it follows from Lemma 3.4 that $\phi_i^-(g) = -\phi_i^+(gs_i^{-1})$. Then

$$\phi_i^+ + \phi_i^- = \phi_i^+ - s_i^{-1}\phi_i^+.$$

Quite similarly,

$$\phi_j^0 = \tfrac{1}{2}(\phi_j^0 - t_j^{-1}\phi_j^0).$$

Thus since $E = D \cap \partial M$ is a fundamental domain of ∂M, we obtain

$$\int_M d\omega = \int_{\partial M} \omega + \sum_{i=1}^{k} (\phi_i^+ - s_i^{-1}\phi_i^+) + \sum_{j=1}^{l} \tfrac{1}{2}(\phi_j^0 - t_j^{-1}\phi_j^0). \qquad \square$$

The following is immediate from Proposition 4.2:

Corollary 4.3 *If M is without boundary, then the integral (4) induces a map*

$$\int_M : \widehat{H}^n(M) \to \ell^\infty(G)_G.$$

By considering n–forms with support in gD, we can easily see that the integral in bounded cohomology is always surjective. We give two supporting examples for Conjecture 1.5:

Proposition 4.4 *Conjecture 1.5 is true for G finite.*

Proof If G is finite, then M is compact, and so $\widehat{H}^n(M)$ coincides with the usual n^{th} de Rham cohomology of M, which is isomorphic with $\mathbb{R}$. On the other hand, by Corollary 2.2, $\ell^\infty(G)_G \cong \mathbb{R}$. So since the integral in bounded cohomology is surjective, as mentioned above, it is actually an isomorphism, as stated. □

Proposition 4.5 *Conjecture 1.5 is true for $M = \mathbb{R}$ and $G = \mathbb{Z}$, where $\mathbb{Z}$ acts on $\mathbb{R}$ by translation.*

Proof We choose the interval $[0,1] \subset \mathbb{R}$ as a fundamental domain. Let $g = 1 \in \mathbb{Z}$. Suppose that

$$\int_{\mathbb{R}} f(x)\,dx = \phi - g\phi \tag{5}$$

for a bounded function $f(x)$ on $\mathbb{R}$ and $\phi \in \ell^\infty(\mathbb{Z})$, where the integral is taken in the sense of (4). Equation (5) is equivalent to the fact that the 1–form $f(x)\,dx$ belongs to the kernel of the integral in bounded cohomology because

$$\phi - g^n\phi = (\phi + g\phi + \cdots + g^{n-1}\phi) - g(\phi + g\phi + \cdots + g^{n-1}\phi).$$

Note that $(\phi - g\phi)(i) = \phi(i) - \phi(i+1)$. Now we define

$$h(x) = \int_0^x f(t)\,dt.$$

To see that the integral in bounded cohomology is injective, it is sufficient to show that $h(x) \in \widehat{\Omega}^0(\mathbb{R})$. Since $dh(x) = f(x)\,dx$, $dh(x)$ is bounded. For $0 \le n \le x < n+1$, we have

$$h(x) = \sum_{i=0}^{n-1} \int_i^{i+1} f(t)\,dt + \int_n^x f(t)\,dt = \phi(0) - \phi(n) + \int_n^x f(t)\,dt.$$

Since $f(x)$ is bounded, $\int_n^x f(t)\,dt$ is bounded too as x and n vary. Then $h(x)$ is bounded for $x \ge 0$. Quite similarly, we can show that $h(x)$ is bounded for $x < 0$ too, and so $h(x) \in \widehat{\Omega}^0(\mathbb{R})$. Thus we obtain the injectivity. Since the integral in bounded cohomology is surjective as mentioned above, it is an isomorphism. □

5 Poincaré–Hopf theorem

In this section, we prove Theorems 1.1 and 1.2. Throughout this section, let M be a connected manifold of dimension n equipped with a cocompact and properly discontinuous action of a discrete group G such that M/G is oriented. The metric of M will be the lift of a metric of M/G.

Let Φ denote a representative of the Thom class of M/G. Then as in [Bott and Tu 1982], the support of Φ is compactly supported, and so Φ is a bounded n–form on $T(M/G)$. Let $\pi\colon M \to M/G$ denote the projection. Then the derivative of π is bounded, and so by Lemma 4.1 we get the induced map $\pi^*\colon \widehat{\Omega}^*(T(M/G)) \to \widehat{\Omega}^*(TM)$. In particular, $\pi^*(\Phi)$ is a bounded n–form on TM. Note that $\pi^*(\Phi)$

represents the Thom class of M in bounded cohomology. Let v be a vector field on M with $|dv|$ bounded. Then by Lemma 4.1, $v^*(\pi^*(\Phi))$ is a bounded n–form on M, and so we can define the index of v by

$$\operatorname{ind}(v) = \int_M v^*(\pi^*(\Phi)) \in \ell^\infty(G)_G.$$

We remark that the index $\operatorname{ind}(v)$ is independent of the choice of a representative Φ of the Thom class of M/G. Indeed, if Ψ is another representative of the Thom class of M/G, then $\Phi - \Psi = d\alpha$ for some compactly supported $(n-1)$–form α on $T(M/G)$, where Ψ is compactly supported. Hence we get $\pi^*(\Phi) - \pi^*(\Psi) = d\pi^*(\alpha)$, where all differential forms are bounded, and so by Corollary 4.3 the indices of v defined by Φ and Ψ are equal. We also remark that by Proposition 2.4, the index of a bounded vector field on M is always zero whenever G is not amenable; see [Weinberger 2009, Theorem 2]

We now show some properties of the index. Let v_0 denote the zero vector field, that is, the zero section $M \to TM$. Then $v_0^*(\pi^*(\Phi))$ is a representative of the Euler class $e(M)$ in bounded cohomology, which was considered by Weinberger [2009].

Proposition 5.1 *There is an equality*

$$\int_M e(M) = \chi(M/G)\mathbb{1}.$$

Proof Let $\bar{v}_0$ denote the zero vector field on M/G, so v_0 is the lift of $\bar{v}_0$. Since the projection $\pi : \operatorname{Int}(gD) \to M/G - \pi(\partial(gD))$ is a diffeomorphism and both $\partial(gD)$ and $\pi(\partial(gD))$ have measure zero,

$$\int_{gD} v_0^*(\pi^*(\Phi)) = \int_{M/G} \bar{v}_0^*(\Phi) = \int_{M/G} e(M/G) = \chi(M/G). \qquad \square$$

Lemma 5.2 *If vector fields v and w on M with $|dv|$ and $|dw|$ bounded are homotopic by a homotopy with bounded derivative, then*

$$\operatorname{ind}(v) = \operatorname{ind}(w).$$

Proof Let $v_t : M \times [0,1] \to TM$ be a homotopy with bounded derivative such that $v_0 = v$ and $v_1 = w$. Since the induced maps $v_t^* : \widehat{\Omega}^*(TM) \to \widehat{\Omega}^*(M \times [0,1])$ and $\pi^* : \widehat{\Omega}^*(T(M/G)) \to \widehat{\Omega}^*(TM)$ commute with the differential,

$$\int_{M\times[0,1]} dv_t^*(\pi^*(\Phi)) = \int_{M\times[0,1]} v_t^*(\pi^*(d\Phi)) = 0,$$

where Φ is a closed n–form representing the Thom class of $T(M/G)$. On the other hand, by Proposition 4.2

$$\int_{M\times[0,1]} dv_t^*(\pi^*(\Phi)) = \int_{M\times 1} w^*(\pi^*(\Phi)) - \int_{M\times 0} v^*(\pi^*(\Phi)). \qquad \square$$

Proposition 5.3 *Let v be a bounded vector field on M. Then we have*

$$\operatorname{ind}(v) = \operatorname{ind}(v_0).$$

Proof Clearly tv is a homotopy from v_0 to v with bounded derivative. So by Lemma 5.2, we are done. $\square$

We consider a mild condition on zeros of a vector field. Let $B_\delta(x)$ denote the open δ–neighborhood of $x \in M$, and let $B_\epsilon(M)$ denote the open ϵ–neighborhood of M in TM.

Definition 5.4 A vector field v on a manifold M is *tame* if there are $\delta > 0$ and $\epsilon > 0$ such that

(i) $B_\delta(x) \cap B_\delta(y) = \varnothing$ for $x \neq y \in \mathrm{Zero}(v)$, and

(ii) $v^{-1}(B_\epsilon(M)) \subset \bigcup_{x \in \mathrm{Zero}(v)} B_\delta(x)$.

Definition 5.5 A vector field v on M is *strongly tame* if it is tame and there is $\delta > 0$ such that for each $x \in \mathrm{Zero}(v)$, we have $B_\delta(x) \subset gD$ for some $g \in G$.

We also define a tame diffeomorphism, in analogy with tame vector fields:

Definition 5.6 A diffeomorphism $f\colon M \to M$ is *tame* if there are $\delta > 0$ and $\epsilon > 0$ such that

(i) $B_\delta(x) \cap B_\delta(y) = \varnothing$ for $x \neq y \in \mathrm{Fix}(f)$, and

(ii) $d(x, f(x)) > \epsilon$ for $x \in M - \bigcup_{y \in \mathrm{Fix}(y)} B_\delta(y)$,

where d stands for the metric of M.

We prove a technical lemma:

Lemma 5.7 *Let $f\colon \mathbb{R}^n \to T\mathbb{R}^n$ be a section of the tangent bundle $T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ such that for some $\delta, \epsilon > 0$, we have $f^{-1}(B_\epsilon(0)) \subset B_\delta(0)$. Let ω be a representative of the Thom class of $T\mathbb{R}^n$ such that* $\mathrm{supp}(\omega) \subset \mathbb{R}^n \times B_{\epsilon/2}(0)$. *Then*

$$\mathrm{ind}_0(f) = \int_{B_\delta(0)} f^*(\omega).$$

Proof Let B denote the closure of $B_\delta(0)$. Then by definition, $\mathrm{ind}_0(f)$ is the mapping degree of the composite

$$\partial B \xrightarrow{f} \mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\}) \xrightarrow{p} \mathbb{R}^n \setminus \{0\} \xrightarrow{q} S^{n-1},$$

where $p\colon T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ denotes the second projection and $q\colon \mathbb{R}^n \setminus \{0\} \to S^{n-1}$ denotes the natural projection onto the unit sphere. There is a function $\rho\colon [0,\infty) \to \mathbb{R}$ such that $\rho(x) = 0$ for x sufficiently close to 0 and $\rho(x) = 1$ for $x \geq \frac{1}{2}\epsilon$. Let α be an $(n-1)$–form on the unit sphere S^{n-1} of $\mathbb{R}^n$ such that $\int_{S^{n-1}} \alpha = 1$. Now we define

$$\eta = p^*(d\rho \wedge q^*(\alpha)).$$

By definition $\mathrm{supp}(\eta) \subset \mathbb{R}^n \times B_{\epsilon/2}(0)$, and since $\int_{\mathbb{R}^n} d\rho \wedge q^*(\alpha) = 1$, η represents the Thom class of $T\mathbb{R}^n$. Then by the uniqueness of the Thom class, there is an $(n-1)$–form τ on $T\mathbb{R}^n$ with vertically compact support such that

$$\omega - \eta = d\tau.$$

We have $\operatorname{supp}(\tau) \subset \mathbb{R}^n \times B_{\epsilon/2}(0)$, implying $f^*(\tau)|_{\partial B} = 0$. So by Stokes' theorem, we get

$$\int_B f^*(\omega) - \int_B f^*(\eta) = \int_B f^*(d\tau) = \int_B df^*(\tau) = \int_{\partial B} f^*(\tau) = 0.$$

Note that $\eta = dp^*(\rho \cdot q^*(\alpha))$ and $\rho(f(\partial B)) = 1$. Then by Stokes' theorem, we have

$$\int_B f^*(\eta) = \int_B df^* \circ p^*(\rho \cdot q^*(\alpha)) = \int_{\partial B} f^* \circ p^*(\rho \cdot q^*(\alpha)) = \int_{\partial B} f^* \circ p^* \circ q^*(\alpha) = \operatorname{ind}_0(f).$$

Thus since $\int_{B_\epsilon(0)} f^*(\omega) = \int_B f^*(\omega)$, the proof is finished. □

We compute the index of a strongly tame bounded vector field:

Proposition 5.8 *Let v be a strongly tame bounded vector field on M. Then*

$$\operatorname{ind}(v)(g) = \sum_{x \in \operatorname{Zero}(v) \cap gD} \operatorname{ind}_x(v).$$

Proof Let δ and ϵ be as in the definition of a tame vector field. As in [Bott and Tu 1982], we may assume that the support of $\pi^*(\Phi)$ is contained in $B_{\epsilon/2}(M)$. Then

$$\operatorname{ind}(v)(g) = \sum_{x \in \operatorname{Zero}(v) \cap gD} \int_{B_\delta(x)} v^*(\pi^*(\Phi))$$

for each $g \in G$. On the other hand, by tameness of v, $\pi^*(\Phi)|_{B_\delta(x)}$ is compactly supported for $x \in \operatorname{Zero}(v)$, and so by Lemma 5.7, we get

$$\int_{B_\delta(x)} v^*(\pi^*(\Phi)) = \operatorname{ind}_x(v)$$

for each $x \in \operatorname{Zero}(v)$. □

Proof of Theorem 1.1 Combine Propositions 5.1, 5.3, and 5.8. □

Proof of Theorem 1.2 As mentioned at the beginning of Section 3, G is finitely generated. Then we can apply the results in Section 2. Let v be a tame bounded vector field on M, and suppose that v has finitely many zeros. We can easily see that v is homotopic to a strongly tame vector field by a homotopy with bounded derivative. Then by Lemma 5.2, we may assume that v itself is strongly tame, and so by Theorem 1.1, we have

$$\operatorname{ind}(v) = \chi(M/G)\mathbb{1}.$$

Since $\chi(M/G) \neq 0$, it follows from Proposition 2.4 that $\chi(M/G)\mathbb{1}$ is nonzero in $\ell^\infty(G)_G$. Then by Proposition 2.3, we obtain that v must have infinitely many zeros, a contradiction. Thus v must have infinitely many zeros. □

Proof of Corollary 1.3 Observe that a tame diffeomorphism is the composite of a tame vector field and the exponential map if it is C^1 close to the identity map. Then the statement follows from Theorem 1.2. □

References

[Attie and Block 1998] **O Attie**, **J Block**, *Poincaré duality for L^p cohomology and characteristic classes*, preprint 98-48, DIMACS (1998) Available at http://dimacs.rutgers.edu/archive/TechnicalReports/TechReports/1998/98-48.ps.gz

[Attie et al. 1992] **O Attie**, **J Block**, **S Weinberger**, *Characteristic classes and distortion of diffeomorphisms*, J. Amer. Math. Soc. 5 (1992) 919–921 MR Zbl

[Block and Weinberger 1992] **J Block**, **S Weinberger**, *Aperiodic tilings, positive scalar curvature and amenability of spaces*, J. Amer. Math. Soc. 5 (1992) 907–918 MR Zbl

[Bott and Tu 1982] **R Bott**, **L W Tu**, *Differential forms in algebraic topology*, Graduate Texts in Math. 82, Springer (1982) MR Zbl

[Brodzki et al. 2012] **J Brodzki**, **G A Niblo**, **N Wright**, *Pairings, duality, amenability and bounded cohomology*, J. Eur. Math. Soc. 14 (2012) 1513–1518 MR Zbl

[Cima and Llibre 1990] **A Cima**, **J Llibre**, *Bounded polynomial vector fields*, Trans. Amer. Math. Soc. 318 (1990) 557–579 MR Zbl

[Kato et al. 2023a] **T Kato**, **D Kishimoto**, **M Tsutaya**, *Homotopy type of the space of finite propagation unitary operators on $\mathbb{Z}$*, Homology Homotopy Appl. 25 (2023) 375–400 MR Zbl

[Kato et al. 2023b] **T Kato**, **D Kishimoto**, **M Tsutaya**, *Homotopy type of the unitary group of the uniform Roe algebra on $\mathbb{Z}^n$*, J. Topol. Anal. 15 (2023) 495–512 MR Zbl

[Kato et al. 2024] **T Kato**, **D Kishimoto**, **M Tsutaya**, *Hilbert bundles with ends*, J. Topol. Anal. 16 (2024) 291–322 MR Zbl

[Roe 1988] **J Roe**, *An index theorem on open manifolds, I*, J. Differential Geom. 27 (1988) 87–113 MR Zbl

[Weinberger 2009] **S Weinberger**, *Fixed-point theories on noncompact manifolds*, J. Fixed Point Theory Appl. 6 (2009) 15–25 MR Zbl

[Whyte 1999] **K Whyte**, *Amenability, bilipschitz equivalence, and the von Neumann conjecture*, Duke Math. J. 99 (1999) 93–112 MR Zbl

Department of Mathematics, Kyoto University
Kyoto, Japan

Faculty of Mathematics, Kyushu University
Fukuoka, Japan

Faculty of Mathematics, Kyushu University
Fukuoka, Japan

tkato@math.kyoto-u.ac.jp, kishimoto@math.kyushu-u.ac.jp, tsutaya@math.kyushu-u.ac.jp

Received: 21 December 2022 Revised: 6 September 2023

Algebraic & Geometric Topology **24**:7 (2024) 3997–4006
DOI: 10.2140/agt.2024.24.3997
Published: 9 December 2024

Smallest nonabelian quotients of surface braid groups

CINDY TAN

We give a sharp lower bound on the size of nonabelian quotients of the surface braid group $B_n(\Sigma_g)$ and classify all quotients that attain the lower bound: depending on n and g, a quotient of minimum order is either a symmetric group or a 2–step nilpotent p–group.

20F65; 57K20

1 Introduction

The Artin braid group B_n arises as the fundamental group of $\mathrm{UConf}_n(\mathbb{D})$, the configuration space of n distinct unordered points on the open disk $\mathbb{D}$. One can generalize this construction to define, for an oriented closed genus-g surface Σ_g, the *surface braid groups*

$$B_n(\Sigma_g) = \pi_1(\mathrm{UConf}_n(\Sigma_g)).$$

It was shown by Kolay [4] that for $n = 3$ or $n \geq 5$, the smallest noncyclic finite quotient of B_n is the symmetric group S_n, in the sense that S_n has minimum order amongst noncyclic quotients of B_n and S_n is the unique noncyclic quotient of B_n of minimum order.

We consider the analogous question for surface braid groups. With our main result we show that whilst S_n is a quotient of $B_n(\Sigma_g)$, it is not in general the smallest nonabelian quotient.

For $g \geq 1$ the inclusion of a disk into Σ_g induces an embedding $B_n \hookrightarrow B_n(\Sigma_g)$ (see Birman [2]); any two such embeddings are conjugate in $B_n(\Sigma_g)$. By a *braid-free* quotient of $B_n(\Sigma_g)$ we mean a finite quotient with (any such embedding of) B_n having cyclic image. Our main result is the following theorem:

Theorem 1 (Smallest nonabelian quotients of $B_n(\Sigma_g)$) *Let $n \geq 5$ and $g \geq 1$. Suppose that G is a finite nonabelian quotient of $B_n(\Sigma_g)$.*

(a) *If G is not braid-free then $|G| \geq n!$ with equality if and only if $G \cong S_n$.*
(b) *If G is braid-free then G is 2–step nilpotent and $|G| \geq p^{2g+j}$, where p is the smallest prime dividing $g+n-1$ and $j = 1$ or 2 according to whether p is odd or 2, respectively. Equality occurs if and only if either $G \cong \mathrm{I}(p^j, g)$ or $G \cong \mathrm{II}(p^j, g)$* (these two groups are nonisomorphic 2–step nilpotent p–groups defined in Construction 10).

In particular, the smallest nonnilpotent quotient of $B_n(\Sigma_g)$ is S_n.

Note that Theorem 1 implies the following qualitative result:

Corollary 2 (a) *Fix $g \geq 1$. For all sufficiently large n, the smallest nonabelian quotients of $B_n(\Sigma_g)$ are 2–step nilpotent p–groups (in particular, the smallest nonabelian quotient is not S_n).*

(b) *Fix $n \geq 5$. For all sufficiently large g, the smallest nonabelian quotient of $B_n(\Sigma_g)$ is S_n. Also, there exists a (small) g for which this not true.*

Remarks 3 (Smaller cases) (a) If $n = 1, 2, 3, 4$ and $g \geq 1$ (with the exception of $(n, g) = (1, 1)$ where $B_n(\Sigma_g) = \pi_1(T^2) = \mathbb{Z}^2$ is abelian) then the symmetric group S_3 is the smallest nonabelian quotient of $B_n(\Sigma_g)$.

(b) If $g = 0$ then $B_n(\Sigma_g)$ is the spherical braid group $B_n(S^2)$, which is an intermediate quotient of the map $B_n \to S_n$; see Fadell and Van Buskirk [3]. It follows from the result of Kolay [4] that the smallest quotient of $B_n(S^2)$ is S_n for $n \geq 5$ and S_3 for $n = 3, 4$. For $n = 1, 2$ we note that $B_n(S^2)$ is abelian.

From Theorem 1 we obtain partial confirmation of a conjecture of Chen [2, Conjecture 1.3]:

Corollary 4 *Let $n \geq 5$ and $m \geq 3$, and let $g, h \geq 0$. If $n > m$ then there are no surjective homomorphisms*

$$B_n(\Sigma_g) \to B_m(\Sigma_h).$$

Proof method Theorem 1(a) follows from Kolay: By mapping a braid to its permutation on points, S_n is a finite quotient of $B_n(\Sigma_g)$. If $B_n \to B_n(\Sigma_g) \to G$ has noncyclic image then $|G| \geq n!$ with the bound attained only by $G \cong S_n$.

Our primary contribution here is Theorem 1(b), which considers the braid-free quotients. We utilize a presentation of $B_n(\Sigma_g)$ (Theorem 13) due to Bellingeri [1] and assume that B_n has cyclic image to reduce the relations, and conclude that a braid-free quotient G must be nilpotent. If we further assume that G is a nonabelian braid-free quotient of minimum order, then G belongs to a class of nilpotent groups called JN2 groups (Definition 5) which were classified by Newman in 1960 [5]. It then suffices to find the smallest JN2 groups which can be realized as a quotient of $B_n(\Sigma_g)$, a straightforward task given the concrete nature of Newman's classification.

Section 2 provides a self-contained exposition of the classification of JN2 groups. In Section 3 we prove Theorem 1(b), as well as Corollary 4.

Acknowledgements I am grateful to my advisor Benson Farb for continued support throughout this project and for detailed comments on many revisions of this paper, as well as for suggesting this problem in the first place. I thank Peter Huxford for useful discussions about braid groups and small p–groups, and for many helpful suggestions during the editing process. I also thank Dan Margalit for taking the time to read and comment on an earlier draft.

2 Just 2–step nilpotent groups

In this section we introduce and classify JN2 groups, a class of nilpotent groups which includes all minimal nonabelian braid-free quotients of $B_n(\Sigma_g)$.

Definition 5 A group G is *just* 2*–step nilpotent* (JN2) if G is 2–step nilpotent (in particular, nonabelian) and every proper quotient of G is abelian.

Finite JN2 groups admit a complete and explicit classification due to Newman [5]: any finite JN2 group can be assigned a unique class (p^j, m) where p is a prime and j and m are positive integers; up to isomorphism, there are precisely two JN2 groups of a given class (p^j, m). We will state and prove this classification theorem in Theorem 11, following the general ideas of [5].

All JN2 groups will hereafter be assumed to be finite. The following proposition will allow us to define the class (p^j, m) of a JN2 group:

Proposition 6 (Characterization of JN2 groups [5, Theorem 1]) *A finite group G is JN*2 *if and only if there exists a prime p such that*

(a) *$G' := [G, G]$ is cyclic of order p,*

(b) *the center ZG is cyclic of order a power of p, and*

(c) *G/ZG is elementary abelian of exponent p.*

*In particular, a JN*2 *group is a p–group.*

Proof ($\Rightarrow$) Let G be a finite JN2 group. For every nontrivial normal subgroup $N \trianglelefteq G$, we have that $G' \le N$ since any proper quotient of G is abelian. Since G is 2–step nilpotent, $G' \le ZG$. Consequently:

(a) G' is abelian and admits no proper nontrivial subgroups so $G' \cong \mathbb{Z}/p\mathbb{Z}$ for some prime p.

(b) ZG cannot be properly decomposed as a direct sum: any nontrivial subgroup of ZG contains G' so no two nontrivial subgroups intersect trivially. Since ZG is finite abelian, it must be cyclic of prime power order. The prime must be p because $G' \le ZG$.

(c) G/ZG is abelian because $G' \le ZG$. For $x, y \in G$, we have that $[x^p, y] = [x, y]^p$ by using the identity

$$[xz, y] = z[x, y]z^{-1}[z, y]$$

and noting that $[x, y]$ is central because $G' \le ZG$. But G' has order p, so in fact $[x^p, y] = 1$. Thus $x^p \in ZG$ for all $x \in G$, which is to say that G/ZG has exponent p.

($\Leftarrow$) Suppose G is a finite group satisfying (a), (b), and (c). Then $G' \ne \{1\}$ by (a) and $G' \le ZG$ by (c), so G is 2–step nilpotent.

If $N \trianglelefteq G$ is a normal subgroup with $G' \not\le N$ then $N \cap G' = \{1\}$ by (a). Since N is normal, $[N, G] \le N \cap G' = \{1\}$ so $N \le ZG$. But $G' \le ZG$, and (a) and (b) imply that any nontrivial subgroup of ZG intersects G' nontrivially. Thus $N = \{1\}$. We conclude that every proper quotient of G is abelian. □

An immediate corollary of (c) is that $V := G/ZG$ has the structure of an $\mathbb{F}_p$–vector space. Note that vector addition in V is written multiplicatively and scalar multiplication of an element $x \bmod ZG \in V$ by a scalar $r \in \mathbb{F}_p$ is written as

$$r(x \bmod ZG) = x^r \bmod ZG.$$

Fix a generator z of ZG. This fixes a generator $z^{p^{j-1}}$ of G', and hence an identification of G' with $\mathbb{F}_p$. Define a pairing

$$V \times V \to G' = \mathbb{F}_p \quad \text{given by } (x \bmod ZG, y \bmod ZG) \mapsto [x, y].$$

This pairing is a well-defined bilinear nondegenerate alternating form, which makes V into a symplectic vector space. In particular, $\dim V$ is even.

Thus associated to each JN2 group G is a *class* (p^j, m) where $|ZG| = p^j$ and $\dim V = 2m$, so G fits into the short exact sequence

$$1 \to \mathbb{Z}/p^j\mathbb{Z} \to G \to (\mathbb{Z}/p\mathbb{Z})^{2m} \to 0.$$

The symplectic structure on central factor groups $V = G/ZG$ is key to the classification theorem because symplectic automorphisms on central factor groups can be used to construct isomorphisms between certain JN2 groups of the same class. The following lemma extracts from a JN2 group a normalized symplectic basis on its associated vector space V:

Lemma 7 *Let G be JN2 of class (p^j, m) where $p^j \neq 2$, with a fixed generator z of ZG. Then there exists a symplectic basis $\mathscr{B} = \{a_i \bmod ZG, b_i \bmod ZG\}_{i=1}^m$ of $V = G/ZG$ such that the representatives $a_i, b_i \in G$ satisfy either*

(I) $a_i^p = b_i^p = 1$ *for all i, or*

(II) $a_1^p = b_1^p = z$ *and* $a_i^p = b_i^p = 1$ *for* $2 \leq i \leq m$.

We will say that $\mathscr{B}$ is type I or II accordingly.

Remark 8 (Nomenclature) For the reader familiar with existing terminology from [5], a "type I (respectively II) basis" as named in our Lemma 7 corresponds to a "canonic normal basis with zero (respectively one) pairs of type II" in the vocabulary of Newman.

Proof Note that $x^p \in ZG$ for all $x \in G$ because G/ZG has exponent p. Let $(ZG)^p = \{u^p : u \in ZG\}$ and identify $ZG/(ZG)^p$ with $\mathbb{F}_p$ by the mapping $z \bmod (ZG)^p \mapsto 1$. Define a map

$$\nu\colon V \to ZG/(ZG)^p = \mathbb{F}_p \quad \text{given by } x \bmod ZG \mapsto x^p \bmod (ZG)^p.$$

Viewing $V = G/ZG$ as a vector space written multiplicatively, ν commutes with scalar multiplication and

$$\nu((x \bmod ZG)(y \bmod ZG)) = (xy)^p \bmod (ZG)^p = [y, x]^{p(p-1)/2} x^p y^p \bmod (ZG)^p$$

for $x, y \in G$, so ν is a linear functional as long as $[y, x]^{p(p-1)/2} = 1 \bmod (ZG)^p$. This holds if $p^j \neq 2$: If p is odd then $p \mid \frac{1}{2}p(p-1)$, so $[y, x]^{p(p-1)/2} = 1$ because G' has order p. If $j \geq 2$ then $G' \lneq ZG$ so $G' \leq (ZG)^p$.

If ν is the trivial linear functional on V, take $\mathscr{B}$ to be any symplectic basis of V. Otherwise there exists a symplectic basis $\mathscr{B}$ of V such that ν written with respect to $\mathscr{B}$ is the row vector

$$\nu = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix},$$

because symplectic automorphisms act transitively on nontrivial vectors.

For each basis vector $x_j \bmod ZG \in \mathscr{B}$,

$$x_j^p \bmod (ZG)^p = \nu(x_j \bmod ZG) = z^{\nu_j} \bmod (ZG)^p,$$

so there exists $u_j \in ZG$ such that $x_j^p = z^{\nu_j} u_j^p$. Then $x_j u_j^{-1} \equiv x_j \bmod ZG$ and $(x_j u_j^{-1})^p = z^{\nu_j}$. Thus $x_j u_j^{-1} \in G$ are representatives of the basis $\mathscr{B}$ satisfying (I) if ν is trivial and (II) otherwise. □

We will now construct two standard nonisomorphic JN2 groups for each given class (p^j, m). The proof of the classification theorem will exhibit an isomorphism from any arbitrary JN2 group to a standard one. The primary method of constructing larger JN2 groups from smaller ones is taking a central product.

Definition 9 (Central product) Let G and H be groups for which $ZG \cong ZH$. Define the *central product* of G and H (with respect to an isomorphism $\varphi : ZG \to ZH$) to be

$$G \odot H = (G \times H)/N \quad \text{where } N = \langle (g, \varphi(g)^{-1}) : g \in ZG \rangle,$$

i.e. identifying $ZG \times 1$ with $1 \times ZH$ by the isomorphism φ. By $G^{\odot n}$ we mean the central product of n copies of G with the identity isomorphism on ZG.

Note that if G and H are JN2 of class (p^j, m_1) and (p^j, m_2), then $G \odot H$ is JN2 of class $(p^j, m_1 + m_2)$ by Proposition 6 since

(a) $(G \odot H)' = G' \times H'/N \cong G' \cong H' \cong \mathbb{Z}/p\mathbb{Z}$,

(b) $Z(G \odot H) \cong ZG \cong ZH$, and

(c) $(G \odot H)/Z(G \odot H) \cong (G/ZG) \times (H/ZG)$.

Construction 10 (Standard JN2 groups) Define the groups

$$\begin{aligned} M(p^j) &= \langle z, a, b : [z,a] = [z,b] = 1; [a,b] = z^{p^{j-1}}; z^{p^j} = a^p = b^p = 1 \rangle, \\ N(p^j) &= \langle z, a, b : [z,a] = [z,b] = 1; [a,b] = z^{p^{j-1}}; z^{p^j} = 1; a^p = b^p = z \rangle, \\ \mathrm{I}(p^j, m) &= M(p^j)^{\odot m}, \\ \mathrm{II}(p^j, m) &= N(p^j) \odot M(p^j)^{\odot(m-1)}. \end{aligned}$$

Observe the following:

(1) $M(p^j)$ and $N(p^j)$ are JN2 (by Proposition 6) of class $(p^j, 1)$, with each center generated by z and $\{a, b\}$ as a symplectic basis of V.

(2) $\mathrm{I}(p^j, m)$ and $\mathrm{II}(p^j, m)$ are JN2 of class (p^j, m) by the remarks following Definition 9.

(3) $\mathrm{I}(p^j, m)$ and $\mathrm{II}(p^j, m)$ are not isomorphic when $p^j \neq 2$: the group $N(p^j)$ has an element of order p^{j+1} (for example, a or b) and therefore so does $\mathrm{II}(p^j, m)$. On the contrary, the group $M(p^j)$, and consequently also $\mathrm{I}(p^j, m) = M(p^j)^{\odot m}$, has exponent at most p^j: the linear functional ν (as in the proof of Lemma 7) is trivial on the symplectic basis $\{a, b\}$ so $M(p^j)^p \leq (ZG)^p$, and hence $M(p^j)^{p^j} \leq (ZG)^{p^j} = 1$.

Note If $p^j = 2$, then $\mathrm{I}(p^j, m)$ and $\mathrm{II}(p^j, m)$ are still nonisomorphic: $M(2)$ is the dihedral group D_8 and $N(2)$ is the quaternion group Q_8, which contain two and six elements of order 4, respectively, and both have centers of order 2. In particular no elements of order 4 are central. The larger groups $\mathrm{I}(p^j, m)$ and $\mathrm{II}(p^j, m)$ can then be distinguished by counting the number of elements of order 4 because only central elements are identified in the central product. We will not require this case.

Theorem 11 (Classification of finite JN2 groups [5, Theorems 5 and 7(c) and Lemma 8(i)]) *Let G be JN2 of class (p^j, m). Suppose that $p^j \neq 2$. Then G is isomorphic to either $\mathrm{I}(p^j, m)$ or $\mathrm{II}(p^j, m)$.*

Proof Let z be a generator of ZG and let $\mathscr{B}$ be the symplectic basis given by Lemma 7. In the notation of Lemma 7, let $H_i = \langle z, a_i, b_i \rangle$. If $\mathscr{B}$ is type I then $H_i = M(p^j)$ for all i. If $\mathscr{B}$ is type II then $H_1 = N(p^j)$ and $H_i = M(p^j)$ for $i \geq 2$.

The subgroups H_i commute pairwise, together generate G, and intersect precisely in their centers $\langle z \rangle$, so $G \cong \bigodot_{i=1}^m H_i$. Hence G is isomorphic to $\mathrm{I}(p^j, m)$ or $\mathrm{II}(p^j, m)$, according to the type of the basis $\mathscr{B}$. □

Remarks 12 (a) **Generalizations** For brevity, we have excluded the case of $p^j = 2$ and specialized to finite groups. With additional work, the $p^j = 2$ case and some infinite JN2 groups (those with a countable symplectic basis) also admit a classification as central products of elementary JN2 groups, see [5].

(b) **Special cases** Note that $M(p)$ and $N(p)$ are the only two groups of order p^3. The group $M(p) = \mathrm{I}(p, 1)$ is isomorphic to the Heisenberg group over $\mathbb{F}_p$. A generalization of the finite Heisenberg groups are the *extraspecial groups*, which are defined to be p–groups G with ZG order p and G/ZG nontrivial elementary abelian. In particular, extraspecial groups are JN2, and it follows from Theorem 11 that there are precisely two distinct extraspecial groups of order p^{1+2m} for each choice of a prime p and positive integer m and that this exhausts all extraspecial groups.

3 Minimal nonabelian quotients of $B_n(\Sigma_g)$

In this section we provide the proof of Theorem 1(b). The strategy of the proof will be to utilize an explicit presentation of the surface braid groups (Theorem 13) to characterize braid-free quotients by the relations that they must satisfy (Lemma 15). We will then show that many JN2 groups are realized as nonabelian braid-free quotients of $B_n(\Sigma_g)$ (Lemma 16) and finally prove that all nonabelian braid-free quotients of minimum order belong to the list of JN2 groups in Lemma 16.

Theorem 13 (Presentation of $B_n(\Sigma_g)$, Bellingeri [1, Theorem 1.2]) *For $g \geq 1$ and $n \geq 2$, the surface braid group $B_n(\Sigma_g)$ admits the presentation given by*

- *generators* $\sigma_1, \dots, \sigma_{n-1}, a_1, \dots, a_g, b_1, \dots, b_g$,
- *braid relations*

$$[\sigma_i, \sigma_j] = 1 \qquad \text{for } 1 \leq i, j \leq n-1 \text{ and } |i-j| \geq 2,$$

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \qquad \text{for } 1 \leq i \leq n-2,$$

and mixed relations

(R1) $$[a_r, \sigma_i] = [b_r, \sigma_i] = 1 \qquad \text{for } 1 \leq r \leq g \text{ and } i \neq 1,$$

(R2) $$[a_r, \sigma_1^{-1} a_r \sigma_1^{-1}] = [b_r, \sigma_1^{-1} b_r \sigma_1^{-1}] = 1 \qquad \text{for } 1 \leq r \leq g,$$

(R3) $$[a_s, \sigma_1 a_r \sigma_1^{-1}] = [b_s, \sigma_1 b_r \sigma_1^{-1}] = 1 \qquad \text{for } 1 \leq s < r \leq g,$$

$$[b_s, \sigma_1 a_r \sigma_1^{-1}] = [a_s, \sigma_1 b_r \sigma_1^{-1}] = 1 \qquad \text{for } 1 \leq s < r \leq g,$$

(R4) $$[a_r, \sigma_1^{-1} b_r \sigma_1^{-1}] = \sigma_1^2 \qquad \text{for } 1 \leq r \leq g,$$

(TR) $$[a_1, b_1^{-1}] \cdots [a_g, b_g^{-1}] = \sigma_1 \sigma_2 \cdots \sigma_{n-1}^2 \cdots \sigma_2 \sigma_1.$$

Remark 14 (Geometric interpretation of the presentation) The Bellingeri generators σ_i can be identified as the images of the Artin braid generators under a choice of embedding $B_n \hookrightarrow B_n(\Sigma_g)$. The remaining generators a_r and b_r can be understood loosely to be the standard generators of $\pi_1(\Sigma_g)$.

More precisely, let $\{p_1, \dots, p_n\} \in \mathrm{UConf}_n(\Sigma_g)$ denote the basepoint of $B_n(\Sigma_g)$ and let $D \subset \Sigma_g$ be an open disk with $p_1 \in \partial D$, with $p_2, \dots, p_n$ in the interior of D. There is an inclusion

$$\pi_1(\Sigma_g - D, p_1) \hookrightarrow B_n(\Sigma_g)$$

which takes a loop γ in $\Sigma_g - D$ to the braid on Σ_g with first strand γ and all other strands trivial. The group $\pi_1(\Sigma_g - D, p_1)$ is free on $2g$ generators and surjects onto $\pi_1(\Sigma_g, p_1)$, which has a standard presentation. The surface braid group generators $a_r, b_r \in B_n(\Sigma_g)$ can then be understood as a choice of a free generating set of $\pi_1(\Sigma_g - D, p_1)$ which lifts the standard generating set of $\pi_1(\Sigma_g, p_1)$. It should be emphasized that the lifts are not canonical and that the presentation depends on the choices; the curious reader may refer to [1] for illustrations of the loops which produce this particular presentation.

Lemma 15 (Characterization of braid-free quotients) *Let $n \geq 3$ and $g \geq 1$. A finite group G is a braid-free quotient of $B_n(\Sigma_g)$ if and only if G admits a generating set $\{\sigma, a_1, b_1, \dots, a_g, b_g\}$ satisfying the relations*

(R1′) $$[a_r, \sigma] = [b_r, \sigma] = 1 \qquad \text{for } 1 \leq r \leq g,$$

(R3′) $$[a_s, a_r] = [b_s, b_r] = [b_s, a_r] = [a_s, b_r] = 1 \qquad \text{for } 1 \leq s < r \leq g,$$

(R4′) $$[a_r, b_r] = \sigma^2 \qquad \text{for } 1 \leq r \leq g,$$

(TR′) $$\sigma^{2(g+n-1)} = 1.$$

Proof A finite quotient of $B_n(\Sigma_g)$ is presented by Theorem 13 with additional relations. The condition that B_n has cyclic image in a quotient is equivalent to adding the relations

$$\sigma_i = \sigma_1, \qquad 1 \le i \le n.$$

If we add these relations and write $\sigma = \sigma_1$, the relation (R2) is made redundant and (R1), (R3), and (R4) respectively reduce to the relations (R1′), (R3′), and (R4′), as in the statement of the lemma. The final relation (TR) reduces to

$$[a_1, b_1^{-1}] \cdots [a_g, b_g^{-1}] = \sigma^{2(n-1)},$$

which is equivalent to (TR′) because from (R4′) we can write $a_r = b_r^{-1}\sigma^{-2}a_r b_r$ so that

$$[a_r, b_r^{-1}] = a_r b_r^{-1} a_r^{-1} b_r \overset{(R4')}{=} (b_r^{-1}\sigma^{-2}a_r b_r) b_r^{-1} a_r^{-1} b_r = b_r^{-1}\sigma^{-2} b_r \overset{(R1')}{=} \sigma^{-2}. \qquad \square$$

The following lemma proves that many JN2 groups are braid-free quotients:

Lemma 16 *Let $n \ge 3$ and $g \ge 1$. Let p be a prime dividing $g+n-1$.*

(a) *If $p = 2$ then $\mathrm{I}(2^2, g)$ and $\mathrm{II}(2^j, g)$ for all $j \ge 2$ are nonabelian braid-free quotients of $B_n(\Sigma_g)$.*

(b) *If p is odd then $\mathrm{I}(p, g)$ and $\mathrm{II}(p^j, g)$ for all $j \ge 1$ are nonabelian braid-free quotients of $B_n(\Sigma_g)$.*

Proof Let p be a prime dividing $g+n-1$. By Lemma 15 we need to exhibit a generating set $\{\sigma, a_r, b_r\}$ of each group satisfying relations (R1′), (R3′), (R4′), and (TR′).

In any of the JN2 groups in the statement of the theorem, fix a generator z of the center and choose $a_1, b_1, \dots, a_g, b_g$ to be the representatives of a symplectic basis of V given by Lemma 7. By Theorem 11 this basis will be type I for $\mathrm{I}(2^2, g)$ and $\mathrm{I}(p, g)$, and type II for $\mathrm{II}(2^j, g)$ and $\mathrm{II}(p^j, g)$. Note that with the given symplectic form, the condition that a basis is symplectic is simply that all basis elements commute except symplectic pairs $[a_r, b_r] = z^{p^{j-1}}$. In particular, (R3′) is satisfied.

We will now choose σ for each group and verify that $\{\sigma, a_r, b_r\}$ generate the group and satisfy (R4′).

(1) $\mathrm{I}(2^2, g)$ is generated by $\sigma = z$ and $\{a_r, b_r\}$. These satisfy (R4′) because $[a_r, b_r] = z^2 = \sigma^2$.

(2) $\mathrm{II}(2^j, g)$, for a given $j \ge 2$, is generated by $\{a_r, b_r\}$ alone because $a_1^p = z$. If we choose $\sigma = z^{2^{j-2}}$ then (R4′) is satisfied because $[a_r, b_r] = z^{2^{j-1}} = \sigma^2$.

(3) $\mathrm{I}(p, g)$ for odd prime p is generated by $\sigma = z^{(p^j+1)/2}$ and the a_r and b_r. Then (R4′) is satisfied because $[a_r, b_r] = z = \sigma^2$.

(4) $\mathrm{II}(p^j, g)$, for given odd prime p and $j \ge 1$, is generated by a_r and b_r alone because $a_1^p = z$. Then set $\sigma = z^{(p^j + p^{j-1})/2}$ so that (R4′) is satisfied because $[a_r, b_r] = z^{p^{j-1}} = \sigma^2$.

In all cases σ was chosen to be central, and hence (R1′) is satisfied.

It remains to check that (TR′) holds, namely that $|\sigma|$ divides $2(g+n-1)$. Recall that we are assuming that $p \mid (g+n-1)$. In cases (1) and (2), we have $p = 2$ and $|\sigma| = 4 = 2p \mid 2(g+n-1)$. In cases (3) and (4), we have $|\sigma| = p \mid (g+n-1)$. $\square$

Proof of Theorem 1(b) Let G be a nonabelian braid-free quotient and let $\{\sigma, a_1, b_1, \dots, a_g, b_g\}$ denote the generating set of G as given by Lemma 15. By (R1′) and (R3′), all pairs of these generators commute except for pairs a_r and b_r, so $G' = \langle \sigma^2 \rangle$ by (R4′). Then G' is central and nontrivial, which is to say that G is 2–step nilpotent.

Assume now that G is of minimum order amongst nonabelian braid-free quotients of $B_n(\Sigma_g)$. Then G has no proper nonabelian quotients and thus is JN2 of some class (p^j, m).

We make three claims:

(1) $p^j \neq 2$,

(2) $m = g$, and

(3) $p \mid (g+n-1)$.

These claims will complete the proof: Since $|G| = p^{2m+j}$, it follows from claims (1) and (2) that $|G| \geq p^{2g+1}$ if p is odd and $|G| \geq 2^{2g+2}$ if $p = 2$. Claims (1) and (3) along with the minimality of G together imply that G is one of (in particular, the smallest of) the quotients constructed in Lemma 16. Explicitly, if $g+n-1$ is even then $p^j = 2^2$. Otherwise $p^j = p$, where p is the smallest prime dividing $g+n-1$. Finally, G must be isomorphic to either $\mathrm{I}(p^j, g)$ or $\mathrm{II}(p^j, g)$ by Theorem 11.

Proof of claims Let $d = |\sigma|$. By (R1′), σ is central so $d \mid p^j$. But $p = |G'| = |\sigma^2|$ so $d \mid 2p$. Thus either p is odd and $p = d$, or $p = 2$ and $d = 4$.

(1) If p is odd then $p^j \neq 2$. If $p = 2$ then $p^j \geq d = 4$ so $p^j \neq 2$.

(2) We will show that $\dim V = 2g$ by proving that

$$\mathscr{B} = \{a_r \bmod ZG, b_r \bmod ZG\}_{r=1}^{g}$$

is a basis of V. Every element $x \in G$ can be written uniquely in the form

$$x = \sigma^k a_1^{i_1} \cdots a_g^{i_g} b_1^{j_1} \cdots b_g^{j_g}$$

using commuting relations (R1′), (R3′), and (R4′) so $\mathscr{B}$ is a generating set. To prove that $\mathscr{B}$ is linearly independent, let

$$y = a_1^{i_1} \cdots a_g^{i_g} b_1^{j_1} \cdots b_g^{j_g} \in G$$

and suppose that $y = 0 \bmod ZG$, which is to suppose that an arbitrary linear combination of elements of $\mathscr{B}$ is trivial in V. Then y is central, so

$$[y, b_1] = [a_1^{i_1}, b_1] = \sigma^{-2i_1} = 1,$$

which implies that $d \mid 2i_1$ and thus $i_1 = 0 \bmod p$: If p is odd then $d = p$, so $p \mid i_1$. If $p = 2$ then $d = 4 \mid 2i_1$ so, $p = 2 \mid i_1$.

Similarly $i_r = j_r = 0 \bmod p$ for all r, which is to say that all coefficients of the linear combination are trivial over the base field $\mathbb{F}_p$. This proves the linear independence of $\mathscr{B}$.

(3) The relation (TR′) imposes the relation $d \mid 2(g+n-1)$. Either $d = p$ is odd or $d = 4$ and $p = 2$; in both cases (TR′) implies that $p \mid (g+n-1)$. □

Proof of Corollary 4 Let $n \geq 5$ and $m \geq 3$, and let $g, h \geq 0$. If there is a surjection $B_n(\Sigma_g) \to B_m(\Sigma_h)$ then the composition $B_n(\Sigma_g) \to B_m(\Sigma_h) \to S_m$ is also surjective. Since S_m is not nilpotent when $m \geq 3$, we must have $m \geq n$. □

Remark 17 (Punctured surfaces, surfaces with boundary) Bellingeri [1] also gives a presentation of the braid group of a genus-g surface with m punctures (equivalently for the purposes of braid groups, m boundary components). The above methods can be used nearly verbatim to prove that the smallest nonabelian quotient of $B_n(\Sigma_{g,m})$ is the smaller of S_n or $\mathrm{I}(2^2, g)$ and $\mathrm{II}(2^2, g)$.

References

[1] **P Bellingeri**, *On presentations of surface braid groups*, J. Algebra 274 (2004) 543–563 MR Zbl

[2] **L Chen**, *Surjective homomorphisms between surface braid groups*, Israel J. Math. 232 (2019) 483–500 MR Zbl

[3] **E Fadell**, **J Van Buskirk**, *On the braid groups of E^2 and S^2*, Bull. Amer. Math. Soc. 67 (1961) 211–213 MR Zbl

[4] **S Kolay**, *Smallest noncyclic quotients of braid and mapping class groups*, Geom. Topol. 27 (2023) 2479–2496 MR Zbl

[5] **M F Newman**, *On a class of nilpotent groups*, Proc. London Math. Soc. 10 (1960) 365–375 MR Zbl

Department of Mathematics, University of Chicago
Chicago, IL, United States

cindy@math.uchicago.edu

Received: 16 January 2023 Revised: 7 June 2023

Algebraic & Geometric Topology **24**:7 (2024) 4007–4060
DOI: 10.2140/agt.2024.24.4007
Published: 9 December 2024

Lattices, injective metrics and the $K(\pi, 1)$ conjecture

THOMAS HAETTEL

Starting with a lattice with an action of $\mathbb{Z}$ or $\mathbb{R}$, we build a Helly graph or an injective metric space. We deduce that the ℓ^∞ orthoscheme complex of any bounded graded lattice is injective. We also prove a Cartan–Hadamard result for locally injective metric spaces. We apply this to show that any Garside group or any FC-type Artin group acts on an injective metric space and on a Helly graph. We also deduce that the natural piecewise ℓ^∞ metric on any Euclidean building of type $\tilde{A}_n$ extended, $\widetilde{B}_n$, $\widetilde{C}_n$ or $\widetilde{D}_n$ is injective, and its thickening is a Helly graph.

Concerning Artin groups of Euclidean types $\tilde{A}_n$ and $\widetilde{C}_n$, we show that the natural piecewise ℓ^∞ metric on the Deligne complex is injective, the thickening is a Helly graph, and it admits a convex bicombing. This gives a metric proof of the $K(\pi, 1)$ conjecture, as well as several other consequences usually known when the Deligne complex has a CAT(0) metric.

20E42, 05B35, 52A35, 06A12

Introduction

Injective metric spaces are geodesic metric spaces where every family of pairwise intersecting balls has a nonempty global intersection. The discrete counterpart of the injective metric space is the Helly graph. Its use in geometric group theory is recent and growing; see notably [Dress 1984; Lang 2013; Huang and Osajda 2021; Chalopin et al. 2020a; Haettel et al. 2023; Osajda and Valiunas 2024; Haettel 2022a]. Roughly speaking, CAT(0) spaces are typically locally Euclidean spaces, whereas injective metric spaces are typically locally ℓ^∞ metric spaces. Injective metric spaces display many nonpositive curvature features observed in CAT(0) spaces. We believe that Helly graphs and injective metric spaces may also prove to be more powerful than CAT(0) spaces for some purposes.

Firstly, it appears that many nonpositively curved groups (notably most CAT(0) groups) have a nice action on an injective metric space, so that the theory encompasses a vast class of groups: CAT(0) cubical groups, hyperbolic groups, relatively hyperbolic groups, uniform lattices in semisimple Lie groups over local fields, braid groups and more generally Garside groups, Artin groups of type FC, mapping class groups and more generally hierarchically hyperbolic groups. For instance, Helly graphs admit a simple combinatorial local characterization (see Theorem 1.12), which makes them potentially as powerful as CAT(0) cube complexes, whereas piecewise Euclidean CAT(0) complexes desperately lack a combinatorial characterization.

Secondly, we can still deduce many of the results that hold true for CAT(0) spaces and CAT(0) groups in the setting of injective spaces and injective groups (groups acting geometrically on injective spaces). For instance, injective metric spaces have a conical geodesic bicombing and are thus contractible, and every isometric bounded group action has a fixed point. Moreover, every injective group is semihyperbolic (see [Alonso and Bridson 1995]) and satisfies the Farrell–Jones conjecture (see [Kasprowski and Rüping 2017]). Sometimes we can even deduce stronger results than in the CAT(0) setting: as a sample result, note that any group acting properly and cocompactly on a Helly graph is biautomatic (see [Chalopin et al. 2020a]), whereas not all CAT(0) groups are biautomatic (see [Leary and Minasyan 2021]). Also note that infinite, finitely generated torsion groups do not act properly on uniformly locally finite Helly graphs (see [Haettel and Osajda 2021]), whereas the analogous statement is open for CAT(0) complexes.

In this article, we will pursue this philosophy. On one hand, we develop results useful to prove that some metric spaces are injective. On the other hand, we apply these results to Euclidean buildings and Deligne complexes of Euclidean Artin groups. We believe, however, that the scope of our results will not be limited to Artin groups of Euclidean type, and could concern much larger classes of groups.

We now present a very simple criterion (already appearing in a restricted form in [Haettel 2022a]) showing how to build a Helly graph or an injective metric space, starting with a lattice and an action of $\mathbb{Z}$ or $\mathbb{R}$.

Theorem A (Theorem 2.1) *Assume that L is a lattice such that each upper bounded subset of L has a join. Assume that there is an order-preserving increasing continuous action $(f_t)_{t\in H}$ of $H=\mathbb{Z}$ or $\mathbb{R}$ on L such that,*

$$\text{for all } x,y\in L, \text{ there exists } t\in H_+ \text{ such that } f_{-t}(x)\leqslant y\leqslant f_t(x).$$

Let us define the following metric d on L:

$$\text{for all } x,y\in L,\quad d(x,y)=\inf\{t\in H_+ \mid f_{-t}(x)\leqslant y\leqslant f_t(x)\}.$$

- *If $H=\mathbb{Z}$, the metric space (L,d) is the vertex set of a Helly graph.*
- *If $H=\mathbb{R}$, the metric space (L,d) is injective.*

With Jingyin Huang, we used the same situation of a lattice with an appropriate action of $\mathbb{Z}$ to define another natural graph which is weakly modular, and this applies to numerous examples (see [Haettel and Huang 2024]).

We applied this criterion in [Haettel 2022a] to prove that the thickening of a Bruhat–Tits Euclidean building of extended type $\tilde{A}_n$ is a Helly graph. It turns out that the Bruhat–Tits restriction is unnecessary.

Theorem B (Theorems 4.4 and 7.4) *The natural ℓ^∞ metric on any Euclidean building of type $\tilde{A}_n$ extended, $\widetilde{B}_n$, $\widetilde{C}_n$ or $\widetilde{D}_n$ is injective. Furthermore, the thickening of its vertex set is a Helly graph.*

Since Euclidean buildings are also endowed with CAT(0) metrics, we may wonder about the importance of this result. For one, it gives a lot of new examples to the theory of injective metric spaces and Helly graphs. Another consequence is another approach to Świątkowski's result [2006] that cocompact lattices in Euclidean buildings are biautomatic, in types $\tilde{A}_n$, $\widetilde{B}_n$, $\widetilde{C}_n$ or $\widetilde{D}_n$. The proof for types $\widetilde{B}_n$, $\widetilde{C}_n$ and $\widetilde{D}_n$ relies on a generalization for graded semilattices; see Section 6.

Another immediate consequence of Theorem A is the following.

Theorem C (Corollary 2.2) *The thickening of the Cayley graph of any Garside group with respect to its simple elements is a Helly graph.*

Note that, in the case of a finite-type Garside group, this is due to Huang and Osajda [2021]. However, our proof is different, and does not rely on the deep local-to-global result for Helly graphs (see Theorem 1.12). Additionally, it also works in the case of a Garside group with infinite set of simple elements.

One application is the study of the orthoscheme complex of a bounded, graded lattice L. Note that simplices of the geometric realization $|L|$ of L correspond to chains in L, so that one can endow each simplex with metric of the standard ℓ^∞ orthosimplex associated to this order on vertices (see Section 1 for details). We endow the geometric realization $|L|$ of L with a lattice structure, and use Theorem A to deduce the following.

Theorem D (Theorem 3.10) *Let L denote a bounded, graded lattice. Then the orthoscheme complex $|L|$ of L, with the piecewise ℓ^∞ metric, is injective.*

When L is a bounded, graded lattice, one may endow its orthoscheme complex with the piecewise Euclidean metric. Deciding whether it is a CAT(0) metric space is a very difficult question. It turns out that for orthoscheme complexes of posets, the CAT(0) property is more restrictive than the injective property; see Theorem 3.11.

One famous conjecturally CAT(0) example is the dual braid complex, defined by Brady and McCammond [2010]. The n–strand braid group B_n has a standard Garside structure, associated to the "half-turn" as a Garside element. The braid group also enjoys a dual presentation introduced by Birman, Ko and Lee [Birman et al. 1998], which corresponds to a dual Garside structure (see [Dehornoy and Paris 1999]). It is associated to the rotation of an n^{th} of a turn as a Garside element. Associated to this dual Garside structure, one may consider the geometric realization and endow it with the piecewise Euclidean orthoscheme metric (see Section 1). Brady and McCammond conjecture that this so-called dual braid complex is CAT(0) for all braid groups, but it has only been proved for $n \leqslant 7$ (see [Brady and McCammond 2010; Haettel et al. 2016; Jeong 2023]). However, an immediate consequence of Theorem D is the following.

Theorem E (Corollary 3.12) *The Garside complex of any Garside group, endowed with the piecewise ℓ^∞ orthoscheme metric, is injective.*

We also obtain another proof of a result by Huang and Osajda stating that FC-type Artin groups are Helly [Huang and Osajda 2021, Theorem 5.8]. We also provide explicit Helly and injective models (see Theorem 7.6 for the precise statement).

Theorem F (Huang–Osajda, Theorem 7.6) *Let A denote an Artin group of type FC. Then a natural simplicial complex X with vertex set A, with the ℓ^∞ metric, is injective. Moreover, a thickening of X is a Helly graph.*

As a particular case of Theorem E, the dual braid complex, endowed with the piecewise ℓ^∞ orthoscheme metric, is injective for all braid groups. This also holds, more generally, for every spherical-type Artin group with some Garside structure. Note that Theorem D proves local injectivity, and one also needs a Cartan–Hadamard result for injective metric spaces in order to conclude. We therefore rely on the local-to-global result for Helly graphs to prove the following generalization of [Miesch 2017] in the nonproper setting (see Section 1 for the definition of semiuniformly locally injective). This result is clearly of independent interest.

Theorem G (Cartan–Hadamard for injective metric spaces, Theorem 1.14) *Let X denote a complete, simply connected, semiuniformly locally injective metric space. Then X is injective.*

Another very promising family of examples is Deligne complexes of Artin groups (see Section 1 for definitions). Note that buildings and Deligne complexes of Artin groups are closely related: in fact in his original article Deligne [1972] called “buildings for generalized braid groups” the complexes later called Deligne complexes. However, to the best of our knowledge, the close relationship between Euclidean buildings and Deligne complexes of Euclidean-type Artin groups has not yet been exploited in the literature. Notably, the automorphism groups of Euclidean buildings do not possess a Garside structure, and the Deligne complexes do not have an apartment system as rich as in the building case. However, one common feature is that they locally look like a lattice, which is the key combinatorial property we are using in this article.

Associated to every Coxeter graph Γ with vertex set S, we may define the Coxeter group $W(\Gamma)$, the Artin group $A(\Gamma)$ and the hyperplane complement $M(\Gamma)$ (see Section 1). The Coxeter group $W(\Gamma)$ acts naturally on $M(\Gamma)$, and $A(\Gamma)$ is the fundamental group of the quotient $W(\Gamma)\backslash M(\Gamma)$. One very natural question is to decide whether it is a classifying space. This is the statement of the following conjecture.

Conjecture H (the $K(\pi,1)$ conjecture) *The hyperplane complement $M(\Gamma)$ is aspherical.*

The $K(\pi,1)$ conjecture has been proved for spherical-type Artin groups by Deligne [1972], and for Euclidean-type Artin groups by Paolini and Salvetti [2021] very recently, even for the type $\widetilde{D}_n$. Another approach, more closely related to the metric approach of this article, was used by Charney and Davis [1995b] to prove the $K(\pi,1)$ conjecture for Artin groups of type FC or of 2–dimensional type. Their proof relies on the use of a simplicial complex, called the Deligne complex $\Delta(\Gamma)$, which is the geometric realization of the poset of cosets of spherical-type parabolic subgroups (see Section 1). This complex (in this form) was defined in [Charney and Davis 1995b], where they proved that $\Delta(\Gamma)$ is homotopy equivalent to the universal cover of $W(\Gamma)\backslash M(\Gamma)$. In other words, the $K(\pi,1)$ conjecture is equivalent to the contractibility of the Deligne complex $\Delta(\Gamma)$.

Charney and Davis's method for proving that the Deligne complex is contractible is to endow it with a CAT(0) metric. This works for Artin groups of type FC or of 2–dimensional type. However, in the general case, the key question is to decide whether the Deligne complex for the braid group is CAT(0) with Moussong's metric. It is only known for the braid group up to four strands only (see [Charney 2004]). However, we will see that the natural piecewise ℓ^∞ metric is injective for all braid groups, up to taking the product with $\mathbb{R}$.

Theorem I (Theorem 4.4) *Let Δ denote the Deligne complex of the Artin group of Euclidean type $\tilde{A}_n$. Then the natural piecewise ℓ^∞ length metric on $\Delta\times\mathbb{R}$ is injective. Moreover, the thickening of $\Delta\times\mathbb{R}$ is a Helly graph.*

The proof consists in applying Theorem D to prove that the Deligne complex is locally injective. One key combinatorial property is that the Deligne complex in type A_n is essentially a lattice; see Section 5. The proof of the lattice property, through the cut-curve lattice, is due to Crisp and McCammond, copied here with their permission.

Note that one may wonder whether it is necessary to consider the direct product with $\mathbb{R}$. In fact, Hoda [2023] proved that the Euclidean Coxeter group of type $\tilde{A}_n$ is not Helly for $n\geqslant 2$, even though its direct product with $\mathbb{Z}$ is. We made a similar distinction for automorphism groups of Euclidean buildings of type $\tilde{A}_n$ in [Haettel 2022a]. We therefore strongly believe that there is no injective metric on the Deligne complex of type $\tilde{A}_n$ itself which is invariant under the Artin group. However, we will see in Theorem M below that there is a convex bicombing on the Deligne complex itself.

In order to deal with the Euclidean type $\widetilde{C}_n$, we first prove a generalization of Theorem D for graded semilattices; see Section 6.

Theorem J (Theorem 7.4) *Let Δ denote the Deligne complex of the Artin group of Euclidean type $\widetilde{C}_n$. Then the natural piecewise ℓ^∞ length metric on Δ is injective. Moreover, the thickening of Δ is a Helly graph.*

An immediate consequence is another proof of the $K(\pi,1)$ conjecture in Euclidean types $\tilde{A}_n$ and $\widetilde{C}_n$, originally due to Okonek [1979]. The novelty is that it is the first metric proof. Moreover, in Charney and Davis's approach to the $K(\pi,1)$ conjecture by showing that Moussong's metric is CAT(0), the main difficulty is to prove that it is locally CAT(0) for the braid groups. It is precisely this statement that we are able to prove in the injective setting.

Corollary K (Okonek) *The Deligne complex Δ of Euclidean type $\tilde{A}_n$ or $\widetilde{C}_n$ is contractible. In particular, the $K(\pi,1)$ conjecture holds in these cases.*

Moreover, several results have been proved, relying on the assumption that one may endow the Deligne complex with a piecewise Euclidean CAT(0) metric. Crisp [2000] studied the fixed-point subgroup under a symmetry group of the Artin system. Godelle [2007] studied the centralizer and normalizer of standard parabolic subgroups. Morris-Wright [2021] studied the intersections of parabolic subgroups. It turns out almost all the arguments merely used the existence of an equivariant geodesic bicombing on the Deligne complex (see the end of Section 3 for a definition of a bicombing). Moreover, Descombes and Lang [2015; 2016] justified the importance of convex geodesic bicombings for themselves. We therefore state the following conjecture, which may be seen as a metric strategy for the proof of the $K(\pi,1)$ conjecture.

Conjecture L *The Deligne complex of any Artin group A has an A–invariant metric that admits a convex, consistent, reversible geodesic bicombing.*

We are able to prove this conjecture in spherical types A_n, B_n and Euclidean types $\tilde{A}_n$, $\widetilde{C}_n$, and we believe that our result represents a major step towards the general case.

Theorem M (Theorem 8.1) *Let Δ denote the Deligne complex of the Artin group of spherical type A_n or B_n or Euclidean type $\tilde{A}_n$ or $\widetilde{C}_n$. There exists a metric on Δ, invariant under the Artin group, which admits a convex, consistent, reversible geodesic bicombing.*

If an Artin group satisfies Conjecture L, then the $K(\pi,1)$ conjecture follows. Moreover, we may also list consequences of [Crisp 2000; Godelle 2007; Morris-Wright 2021; Cumplido et al. 2019] that rely on the assumption that the Deligne complex has a CAT(0) metric. However, most of the arguments only use the geodesic bicombing. The following results were only essentially known for Artin groups of type FC or of 2–dimensional type. Note that concerning Artin groups of type FC, only intersections of spherical-type parabolic subgroups are known to be parabolic (see [Morris-Wright 2021]). See also [Möller et al. 2024].

Corollary N (Corollaries 8.3, 8.4 and 8.5) *Let A denote the Artin group of Euclidean type $\tilde{A}_n$ or $\widetilde{C}_n$.*

- *The intersection of any parabolic subgroups of A is a parabolic subgroup.*
- *A satisfies Properties $(\star)$, $(\star\star)$ and $(\star\star\star)$ from [Godelle 2007], notably: for any subset $X \subset S$, we have*
$$\mathrm{Com}_A(A_X) = N_A(A_X) = A_X \cdot QZ_A(X),$$
where the quasicentralizer of X is $QZ_A(X) = \{g \in A \mid g \cdot X = X\}$.

- *For any group G of symmetries of the Artin system, the fixed-point subgroup A^G is isomorphic to an Artin group.*

In view of Conjecture L, looking for other consequences of the existence of a convex bicombing on the Deligne complex may prove to be fruitful.

Structure of the article

In Section 1, we review basic definitions of posets, lattices, Artin groups, Deligne complexes, injective metric spaces and Helly graphs. We also prove the Cartan–Hadamard theorem for injective metric spaces. In Section 2, we prove the central simple criterion showing how to produce an injective metric space or a Helly graph starting from a lattice with an action of $\mathbb{Z}$ or $\mathbb{R}$. In Section 3, we apply this criterion to prove that the orthoscheme complex of a bounded, graded lattice is injective. In Section 4, we use this to prove that, for Euclidean buildings and the Deligne complex in Euclidean type $\tilde{A}_n$, the natural piecewise ℓ^∞ metric is injective. In Section 6, we show how to adapt the criterion to a mere semilattice with some extra property. We then apply it in Section 7 to prove that, for Euclidean buildings and the Deligne complex in Euclidean type $\widetilde{C}_n$, the natural piecewise ℓ^∞ metric is injective. Finally, in Section 8, we use the convex bicombing on the Deligne complexes to deduce many corollaries about parabolic subgroups of Artin groups.

Acknowledgments

We would like to thank very warmly Jingyin Huang, Jon McCammond, Damian Osajda and Luis Paris for very interesting discussions. We are indebted to Jon McCammond and John Crisp for their authorization to copy their work about the lattice of cut-curves. We also would like to thank the organizers of the ICMS Edinburgh 2021 meeting *Perspectives on Artin groups*, where we had fruitful discussions. We also would like to thank Jérémie Chalopin, Victor Chepoi, María Cumplido, Anthony Genevois, Hiroshi Hirai, Alexandre Martin and Bert Wiest for various discussions. We also thank an anonymous referee for many helpful comments.

We acknowledge support from French project ANR-16-CE40-0022-01 AGIRA.

1 Review of posets, Artin groups and injective metric spaces

1.1 Posets and lattices

We recall briefly basic definitions related to posets and lattices.

Definition 1.1 (poset) Let L denote a poset. A *chain* of L is a totally ordered nonempty subset of L. A *maximal chain* is a chain that is maximal with respect to inclusion. A finite chain of $n+1$ elements $x_0 < x_1 < \cdots < x_n$ is called of *length* n. The poset L is called *bounded below* (resp. above) if it has a global minimum denoted by 0 (resp. global maximum denoted by 1). The poset L is called *bounded* if it is both bounded above and below. The poset L has *rank* n if it is bounded and all maximal chains have length n.

Definition 1.2 (interval) Given two elements $x \leqslant y$ in a poset L, we define the *interval*

$$I(x,y) = \{z \in L \mid x \leqslant z \leqslant y\}.$$

In the case of possible confusion, we may also denote the interval by $I_L(x,y)$ to emphasize that we are considering the interval in the poset L. The poset L is called *graded* if every interval of L has a rank. Let x denote an element in a graded poset L that is bounded below; the *rank* of x is the rank of the interval $I(0,x)$.

Definition 1.3 (lattice) Given elements x, y in a poset L, if there exists a unique maximal lower bound to $\{x,y\}$, it is called the *meet* of x and y and denoted by $x \wedge y$. Similarly, if there exists a unique minimal upper bound to $\{x,y\}$, it is called the *join* of x and y and denoted by $x \vee y$. If any two elements of L have a meet (resp. a join), the poset L is called a *meet-semilattice* (resp. join-semilattice). If any two elements of L have a meet and a join, the poset L is called a *lattice*.

Examples • The Boolean lattice L of rank n is the poset of subsets of $E = \{1,\dots,n\}$, partially ordered by inclusion. The join of $A, B \in L$ is $A \cup B$ and their meet is $A \cap B$.

• Consider a CAT(0) cube complex X, with a base vertex v_0. Order the set V of vertices of X by declaring that $v \leqslant w$ if some combinatorial geodesic from v_0 to w passes through v. Then V is a graded meet-semilattice, with minimum v_0. The meet of two vertices $v, w \in V$ is the median of v_0, v, w.

• The partition lattice L of $E = \{1,\dots,n\}$ is the poset of partitions of E, partially ordered by declaring that $A \leqslant B$ if every element of A is contained in an element of B. This lattice has rank $n-1$, its minimum is the partition $\{\{1\},\{2\},\dots,\{n\}\}$ into singletons and its maximum is the partition $\{E\}$ into one element.

We will now describe a very simple criterion due to [Brady and McCammond 2010] to decide when a bounded graded poset is a lattice.

Definition 1.4 (bowtie) In a poset L, a *bowtie* consists of four distinct elements a, b, c, d such that a, c are minimal upper bounds of b, d, and b, d are maximal lower bounds of a, c.

Proposition 1.5 [Brady and McCammond 2010, Proposition 1.5] *Let L denote a bounded graded poset. Then L is a lattice if and only if L does not contain a bowtie.*

1.2 Coxeter groups, Artin groups and Deligne complexes

We recall the definitions of Coxeter groups, Artin groups, and their associated Deligne complexes.

For every finite simple graph Γ with vertex set S and with edges labeled by some integer in $\{2,3,\dots\}$, one associates the Coxeter group $W(\Gamma)$ with the following presentation:

$$W(\Gamma) = \langle S \mid \text{for all } \{s,t\} \in \Gamma^{(1)}, \text{ for all } s \in S,\ s^2 = 1,\ [s,t]_m = [t,s]_m \text{ if the edge } \{s,t\} \text{ is labeled } m\rangle,$$

where $[s,t]_m$ denotes the word $ststs\dots$ of length m.

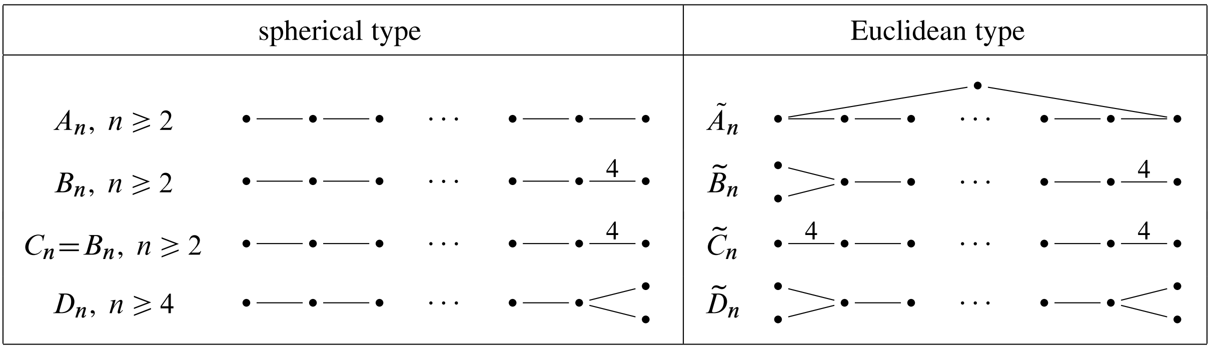


Table 1: Spherical and Euclidean irreducible diagrams of type A, B, C and D.

The associated Artin group $A(\Gamma)$ is defined by a similar presentation:

$$A(\Gamma) = \langle S \mid \text{for all } \{s,t\} \in \Gamma^{(1)},\ [s,t]_m = [t,s]_m \text{ if the edge } \{s,t\} \text{ is labeled } m \rangle.$$

The groups $A(\Gamma)$ are also called Artin–Tits groups, since they were defined by Tits [1966].

Note that only the relations $s^2 = 1$ have been removed, so that there is a natural surjective morphism from $A(\Gamma)$ to $W(\Gamma)$. Also note that if $m = 2$, then s and t commute, and if $m = 3$, then s and t satisfy the classical braid relation $sts = tst$.

The knowledge of general Artin groups is extremely limited (see notably [McCammond 2017; Charney 2008; Godelle and Paris 2012]). In particular, we do not know whether the word problem is solvable in general Artin groups, nor whether they are torsion-free.

Most results about Artin–Tits groups concern particular classes. The Artin group $A(\Gamma)$ is called:

- of *spherical type* if its associated Coxeter group $W(\Gamma)$ is finite, ie may be realized as a reflection group of a sphere.
- of *Euclidean type* if its associated Coxeter group $W(\Gamma)$ may be realized as a reflection group of a Euclidean space.
- of *FC type* if for any complete subset $T \subset S$ the parabolic subgroup $A_T = \langle T \rangle$ is spherical.

We recall in Table 1 the classification of the four infinite families of spherical and Euclidean irreducible diagrams; see [Bourbaki 2002] for the full classification. We only present those because we will only consider these types in this article. Note that we use in this table the convention of Dynkin diagrams: vertices that are not joined by an edge commute, and we drop the label 3 from edges.

Artin groups are closely related to hyperplane complements, which can be presented in a simple way in spherical and Euclidean types. Fix a Coxeter group $W = W(\Gamma)$ of spherical type or Euclidean type acting by reflections on a sphere $\mathbb{S}^{n-1}$ or a Euclidean space $\mathbb{R}^n$. In the case W is of spherical type, consider the associated linear action on $\mathbb{R}^n$. A conjugate of an element of the standard generating set S is called a *reflection*. Let $\mathcal{R}$ denote the set of reflections in W. Consider the family of affine hyperplanes of $\mathbb{R}^n$

$$\mathcal{H} = \{H_r \mid r \in \mathcal{R}\},$$

where $H_r \subset \mathbb{R}^n$ denotes the fixed-point set of the reflection r.

The complement of the complexified hyperplane arrangement is

$$M(\Gamma) = \mathbb{C}^n \setminus \bigcup_{r \in \mathcal{R}} (\mathbb{C} \otimes H_r).$$

Note that W acts naturally on M, and we have the following (see [van der Lek 1983]):

$$\pi_1(W(\Gamma)\backslash M(\Gamma)) \simeq A(\Gamma).$$

So the Artin group $A(\Gamma)$ appears as the fundamental group of (a quotient of) the complement of a complexified hyperplane arrangement. One very natural question is to decide whether it is a classifying space. This is the statement of the following conjecture.

Conjecture (the $K(\pi,1)$ conjecture) *The space $M(\Gamma)$ is aspherical.*

This conjecture was proved for spherical-type Artin groups by Deligne [1972], and for Euclidean-type Artin groups by Paolini and Salvetti [2021] very recently, even for the type $\widetilde{D}_n$. Another approach, more closely related to the content of this article, was used by Charney and Davis [1995a] to prove the $K(\pi,1)$ conjecture for Artin groups of type FC or of 2–dimensional type. Their proof relies on the use of a simplicial complex, called the Deligne complex, and they endow it with a particular metric to show that it is contractible.

We will now recall the definition of the Deligne complex of an Artin group $A = A(\Gamma)$. A *standard parabolic* subgroup of A is the subgroup $A_T = \langle T \rangle$ generated by a subset T of S, the standard generating set of A. A *parabolic* subgroup denotes any conjugate of a standard parabolic subgroup. Let us define

$$\mathcal{S}_f = \{T \subset S \mid W_T \text{ is finite}\}.$$

The *Deligne complex* $\Delta = \Delta(\Gamma)$ is the order complex of the set of cosets of parabolic subgroups

$$\{gA_T \mid g \in A,\ T \in \mathcal{S}_f\},$$

where the partial order is given by the inclusion $gA_T \subset g'A_{T'}$ of cosets.

One key property of the Deligne complex is that it has the same homotopy type as the universal cover of the hyperplane complement:

Theorem 1.6 [Charney and Davis 1995b, Theorem 1.5.1] *The Deligne complex $\Delta(\Gamma)$ is homotopy equivalent to the universal cover of the quotient of the hyperplane complement $W(\Gamma)\backslash M(\Gamma)$.*

In particular, the $K(\pi,1)$ conjecture amounts to proving that the Deligne complex is contractible.

1.3 Injective metrics and Helly graphs

We briefly recall basic definitions of injective metric spaces and Helly graphs. We will also state the local-to-global result for Helly graphs and deduce the analogous Cartan–Hadamard result for injective metric spaces.

A geodesic metric space is called *injective* (one may also say *hyperconvex*, or *absolute* 1*–Lipschitz retract*) if any family of pairwise intersecting closed balls has a nonempty global intersection, the so-called *Helly property*. We refer the reader to [Lang 2013] for a presentation of injective metric spaces.

Examples • The normed vector space $(\mathbb{R}^n, \ell^\infty)$ is injective for all $n \geqslant 1$. In fact, it is up to isometry the only injective norm on $\mathbb{R}^n$ (see [Nachbin 1950]).

- Any tree is injective.
- Any finite-dimensional CAT(0) cube complex, endowed with the standard piecewise ℓ^∞ metric, is injective (see [Miesch 2014] and [Bowditch 2020]).

A connected graph X is called *Helly* if any family of pairwise intersecting combinatorial balls of X has a nonempty global intersection, in other words the combinatorial balls satisfy the Helly property. We refer the reader to [Chalopin et al. 2020a] for a presentation of Helly graphs and Helly groups. Many examples of Helly graphs come from thickening of complexes, which we define now (see [Chalopin et al. 2020a; Huang and Osajda 2021]).

Definition 1.7 (thickening) Let X denote a cell complex. The *thickening* of X (with respect to the cell structure) is the graph with vertex set $X^{(0)}$, with an edge between two vertices if and only if they are contained in a common cell of X.

Examples • For each $n \geqslant 1$, the graph with vertex set $\mathbb{Z}^n$, with an edge between v, w if $d_\infty(v, w) = 1$, is a Helly graph.

- Any simplicial tree is a Helly graph.
- The thickening of the vertex set of a CAT(0) cube complex is a Helly graph.

In order to endow a simplicial complex with a potentially injective (or CAT(0)) metric, it is natural to ask for metric simplices which may tile the Euclidean space $\mathbb{R}^n$. One choice is to consider the barycentric subdivision of the standard cubical tiling of $\mathbb{R}^n$, whose simplices are orthosimplices, which we now formally define.

Definition 1.8 (orthosimplex) The *standard orthosimplex* of dimension n is the simplex of $\mathbb{R}^n$ with vertices $(0, \dots, 0), (1, 0, \dots, 0), \dots, (1, 1, \dots, 1)$ (see Figure 1). One may endow the simplex with the standard ℓ^p metric on $\mathbb{R}^n$ for any $p \in [1, \infty]$. Throughout this article (except in Theorem 3.11), we will

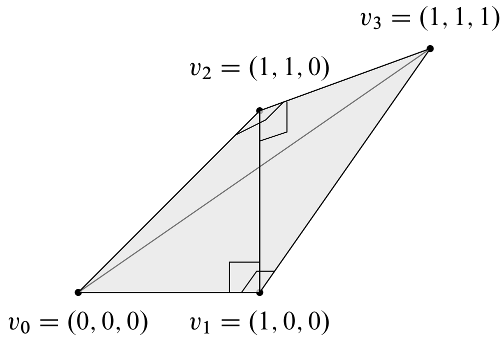


Figure 1: The standard 3–dimensional orthosimplex.

only consider the ℓ^∞ metric, called the ℓ^∞ orthosimplex. Note that any n–simplex with a total order on its vertices $v_0 < v_1 < \cdots < v_n$ may be identified uniquely with the ℓ^∞ orthosimplex of dimension n, where each v_i is identified with the vertex $(1, \ldots, 1, 0, \ldots, 0)$ with i ones and $n-i$ zeros. Also note that reversing the total order on the vertices gives rise to an isometry of the orthosimplex.

Definition 1.9 (simplicial complex with ordered simplices and maximal edges) We say X is a simplicial complex X *with ordered simplices* if each simplex of X has a total order on its vertex set, which is consistent with respect to inclusions of simplices of X. Moreover, we say that X has *maximal edges* if, given any simplex σ of X, there exist adjacent vertices a, b in X such that $\sigma \cup \{a, b\}$ is a simplex of X, and such that any vertex c of X adjacent to both a and b satisfies $a \leqslant c \leqslant b$. Such an edge $\{a, b\}$ is then called a maximal edge of X.

Definition 1.10 (orthoscheme complex) Let X denote a finite-dimensional simplicial complex with ordered simplices. Then one may endow each simplex of X with the associated ℓ^∞ orthoscheme metric. Then the geometric realization of $|X|$, endowed with the length metric associated with the ℓ^∞ orthoscheme metric on each simplex, is called the ℓ^∞ *orthoscheme complex* of X.

As a particular case, if L is a poset with a bound on the length of its chains, then the geometric realization of $|L|$ satisfies the assumptions, so we may talk about the ℓ^∞ orthoscheme complex of L. The following is an immediate adaptation of [Bridson and Haefliger 1999, Theorem 7.13].

Theorem 1.11 *Let X denote a finite-dimensional simplicial complex with ordered simplices. Then its ℓ^∞ orthoscheme complex is a complete length space.*

Proof Note that since X has dimension n, the ℓ^∞ orthoscheme complex of X has finitely many isometry types of cells: the standard orthoscheme k–simplices, where $k \leqslant n$. The proof of [Bridson and Haefliger 1999, Theorem 7.13] adapts without change to this situation. □

One key property in the study of Helly graphs is the following local-to-global statement (see [Chalopin et al. 2020b]). Recall that a *clique* of a graph is a complete subgraph. A graph is called *clique-Helly* if its family of maximal cliques satisfies the Helly property (ie any family of pairwise intersecting maximal cliques has nonempty intersection). The triangle complex of a simplicial graph X is the simplicial 2–complex whose 1–skeleton is X, and whose 2–simplices correspond to triangles in X.

Theorem 1.12 [Chalopin et al. 2020b, Theorem 3.5] *Let X denote a graph. Then X is Helly if and only if X is clique-Helly and its triangle complex is simply connected.*

In order to transfer this local-to-global property to injective metric spaces, we will need the following technical lemma. Say that a metric space is *ε–coarsely injective* (for some $\varepsilon \geqslant 0$) if, for any families $(x_i)_{i \in I}$ in X and $(r_i)_{i \in I}$ in $\mathbb{R}_+$ such that, for all $i, j \in I$, $d(x_i, x_j) \leqslant r_i + r_j$, we have $\bigcap_{i \in I} B_X(x_i, r_i + \varepsilon) \neq \varnothing$. Note that when $\varepsilon = 0$, we recover the definition of an injective metric space.

Lemma 1.13 *Let X denote a complete metric space that is ε–coarsely injective for every $\varepsilon > 0$. Then X is injective.*

Proof We will prove that X is injective: consider a family $(B_X(x_i, r_i))_{i \in I}$ of pairwise intersecting balls in X. We know that, for all $i, j \in I$, $d(x_i, x_j) \leqslant r_i + r_j$. For any $\varepsilon > 0$, let us define $A_\varepsilon = \bigcap_{i \in I} B_X(x_i, r_i + \varepsilon)$, which is nonempty by the assumption of ε–coarse injectivity.

Fix $0 < \varepsilon \leqslant \varepsilon'$. We will prove that the Hausdorff distance between A_ε and $A_{\varepsilon'}$ is at most $\varepsilon + \varepsilon'$. Note that $A_\varepsilon \subset A_{\varepsilon'}$. Fix $x_0 \in A_{\varepsilon'}$. We will prove that $d(x_0, A_\varepsilon) \leqslant \varepsilon + \varepsilon'$. Let $I_0 = I \sqcup \{0\}$, and let $r_0 = \varepsilon'$. Consider the families $(x_i)_{i \in I_0}$ in X and $(r_i)_{i \in I_0}$ in $\mathbb{R}_+$. For each $i, j \in I_0$, we know that $d(x_i, x_j) \leqslant r_i + r_j$: indeed, for any $i \in I$, we have $x_0 \in B_X(x_i, r_i + \varepsilon')$. By ε–coarse injectivity, we deduce that the intersection $\bigcap_{i \in I_0} B_X(x_i, r_i + \varepsilon)$ is not empty. In particular, the ball $B_X(x_0, \varepsilon + \varepsilon')$ intersects $A_\varepsilon = \bigcap_{i \in I} B_X(x_i, r_i + \varepsilon)$. This implies that $d(x_0, A_\varepsilon) \leqslant \varepsilon + \varepsilon'$. So we have proved that the Hausdorff distance between A_ε and $A_{\varepsilon'}$ is at most $\varepsilon + \varepsilon'$.

For each $n \in \mathbb{N}$, consider by induction $x_n \in A_{2^{-n}}$ such that, for all $n \geqslant 0$,

$$d_X(x_{n+1}, x_n) \leqslant 2^{-n} + 2^{-(n+1)} \leqslant 2^{-n+1}.$$

For each $0 \leqslant n \leqslant m$, we have $d_X(x_n, x_m) \leqslant 2^{-n+2}$; hence the sequence $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in X. Since X is complete, it has a limit $y \in X$. For each $n \in \mathbb{N}$, we have $y \in A_{2^{-n}}$, so for each $i \in I$ we have $d_X(y, x_i) \leqslant r_i + 2^{-n}$. We deduce that, for each $i \in I$, we have $d_X(y, x_i) \leqslant r_i$. In other words, y belongs to the intersection $\bigcap_{i \in I} B(x_i, r_i)$: we have proved that X is injective. □

Say that a metric space is *uniformly locally injective* if there exists $\varepsilon > 0$ such that each ball of radius ε is injective. Say that a metric space is *semiuniformly locally injective* if there exists $\varepsilon > 0$ such that each ball of radius ε is uniformly locally injective. For instance, any locally compact, locally injective metric space is semiuniformly locally injective.

As a concrete example, if X denotes the injective hull of the hyperbolic plane $\mathbb{H}^2_\mathbb{R}$ and Γ is a nonuniform lattice in $\mathrm{PGL}(2, \mathbb{R}) = \mathrm{Isom}(\mathbb{H}^2_\mathbb{R})$, then the quotient $\Gamma \backslash X$ is semiuniformly locally injective, but not uniformly locally injective. There are similar examples in higher rank; see [Haettel 2022a].

There are also nonlocally compact examples: let X denote a metric simplicial graph such that the systole at each vertex of X is bounded below. Then X is locally uniformly locally injective. However, if the systole of X is 0, then X is not uniformly locally injective.

We now prove a Cartan–Hadamard theorem for such injective metric spaces, relying on the local-to-global property for Helly graphs, Theorem 1.12. Note that this statement generalizes [Miesch 2017, Theorem 1.2] without the local compactness assumption.

Theorem 1.14 (Cartan–Hadamard for injective metric spaces) *Let X denote a complete, simply connected, semiuniformly locally injective metric space. Then X is injective.*

Proof We will first prove the statement when X is uniformly locally injective. Fix $\varepsilon > 0$ small enough such that balls in X of radius at most 2ε are injective. Consider the graph Γ_ε with vertex set X and with an edge between $x, y \in X$ if $d(x,y) \leqslant \varepsilon$. Since X is geodesic, Γ_ε is a connected graph. Also note that, for any $x \in X$ and $n \in \mathbb{N}$, we have

$$B_{\Gamma_\varepsilon}(x,n) = B_X(x, n\varepsilon).$$

We will prove that, for each $\varepsilon > 0$, the graph Γ_ε is a Helly graph by applying Theorem 1.12.

We first prove that the family of combinatorial 1–balls in Γ_ε satisfies the Helly property. Fix a family of vertices $(x_i)_{i\in I}$ of Γ_ε such that, for all $i, j \in I$, $d_{\Gamma_\varepsilon}(x_i, x_j) \leqslant 2$. We want to prove that these balls intersect in Γ_ε.

The family of metric balls $(B_X(x_i,\varepsilon))_{i\in I}$ in X pairwise intersects: since such balls have the Helly property by the assumption on X, we deduce that there exists $y \in X$ such that, for all $i \in I$, $d_X(x_i, y) \leqslant \varepsilon$. In other words, the vertex $y \in \Gamma_\varepsilon$ lies in the intersection of all combinatorial 1–balls $(B_{\Gamma_\varepsilon}(x_i,1))_{i\in I}$.

We now deduce that Γ_ε is clique-Helly: fix a family of pairwise intersecting maximal cliques $(\sigma_i)_{i\in I}$ of Γ_ε. Then the family of combinatorial 1–balls centered at each vertex of each clique σ_i, for $i \in I$, pairwise intersects: according to the previous paragraph, we deduce that there exists a vertex $y \in \Gamma_\varepsilon$ adjacent to each vertex of each clique σ_i for $i \in I$. Since each clique σ_i is maximal, we deduce that y belongs to the intersection of all σ_i for $i \in I$. The graph Γ_ε is clique-Helly.

We now prove that the triangle complex of Γ_ε is simply connected. Fix a combinatorial loop ℓ in Γ_ε. Since X is simply connected, there exists a disk D in X bounding ℓ. Consider a triangulation T of D such that triangles have diameter for d_X at most ε. Then the vertex set of each triangle of T is a clique in Γ_ε; therefore ℓ is null-homotopic in the triangle complex of Γ_ε. So the triangle complex of Γ_ε is simply connected.

The graph Γ_ε is clique-Helly and has a simply connected triangle complex, so according to Theorem 1.12, we deduce that Γ_ε is a Helly graph.

Note that, for any $\varepsilon > 0$, we have $d_X \leqslant \varepsilon d_{\Gamma_\varepsilon} \leqslant d_X + \varepsilon$, and balls for the metric d_{Γ_ε} with integral radius satisfy the Helly property.

We will show that the metric space X is ε–coarsely injective: consider families $(x_i)_{i\in I}$ in X and $(r_i)_{i\in I}$ in $\mathbb{R}_+$ such that, for all $i,j\in I$, $d_X(x_i,x_j)\leqslant r_i+r_j$. For each $i\in I$, let $n_i\in\mathbb{N}$ such that $n_i\varepsilon\leqslant r_i<(n_i+1)\varepsilon$. For each $i,j\in I$, since the balls $B_X(x_i,r_i)$ and $B_X(x_j,r_j)$ intersect, we deduce that $d_{\Gamma_\varepsilon}(x_i,x_j)\leqslant n_i+n_j+2$. So, in the Helly graph Γ_ε, the balls $(B_{\Gamma_\varepsilon}(x_i,n_i+1))_{i\in I}$ pairwise intersect: we deduce that there exists $y\in\Gamma_\varepsilon$ such that, for all $i\in I$, $d_{\Gamma_\varepsilon}(y,x_i)\leqslant n_i+1$. In particular, for any $i\in I$, we have $d_X(y,x_i)\leqslant(n_i+1)\varepsilon\leqslant r_i+\varepsilon$. Hence y belongs to each ball $B_X(x_i,r_i+\varepsilon)$, for $i\in I$: this proves that X is ε–coarsely injective.

Since this holds for any small enough $\varepsilon>0$, according to Lemma 1.13, we conclude that X is injective.

We now turn to the general case, when X is only semiuniformly locally injective: there exists $\varepsilon>0$ such that each ball of radius ε in X is uniformly locally injective. According to the uniformly locally injective case, we deduce that each ball of radius ε is injective: this means that X is uniformly locally injective. According to the uniformly locally injective case again, we deduce that X is injective. □

We now see that, under a mild assumption on a simplicial complex with ordered simplices, saying that the ℓ^∞ orthoscheme realization is injective is equivalent to saying that the thickening of the 1–skeleton is Helly. Note that, if we refer to the definition of thickening as in Definition 1.7, the corresponding cell structure is not the simplicial one, but a coarser cell structure whose cells correspond to intervals.

Theorem 1.15 *Let X denote a finite-dimensional simplicial complex with ordered simplices, and with maximal edges. Let Γ denote the graph with vertex set $X^{(0)}$, and with an edge between $x,y\in X^{(0)}$ if there exist $a,b\in X^{(0)}$ and ordered triangles $a\leqslant x\leqslant b$ and $a\leqslant y\leqslant b$ in X. If the ℓ^∞ orthoscheme complex of X is injective, then the thickening Γ of X is a Helly graph.*

Proof We see that maximal cliques in Γ correspond to intervals

$$I_{ab}=\{x\in X^{(0)}\mid a\leqslant x\leqslant b \text{ is an ordered triangle in } X\}$$

for any maximal edge $a\leqslant b$ in X. Let $m_{ab}\in|X|$ denote the midpoint of the maximal edge $a\leqslant b$. By assumption, any simplex of X containing a and b has for its vertex set a chain from a to b. Then $B_{|X|}\bigl(m_{ab},\frac{1}{2}\bigr)$ is a subcomplex of $|X|$ with vertex set $B_{|X|}\bigl(m_{ab},\frac{1}{2}\bigr)\cap X^{(0)}=I_{ab}$.

We will prove that Γ is clique-Helly: let $(I_{a_ib_i})_{i\in I}$ denote a family of pairwise intersecting maximal cliques in Γ. Since $|X|$ is injective, there exists $z\in\bigcap_{i\in I}B_{|X|}\bigl(m_{a_ib_i},\frac{1}{2}\bigr)$. Since each such ball is a subcomplex of $|X|$, we may assume that z is a vertex of X. We deduce that z belongs to each clique $I_{a_ib_i}$ for $i\in I$. So Γ is clique-Helly.

We will now prove that the triangle complex of Γ is simply connected. Let ℓ denote a combinatorial loop in the 1–skeleton of Γ. Up to homotopy in the triangle complex of Γ, we may assume that ℓ lies in the 1–skeleton of X. Since $|X|$ is injective, it is contractible, so the 2–skeleton of X is simply connected. As the 2–skeleton of X is contained in the triangle complex of Γ, we conclude that ℓ is null-homotopic in the triangle complex of Γ.

According to Theorem 1.12, we deduce that Γ is Helly. □

2 The thickening of a lattice

We will explain a very simple construction of Helly graphs and injective metric spaces, starting with a lattice endowed with an action of the group $\mathbb{Z}$ or $\mathbb{R}$.

Assume that L is a lattice such that each upper bounded subset of L has a join. Assume that there is an order-preserving, increasing, continuous (with respect to the order topology on L) action $(f_t)_{t\in H}$ of $H = \mathbb{Z}$ or $\mathbb{R}$ on L, such that,

$$\text{for all } x, y \in L, \text{there exists } t \in H_+ \text{ such that } f_{-t}(x) \leqslant y \leqslant f_t(x).$$

Let us define the following metric d on L:

$$\text{for all } x, y \in L, \quad d(x, y) = \inf\{t \in H_+ \mid f_{-t}(x) \leqslant y \leqslant f_t(x)\}.$$

Theorem 2.1 • *If $H = \mathbb{Z}$, then (L, d) is a Helly graph with the combinatorial distance.*

• *If $H = \mathbb{R}$, then (L, d) is injective.*

Proof We start by proving that d is indeed a metric on L. If $x, y \in L$ and $t \in H_+$ are such that $f_{-t}(x) \leqslant y \leqslant f_t(x)$, then by applying f_t and f_{-t} we deduce that $f_{-t}(y) \leqslant x \leqslant f_t(y)$: the metric d is symmetric.

We will now prove the triangle inequality: let $x, y, z \in L$, and for $\varepsilon > 0$ consider $t, s \in H_+$ such that $d(x, y) \leqslant t < d(x, y) + \varepsilon$ and $d(y, z) \leqslant s < d(y, z) + \varepsilon$. We have

$$f_{-t}(x) \leqslant y \leqslant f_t(x) \quad \text{and} \quad f_{-s}(y) \leqslant z \leqslant f_s(y).$$

Hence

$$f_{-t-s}(x) \leqslant f_{-s}(y) \leqslant z \leqslant f_s(y) \leqslant f_{s+t}(x),$$

so $d(x, z) \leqslant t + s \leqslant d(x, y) + d(y, z) + 2\varepsilon$. This holds for any $\varepsilon > 0$; hence $d(x, z) \leqslant d(x, y) + d(y, z)$.

We will now prove that the metric is positive: assume that $x, y \in L$ are distinct, we will prove that $d(x, y) > 0$.

• Assume first that x, y are comparable, for instance $x < y$. Since $\{z \in L \mid z < y\}$ is open and contains x, by continuity of the action there exists $\varepsilon > 0$ such that, for any $|t| \leqslant \varepsilon$, we have $f_t(x) < y$. In particular $d(x, y) \geqslant \varepsilon$.

• Assume now that x, y are not comparable. Since $U = \{z \in L \mid x \wedge y < z < x \vee y\}$ is open and contains x, by continuity of the action there exists $\varepsilon > 0$ such that, for any $|t| \leqslant \varepsilon$, we have $f_t(x) \in U$. For any $0 \leqslant t < \varepsilon$, since $x \leqslant f_t(x) < x \vee y$, we have $f_t(x) \ngeqslant y$. In particular, $d(x, y) \geqslant \varepsilon$.

So we have proved that d is a metric on L. We will now prove that balls in L satisfy the Helly property.

We will first consider the case $H=\mathbb{Z}$. Let us define a graph $\widehat{L}$ with vertex set L, and with an edge between $x,y\in L$ if $f_{-1}(x)\leqslant y\leqslant f_1(x)$: we will prove that the graph $\widehat{L}$ is Helly. We will first prove, by induction on $k\geqslant 0$, that for any $x\in L$ the ball $B_{\widehat{L}}(x,k)$ in the graph $\widehat{L}$ coincides with the interval $I(f_{-k}(x),f_k(x))=\{y\in L\mid f_{-k}(x)\leqslant y\leqslant f_k(x)\}$.

For $k\leqslant 1$ it is the definition of the edges of $\widehat{L}$, so fix $k\geqslant 2$ and assume that the statement holds for $k-1$. Fix $y\in I(f_{-k}(x),f_k(x))$, we will prove that $y\in B_{\widehat{L}}(x,r)$. Since $y\geqslant f_{-k}(x)$, we deduce that $f_1(y)\geqslant f_{-k+1}(x)$, and also since $y\leqslant f_k(x)$ we deduce that $f_{-1}(y)\leqslant f_{k-1}(x)$. So we have that both $f_1(y)$ and $f_{k-1}(x)$ are superior to both $f_{-1}(y)$ and $f_{-k+1}(x)$: since L is a lattice, there exists some element $z\in L$ in the intersection $I(f_{-1}(y),f_1(y))\cap I(f_{-k+1}(x),f_{k-1}(x))$. In particular, y and z are adjacent in $\widehat{L}$, and by induction we know that $d_{\widehat{L}}(z,x)\leqslant k-1$, so $d_{\widehat{L}}(x,y)\leqslant k$. Conversely, it is clear that the ball $B_{\widehat{L}}(x,k)$ is included in the interval $I(f_{-k}(x),f_k(x))$. So we have $B_{\widehat{L}}(x,k)=I(f_{-k}(x),f_k(x))$.

By assumption on the lattice L, we deduce that the graph $\widehat{L}$ is connected, and furthermore that $d_{\widehat{L}}=d$.

Note now that intervals in the lattice L satisfy the Helly property. Fix any collection $(I(x_i,y_i))_{i\in I}$ of pairwise intersecting intervals of L. Fix $j_0\in I$. For any $i\in I$, we have $x_i\leqslant y_{j_0}$, so the set $\{x_i\mid i\in I\}$ is upper bounded. By assumption, the set $\{x_i\mid i\in I\}$ has a join $z\in L$ such that $z\leqslant y_{j_0}$. This holds for any $j_0\in I$, so z belongs to the intersection of all intervals $(I(x_i,y_i))_{i\in I}$.

Hence the graph $\widehat{L}$ is connected, and its balls satisfy the Helly property: it is a Helly graph.

We now turn to the case $H=\mathbb{R}$, and we will prove that (L,d) is an injective metric space. First note that balls in (L,d) are intervals in L, so according to the previous argument, we know that balls in (L,d) satisfy the Helly property. In order to prove that (L,d) is injective, according to the definition of hyperconvex metric spaces (see for instance [Lang 2013]), it is sufficient to prove that if $x,y\in L$ and $r,s\geqslant 0$ are such that $d(x,y)\leqslant r+s$, then the balls $B(x,r)$ and $B(y,s)$ intersect. In other words, it is enough to prove that (L,d) is weakly geodesic, ie for any $x,y\in L$ and any $0\leqslant r\leqslant d(x,y)$, there exists $z\in B(x,r)\cap B(y,d(x,y)-r)$.

For each $k\in\mathbb{N}\backslash\{0\}$, let us consider the action of $\frac{1}{k}\mathbb{Z}\subset\mathbb{R}$ on L, and the associated Helly graph distance,

$$\text{for all } x,y\in L,\quad d_k(x,y)=\inf\Bigl\{t\in\frac{1}{k}\mathbb{N}\;\Bigm|\;f_{-t}(x)\leqslant y\leqslant f_t(x)\Bigr\}.$$

Fix $x,y\in L$, and $0\leqslant r\leqslant d(x,y)$. For each $k\in\mathbb{N}\backslash\{0\}$, there exists $z_k\in L$ such that $d_k(x,z_k)\leqslant r+\frac{1}{k}$ and $d_k(z_k,y)\leqslant d_k(x,y)-r+\frac{1}{k}\leqslant d(x,y)-r+\frac{2}{k}$. So the intervals $I_k=I(f_{-r-1/k}(x),f_{r+1/k}(x))$ and $J_k=I(f_{-d(x,y)+r-2/k}(y),f_{d(x,y)-r+2/k}(y))$ in L intersect. If $k\leqslant k'$, we know that $I_{k'}\subset I_k$ and $J_{k'}\subset J_k$. We deduce that the family of intervals $\{I_k\}_{k\in\mathbb{N}\backslash\{0\}}\cup\{J_k\}_{k\in\mathbb{N}\backslash\{0\}}$ pairwise intersects. By the Helly property for intervals, the global intersection is nonempty: let us denote by z some element in the intersection.

We know that $\lim_{k\to+\infty}d_k(x,z)=d(x,z)\leqslant r$ and $\lim_{k\to+\infty}d_k(y,z)=d(y,z)\leqslant d(x,y)-r$, so we have proved that (L,d) is a weakly geodesic metric space. Since balls in (L,d) satisfy the Helly property, we conclude that (L,d) is injective. □

An immediate consequence concerns Garside groups (see [Dehornoy et al. 2015; McCammond 2006; Haettel and Huang 2024]). Recall that a group G is called *Garside* if there exists a subset $S \subset G$ and an element $\delta \in G$ such that the following hold:

- S spans the group G.
- For each element in the semigroup $\langle S\rangle^+$, there is a bound on the length of factorizations over S.
- The element δ belongs to the semigroup $\langle S\rangle^+$, and δ is *balanced*: the set of prefixes of δ coincides with the set of suffixes of δ.
- The poset of prefixes of δ in $\langle S\rangle^+$ is a lattice.

Note that G is endowed with two natural orders, the left order $\leqslant_L$ and the right order $\leqslant_R$:

- $g \leqslant_L h$ if and only if $g^{-1}h \in \langle S\rangle^+$.
- $g \leqslant_R h$ if and only if $hg^{-1} \in \langle S\rangle^+$.

Authors sometimes add the requirement that S is finite, in which case G may also be called a Garside group of finite type. Also note that, given a Garside group (G, S, δ), for any $g \in G$, there exists $n \in \mathbb{N}$ such that $\delta^{-n} \leqslant_L g \leqslant_L \delta^n$.

Fix a Garside group (G, S, δ). Let X denote the graph with vertex set G, with an edge between $g, h \in G$ if $g\delta^{-1} \leqslant_L h \leqslant_L g\delta$. The graph X is called the *thickening* of G. In relation to Definition 1.7, it corresponds to the cell complex with vertex set G, whose maximal cells are translates $(g[\Delta^{-1}, \Delta]_{\leqslant_L})_{g\in G}$ of the interval $[\Delta^{-1}, \Delta]$.

Corollary 2.2 *The thickening of any Garside group is a Helly graph.*

Proof The left order on G is a lattice order (see [Dehornoy et al. 2015]). Furthermore, consider the action of $\mathbb{Z}$ on G by right multiplication by δ. For any $g \in G$, there exists $n \in \mathbb{N}$ such that $\delta^{-n} \leqslant_L g \leqslant_L \delta^n$. So this action satisfies the assumptions of Theorem 2.1: we deduce that the graph X is Helly. □

Note that this applies, in particular, to Garside groups of infinite type, such as crystallographic Garside groups (see [McCammond and Sulway 2017]). However, we do not have yet an application of this simply transitive action of a Garside group on a locally infinite Helly graph.

In the case of Garside groups of finite type, we recover a particularly simple proof of the following result by Huang and Osajda [2021]. In particular, our proof does not rely on the deep local-to-global result for Helly graphs (Theorem 1.12, see [Chalopin et al. 2020b]).

Corollary 2.3 (Huang–Osajda) *Any Garside group of finite type is a Helly group.*

In particular, this leads to a particularly simple proof that braid groups are Helly, relying only on some Garside structure.

3 The affine version of a lattice

In this section, we will prove Theorem 3.10 stating that the orthoscheme complex of a bounded graded lattice, endowed with the orthoscheme ℓ^∞ metric, is injective. In order to do so, we will apply results from the previous section, and endow the geometric realization of a lattice with a partial order, which is a lattice.

Assume that L is a bounded, graded lattice of rank n. Let $\leqslant_L$ denote the order on L. Let H denote either a cyclic subgroup of $(\mathbb{R},+)$ or $H=\mathbb{R}$. We will define a new poset M_H, which will be called the *affine version* of L over H. If there is no ambiguity about H, we will simply write $M=M_H$. Let $C(L)$ denote the set of maximal chains $c_{0,1}=0<_L c_{1,2}<_L\cdots<_L c_{n-1,n}<_L c_{n,n+1}=1$ in L. We will use the convention that the element denoted by $c_{i,i+1}$ has rank i.

Let us consider the subspace

$$\sigma=\{u\in H^n \mid u_1\leqslant u_2\leqslant\ldots\leqslant u_n\}$$

of H^n.

For each maximal chain $c\in C(L)$, let σ_c denote a copy of σ.

Let us consider the space

$$M=\bigcup_{c\in C(L)}\sigma_c/{\sim},$$

where for each $c,c'\in C(L)$, if we let $I=\{1\leqslant i\leqslant n-1 \mid c_{i,i+1}\neq c'_{i,i+1}\}$, we identify σ_c and $\sigma_{c'}$ along the subspaces

$$\{u\in\sigma_c \mid \text{for all } i\in I,\ u_i=u_{i+1}\}\simeq\{u\in\sigma_{c'} \mid \text{for all } i\in I,\ u_i=u_{i+1}\}.$$

We can describe the set of elements of M as a quotient of the space $M_0=C(L)\times\sigma$.

Example One illustrating example is the following: consider the Boolean lattice L of rank n, ie the lattice of subsets of the finite set $\{1,\ldots,n\}$, with the inclusion order. Maximal chains in L correspond to permutations of $\{1,\ldots,n\}$. The space M_H may be identified with H^n, where, for each permutation w of $\{1,\ldots,n\}$, the subspace σ_w is

$$\sigma_w=\{x\in H^n \mid x_{w(1)}\leqslant x_{w(2)}\leqslant\ldots\leqslant x_{w(n)}\}.$$

If $c\in C(L)$ and $u\in\sigma$, let us denote by $[c,u]$ the equivalence class of $(c,u)\in M_0$ in M.

For each $c\in C(L)$, let us endow σ_c with the partial order from $H^n\subset\mathbb{R}^n$: $u\leqslant v$ if, for all $1\leqslant i\leqslant n$, $u_i\leqslant v_i$. Let us endow M with the induced partial order: we have $\alpha\leqslant\beta$ in M if there exists a sequence $\alpha_0=\alpha,\alpha_1,\ldots,\alpha_m=\beta$ in M such that, for each $0\leqslant i\leqslant m-1$, there exists $c\in C(L)$ and $x\leqslant y$ in σ_c such that $\alpha_i=[c,x]$ and $\alpha_{i+1}=[c,y]$.

Let $\leqslant_M$ denote the order on M. We will prove the following.

Theorem 3.1 *If H is a discrete subgroup of $\mathbb{R}$, the poset M_H is a lattice.*

Before proving Theorem 3.1, we will gather some preliminary results. Without loss of generality, assume that $H=\mathbb{Z}$. To simplify notation, we will let $M=M_H$.

First notice that if $(a,u)\sim(b,v)$, then $u=v$. Therefore the second projection $M_0\to\sigma$ defines a projection $\pi\colon M\to\sigma$. Fix $\beta=[b,v]$, $\gamma=[c,w]\in M$, and fix $\alpha\leqslant_M \beta,\gamma$ in M. We will prove that β,γ have a join, and α will play an auxiliary role.

We say that β is *elementarily superior* to α if there are representatives $\alpha=[a,u]$ and $\beta=[b,v]$ such that $a=b$ and there exist $1\leqslant i\leqslant j\leqslant n$ such that

- $u_i=u_{i+1}=\cdots=u_j$,
- $v_i=v_{i+1}=\cdots=v_j=u_j+1$, and
- for all $k\notin[i,j]$, $u_k=v_k$.

Lemma 3.2 *Fix $\alpha=[a,u]\in M$. Any element $\beta=[b,v]$ of M elementarily superior to α is uniquely determined by*:

- *integers $1\leqslant i\leqslant j\leqslant n$, such that $u_i=u_{i+1}=\cdots=u_j$, and $u_j<u_{j+1}$ if $j<n$,*
- *some element $b_{i-1,i}$ of rank $i-1$ in the interval $I(a_{i_0-1,i_0},a_{j,j+1})$, where $i_0\in[1,i]$ is minimal such that $u_{i_0}=u_i$.*

Let us set $\beta=\alpha[i,j,b_{i-1,i}]$.

Proof Let us define the element $\beta=[b,v]$ in M by

$$\begin{aligned}
&\text{for all } k\notin[i,j], && v_k=u_k, && \text{for all } k\leqslant i_0-1, && b_{k,k+1}=a_{k,k+1},\\
&\text{for all } k\in[i,j], && v_k=u_j+1, && \text{for all } k\geqslant j, && b_{k,k+1}=a_{k,k+1}.
\end{aligned}$$

Note that v is nondecreasing; hence $v\in\sigma$. Furthermore, since $v_{i_0}=v_{i_0+1}=\cdots=v_{i-1}$ and $v_i=v_{i+1}=\cdots=u_j+1\leqslant v_{j+1}$, it is enough to define $b_{k,k+1}$ for $k\leqslant i_0-1$, $k=i-1$ and $k\geqslant j$. Hence β is well-defined, and it is elementarily superior to α. It is clear that β is the only such element in M. □

Lemma 3.3 *Given $\alpha\leqslant_M\beta$ in M, there exist $m\geqslant 0$ and a sequence $\beta_0=\alpha,\beta_1,\dots,\beta_m=\beta$ for which, for each $0\leqslant i\leqslant m-1$, β_{i+1} is elementarily superior to β_i.*

Proof According to the definition of the order on M, it is sufficient to prove the statement for $\alpha=[a,u]$ and $\beta=[b,v]$ such that $a=b$. We have $u\leqslant v$ in $\mathbb{Z}^n\subset\mathbb{R}^n$. Then consider a sequence $u^0=u\leqslant u^1\leqslant\ldots\leqslant u^m=v$ such that, for each $0\leqslant k\leqslant m-1$, there exist $1\leqslant i_k\leqslant j_k\leqslant n$ such that

- $u^k_{i_k}=u^k_{i_k+1}=\cdots=u^k_{j_k}$,
- $u^{k+1}_{i_k}=u^{k+1}_{i_k+1}=\cdots=u^{k+1}_{j_k}=u^k_{j_k}+1$, and
- for all $\ell\notin[i_k,j_k]$, $u^k_\ell=u^{k+1}_\ell$.

For each $0\leqslant k\leqslant m-1$, the element $\beta_{k+1}=[a,u^{k+1}]$ is elementarily superior to $\beta_k=[a,u^k]$. □

Lemma 3.4 *We have $\alpha=[a,u]\leqslant\beta=[b,v]$ in M if and only if $u\leqslant v$ and for every $0\leqslant j\leqslant n-1$ such that $u_j<u_{j+1}$. If we denote by $i\in\{0,\dots,j\}$ the minimal element such that $v_{i+1}\geqslant u_{j+1}$, we have $b_{i,i+1}\leqslant_L a_{j,j+1}$.*

Proof Let us denote by $\prec$ this relation. We first show that $\prec$ is transitive and antisymmetric.

Assume that $\alpha=[a,u]\preccurlyeq\beta=[b,v]$ and $u=v$; we will prove that $\alpha=\beta$. Fix $0\leqslant j\leqslant n-1$, and assume that $a_{j,j+1}\neq b_{j,j+1}$. Since these two elements of L have the same rank j, we deduce that they are not comparable: $b_{j,j+1}\not\leqslant_L a_{j,j+1}$ and $a_{j,j+1}\not\leqslant_L b_{j,j+1}$. We will prove that $u_j=u_{j+1}$: by contradiction, if $u_j<u_{j+1}$, then the minimal $i\in\{0,\dots,j\}$ such that $v_{i+1}\geqslant u_{j+1}$ is equal to j, so $b_{j,j+1}\leqslant_L a_{j,j+1}$. So $u_j=u_{j+1}$. This proves that $\alpha=\beta$.

This implies in particular that $\prec$ is antisymmetric: indeed assume that $\alpha\preccurlyeq\beta$ and $\beta\preccurlyeq\alpha$. Then $u\leqslant v$ and $v\leqslant u$, so $u=v$; hence $\alpha=\beta$.

We now prove that $\prec$ is transitive: assume that $\alpha=[a,u]\preccurlyeq\beta=[b,v]\preccurlyeq\gamma=[c,w]$. Then $u\leqslant v\leqslant w$, so $u\leqslant w$. Let $0\leqslant k\leqslant n-1$ such that $u_k<u_{k+1}$, and let $i\in\{0,\dots,k\}$ be minimal such that $w_{i+1}\geqslant u_{k+1}$: we want to prove that $c_{i,i+1}\leqslant_L a_{k,k+1}$. Let $j\in\{0,\dots,k\}$ be minimal such that $v_{j+1}\geqslant u_{k+1}$: we know that $b_{j,j+1}\leqslant_L a_{k,k+1}$. Since j is minimal, we know that either $j=0$ or $v_j<u_{k+1}$.

If $j=0$, then since $i\leqslant j$ we know that $i=0$ and so $c_{i,i+1}=0=b_{j,j+1}\leqslant_L a_{k,k+1}$.

If $v_j<u_{k+1}$ then $v_j<v_{j+1}$, and let $i'\in\{0,\dots,j\}$ be minimal such that $w_{i'+1}\geqslant v_{j+1}$. We know that $c_{i',i'+1}\leqslant_L b_{j,j+1}$. But since $w_i<u_{k+1}\leqslant v_{j+1}$, we have $i\leqslant i'$, so $c_{i,i+1}\leqslant_L c_{i',i'+1}\leqslant_L b_{j,j+1}\leqslant_L a_{k,k+1}$.

Hence $\alpha\preccurlyeq\gamma$.

We will now prove that if $\beta=[b,v]$ is elementarily superior to $\alpha=[a,u]$, we have $\alpha\preccurlyeq\beta$. Let us set $\beta=\alpha[i,j,b_{i-1,i}]$. First note that $u\leqslant v$. If $0\leqslant k\leqslant n-1$ is such that $u_k<u_{k+1}$, then either $k\leqslant i-1$ or $k\geqslant j$. Let $\ell\in\{0,\dots,k\}$ denote the minimal element such that $v_{\ell+1}\geqslant u_{k+1}$; we will prove that $b_{\ell,\ell+1}\leqslant_L a_{k,k+1}$.

- If $k\leqslant i-1$, then $\ell=k$ and $b_{\ell,\ell+1}=a_{k,k+1}$.
- If $k\geqslant j$ and $v_{k+1}=v_j$, then $\ell=i-1$ and $b_{\ell,\ell+1}=b_{i-1,i}\leqslant_L a_{j,j+1}\leqslant_L a_{k,k+1}$.
- If $k\geqslant j$ and $v_{k+1}>v_j$, then $\ell\geqslant j$ and $b_{\ell,\ell+1}=a_{\ell,\ell+1}\leqslant_L a_{k,k+1}$.

According to Lemma 3.3, we deduce that if $\alpha\leqslant_M\beta$, then $\alpha\preccurlyeq\beta$.

We will now prove that if $\alpha=[a,u]\preccurlyeq\beta=[b,v]$ with $\alpha\neq\beta$, there exists γ elementarily superior to α such that $\gamma\preccurlyeq\beta$. If $u=v$, we have seen that $\alpha=\beta$, so let us assume that $u<v$. Let $1\leqslant i\leqslant n$ be minimal such that $u_i<v_i$. Let $j\in\{i,\dots,n\}$ be maximal such that $u_i=u_j$.

Let $i_0\in\{1,\dots,i\}$ be minimal such that $u_{i_0}=u_i$. Since for all $k\leqslant i-1$ we have $u_k=v_k$, there exist representatives $\alpha=[a,u]$ and $\beta=[b,v]$ such that, for any $1\leqslant k\leqslant i-1$, we have $a_{k-1,k}=b_{k-1,k}$.

We will also prove that we may assume that $a_{i-1,i}=b_{i-1,i}$.

Assume first that $u_{i-1} < u_i$. Let $p \in \{0, \dots, i-1\}$ be minimal such that $v_{p+1} \geqslant u_i$. Since $v_{i-1} = u_{i-1} < u_i$, we deduce that $p = i-1$. Since $\alpha \preccurlyeq \beta$, we have $b_{i-1,i} \leqslant_L a_{i-1,i}$. As both elements have the same rank $i-1$ in L, we have $a_{i-1,i} = b_{i-1,i}$.

Assume now that $u_{i-1} = u_i$. In the case $j < n$, we have $u_j < u_{j+1}$: let $p \in \{0, \dots, j\}$ be minimal such that $v_{p+1} \geqslant u_{j+1}$. Since $v_{i-1} = u_{i-1} < u_{j+1}$, we deduce that $p \geqslant i-1$. Since $\alpha \preccurlyeq \beta$, we have $b_{i-1,i} \leqslant_L b_{p,p+1} \leqslant_L a_{j,j+1}$. Since $u_{i-1} = u_i = \dots = u_j$, we may choose a representative of α such that $a_{i-1,i} = b_{i-1,i}$. In the case $j = n$, we have $u_{i-1} = u_i = \dots = u_n$; we may also choose a representative of α such that $a_{i-1,i} = b_{i-1,i}$.

We have proved that we may assume that $a_{i-1,i} = b_{i-1,i}$.

In the case $j = n$, since $u_i = \dots = u_n$, we may choose a representative of α such that $a = b$. In this case, the element $\gamma = \alpha[i, n, b_{i-1,i}]$ is elementarily superior to α, and $\gamma \preccurlyeq \beta$.

For the rest of the proof, assume that $j < n$.

We know that $b_{i-1,i} \geqslant_L b_{i_0-1,i_0} = a_{i_0-1,i_0}$. Since $u_j < u_{j+1}$ and $\alpha \preccurlyeq \beta$, if we denote by $i' \in \{0, \dots, j\}$ the minimal element such that $v_{i'+1} \geqslant u_{j+1}$, we have $b_{i',i'+1} \leqslant_L a_{j,j+1}$. Since $v_{i-1} = u_{i-1} < u_{j+1}$, we have $i' > i-2$, so $b_{i-1,i} \leqslant_L b_{i',i'+1} \leqslant_L a_{j,j+1}$.

So we know that $b_{i-1,i} \in I(a_{i_0-1,i_0}, a_{j,j+1})$: we can define $\gamma = \alpha[i, j, b_{i-1,i}] = [w, c]$. We will now prove that $\gamma \preccurlyeq \beta$. Fix $0 \leqslant k \leqslant n-1$ such that $w_k < w_{k+1}$, and denote by $\ell \in \{0, \dots, k\}$ the minimal element such that $v_{\ell+1} \geqslant w_{k+1}$; we will prove that $b_{\ell,\ell+1} \leqslant_L c_{k,k+1}$. Recall that either $k \leqslant i-1$ or $k \geqslant j$.

- If $k \leqslant i-2$, then $w_{k+1} = u_{k+1}$. Since $\alpha \preccurlyeq \beta$, we know that $b_{\ell,\ell+1} \leqslant_L a_{k,k+1} = c_{k,k+1}$.
- If $k = i-1$, then $w_{k+1} = w_i = u_i + 1 > u_i$. Then $v_{i-1} = u_{i-1} \leqslant u_i < w_{k+1}$, so $\ell = i-1 = k$. Hence $b_{\ell,\ell+1} = b_{i-1,i} = a_{i-1,i} = c_{i-1,i} = c_{k,k+1}$.
- If $k \geqslant j$, then $w_{k+1} = u_{k+1}$. Since $\alpha \preccurlyeq \beta$, we know that $b_{\ell,\ell+1} \leqslant_L a_{k,k+1}$. We also have $a_{k,k+1} = c_{k,k+1}$, so $b_{\ell,\ell+1} \leqslant_L c_{k,k+1}$.

So we conclude that $\gamma \preccurlyeq \beta$.

Now fix any $\alpha = [a, u] \preccurlyeq \beta = [b, v]$. By induction on $\|v - u\|_1$, we see that there is a bound on sequences of elementarily superior elements starting from α which all are $\preccurlyeq \beta$. Therefore we conclude that $\alpha \leqslant_M \beta$.

In conclusion, the two orders $\leqslant_M$ and $\prec$ coincide. □

Given $\alpha \leqslant_M \beta$ in M, let us denote by $D(\alpha, \beta)$ the minimal number $m \geqslant 0$ such that there exists a sequence $\beta_0 = \alpha, \beta_1, \dots, \beta_m = \beta$ for which, for each $0 \leqslant i \leqslant m-1$, β_{i+1} is elementarily superior to β_i.

We will prove, by induction on $D(\alpha, \beta) + D(\alpha, \gamma)$, that β and γ have a join.

Lemma 3.5 *Assume that $\alpha, \beta, \gamma \in M$ are such that $\alpha \leqslant_M \beta$, $\alpha \leqslant_M \gamma$ and $D(\alpha, \beta) = D(\alpha, \gamma) = 1$. Then β, γ have a join δ such that $D(\beta, \delta) \leqslant 1$ and $D(\gamma, \delta) \leqslant 1$.*

Proof Consider representatives $\alpha=[a,u]$, $\beta=[b,v]$ and $\gamma=[c,w]$ of α, β and γ respectively. According to Lemma 3.2, there exist $1\leqslant i\leqslant j\leqslant n$ and $b_{i,i+1}$ such that $\beta=\alpha[i,j,b_{i,i+1}]$, and there exist $1\leqslant i'\leqslant j'\leqslant n$ and $c_{i',i'+1}$ such that $\gamma=\alpha[i',j',c_{i',i'+1}]$.

First case Assume that the intervals $[i,j]$ and $[i',j']$ are disjoint, for instance $j<i'$. Let us define $\delta=\beta[i',j',c_{i',i'+1}]$: we will see that δ is well-defined and that $\delta=\gamma[i,j,b_{i,i+1}]$.

First note that $v_{j'}=u_{j'}<u_{j'+1}=v_{j'+1}$. Furthermore, let $i'_0\in[1,i']$ be minimal such that $u_{i'_0}=u_{i'}$. Since $u_j<u_{j+1}$, we know that $j+1\leqslant i'_0$. So we have $v_{i'_0}=u_{i'_0}=u_{i'}=v_{i'}$. So, if we denote by $i''_0\in[1,i']$ the minimal integer such that $v_{i''_0}=v_{i'}$, we have $i''_0\leqslant i'_0$.

By definition we have $c_{i',i'+1}\in I(a_{i'_0-1,i'_0},a_{j',j'+1})$. Since $j<j'$, we have $b_{j',j'+1}=a_{j',j'+1}$, so $c_{i',i'+1}\leqslant_L b_{j',j'+1}$. And as $b_{i''_0-1,i''_0}\leqslant_L b_{i'_0-1,i'_0}=a_{i'_0-1,i'_0}$, we deduce that $c_{i',i'+1}\geqslant_L a_{i'_0-1,i'_0}\geqslant_L b_{i''_0-1,i''_0}$. So we have $c_{i',i'+1}\in I(b_{i''_0-1,i''_0},b_{j',j'+1})$. Hence $\delta=\beta[i',j',c_{i',i'+1}]$ is well-defined, and it is elementarily superior to β.

Following the same argument, the element $\delta'=\gamma[i,j,b_{i,i+1}]$ is well-defined. We will prove that $\delta=\delta'$. Let $i_0\in[1,i]$ be minimal such that $u_{i_0}=u_i$. According to the proof of Lemma 3.2, we can see that $\delta=\delta'=[d,x]$ are explicitly equal to the following:

$$\begin{aligned}
\text{for all } k\notin[i,j]\cup[i',j'],\quad & x_k=u_k,\\
\text{for all } k\in[i,j],\quad & x_k=u_j+1,\\
\text{for all } k\in[i',j'],\quad & x_k=u_{j'}+1,\\
\text{for all } k\leqslant i_0-1,\quad & d_{k,k+1}=a_{k,k+1},\\
& d_{i,i+1}=b_{i,i+1},\\
\text{for all } k\in[j,i'_0-1],\quad & d_{k,k+1}=a_{k,k+1},\\
& d_{i',i'+1}=c_{i',i'+1},\\
\text{for all } k\geqslant j',\quad & d_{k,k+1}=a_{k,k+1}.
\end{aligned}$$

So the element $\delta=\delta'$ is elementarily superior to both β and γ.

We will now prove that δ is the minimal element of M superior to both β and γ. Fix $\theta=[e,y]\in M$ as any element superior to both β and γ; we will prove that $\delta\leqslant\theta$. Since $y\geqslant v$ and $y\geqslant w$, we deduce that $y\geqslant x$. Fix any $0\leqslant k\leqslant n-1$ such that $x_k<x_{k+1}$, and let $\ell\in\{0,\dots,k\}$ denote the minimal element such that $y_{\ell+1}\geqslant x_{k+1}$. We will prove that $e_{\ell,\ell+1}\leqslant_L d_{k,k+1}$.

- Assume that $k\leqslant j-1$. Then $x_{k+1}=v_{k+1}$, and since $\beta\leqslant\theta$, we deduce by Lemma 3.4 that $e_{\ell,\ell+1}\leqslant_L b_{k,k+1}=d_{k,k+1}$.
- Assume that $k\geqslant j$. Then $x_{k+1}=w_{k+1}$, and since $\gamma\leqslant\theta$, we deduce by Lemma 3.4 that $e_{\ell,\ell+1}\leqslant_L c_{k,k+1}=d_{k,k+1}$.

According to Lemma 3.4, we deduce that $\delta \leqslant_M \theta$: δ is the minimal element of M superior to both β and γ. Hence $\delta = \beta \vee_M \gamma$. Furthermore, we have noticed that δ is elementarily superior to both β and γ, so $D(\beta,\delta) = D(\gamma,\delta) = 1$.

Second case Assume now that the intervals $[i, j]$ and $[i', j']$ intersect. Without loss of generality, assume that $i \leqslant i'$. Since $u_j < u_{j+1}$, we deduce that $j = j'$. Let $1 \leqslant i_0 \leqslant n$ be minimal such that $u_{i_0} = u_i$. The elements $b_{i-1,i}$ and $c_{i'-1,i'}$ both belong to the interval $I(a_{i_0-1,i_0}, a_{j,j+1})$.

If $b_{i-1,i} = c_{i'-1,i'}$, then $\beta = \gamma$ and they have a trivial join $\delta = \beta = \gamma$. So we may assume that $b_{i-1,i} \neq c_{i'-1,i'}$.

If $b_{i-1,i} <_L c_{i'-1,i'}$, we have $\beta \leqslant \gamma$, so β and γ have a join $\delta = \gamma$ which satisfies $D(\beta,\delta) = 1$ and $D(\gamma,\delta) = 0$. Let us assume now that $b_{i-1,i} \not\leqslant_L c_{i'-1,i'}$.

Consider the meet $g = b_{i-1,i} \wedge_L c_{i'-1,i'} \in L$: its rank $r-1$ is such that $i_0 \leqslant r < i, i'$. Let us define $\delta = \alpha[r, j, g] \in M$. We see that $\delta = \beta[r, i-1, g] = \gamma[r, i'-1, g]$, so δ is elementarily superior to β and γ.

We will now prove that $\delta = [d, x]$ is the minimal element of M superior to both β and γ. Fix $\theta = [e, y] \in M$ as any element superior to both β and γ; we will prove that $\delta \leqslant \theta$.

We will first prove that $x \leqslant y$. Since $\beta, \gamma \leqslant_M \theta$, we deduce that $v, w \leqslant y$. In particular, for any $m < r$ or $m \geqslant i$, we have $y_m \geqslant b_m = x_m$. And for $r \leqslant m \leqslant i-1$, we have $y_m \geqslant a_m = x_m - 1$. Assume by contradiction that there exists $m \in \{r, \dots, i-1\}$ such that $y_m = x_m - 1$, and choose such m maximal. Since $x_r = x_{r+1} = \dots = x_j$, we have $y_{m+1} \geqslant x_{m+1} = x_i = v_i = x_{i'} = w_{i'}$. Since $\beta \leqslant_L \theta$ and $\gamma \leqslant_L \theta$, according to Lemma 3.4, we know that $e_{m,m+1} \leqslant_L b_{i-1,i}$ and $e_{m,m+1} \leqslant_L c_{i'-1,i'}$. In particular, we deduce that $e_{m,m+1} \leqslant_L b_{i-1,i} \wedge_L c_{i'-1,i'} = g$. Note that the rank of $e_{m,m+1}$ is m, whereas the rank of g is $r-1$. Since $m > r-1$, this is a contradiction. Hence $x \leqslant y$.

Fix any $0 \leqslant k \leqslant n-1$ such that $x_k < x_{k+1}$, and let $\ell \in \{0, \dots, k\}$ denote the minimal element such that $y_{\ell+1} \geqslant x_{k+1}$. We will prove that $e_{\ell,\ell+1} \leqslant_L d_{k,k+1}$.

- Assume that $k \leqslant r-2$. Then $x_{k+1} = v_{k+1}$, and since $\beta \leqslant \theta$ we deduce by Lemma 3.4 that $e_{\ell,\ell+1} \leqslant_L b_{k,k+1} = d_{k,k+1}$.
- Assume that $r-1 \leqslant k \leqslant j-1$. Then $x_{k+1} = v_i = w_{i'}$. Since $\beta \leqslant \theta$, we deduce by Lemma 3.4 that $e_{\ell,\ell+1} \leqslant_L b_{i-1,i}$. And since $\gamma \leqslant \theta$, we also deduce that $e_{\ell,\ell+1} \leqslant_L c_{i'-1,i'}$. Hence $e_{\ell,\ell+1} \leqslant_L b_{i-1,i} \wedge_L c_{i'-1,i'} = g = d_{r-1,r} \leqslant_L d_{k,k+1}$.
- Assume that $k \geqslant j$. Then $x_{k+1} = v_{k+1}$, and since $\beta \leqslant \theta$ we deduce by Lemma 3.4 that $e_{\ell,\ell+1} \leqslant_L b_{k,k+1} = d_{k,k+1}$.

According to Lemma 3.4, we deduce that $\delta \leqslant_M \theta$: δ is the minimal element of M superior to both β and γ. Hence $\delta = \beta \vee_M \gamma$. Furthermore, we have noticed that δ is elementarily superior to both β and γ, so $D(\beta,\delta) = D(\gamma,\delta) = 1$. □

Lemma 3.6 *Assume that $\alpha,\beta,\gamma\in M$ are such that $\alpha\leqslant_M\beta$, $\alpha\leqslant_M\gamma$, $D(\alpha,\beta)=m$ and $D(\alpha,\gamma)=m'$ for some $m,m'\in\mathbb{N}$. Then β,γ have a meet δ such that $D(\beta,\delta)\leqslant m'$ and $D(\gamma,\delta)\leqslant m$.*

Proof We proceed by induction on $m+m'$: when $m+m'\leqslant 2$, the statement holds by Lemma 3.5. Now fix $k\geqslant 3$, and assume that the statement holds when $m+m'<k$. Fix m,m' such that $m+m'=k$, and without loss of generality assume that $m\geqslant 2$. Choose $\beta_0=\alpha,\beta_1,\dots,\beta_m=\beta$ an elementary sequence from α to β, with $m=D(\alpha,\beta)$. We have $D(\alpha,\beta_1)+D(\alpha,\gamma)=1+m'<k$, so by induction there exists $\delta'=\beta_1\vee_M\gamma$ with $D(\beta_1,\delta')\leqslant m'$ and $D(\gamma,\delta')\leqslant 1$. Since $D(\beta_1,\beta)+D(\beta_1,\delta')\leqslant m-1+m'<k$, by induction there exists $\delta=\beta\vee_M\delta'$ such that $D(\beta,\delta)\leqslant m'$ and $D(\delta',\delta)\leqslant m-1$. So we deduce that $D(\gamma,\delta)\leqslant m$.

We will now prove that δ is the meet of β and γ. We have $\delta\geqslant_M\beta$ and $\delta\geqslant_M\delta'\geqslant_M\gamma$. Furthermore, consider any $\theta\in M$ such that $\theta\geqslant_M\beta$ and $\theta\geqslant_M\gamma$. As $\beta\geqslant_M\beta_1$, we deduce that $\theta\geqslant_M\beta_1\vee_M\gamma=\delta'$. And we deduce that $\theta\geqslant_M\beta\vee_M\delta'=\delta$. So we have proved that $\delta=\beta\vee_M\gamma$. □

Proof of Theorem 3.1 Fix any $\beta=[b,v]$, $\gamma=[c,w]\in M$. Let $k\geqslant 0$ such that $v_n-k<w_1$. Let $u=(v_1-k,v_2-k,\dots,v_n-k)$. Then $\alpha=[b,u]\in M$ is inferior to β, and we will see that it is also inferior to γ. Indeed let $\gamma'=[c,w']$, where $w'=(w_1,w_1,\dots,w_1)$. Since $w'\leqslant w$, we have $\gamma'\leqslant_M\gamma$. On the other hand, since $\gamma'=[b,w']$ and $u\leqslant w'$, we have $\alpha\leqslant_M\gamma'$, so $\alpha\leqslant_M\gamma$.

We can now apply Lemma 3.6 to deduce that β and γ have a meet in M. By symmetry of the construction, β and γ also have a join in M. So M is a lattice. □

If $H=\mathbb{R}$, the affine version $M_{\mathbb{R}}$ of L over $\mathbb{R}$ is a gluing of subspaces $\sigma\subset\mathbb{R}^n$. We may therefore endow $M_{\mathbb{R}}$ with the piecewise length metric $d_{\mathbb{R}}$ induced by the standard ℓ^∞ metric on each $\sigma\subset\mathbb{R}^n$.

Let us define an action of $\mathbb{R}$ on $M_{\mathbb{R}}$ as

$$\mathbb{R}\times M_{\mathbb{R}}\to M_{\mathbb{R}},\quad (t,[a,u])\mapsto t\cdot[a,u]=[a,(u_1+t,u_2+t,\dots,u_n+t)].$$

This action is well-defined, preserves the order $\leqslant_M$, is increasing and continuous. Moreover, we have the following property.

Lemma 3.7 *For any $\alpha,\beta\in M_{\mathbb{R}}$, there exists $t>0$ such that $(-t)\cdot\alpha\leqslant_M\beta\leqslant_M t\cdot\alpha$.*

Proof Consider representatives $\alpha=[a,u]$ and $\beta=[b,v]$ of α and β respectively. Let $t>0$ such that $v_n\leqslant u_1+t$ and $u_n\leqslant v_1+t$. Then, if we let $\gamma=[a,(v_n,\dots,v_n)]=[b,(v_n,\dots,v_n)]$, we have $\beta\leqslant_M\gamma\leqslant_M t\cdot\alpha$; hence $\beta\leqslant_M t\cdot\alpha$. Similarly, we have $(-t)\cdot\alpha\leqslant_M\beta$. □

Theorem 3.8 *If $H=\mathbb{R}$, the poset $M_{\mathbb{R}}$ is a lattice, and the metric space $(M_{\mathbb{R}},d_{\mathbb{R}})$ is injective.*

Proof Note that, for each $\theta>0$, the space $M_{\theta\mathbb{Z}}$ may be realized naturally as a closed subspace of $M_{\mathbb{R}}$. Furthermore, the sequence of closed subsets $M_{\theta\mathbb{Z}}$ of $M_{\mathbb{R}}$ converges to $M_{\mathbb{R}}$ as $\theta\to 0$. We will use this convergence to prove that $M_{\mathbb{R}}$ is a lattice. We will then prove that the assumptions of Theorem 2.1 are satisfied.

We will now prove that any pair of elements in $M_{\mathbb{R}}$ have a join. Fix $\alpha, \beta \in M_{\mathbb{R}}$; we will define a common upper bound γ for α and β.

Consider maximal chains a, b in L such that $\alpha \in \sigma_a$ and $\beta \in \sigma_b$. For any $\theta > 0$, note that, for any $x \in \mathbb{R}$, there exists $x_\theta \in \theta\mathbb{Z}$ such that $x_\theta - \theta \leqslant x \leqslant x_\theta + \theta$. So we may consider $\alpha_\theta \in M_{\theta\mathbb{Z}} \cap \sigma_a$ and $\beta_\theta \in M_{\theta\mathbb{Z}} \cap \sigma_b$ such that $(-\theta)\cdot\alpha_\theta \leqslant_M \alpha \leqslant_M \theta\cdot\alpha_\theta$, and similarly $(-\theta)\cdot\beta_\theta \leqslant_M \beta \leqslant_M \theta\cdot\beta_\theta$.

According to Theorem 3.1, the poset $M_{\theta\mathbb{Z}}$ is a lattice: consider $\gamma_\theta = \alpha_\theta \vee_{M_{\theta\mathbb{Z}}} \beta_\theta$. Let $C_{a,b} \subset L$ denote the smallest subset of L containing a, b, and which is stable under meets. Since L is bounded and graded, $C_{a,b}$ is finite. According to the proof of Theorem 3.1, we see that, for every $\theta > 0$, there exists a maximal chain $c_\theta \subset C_{a,b}$ such that $\gamma_\theta \in \sigma_{c_\theta}$. Since $C_{a,b}$ is finite, σ is locally compact and $(\gamma_\theta)_{\theta>0}$ is bounded, there exists a sequence $\theta_k \xrightarrow{k\to+\infty} 0$ such that the sequence $(\gamma_{\theta_k})_{k\in\mathbb{N}}$ converges to some $\gamma \in M_{\mathbb{R}}$. Note that, for any $\theta > 0$, we have $\gamma_\theta \geqslant_M \alpha_\theta \geqslant_M (-\theta)\cdot\alpha$. Since the sequence $(\gamma_{\theta_k})_{k\in\mathbb{N}}$ converges to γ, and the sequence $((-\theta_k)\cdot\alpha)_{k\in\mathbb{N}}$ converges to α by continuity of the action, we deduce that $\gamma \geqslant_M \alpha$. Similarly $\gamma \geqslant_M \beta$.

So γ is a common upper bound for α and β. We will now prove that γ is a minimal upper bound, which will prove that γ is the join of α and β. Let us consider an upper bound $\delta \in M_{\mathbb{R}}$ of α and β; we will prove that $\gamma \leqslant_M \delta$. For any $\theta > 0$, fix $\delta_\theta \in M_{\theta\mathbb{Z}}$ such that $(-\theta)\cdot\delta_\theta \leqslant_M \delta \leqslant_M \theta\cdot\delta_\theta$. In particular $(2\theta)\cdot\delta_\theta \geqslant_M \theta\cdot\delta \geqslant_M \theta\cdot\alpha \geqslant_M \alpha_\theta$, and similarly $(2\theta)\cdot\delta_\theta \geqslant_M \beta_\theta$. We deduce that $(2\theta)\cdot\delta_\theta \geqslant_M \alpha_\theta \vee_{M_{\theta\mathbb{Z}}} \beta_\theta = \gamma_\theta$. Considering the limit along $(\theta_k)_{k\in\mathbb{N}}$ as $k \to +\infty$, we deduce that $\delta \geqslant_M \gamma$.

So we have proved that α and β have a join γ in $M_{\mathbb{R}}$. By symmetry of the construction, they also have a meet, so $M_{\mathbb{R}}$ is a lattice.

We now turn to the assumptions of Theorem 2.1: we will first prove that every upper bounded subset of $M_{\mathbb{R}}$ has a join. Since $M_{\mathbb{R}}$ is a lattice, it is enough to prove that every bounded, increasing sequence is convergent. Fix an increasing sequence $(\alpha_k)_{k\geqslant 0}$ in $M_{\mathbb{R}}$, bounded above by some $\alpha \in M_{\mathbb{R}}$.

We will prove an intermediate result concerning ℓ^1 metrics. Let us endow $M_{\mathbb{R}}$ with the length metric d_1 associated to the standard ℓ^1 metric on each sector $\sigma_c \subset \mathbb{R}^n$. Let us also denote by d_1 the standard metric on $\mathbb{R}^n$. We claim that if $\beta = [b, v] \leqslant_M \gamma = [c, w]$, then $d_1(\beta, \gamma) = d_1(v, w)$. First notice that the second projection $M_{\mathbb{R}} \to \mathbb{R}$ is 1–Lipschitz with respect to the metrics d_1; hence we have $d_1(\beta, \gamma) \geqslant d_1(v, w)$. By the definition of the order relation $\leqslant_M$, there exists a sequence $\beta_0 = \beta \leqslant_M \beta_1 \leqslant_M \cdots \leqslant_M \beta_p = \gamma$ such that, for each $0 \leqslant i \leqslant p-1$, the points β_i, β_{i+1} lie in a common sector σ_{c_i}. Let us write representatives $\beta_i = [b_i, v_i]$, with $b_i \in L$ and $v_i \in \sigma$, for $0 \leqslant i \leqslant p$. We deduce that, for each $0 \leqslant i \leqslant p-1$ we have $d_1(\beta_i, \beta_{i+1}) = d_1(v_i, v_{i+1})$; hence

$$d_1(\beta, \gamma) \leqslant \sum_{i=0}^{p-1} d_1(\beta_i, \beta_{i+1}) \leqslant \sum_{i=0}^{p-1} d_1(v_i, v_{i+1}) = d_1(v, w).$$

So we have proved that if $\beta = [b, v] \leqslant_M \gamma = [c, w]$, then $d_1(\beta, \gamma) = d_1(v, w)$.

We now return to the increasing sequence $(\alpha_k)_{k\geqslant 0}$ in $M_{\mathbb{R}}$, bounded above by some $\alpha \in M_{\mathbb{R}}$. For each $k \in \mathbb{N}$, let us consider a representative $\alpha_k = [a_k, u_k]$ of α_k, and a representative $\alpha = [a, u]$ of α. Since the sequence $(\alpha_k)_{k\geqslant 0}$ is increasing in $M_{\mathbb{R}}$, we deduce that the sequence $(u_k)_{k\geqslant 0}$ is increasing in $\mathbb{R}^n$, and bounded above by u. Since increasing sequences in $\mathbb{R}^n$ are geodesics for the metric d_1, we deduce that, for each $0 \leqslant j \leqslant k$, we have $d_1(u_j, u_k) + d_1(u_k, u) = d_1(u_j, u)$.

Then, for each $0 \leqslant j \leqslant k$, according to the claim about the metric d_1, we have $d_1(\alpha_j, \alpha_k) + d_1(\alpha_k, \alpha) = d_1(\alpha_j, \alpha)$. In particular, the sequence $(\alpha_k)_{k\geqslant 0}$ is a Cauchy sequence in $(M_{\mathbb{R}}, d_1)$. Since $M_{\mathbb{R}}$ has finitely many shapes, the proof of [Bridson and Haefliger 1999, Theorem 7.13] applies to show that the metric space $(M_{\mathbb{R}}, d_1)$ is complete; hence the sequence $(\alpha_k)_{k\geqslant 0}$ converges in $M_{\mathbb{R}}$. Equivalently, we can apply [Bridson and Haefliger 1999, Theorem 7.13] to the metric space $M_{\mathbb{R}}$ endowed with the length metric d_2 associated to the standard ℓ^2 metric on each sector $\sigma_c \subset \mathbb{R}^n$. Since d_1 and d_2 are bi-Lipschitz, this also implies that the metric space $(M_{\mathbb{R}}, d_1)$ is complete.

So every upper bounded subset of $M_{\mathbb{R}}$ has a join.

We have proved that $M_{\mathbb{R}}$ is a lattice such that each upper bounded subset has a join. There is an increasing action of $\mathbb{R}$ on $M_{\mathbb{R}}$ satisfying the assumptions of Theorem 2.1 according to Lemma 3.7. As a consequence, we deduce that the metric space $(M_{\mathbb{R}}, d)$ is injective, with respect to the metric,

$$\text{for all } x, y \in M_{\mathbb{R}}, \quad d(x, y) = \inf\{t \geqslant 0 \mid (-t)\cdot x \leqslant y \leqslant t \cdot x\}.$$

Note that the metric d is geodesic, and it restricts on each $\sigma_c \subset M_{\mathbb{R}}$, for $c \in C(L)$, to the natural ℓ^∞ metric on $\sigma_c \subset \mathbb{R}^n$. Therefore d coincides with the length metric $d_{\mathbb{R}}$.

So we conclude that $(M_{\mathbb{R}}, d_{\mathbb{R}})$ is injective. □

We saw in the Introduction that the existence of a bicombing may be extremely useful, notably in the case of Deligne complexes of Artin groups. Let us recall that a *geodesic bicombing* on a metric space X is a map $\sigma\colon X \times X \times [0,1] \to X$ such that, for all $x, y \in X$, the map $t \in [0,1] \mapsto \sigma(x, y, t)$ is a constant-speed geodesic from x to y.

The bicombing σ is called

- *reversible* if for all $x, y \in X$, for all $t \in [0,1]$, $\sigma(x, y, t) = \sigma(y, x, 1-t)$,
- *consistent* if for all $x, y \in X$, for all $r, s, t \in [0,1]$, $\sigma(\sigma(x, y, r), \sigma(x, y, s), t) = \sigma(x, y, (1-t)r + ts)$,
- *conical* if for all $x, x', y, y' \in X$, for all $t \in [0,1]$, $d(\sigma(x, y, t), \sigma(x', y', t)) \leqslant (1-t)d(x, x') + t d(y, y')$,
- *convex* if for all $x, x', y, y' \in X$, the map $t \in [0,1] \mapsto d(\sigma(x, y, t), \sigma(x', y', t))$ is convex.

Note that any consistent, conical bicombing is convex.

Theorem 3.9 *The metric space $(M_{\mathbb{R}}, d_{\mathbb{R}})$ has a unique convex, consistent, reversible geodesic bicombing.*

Proof Let us write $X = M_{\mathbb{R}}$ for simplicity. Given any $x, y \in X$, let us define

$$D(x, y) = \inf\{t \in \mathbb{R} \mid x \leqslant t \cdot y\} \in \mathbb{R}.$$

Since the action of $\mathbb{R}$ on $M_{\mathbb{R}}$ is continuous, this infimum is attained; hence $x \leqslant D(x, y) \cdot y$. Note that this quantity is not symmetric with respect to x and y, and we have,

$$\text{for all } x, y \in X, \quad d(x, y) = \max(|D(x, y)|, |D(y, x)|).$$

We will start by defining a conical bicombing σ on X with a nice property which we call lower consistency. We then show that this is sufficient to bypass the use of properness of X in [Basso 2024, Theorem 1.4] applied to σ.

Fix $x, y \in X$, $t \in [0, 1]$, and let $D = d(x, y)$. For any $a \in X$, since $-D \cdot x \leqslant y \leqslant D \cdot x$, we have $|D(x, a) - D(y, a)| \leqslant D$, and so

$$\begin{aligned} -tD \cdot x &\leqslant (-tD(x, a) + tD(y, a)) \cdot x \\ &\leqslant (-tD(x, a) + tD(y, a)) \cdot (D(x, a) \cdot a) \\ &\leqslant ((1-t)D(x, a) + tD(y, a)) \cdot a. \end{aligned}$$

Since every nonempty subset of X with a lower bound has a meet, we may thus define

$$\sigma(x, y, t) = \bigwedge_{a \in X} ((1-t)D(x, a) + tD(y, a)) \cdot a.$$

We will first prove that it defines a geodesic bicombing, ie that $d(x, \sigma(x, y, t)) = tD$. We have proved that, for every $a \in X$, we have $-tD \cdot x \leqslant ((1-t)D(x, a) + tD(y, a)) \cdot a$; hence $-tD \cdot x \leqslant \sigma(x, y, t)$. Conversely, when $x = a$, we know that

$$\sigma(x, y, t) \leqslant ((1-t)D(x, x) + tD(y, x)) \cdot x \leqslant tD(y, x) \cdot x \leqslant tD \cdot x.$$

We conclude that $d(x, \sigma(x, y, t)) \leqslant tD$. By symmetry, we also have $d(\sigma(x, y, t), y) \leqslant (1-t)D$. Since $d(x, y) = D$, we conclude that $d(x, \sigma(x, y, t)) = tD$. So σ is a geodesic bicombing. It is clear that σ is reversible.

We will now prove that σ is conical. Fix $x, y, z \in X$, and $t \in [0, 1]$. For any $a \in X$, we have $|D(y, a) - D(z, a)| \leqslant d(y, z)$; hence $D(y, a) \leqslant D(z, a) + d(y, z)$. We deduce that

$$\begin{aligned} \sigma(x, y, t) &= \bigwedge_{a \in X} ((1-t)D(x, a) + tD(y, a)) \cdot a \\ &\leqslant \bigwedge_{a \in X} ((1-t)D(x, a) + tD(z, a) + td(y, z)) \cdot a \leqslant td(y, z) \cdot \sigma(x, z, t). \end{aligned}$$

By symmetry, we also have $\sigma(x, z, t) \leqslant td(y, z) \cdot \sigma(x, y, t)$; hence $d(\sigma(x, y, t), \sigma(x, z, t)) \leqslant td(y, z)$. So the bicombing σ is conical.

We will now prove that σ is what we will call *lower consistent*, which is one part of the inequality of the consistency equality. For each $x, y \in X$ and $s, t \in [0, 1]$, we will prove that

$$\sigma(x, \sigma(x, y, t), s) \leqslant \sigma(x, y, st).$$

Let us set $z=\sigma(x,y,t)$ and $w=\sigma(x,\sigma(x,y,t),s)$: we want to prove that $w\leqslant\sigma(x,y,st)$. For each $a\in X$, we have

$$\begin{aligned}D(w,a)&\leqslant(1-s)D(x,a)+sD(z,a)\\&\leqslant(1-s)D(x,a)+s((1-t)D(x,a)+tD(y,a))\\&\leqslant(1-st)D(x,a)+stD(y,a).\end{aligned}$$

Hence we deduce that $w\leqslant D(w,a)\cdot a\leqslant(1-st)D(x,a)\cdot a+stD(y,a)\cdot a$. Since this holds for any $a\in X$, we conclude that $w\leqslant\sigma(x,y,st)$. Hence σ is lower consistent.

According to [Basso 2024, Lemma 5.2], given any $x,y\in X$ and $n\geqslant 1$, there exist unique elements $\sigma_{xy}(n,i)$, for $0\leqslant i\leqslant n$, such that $\sigma_{xy}(n,0)=x$, $\sigma_{xy}(n,n)=y$, and,

$$\text{for all } 1\leqslant i\leqslant n-1,\quad \sigma_{xy}(n,i)=\sigma\big(\sigma_{xy}(n,i-1),\sigma_{xy}(n,i+1),\tfrac{1}{2}\big).$$

Note that even though [Basso 2024, Lemma 5.2] is stated for a proper metric space, the uniqueness part only requires that σ is a conical bicombing. And also the remark after the proof tells us that the only property needed for the existence is that the space is complete. We will actually give a proof below for the existence part.

Fix $n\geqslant 1$, we will prove that the elements $\sigma_{xy}(n,i)$ exist. For each $0\leqslant i\leqslant n$, let us define $x_i^0=\sigma\big(x,y,\frac{i}{n}\big)$. For each $k\in\mathbb{N}$, let us define $x_0^k=x$ and $x_n^k=y$. For each $k\in\mathbb{N}$ and $1\leqslant i\leqslant n-1$, let us define inductively $x_i^{k+1}=\sigma\big(x_{i-1}^k,x_{i+1}^k,\frac{1}{2}\big)$. Since σ is lower consistent, we see by induction that, for each $0\leqslant i\leqslant n$, the sequence $(x_i^k)_{k\in\mathbb{N}}$ is nonincreasing in X. Moreover, since each x_i^k lies on a geodesic from x to y, we have $x_i^k\geqslant(-d(x,y))\cdot x$ for every $k\in\mathbb{N}$ and $0\leqslant i\leqslant n$.

Hence for each $0\leqslant i\leqslant n$, since the sequence $(x_i^k)_{k\in\mathbb{N}}$ has a lower bound, we may define its meet $\sigma_{xy}(n,i)=\bigwedge_{k\geqslant 0}x_i^k$. In fact, the sequence $(x_i^k)_{k\geqslant 0}$ actually converges to $\sigma_{xy}(n,i)$. By the continuity of σ, we deduce that the elements $(\sigma_{xy}(n,i))_{0\leqslant i\leqslant n}$ satisfy the required property.

Note that [Basso 2024, Theorem 1.4] is stated for a proper metric space, but the properness assumption is used in precisely two arguments: first in [Basso 2024, Lemma 5.2] to prove the existence of the elements $\sigma_{xy}(n,i)$, which we obtained using specific properties of X.

Properness of X is used again, although not explicitly stated, in the proof of [Basso 2024, Theorem 1.4] to ensure the pointwise convergence of a sequence of bicombings with respect to some ultrafilter. Instead of using ultrafilters to ensure convergence, we will rather use the lower consistency of the bicombing.

For each $n\geqslant 1$, [Basso 2024, Lemma 5.2] states that the function $\sigma^{(n)}\colon X\times X\times[0,1]\to X$, defined by

$$\sigma^{(n)}\left(x,y,(1-\lambda)\frac{i}{n}+\lambda\frac{i+1}{n}\right)=\sigma(\sigma_{xy}(n,i),\sigma_{xy}(n,i+1),\lambda)$$

for all $x,y\in X$, $\lambda\in[0,1]$ and $0\leqslant i\leqslant n-1$, is a conical bicombing.

First note that, since σ is lower consistent, and by the uniqueness of the points $\sigma_{xy}(n,i)$, we have for all $x, y \in X$, $n \geqslant 1$, $0 \leqslant i \leqslant n-1$ and $p \geqslant 1$ that $\sigma_{xy}(np, ip) \leqslant \sigma_{xy}(n,i)$. For each $x, y \in X$ and $t \in [0,1]$, let us define

$$\gamma(x,y,t) = \bigwedge_{n \geqslant 1} \sigma_{xy}(n, \lceil tn \rceil).$$

According to the previous property, we deduce that

$$\gamma(x,y,t) = \lim_{n \to +\infty} \sigma_{xy}(n!, \lceil tn! \rceil).$$

Since each $\sigma^{(n)}$ is a conical bicombing, one also deduces that γ is conical.

We will prove that γ is lower consistent: Let $x, y \in X$ and $s, t \in [0,1]$. We have

$$\begin{aligned}
\gamma(x, \gamma(x,y,t), s) &\leqslant \lim_{n \to +\infty} \gamma(x, \sigma_{xy}(n!, \lceil tn! \rceil), s) \\
&\leqslant \lim_{n,m \to +\infty} \sigma_{xy}(n!\,m!, \lceil s \lceil tn! \rceil m! \rceil) \\
&\leqslant \lim_{n,m \to +\infty} \sigma_{xy}(n!\,m!, \lceil stn!\,m! \rceil) \leqslant \gamma(x,y,st)
\end{aligned}$$

as

$$d(\sigma_{xy}(n!\,m!, \lceil s \lceil tn! \rceil m! \rceil), \sigma_{xy}(n!\,m!, \lceil stn!\,m! \rceil)) \leqslant \frac{\lceil sm! \rceil}{n!\,m!} d(x,y) \to 0$$

as $n \to +\infty$. So we deduce that γ is a reversible, conical, lower consistent geodesic bicombing such that $\gamma \leqslant \sigma$.

As a consequence, if we start with a reversible, conical, lower consistent geodesic bicombing σ' which is minimal, we have $\gamma' = \sigma'$; hence σ' satisfies the following consistency property: for all $x, y \in X$, for all $s, t \in [0,1]$, $\sigma'(x, \sigma'(x,y,t), s) = \sigma'(x,y,st)$. Since σ' is reversible, we deduce that σ' is actually consistent. Since σ' is conical, it is also convex.

According to [Descombes and Lang 2015, Theorem 1.2], since X has finite combinatorial dimension, we conclude that σ' is the unique convex consistent reversible geodesic bicombing of X. □

Theorem 3.10 *Let L denote a bounded, graded lattice. The orthoscheme realization $|L|$ of L, endowed with the piecewise ℓ^∞ metric, is injective. Moreover, $|L|$ has a unique convex reversible consistent geodesic bicombing.*

Proof Consider the affine version $M = M_{\mathbb{R}}$ of L over $\mathbb{R}$. For some maximal chain $c \in C(L)$, consider the elements $0_M = [(0,\dots,0), c] \in M$, $\mu_M = \left[\left(\frac{1}{2},\dots,\frac{1}{2}\right), c\right] \in M$ and $1_M = [(1,\dots,1), c] \in M$: note that 0_M, μ_M and 1_M do not depend on c.

Note that the interval $I(0_M, 1_M)$ coincides with the ball $B\left(\mu_M, \frac{1}{2}\right)$ in M for the metric $d_{\mathbb{R}}$. According to Theorem 3.8, $(M, d_{\mathbb{R}})$ is injective. So the ball $B\left(\mu_M, \frac{1}{2}\right) = I(0_M, 1_M)$ is injective.

We remark that the interval $I(0_M, 1_M)$ of M, endowed with the metric $d_{\mathbb{R}}$, is isometric to the orthoscheme realization of L, endowed with the piecewise ℓ^∞ metric. Indeed, for each maximal chain $c \in C(L)$, notice that $\sigma_C \cap I(0_M, 1_M)$ identifies with the standard orthoscheme $\{x \in \mathbb{R}^n \mid 0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1\}$, where n denotes the rank of L. We conclude that the orthoscheme realization of L is injective.

According to Theorem 3.9, we know that $M_{\mathbb{R}}$ has a unique convex reversible consistent geodesic bicombing σ. Since the ball $|L| = B\big(\mu_M, \frac{1}{2}\big)$ is stable under σ, we deduce that $|L|$ has a convex reversible consistent geodesic bicombing. Since $|L|$ has combinatorial dimension at most n, according to [Descombes and Lang 2015, Theorem 1.2], we deduce that σ is the only convex bicombing on $|L|$. □

Note that knowing when the orthoscheme complex of a lattice, endowed with the piecewise Euclidean metric, is CAT(0) is a very subtle question (see the Introduction). We can prove the following.

Theorem 3.11 *Let L denote a bounded, graded poset, let $|L|$ denote the geometric realization of L, and let d_p denote the ℓ^p orthoscheme metric on $|L|$ for $p \in \{2, \infty\}$. If $(|L|, d_2)$ is CAT(0), then $(|L|, d_\infty)$ is injective. The converse is false.*

Proof Assume that L is a bounded, graded poset such that $(|L|, d_\infty)$ is not injective. According to Theorem 3.10, we deduce that L is not a lattice. According to Proposition 1.5, there exists a bowtie in L: consider $x_1, x_2, x_3, x_4 \in L$ such that x_1 and x_3 are maximal elements inferior to both x_2 and x_4, and such that x_2 and x_4 are minimal elements superior to both x_1 and x_3. In the link $L_{[0,1]}$ of the diagonal edge $[0, 1]$ in $(|L|, d_2)$, consider the piecewise geodesic loop ℓ going through x_1, x_2, x_3, x_4, x_1.

According to [Brady and McCammond 2010, Proposition 4.8], each geodesic segment $[x_i, x_{i+1}]$ has length smaller than $\frac{\pi}{2}$; hence ℓ has length smaller than 2π. On the other hand, according to Lemma 7.2 of that work the loop ℓ is locally geodesic in $L_{[0,1]}$. Hence $L_{[0,1]}$ is not CAT(1), and $(|L|, d_2)$ is not CAT(0).

We will now prove that the converse does not hold. Consider the Coxeter symmetric group $W = \mathfrak{S}_4$, with standard generators $S = \{s_1, s_2, s_3\}$ and standard Garside longest element $\Delta = s_1 s_2 s_3 s_1 s_2 s_1$. Let L denote the poset W, with order relation "being a left prefix for a shortest representative in S". Then L is a lattice, and so $(|L|, d_\infty)$ is injective according to Theorem 3.10.

On the other hand, consider the piecewise geodesic loop ℓ in the link $L_{[0,1]}$ of the diagonal edge $[0, 1]$ in $(|L|, d_2)$ going through the vertices s_1, $s_1 s_2 s_1$, s_2, $s_2 s_3 s_2$, s_3, $s_3 s_1$, s_1. According to [Brady and McCammond 2010, Proposition 4.8], its length is

$$2 \arccos\Big(\sqrt{\tfrac{1}{2}\tfrac{4}{5}}\Big) + 4 \arccos\Big(\sqrt{\tfrac{1}{3}\tfrac{3}{5}}\Big) \simeq 0.987(2\pi) < 2\pi.$$

Since $L_{[0,1]}$ has the homotopy type of a circle, ℓ is not nullhomotopic in $L_{[0,1]}$. So $L_{[0,1]}$ is not CAT(1), and $(|L|, d_2)$ is not CAT(0). □

We also deduce an immediate consequence concerning the Garside complex of a Garside group. Fix a Garside group (G, S, δ). Let X denote the *Garside complex* of G, ie the simplicial complex with vertex set G, and with simplices corresponding to chains $g_1 <_L g_2 <_L \cdots <_L g_n$ such that $g_n \leqslant_L g_1 \delta$. Since simplices of X have a total order on their vertices, X is a simplicial complex with ordered simplices, so we may endow X with the piecewise ℓ^∞ orthoscheme metric.

Corollary 3.12 *The Garside complex of any Garside group, endowed with the piecewise ℓ^∞ orthoscheme metric, is injective.*

This applies, in particular, to the dual braid complex studied by Brady and McCammond [2010]. This complex, endowed with the piecewise orthoscheme Euclidean metric, is conjectured to be CAT(0), but it is only known for a very small number of strands (see [Brady and McCammond 2010; Haettel et al. 2016; Jeong 2023]). On the other hand, if we endow the dual braid complex with the piecewise orthoscheme ℓ^∞ metric, we see that it is injective.

4 Application to Euclidean buildings and the Deligne complex of type $\tilde{A}_n$

Let us consider the Euclidean Coxeter group $W \simeq \mathfrak{S}_n \ltimes \mathbb{Z}^{n-1}$ of type $\tilde{A}_{n-1}$. Its Coxeter complex may be identified with

$$\Sigma = \{x \in \mathbb{R}^n \mid x_1 + \cdots + x_n = 0\}.$$

Up to homothety, we may choose the following affine hyperplanes to define Σ:

$$\{x_i - x_j = k \mid 1 \leqslant i \neq j \leqslant n,\ k \in \mathbb{Z}\},$$

so that maximal simplices of Σ identify with the W–orbit of

$$K = \{x \in \Sigma \mid x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \leqslant x_1 + 1\}.$$

The vertex set of Σ identifies with $\left(\frac{1}{n}\mathbb{Z}\right)^n \cap \Sigma$. Since the simplex K is a strict fundamental domain for the action of W on Σ, one may define a W–invariant type function τ on the vertex set of Σ:

$$\tau\colon \Sigma^{(0)} \to \mathbb{Z}/n\mathbb{Z}, \quad x \in W \cdot v_i \mapsto i,$$

where $v_i = \left(\frac{n-i}{n}, \ldots, \frac{n-i}{n}, -\frac{i}{n}, \ldots, -\frac{i}{n}\right)$ is the vertex of K whose first i coordinates equal $\frac{n-i}{n}$ and the $n-i$ last coordinates equal $-\frac{i}{n}$.

This type of function is such that adjacent vertices have distinct types.

Let us define the *extended* Coxeter complex

$$\hat{\Sigma} = \Sigma \times \mathbb{R} = \mathbb{R}^n,$$

where the action of the standard generators $w_1, \ldots, w_n$ of W on $\hat{\Sigma}$ is given by

$$\text{for all } 1 \leqslant i \leqslant n-1, \quad w_i \cdot (x_1, \ldots, x_i, x_{i+1}, \ldots, x_n) = (x_1, \ldots, x_{i+1}, x_i, \ldots, x_n)$$
$$\text{and } w_n \cdot (x_1, \ldots, x_n) = (x_n - 1, x_2, \ldots, x_{n-1}, x_1 + 1).$$

Also note that $\hat{\Sigma}$ has a natural simplicial complex structure, with vertex set $\mathbb{Z}^n$, with maximal simplices corresponding to the W–orbits of the simplices

$$\{k \leqslant x_i \leqslant x_{i+1} \leqslant x_{i+2} \leqslant \ldots \leqslant x_{i+n-1} \leqslant k+1 \mid k \in \mathbb{Z},\ i \in \mathbb{Z}/n\mathbb{Z}\},$$

where indices in $\mathbb{R}^n$ are considered modulo n. Note that the action of the Coxeter group W preserves τ.

A fundamental domain for the action of W on Σ is the simplex

$$K = \{x \in \Sigma \mid x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \leqslant x_1 + 1\},$$

and a fundamental domain for the action of W on $\hat{\Sigma}$ is the column

$$\hat{K} = \{x \in \hat{\Sigma} \mid x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n \leqslant x_1 + 1\}.$$

Note that such columns have been studied by Brady and McCammond [2010], and by Dougherty, McCammond and Witzel [Dougherty et al. 2020].

We will endow $\hat{\Sigma}$ with the standard ℓ^∞ metric from $\mathbb{R}^n$.

Let us consider a simplicial complex X such that

- either X is a Euclidean building of type $\tilde{A}_{n-1}$,
- or X is the Deligne complex of the Euclidean Artin group A of type $\tilde{A}_{n-1}$, with a coarser simplicial structure.

We will define the *extended* version of X, denoted by $\hat{X}$. It is a simplicial complex whose geometric realization is homeomorphic to $X \times \mathbb{R}$.

Before giving the precise definition of $\hat{X}$, let us first recall briefly the definition of the Deligne complex Δ of the Euclidean Artin group A of type $\tilde{A}_{n-1}$ (see also Section 1.2). Let $S \simeq \mathbb{Z}/n\mathbb{Z}$ denote the standard generating set of A. Consider the set

$$\{gA_T \mid g \in A,\ T \subsetneq S\},$$

endowed with the following partial order: $gA_T \leqslant g'A_{T'}$ if $gA_T \subset g'A_{T'}$. Then the Deligne complex Δ of A is the geometric realization of this poset. We will define a coarser simplicial structure X on Δ. Note that, for any minimal vertex $gA_\varnothing \in \Delta$ (where $g \in A$), the 1–neighborhood of $gA_\varnothing$ in Δ is

$$\{gA_T \mid T \subsetneq S\},$$

which is precisely the barycentric subdivision of the simplex with vertex set

$$\{gA_T \mid T \subsetneq S,\ |S \backslash T| = 1\}$$

consisting of (the g–translates of) all maximal proper standard parabolic subgroups of A.

Therefore we may define the simplicial complex X with vertex set $\{gA_T \mid g \in A,\ T \subsetneq S,\ |S\backslash T| = 1\}$, and such that $g_1 A_{T_1}, \ldots, g_k A_{T_k}$ span a simplex in X if and only if $\bigcap_{i=1}^k g_i A_{T_i} \neq \varnothing$. Then Δ identifies with the barycentric subdivision of X, and also the geometric realizations of Δ and X are homeomorphic.

Note that, in both cases (Euclidean building or Deligne complex), there is a well-defined type function $\tau\colon X^{(0)} \to \mathbb{Z}/n\mathbb{Z}$ such that adjacent vertices have different types.

- In the case X is a Euclidean building of type $\tilde{A}_{n-1}$, each apartment is identified with the Coxeter complex Σ; we may define the type of a vertex v of X to be its type in any apartment containing v. Since the Coxeter group W preserves the type, this definition does not depend on the choice of apartment.

• In the case X is the Deligne complex of type $\tilde{A}_{n-1}$, one may either use the projection onto the Coxeter complex, or use a direct definition: if gA_T is a vertex of X, with $g \in A$ and $T = S \backslash \{i\}$, its type is i.

The complex X may also be defined as a particular gluing of copies of the Coxeter complex Σ. Roughly speaking, $\widehat{X}$ will be the same gluing of copies of $\widehat{\Sigma}$.

More precisely, let $\widehat{X} = X \times \mathbb{R}$, and we will define a simplicial complex structure on $\widehat{X}$. The vertex set of $\widehat{X}$ is $\{(x,i) \in X^{(0)} \times \mathbb{Z} \mid \tau(x) = i + n\mathbb{Z}\}$, where $X^{(0)}$ denotes the vertex set of X. Using the type function τ, remark that vertices of maximal simplices of X have a well-defined cyclic ordering in $\mathbb{Z}/n\mathbb{Z}$. The maximal simplices of $\widehat{X}$ are

$$\{(x_i, kn+i), (x_{i+1}, kn+i+1), \dots, (x_n, kn+n), (x_1, kn+n+1), (x_2, kn+n+2), \dots, (x_i, kn+n+i)\}$$

for any $k \in \mathbb{Z}$, any $1 \leqslant i \leqslant n$ and any maximal simplex $(x_1, x_2, \dots, x_n)$ of X with for all $1 \leqslant i \leqslant n$, $\tau(x_i) = i$. We endow each such simplex with the ℓ^∞ orthoscheme metric for the given ordering. Endow $\widehat{X}$ with the associated length metric.

Note that the translation action on the $\mathbb{R}$ factor of $\widehat{X} = X \times \mathbb{R}$ defines an isometric action denoted by θ. Also note that θ does not preserve the simplicial structure of $\widehat{X}$, but only its restriction to the subgroup $n\mathbb{Z}$ of $\mathbb{R}$.

If X is a Euclidean building of type $\tilde{A}_{n-1}$, then $\widehat{X}$ is called a Euclidean building of extended type $\tilde{A}_{n-1}$ (see [Bruhat and Tits 1972], and also [Haettel 2022a]).

If X is the Deligne complex of the Euclidean Artin group A of type $\tilde{A}_{n-1}$, we will give another description of $\widehat{X}$. Recall that the classical Deligne complex X may be defined as

$$X = (A \times K)/\sim,$$

where $(g,x) \sim (g',x')$ if $x = x'$ and, if the stabilizer of x in Σ equals W_T for some $T \subsetneq S$, then $g^{-1}g' \in A_T$. Then the extended Deligne complex may also be defined similarly as

$$\widehat{X} = (A \times \widehat{K})/\sim,$$

where $(g,x) \sim (g',x')$ if $x = x'$ and, if the stabilizer of x in $\widehat{\Sigma}$ equals W_T for some $T \subsetneq S$, then $g^{-1}g' \in A_T$.

Fix any vertex $x \in X$, and let $L_{x,0}$ denote the set of vertices of X adjacent to x. Without loss of generality, we may assume that x has type $\tau(x) = 0$. Note that vertices in $L_{x,0}$ have a type in $\mathbb{Z}/n\mathbb{Z} \backslash \{\tau(x)\} = \{1, \dots, n\}$, which is an interval. Since maximal simplices of X have a natural cyclic ordering in $\mathbb{Z}/n\mathbb{Z}$, there is a natural induced order on $L_{x,0}$ that is consistent with the type function $\tau \colon L_{x,0} \to \{1, \dots, n\}$. Consider the poset $L_x = L_{x,0} \cup \{0, 1\}$, where 0 and 1 are defined to be the minimum and the maximum of L_x respectively.

Proposition 4.1 *Consider any $p \in \widehat{X}$ whose projection p_X onto X is contained in the open star of the vertex x. Then $\widehat{X}$ is locally isometric at p to a neighborhood of a point in the ℓ^∞ orthoscheme realization of the poset L_x.*

Proof Remember that there is a diagonal, isometric action θ of $\mathbb{R}$ on $\widehat{X}$, whose quotient is X. So, up to isometry, we may assume that p lies in the open star of the diagonal edge joining the vertices $(x,0)$ and (x,n) of $\widehat{X}$. Note that each simplex of $\widehat{X}$ containing p corresponds to a chain in L_x containing $(x,0)$ (identified with $0 \in L_x$) and (x,n) (identified with $1 \in L_x$). As a consequence, a neighborhood of p in $\widehat{X}$ is isometric to a neighborhood of a point in the orthoscheme realization of L_x. □

We now prove that, in the case of a Euclidean building, the poset L_x is a lattice.

Proposition 4.2 *Let L_0 denote the vertex set of a (possibly nonthick) spherical building of type A_{n-1} for some $n \geqslant 1$, and let $L = L_0 \cup \{0,1\}$. In other words, L is the poset of linear subspaces of a projective space of dimension $n-1$. Then L is a bounded, graded lattice of rank $n-1$.*

Proof The lattice property is obvious: if S, S' are linear subspaces of a projective space X, then $S \wedge S' = S \cap S'$ and $S \vee S'$ is the linear subspace spanned by $S \cup S'$. □

As a consequence, if X is a Euclidean building of type $\tilde{A}_{n-1}$, for any vertex $x \in X$, the poset L_x is a bounded graded lattice.

We now turn to the case of the Deligne complex of type $\tilde{A}_{n-1}$. We defer the proof of the following result to Section 5.

Theorem 4.3 (Crisp–McCammond) *Let X denote the Deligne complex of type $\tilde{A}_{n-1}$. For any vertex $x \in X$, the poset L_x is a bounded, graded lattice.*

We may now deduce the main result.

Theorem 4.4 *Any Euclidean building of extended type $\tilde{A}_{n-1}$, or the extended Deligne complex of type $\tilde{A}_{n-1}$, endowed with the piecewise ℓ^∞ metric, is injective.*

Proof According to Proposition 4.1 and Theorem 4.3, we know that $\widehat{X}$ is locally isometric to the ℓ^∞ orthoscheme complex of a bounded graded lattice. According to Theorem 3.10, we deduce that $\widehat{X}$ is uniformly locally injective.

If X is a Euclidean building, X and $\widehat{X}$ are contractible, and in particular simply connected. If X is a Deligne complex, according to Theorem 1.6, $\widehat{X}$ is simply connected.

According to Theorem 1.11, $\widehat{X}$ is complete. We deduce with Theorem 1.14 that $\widehat{X}$ is injective. □

We can also deduce that the thickening is a Helly graph. Note that this thickening, described more precisely in Theorem 1.15, corresponds to a coarser cell structure on the building or the Deligne complex.

Corollary 4.5 *The thickening of the vertex set of any Euclidean building of extended type $\tilde{A}_{n-1}$, or the extended Deligne complex of type $\tilde{A}_{n-1}$, is a Helly graph.*

Proof Let X denote either a Euclidean building of extended type $\tilde{A}_{n-1}$ or the extended Deligne complex of type $\tilde{A}_{n-1}$. We will see that X satisfies the assumptions of Theorem 1.15.

Since simplices of X have a well-defined order, we see that X is a simplicial complex with ordered simplices. For each maximal simplex σ of X, the minimal and maximal vertices of σ form a maximal edge in X. Therefore, X has maximal edges as in Definition 1.9. According to Theorem 1.15, we deduce that the thickening of X is a Helly graph. $\square$

We will now deduce a bicombing on X, considered as the quotient of $(\widehat{X}, d_{\widehat{X}})$ by the diagonal action θ of $\mathbb{R}$ on $\widehat{X}$. Let us define the quotient metric d_X on X:

$$\text{for all } x, y \in X, \quad d_X(x, y) = \inf_{t\in\mathbb{R}} d_{\widehat{X}}(x, \theta(t)\cdot y).$$

Theorem 4.6 *Any Euclidean building X of type $\tilde{A}_{n-1}$, or the Deligne complex X of type $\tilde{A}_{n-1}$, has a metric d_X that admits a convex, consistent, reversible geodesic bicombing σ. Moreover, d_X and σ are invariant under the group of type-preserving automorphisms of X.*

Proof We will prove it locally, ie for the star Y of a vertex v of X. According to Theorem 3.10, there exists a unique convex, consistent, reversible bicombing $\hat{\sigma}$ on $\widehat{Y}$. For each $x, y \in Y$, choose lifts $\hat{x}, \hat{y} \in \widehat{Y}$ such that $d_{\widehat{X}}(\hat{x}, \hat{y}) = d_X(x, y)$. For each $t \in [0, 1]$, let us define $\sigma(x, y, t) = \theta(\mathbb{R})\cdot\hat{\sigma}(\hat{x}, \hat{y}, t) \in Y$.

We will first see that σ is well-defined: indeed fix $x, y \in Y$, and consider two pairs of lifts $\hat{x}, \hat{x}', \hat{y}, \hat{y}' \in \widehat{Y}$ such that $d_{\widehat{X}}(\hat{x}, \hat{y}) = d_{\widehat{X}}(\hat{x}', \hat{y}') = d_X(x, y)$. Note that, up to the action of θ, we may assume that $\hat{x} = \hat{x}'$. If $\hat{y} \neq \hat{y}'$, let $a \in \mathbb{R}\backslash\{0\}$ such that $\hat{y}' = \theta(a)\cdot\hat{y}$. We then have $d_{\widehat{X}}\big(\hat{x}, \theta\big(\frac{a}{2}\big)\cdot\hat{y}\big) < d(x, y)$, which is a contradiction. So $\hat{y} = \hat{y}'$, and σ is well-defined.

Moreover, it is easy to see that σ is a reversible, consistent, geodesic bicombing.

We will now see that σ is convex. Consider $x, x', y, y' \in Y$, and choose lifts $\hat{x}, \hat{x}', \hat{y}, \hat{y}' \in \widehat{Y}$ such that $d_{\widehat{X}}(\hat{x}, \hat{x}') = d_X(x, x')$ and $d_{\widehat{X}}(\hat{y}, \hat{y}') = d_X(y, y')$. Fix any $t \in [0, 1]$ and $s \in \mathbb{R}$. We have

$$\begin{aligned} d_X(\sigma(x, y, t), \sigma(x', y', t)) &= \inf_{s\in\mathbb{R}} d_{\widehat{X}}(\hat{\sigma}(\hat{x}, \hat{y}, t), \theta(s)\cdot\hat{\sigma}(\hat{x}, \hat{y}, t)) \\ &\leqslant d_{\widehat{X}}(\hat{\sigma}(\hat{x}, \hat{y}, t), \hat{\sigma}(\hat{x}, \hat{y}, t)) \\ &\leqslant (1-t)d_{\widehat{X}}(\hat{x}, \hat{x}') + t d_{\widehat{X}}(\hat{y}, \hat{y}') \\ &\leqslant (1-t)d_X(x, x') + t d_X(y, y'), \end{aligned}$$

so σ is a conical bicombing. Since σ is also consistent, it is a convex bicombing.

We have seen that (X, d_X) is locally convexly reversibly consistently bicombable. According to [Miesch 2017, Theorem 1.1], we deduce that (X, d_X) has a unique global convex reversible geodesic consistent bicombing σ that is consistent with local bicombing. Since the local bicombing only depends on the local combinatorics of X and the type, we deduce that σ is invariant under type-preserving automorphisms of X. $\square$

Note that we strengthened this result in [Haettel 2022b], by proving that this convex geodesic bicombing is actually unique. Our result also extends to simplicial complexes which are much more general than buildings.

5 The lattice of maximal parabolic subgroups of braid groups

This whole section is unpublished work of John Crisp and Jon McCammond, copied here with their permission. The results concerning the lattice of cut-curves are contained in [Bessis 2006] about Garside structures on free groups. For consistency, we choose to follow Crisp and McCammond's presentation instead.

5.1 The lattice of cut-curves

Let $\mathbb{D}^2$ denote the unit disk in $\mathbb{R}^2$, and let $\{p_1, \ldots, p_n\}$ denote a set of n distinct points in $\mathbb{D}^2$, as shown in Figure 2. We write $\mathbb{D}_* = \mathbb{D}^2 \backslash \{p_1, \ldots, p_n\}$. Let $a = (0, 1)$ and $b = (0, -1)$ denote the top and bottom points of the boundary $\partial\mathbb{D}^2$.

Definition 5.1 By a *cut-curve* or *curve* on $\mathbb{D}_*$ we shall mean a smoothly embedded curve c in $\mathbb{D}_*$ which meets $\partial\mathbb{D} = \partial\mathbb{D}^2$ precisely at its endpoints, and which separates the boundary points a and b.

We shall say that two cut-curves are isotopic if they are isotopic in $\mathbb{D}_*$ relative to $\{a, b\}$. We denote by $[c]$ the isotopy class of a curve c, and write $\mathcal{C}$ for the set of all isotopy classes of cut-curves in $\mathbb{D}_*$.

Observe that any cut-curve c separates $\mathbb{D}_*$ into two regions, an upper region containing a and a lower region containing b, and induces, in particular, a partition of the points $\{p_1, \ldots, p_n\}$ into two sets. In general we shall say that the contents of the region containing a lie *above* c and the contents of the region containing b lie *below* c. For each curve c, we write $\deg(c)$ for the number of points p_i which lie below c. Clearly this number is invariant under isotopy, and so defines a degree function on $\mathcal{C}$ by $\deg([c]) = \deg(c)$.

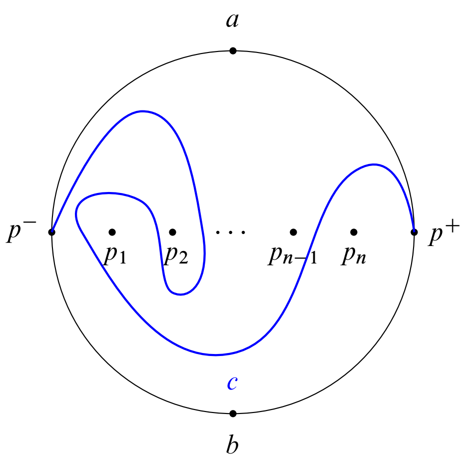


Figure 2: An example of cut-curve c in the punctured disk $\mathbb{D}_*$.

Let c_1, c_2 be two curves. We say that c_1 and c_2 are *in minimal position* with respect to one another if they do not cobound any disk regions ("bigons") in $\mathbb{D}_*$. This includes triangular regions against the boundary. Any such disk regions can always be removed by modifying just one of the two curves in its isotopy class without changing the other. That is to say that, for any two curves c_1, c_2, we can always find c_1' such that $[c_1'] = [c_1]$ and c_1' is in minimal position with respect to c_2.

Here is another argument about minimal position using hyperbolic geometry. Such an argument has been used in [Bessis 2006]. If $n \geqslant 2$, one may also endow $\mathbb{D}_*$ with a fixed complete hyperbolic metric. Fix $p^+ = (1,0)$, $p^- = (-1,0) \in \partial\mathbb{D}$ as points on each connected component of $\partial\mathbb{D}\backslash\{a,b\}$. Then, for each isotopy class $[c]$, we may consider the unique geodesic line in $\mathbb{D}_*$ in the isotopy class $[c]$ with endpoints $p^+, p^- \in \partial\mathbb{D}$. Then, for any $[c_1], [c_2] \in \mathcal{C}$, the geodesic representatives are in minimal position.

Definition 5.2 Let c_1, c_2 be two curves. We say that $[c_1] \leqslant [c_2]$ if c_1 is isotopic to a curve which lies below c_2. It is easily checked that this defines a partial order on the set $\mathcal{C}$ of cut-curves classes.

Note that if c_1 and c_2 are in minimal position with respect to one another then $[c_1] \leqslant [c_2]$ if and only if c_1 lies below c_2 (in particular they are disjoint). Note also that the function $\deg\colon \mathcal{C} \to \{1,\dots,n\}$ is a strict order-preserving map, a grading on the poset $(\mathcal{C}, \leqslant)$.

Theorem 5.3 [Bessis 2006, Theorem 2.6] *The graded poset $(\mathcal{C}, \leqslant)$ is a lattice.*

Proof Let $x = [c_1]$, $y = [c_2]$ be arbitrary elements of $\mathcal{C}$. Suppose that the curves c_1 and c_2 are in minimal position with respect to one another. Now consider how the union of c_1 and c_2 cut $\mathbb{D}_*$ up into connected regions. There is a unique lowermost such region R which lies below both c_1 and c_2 (and contains the point b), and another uppermost such region R' which lies above both c_1 and c_2 (and contains the point a). Let c, resp. c', denote the curves which skirt along the part of the boundary of R, resp. R', lying in the interior of $\mathbb{D}_*$ (ie not along $\partial\mathbb{D}$). Then we claim that $x \wedge y = [c]$ and $x \vee y = [c']$.

We check just the first of these two claims. Suppose that c_0 represents a common lower bound for x and y. Then, by a sequence of bigon-removing isotopies (or by considering the geodesic representative), we may choose the representative c_0 to be in minimal position with respect to both c_1 and c_2. Since c_0 represents a common lower bound, c_0 lies below both c_1 and c_2, and hence lies below the curve c. □

Recall that the n–strand braid group B_n is isomorphic to the mapping class group $\mathrm{MCG}(\mathbb{D}_*, \partial\mathbb{D})$. As a consequence, B_n acts naturally on the set $\mathcal{C}$ of isotopy classes of cut-curves.

Lemma 5.4 *The action of B_n on $\mathcal{C}$ preserves the order and the degree.*

Proof Let $[c] \in \mathcal{C}$ be a curve of degree $1 \leqslant k \leqslant n-1$ and $g \in B_n$. Then k points among $\{p_1,\dots,p_n\}$ lie below c. Since the action of B_n fixes the boundary of $\mathbb{D}_*$ pointwise, we know that k points among $\{p_1,\dots,p_n\}$ lie below $g(c)$. Hence $\deg g([c]) = \deg[c]$.

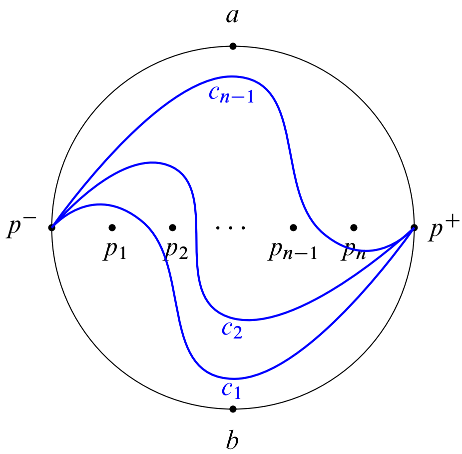


Figure 3: The base maximal chain $\alpha = (c_1, c_2, \ldots, c_{n-1})$ of $\mathcal{C}$.

Similarly, consider two curves $[c_1], [c_2] \in \mathcal{C}$ such that $[c_1] \leqslant [c_2]$. Consider c_1, c_2 in minimal position, so that c_1 is below c_2. For any $g \in B_n$, we have that $g(c_1)$ is below $g(c_2)$; hence $g([c_1]) \leqslant g([c_2])$. We conclude that the action of B_n preserves the order on $\mathcal{C}$. □

Note that B_n acts transitively on the set of cut-curves with fixed degree.

5.2 Cosets in braid groups

We will show how the cut-curve lattice $(\mathcal{C}, \leqslant)$ can be reinterpreted in purely algebraic terms.

Definition 5.5 Recall that the braid group B_n is generated by the set of standard generators $S = \{\sigma_1, \sigma_2, \ldots, \sigma_{n-1}\}$. For each $1 \leqslant k \leqslant n-1$, let P_k denote the maximal parabolic subgroup of B_n generated by $S \setminus \{\sigma_k\}$. Thus P_k is isomorphic to a product $B_k \times B_{n-k}$. Note also that these subgroups are distinct for distinct values of k. We define the following augmented collection of cosets in B_n:

$$\mathcal{B} = \{gP_k \mid g \in B_n,\ 1 \leqslant k \leqslant n-1\} \cup \{0, 1\}.$$

For $b \in \mathcal{B}$, we write $\deg(b) = k$ if $b = gP_k$ for $1 \leqslant k \leqslant n-1$, and $\deg(0) = 0$, $\deg(1) = n$. We define an order relation $\leqslant_{\mathcal{B}}$ on $\mathcal{B}$ as follows. First, define $0 \leqslant b \leqslant 1$ for all $b \in \mathcal{B}$. Otherwise, for $g_1, g_2 \in B_n$ and $1 \leqslant k_1, k_2 \leqslant n-1$, we write $g_1 P_{k_1} \leqslant_{\mathcal{B}} g_2 P_{k_2}$ if $k_1 \leqslant k_2$ and $g_1 P_{k_1} \cap g_2 P_{k_2} \neq \varnothing$.

For simplicity, we shall henceforth identify B_n with the mapping class group of $\mathbb{D}_*$ (relative to $\partial\mathbb{D}$). Let us consider the base maximal chain $\alpha = (c_1, c_2, \ldots, c_{n-1})$ of $\mathcal{C}$ depicted in Figure 3.

Definition 5.6 Define a map $\Phi\colon \mathcal{C} \to \mathcal{B}$ by setting $\Phi(c) = 0$ if $\deg(c) = 0$, $\Phi(c) = 1$ if $\deg(c) = n$, and otherwise

$$\Phi(c) = \{g \in B_n \mid g(\alpha) \text{ contains } c\}.$$

Lemma 5.7 *If $c \in \alpha$ and $1 \leqslant \deg(c) = k \leqslant n-1$, then $\Phi(c) = P_k$.*

Proof The curve c is a base curve such that $p_1, \dots, p_k$ lie below c and $p_{k+1}, \dots, p_n$ lie above c. Let R^- denote the connected component of $\mathbb{D}^* \backslash c$ below c, and R^+ denote the connected component of $\mathbb{D}^* \backslash c$ above c. The stabilizer of c in the mapping class group $B_n = \mathrm{MCG}(\mathbb{D}_*, \partial\mathbb{D})$ is the direct product of $\mathrm{MCG}(R^-, \partial R^-) = \langle \sigma_1, \dots, \sigma_{k-1} \rangle$ and $\mathrm{MCG}(R^+, \partial R^+) = \langle \sigma_{k+1}, \dots, \sigma_{n-1} \rangle$. Hence the stabilizer of c in B_n equals P_k. □

Lemma 5.8 *The map $\Phi \colon \mathcal{C} \to \mathcal{B}$ is well-defined, surjective, and respects degrees.*

Proof Fix $c \in \mathcal{C}$ with $1 \leqslant \deg(c) = k \leqslant n-1$. Fix some $g \in B_n$ such that the curve c lies in $g(\alpha)$. Let $c_0 = g^{-1}(c)$, the degree-k curve in α. For any $h \in B_n$, we have

$$h \in \Phi(c) \iff c \in h(\alpha) \iff g(c_0) = h(c_0).$$

According to Lemma 5.7, this is equivalent to $gP_k = hP_k$. Hence $\Phi(c) = gP_k$ is well-defined. Moreover it is clear that Φ is surjective, and that for all $c \in \mathcal{C}$ we have $\deg(\Phi(c)) = \deg(c)$. □

Lemma 5.9 *The map Φ is injective, so it is a bijection between $\mathcal{C}$ and $\mathcal{B}$.*

Proof Let $c_1, c_2 \in \mathcal{C}$ such that $\Phi(c_1) = \Phi(c_2)$. We may suppose that c_1 lies in $g_1(\alpha)$ and c_2 in $g_2(\alpha)$ for some $g_1, g_2 \in B_n$. According to Lemma 5.8, we deduce that c_1 and c_2 have the same degree $0 \leqslant k \leqslant n$. Since $\mathcal{C}$ and $\mathcal{B}$ have unique elements of degree 0 and n, we may restrict to the case $1 \leqslant k \leqslant n$.

Since $\Phi(c_1) = \Phi(c_2)$, we deduce that $g_1 P_k = g_2 P_k$, so ${g_1}^{-1} g_2 \in P_k$. Let $h \in P_k$ such that $g_2 = g_1 h$. Let c_0 denote the degree-k curve of α. Then $c_1 = g_1(c_0)$ and $c_2 = g_2(c_0)$. Since $h \in P_k$, h fixes c_0 so $c_1 = c_2$. □

Theorem 5.10 (Crisp–McCammond) *The bijection $\Phi \colon \mathcal{C} \to \mathcal{B}$ is an order isomorphism between $(\mathcal{C}, \leqslant)$ and $(\mathcal{B}, \leqslant_{\mathcal{B}})$. As a consequence, $(\mathcal{B}, \leqslant_{\mathcal{B}})$ is a lattice.*

Proof To prove the theorem, we need to show that, for $c_1, c_2 \in \mathcal{C}$, we have $c_1 \leqslant c_2$ if and only if $\Phi(c_1) \leqslant_{\mathcal{B}} \Phi(c_2)$. Write $k_i = \deg(c_i)$ for $i = 1, 2$, and suppose, without loss of generality, that $0 \leqslant k_1 \leqslant k_2 \leqslant n$. Then $\Phi(c_1) \leqslant_{\mathcal{B}} \Phi(c_2)$ if and only if $\Phi(c_1) \cap \Phi(c_2) \neq \varnothing$ if and only if there exists a maximal chain $\alpha' \subset \mathcal{C}$ which contains both c_1 and c_2, if and only if $c_1 \leqslant c_2$ (since we already know that $\deg(c_1) \leqslant \deg(c_2)$). □

We deduce the proof of Theorem 4.3 that each local poset L_x in the Deligne complex of type $\tilde{A}_{n-1}$ is a lattice.

Proof of Theorem 4.3 Let x denote a vertex of the Deligne complex X of type $\tilde{A}_{n-1}$. Without loss of generality, we may assume that x corresponds to the maximal proper parabolic subgroup $A(A_{n-1}) = \langle \sigma_1, \dots, \sigma_{n-1} \rangle$, which is isomorphic to the Artin group of type A_{n-1}, ie the n–strand braid group. Since $A(A_{n-1})$ is a maximal parabolic subgroup of spherical type of the Euclidean Artin group $A(\tilde{A}_{n-1})$, the star of x in X identifies with the Deligne complex of the Artin group $A(A_{n-1})$. In particular, vertices in X adjacent to x may be identified with cosets of proper maximal parabolic subgroups of the braid group $A(A_{n-1})$. Furthermore, given two such cosets gP_i and hP_j for $g, h \in A(A_{n-1})$ and $1 \leqslant i, j \leqslant n-1$, the corresponding vertices of X are adjacent if and only if $gP_i \cap hP_j \neq \varnothing$. As a consequence, the poset $L_{x,0}$ is isomorphic to the poset $\mathcal{B}\backslash\{0,1\}$, and L_x is isomorphic to the poset $\mathcal{B}$. According to Theorem 5.10, we deduce that L_x is a lattice. □

6 The thickening of a semilattice

We will now consider a generalization of Theorem 3.10 to the case of a semilattice. This will be useful to consider Euclidean buildings, Deligne complexes in type $\widetilde{C}_n$ and Artin groups of type FC. We start by recalling the definition of a flag poset.

Definition 6.1 A poset L is called *flag* if any three elements which are pairwise upper bounded have an upper bound.

Lemma 6.2 *Let L denote a graded flag meet-semilattice with bounded rank. Then any family of elements of L which are pairwise upper bounded have a join.*

Proof We first prove that every finite family of k elements $\{a_i\}_{1\leqslant i\leqslant k}$ of L which are pairwise upper bounded have a join, by induction on the number of elements. By assumption, the property is true for $k=3$. Fix $k\geqslant 4$, assume that the property is true for $k-1$, and consider k elements $\{a_i\}_{1\leqslant i\leqslant k}$ of L which are pairwise upper bounded. Let b denote the join of a_{k-1} and a_k. The family $\{a_1, a_2, \dots, a_{k-2}, b\}$ is pairwise upper bounded, so by assumption it has a join $c \in L$. Then c is the join of $\{a_i\}_{1\leqslant i\leqslant k}$, which proves the induction.

Now consider an arbitrary family A of elements of L which are pairwise upper bounded. Since L has bounded rank, we may consider a finite subset $F\subset A$ such that the rank of the join b of F is maximal among all joins of finite subfamilies of A. For any $a\in A$, the join b_a of $F\cup\{a\}$ satisfies $b\leqslant b_a$, and the rank of b_a is at most the rank of b; hence $b_a=b$. We deduce that b is the join of A. □

Theorem 6.3 *Let L denote a graded poset with minimum* 0 *and with bounded rank such that*:

- *L is a meet-semilattice.*
- *L is flag.*

Then the ℓ^∞ orthoscheme realization of $|L|$ is injective.

Proof Let us consider $\bar{1} = L\cup\{1\}$: it is a bounded graded lattice. Let us denote by $\bar{1} = |\bar{1}|$ the geometric realization of $\bar{1}$, and by $M=|L|\subset\bar{1}$ the geometric realization of L. Note that we consider $\bar{1}$ as a subset of $\bar{1}$, its vertex set. The orthoscheme realization of M is endowed with the induced length metric as a subspace of the ℓ^∞ orthoscheme realization of $\bar{1}$.

If we consider the affine version $\bar{1}_{\mathbb{R}}$ of $\bar{1}$ over $\mathbb{R}$, then the elements $0_M = [(0,\dots,0),c]\in\bar{1}_{\mathbb{R}}$ and $1_M=[(1,\dots,1),c]\in\bar{1}_{\mathbb{R}}$ are such that $\bar{1}$ naturally identifies with the interval $I_{\bar{1}_{\mathbb{R}}}(0_M,1_M)$ in $\bar{1}_{\mathbb{R}}$, as in the proof of Theorem 3.10. In particular, M and $\bar{1}$ are also posets.

According to Theorem 3.8, the space $\bar{1}_{\mathbb{R}}$, with its standard ℓ^∞ metric $\bar{d}$, is injective. According to Theorem 3.10, its interval $\bar{1} = I_{\bar{1}_{\mathbb{R}}}(0_M, 1_M)$ is injective. Let us denote by d the length metric of M.

Fix $\varepsilon > 0$, and consider the graph Γ_ε with vertex set M, and with an edge between $x, y \in M$ if $\bar{d}(x, y) \leqslant \varepsilon$. We will prove that Γ_ε is Helly, by proving that it is clique-Helly and that its triangle complex is simply connected.

Let $\sigma \subset \Gamma_\varepsilon$ denote a maximal clique: we claim that there exist $a, b \in \bar{1}$ such that σ is the intersection of an interval $I_{\bar{1}}(a, b)$ in $\bar{1}$ with M. Since σ is a subset of $\bar{1}_{\mathbb{R}}$ of diameter at most ε, there exist $a_0, b_0 \in \bar{1}_{\mathbb{R}}$ such that $\sigma \subset I_{\bar{1}_{\mathbb{R}}}(a_0, b_0)$ and $\bar{d}(a_0, b_0) \leqslant \varepsilon$. Since $\bar{1}$ is an interval, we deduce that there exist $a, b \in \bar{1} \cap I_{\bar{1}_{\mathbb{R}}}(a_0, b_0)$ such that $\sigma \subset I_{\bar{1}_{\mathbb{R}}}(a, b) = I_{\bar{1}}(a, b)$. In particular, we also have $\bar{d}(a, b) \leqslant \varepsilon$. Since σ is a maximal clique, we deduce that $\sigma = I_{\bar{1}}(a, b) \cap M$.

We will prove that Γ_ε is clique-Helly. Consider a family of pairwise intersecting maximal cliques $(\sigma_i)_{i \in I}$ in Γ_ε. For each $i \in I$, according to the previous paragraph, there exists $a_i \leqslant b_i$ in $\bar{1}$ such that $\sigma_i = I_{\bar{1}}(a_i, b_i) \cap M$. For each $i \in I$, let $m_i \in L$ minimal such that $a_i \leqslant m_i$. For each $i \neq j$ in I, since σ_i and σ_j intersect, the intersection $I_{\bar{1}}(a_i, b_i) \cap I_{\bar{1}}(a_j, b_j) \cap M$ is nonempty. In particular, the elements m_i and m_j have a common upper bound in L.

The family $(m_i)_{i \in I}$ is pairwise upper bounded. Since L is a graded flag semilattice with bounded rank, according to Lemma 6.2, this family has a join $m = \bigvee_{i \in I} m_i$ in L. In particular, the family $(\sigma_i \cap I_{\bar{1}}(0, m))_{i \in I}$ of intervals in the lattice $I_{\bar{1}}(0, m) = I_M(0, m)$ has a nonempty intersection. So the graph Γ_ε is clique-Helly.

We will now prove that the triangle complex of Γ_ε is simply connected. For each $t \in [0, 1]$, consider the map $\pi_t : M \to M$ sending each $x \in I_{\bar{1}}(0, m)$ to the point on the affine segment joining 0 to x at distance $t d(0, x)$ from 0. Note that π_t is 1–Lipschitz with respect to the distance d. Furthermore, if $t, t' \in [0, 1]$ are such that $|t - t'| \leqslant \varepsilon$, then for each $x \in M$ we have $d(\pi_t(x), \pi_{t'}(x)) \leqslant \varepsilon$. As a consequence, any combinatorial loop γ in Γ_ε may be homotoped in the triangle complex of Γ_ε to the loop $\pi_0(\gamma)$, which is the constant loop at 0. So we have proved that the triangle complex of Γ_ε is simply connected.

According to Theorem 1.12, we deduce that the graph Γ_ε is Helly. In particular, for each $\varepsilon > 0$, the metric space M is ε–coarsely injective. Since M is complete according to Theorem 1.11, we deduce by Lemma 1.13 that M is injective. □

We will now give natural examples of such semilattices, which will be used in the sequel for Euclidean buildings and Deligne complexes of Euclidean type different from $\tilde{A}_n$.

Proposition 6.4 *Let L_0 denote the vertex set of a (possibly nonthick) spherical building of type B_n for some $n \geqslant 1$, and let $L = L_0 \cup \{0\}$. In other words, L is the poset of subspaces of a polar space of projective dimension $n - 1$. Then L is a graded semilattice of rank n with minimum 0 such that any pairwise upper bounded subset of L has a join.*

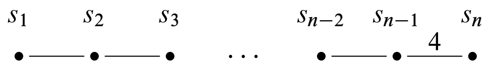


Figure 4: The Dynkin diagram of type B_n.

Proof The semilattice property is obvious: if S, S' are subspaces of a polar space X, then $S \wedge S' = S \cap S'$. If $A \subset L$ are pairwise upper bounded in L, let $A_X = \bigcup_{S \in A} S \subset X$. For any $x, y \in A_X$, there exists a subspace of X containing x and y. According to [Tits 1974, 7.2.1], the subset A_X is contained in a subspace S_0 of X. Hence, for any $S \in A$, we have $S \subset S_0$. So the intersection of all subspaces of X containing $\bigcup_{S \in A} S$ is the join of A. □

Another application concerns the Artin group $A(B_n)$ of spherical type B_n, with the following Dynkin diagram (see Figure 4).

Let $s_1, \dots, s_n$ denote the standard generators of $A(B_n)$, with $s_{n-1}s_n s_{n-1}s_n = s_n s_{n-1} s_n s_{n-1}$. For each $1 \leqslant i \leqslant n$, let P_i denote the maximal proper standard parabolic subgroup of $A(B_n)$

$$P_i = \langle s_1, \dots, s_{i-1} \rangle \times \langle s_{i+1}, \dots, s_n \rangle.$$

Let $L_0 = \{gP_i \mid g \in A(B_n),\ 1 \leqslant i \leqslant n\}$ and $L = L_0 \cup \{0\}$. We define an order relation on L as follows. First, define $0 \leqslant gP_i$ for each $gP_i \in L$. Otherwise, for $g_1, g_2 \in A(B_n)$ and $1 \leqslant k_1, k_2 \leqslant n$, define $g_1 P_{k_1} \leqslant g_2 P_{k_2}$ if $k_1 \leqslant k_2$ and $g_1 P_{k_1} \cap g_2 P_{k_2} \neq \varnothing$.

Lemma 6.5 *The relation $\leqslant$ is a partial order on L.*

Proof First note that $\leqslant$ is antisymmetric: if $g_1 P_{k_1}, g_2 P_{k_2} \in L_0$ are such that $g_1 P_{k_1} \leqslant g_2 P_{k_2}$ and $g_2 P_{k_2} \leqslant g_1 P_{k_1}$, then $k_1 = k_2$, and $g_1 P_{k_1} \cap g_2 P_{k_1} \neq \varnothing$ implies that $g_1 P_{k_1} = g_2 P_{k_1}$.

Now we show that $\leqslant$ is transitive: assume that $g_1 P_{k_1}, g_2 P_{k_2}, g_3 P_{k_3} \in L_0$ are such that $g_1 P_{k_1} \leqslant g_2 P_{k_2}$ and $g_2 P_{k_2} \leqslant g_3 P_{k_3}$. We deduce that $k_1 \leqslant k_2 \leqslant k_3$, $g_1 \in g_2 P_{k_2} P_{k_1}$ and $g_2 \in g_3 P_{k_3} P_{k_2}$, so $g_1 \in g_3 P_{k_3} P_{k_2} P_{k_1}$. Note that $\langle s_1, \dots, s_{k_2-1} \rangle \subset P_{k_3}$ and $\langle s_{k_2+1}, \dots, s_n \rangle \subset P_{k_1}$, so $P_{k_3} P_{k_2} P_{k_1} = P_{k_3} P_{k_1}$.

In particular $g_1 \in g_3 P_{k_3} P_{k_1}$ and $k_1 \leqslant k_3$, so $g_1 P_{k_1} \leqslant g_3 P_{k_3}$. □

Note that we are grateful to Luis Paris for his help in the following proof, notably the use of normal forms.

Proposition 6.6 *L is a graded semilattice of rank n with minimum 0 such that any pairwise upper bounded subset of L has a join.*

Proof Let $t_1, \dots, t_{2n-1}$ be the standard generators of the braid group $A(A_{2n-1})$. Consider the morphism

$$\phi \colon A(B_n) \to A(A_{2n-1}),$$
$$\text{for all } 1 \leqslant i \leqslant n-1, \quad s_i \mapsto t_i t_{2n-i}, \qquad s_n \mapsto t_n.$$

This morphism ϕ is injective; see for instance [Dehornoy and Paris 1999; Michel 1999; Crisp 2000]. Let σ denote the involution of $A(A_{2n-1})$ defined for all $1 \leqslant i \leqslant 2n-1$ by $\sigma(t_i) = t_{2n-i}$. According to [Crisp 2000, Theorem 4], we know furthermore that the image of ϕ coincides with the fixed-point set of σ.

For each $1 \leqslant i \leqslant 2n-1$, consider the standard proper maximal parabolic subgroup

$$Q_i = \langle t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_{2n-1} \rangle$$

of $A(A_{2n-1})$. Let M_0 denote the poset of cosets of maximal proper parabolic subgroups of $A(A_{2n-1})$, and $M = M_0 \cup \{0, 1\}$. We will define a poset map $\psi\colon L_0 \to M_0$.

For each $g \in A(B_n)$ and each $1 \leqslant i \leqslant n$, let us define $\psi(gP_i) = \phi(g)Q_i$. Since $\phi(P_i) \subset Q_i$, the map $\psi\colon L_0 \to M_0$ is well-defined. Assume that $gP_i \leqslant hP_j$, ie $i \leqslant j$ and $h^{-1}g \in P_jP_i$. Then $\phi(h)^{-1}\phi(g) \in \phi(P_j)\phi(P_i) \subset Q_jQ_i$, so $\phi(g)Q_i \leqslant \phi(h)Q_j$. As a consequence, ψ is a rank-preserving injective poset map.

Note that the involution σ extends naturally to an order-reversing involution on M by letting $\sigma(gQ_i) = \sigma(g)Q_{2n-i}$. And for each $g \in A(B_n)$ and each $1 \leqslant i \leqslant n$, we have $\sigma(\psi(gP_i)) = \phi(g)Q_{2n-i}$.

We will now prove that L is a meet-semilattice. Fix $a, b \in A(B_n)$ and $1 \leqslant i, j \leqslant n$; we will prove that aP_i and bP_j have a meet in L. Consider the elements $\phi(a)Q_i$ and $\phi(b)Q_j$ in M: they have a meet γQ_k for some $\gamma \in A(A_{2n-1})$ and $0 \leqslant k \leqslant i, j$. Then as $\phi(a)Q_i \leqslant \phi(a)Q_{2n-i}$ and $\phi(b)Q_j \leqslant \phi(b)Q_{2n-j}$, we deduce that $\gamma Q_k \leqslant \sigma(\gamma)Q_{2n-k}$.

So, up to the choice of $\gamma \in \gamma Q_k$, we may assume that $\sigma(\gamma) \in \gamma Q_{2n-k}$: let $q \in Q_{2n-k}$ such that $\sigma(\gamma) = \gamma q$. Since σ is an involution, we have $\gamma = \gamma q \sigma(q)$, so $q = \sigma(q)^{-1} \in Q_k \cap Q_{2n-k}$.

We claim that if $g \in A(A_{2n-1})$ is such that $\sigma(g) = g^{-1}$, there exists $h \in A(A_{2n-1})^+$ such that $g = h\sigma(h)^{-1}$. According to [Charney 1995], there exist unique $h, h' \in A(A_{2n-1})^+$ such that $g = hh'^{-1}$ and the right greatest common divisor of h and h' in the Garside monoid $A(A_{2n-1})^+$ is 1. Note that σ preserves the monoid $A(A_{2n-1})^+$; hence we have $g^{-1} = h'h^{-1}$ on one side, and $g^{-1} = \sigma(g) = \sigma(h)\sigma(h')^{-1}$ on the other side. By the uniqueness of h, h', we deduce that $h' = \sigma(h)$. Hence $g = h\sigma(h)^{-1}$.

Assume furthermore that $g \in Q_k \cap Q_{2n-k} = \langle t_1, \dots, t_{k-1} \rangle \times \langle t_k + 1, \dots, t_{2n-k-1} \rangle \times \langle t_{2n-k+1}, \dots, t_{2n-1} \rangle$. Then we can decompose g as a fraction $g = hh'^{-1}$ inside the parabolic subgroup $Q_k \cap Q_{2n-k}$: by the uniqueness of h, h', we deduce that $h, h' \in Q_k \cap Q_{2n-k}$.

According to the claim, there exists $q' \in Q_k \cap Q_{2n-k}$ such that $q = q'\sigma(q')^{-1}$. So, up to replacing γ with $\gamma' = \gamma q' \in \gamma Q_k$, we have

$$\sigma(\gamma') = \sigma(\gamma)\sigma(q') = \gamma q \sigma(q') = \gamma q' = \gamma'.$$

So we may assume that γ' is fixed by σ: according to [Crisp 2000, Theorem 4], we may consider $c \in A(B_n)$ such that $\phi(c) = \gamma'$. So we deduce that $cP_k \leqslant aP_i, bP_j$. Conversely, for any $c' \in A(B_n)$ and $k' \leqslant i, j$ such that $c'P_{k'} \leqslant aP_i, bP_j$, we have $\phi(c')Q_{k'} \leqslant \phi(a)Q_i \wedge \phi(b)Q_j = \phi(c)Q_k$, so $c'P_{k'} \leqslant cP_k$. We conclude that cP_k is the meet of aP_i and bP_j in L.

We will now prove that any three pairwise upper bounded elements of L have an upper bound. Fix $g_1, g_2, g_3 \in A(B_n)$ and $1 \leqslant k_1, k_2, k_3 \leqslant n$ such that, for each $i \neq j$, the elements $g_i P_{k_i}$ and $g_j P_{k_j}$ have a join $g_{ij} P_{k_{ij}}$ in L. Consider the elements $\phi(g_{ij}) Q_{k_{ij}}$ of M: they have a join γQ_k in M. Since, for each $1 \leqslant i, j \leqslant 3$, we have $\phi(g_i) Q_i \leqslant \phi(g_{ij}) Q_{k_{ij}} \leqslant \phi(g_{ij}) Q_{2n-k_{ij}} \leqslant \phi(g_j) Q_{2n-k_j}$, we deduce that γQ_k is inferior to the meet $\sigma(\gamma) Q_{2n-k}$ of the elements $\phi(g_{ij}) Q_{2n-k_{ij}}$. We deduce that $k \leqslant n$. Also, as in the previous paragraph, we deduce that we can choose $\gamma \in A(A_{2n-1})$ such that $\gamma = \phi(c)$ for some $c \in A(B_n)$. Hence $c P_k$ is a common upper bound for the elements $(g_i P_{k_i})_{1 \leqslant i \leqslant 3}$.

Since L is graded with finite rank, we deduce that any family of pairwise upper bounded elements of L have a join. □

We believe that a similar statement holds for the Artin group of spherical type D_n, but we have no embedding into an Artin group of type A in order to use a similar proof.

7 Application to Euclidean buildings and the Deligne complex of other Euclidean types

In this section, we will prove that we can deduce from Theorem 4.4 an injective metric on Euclidean buildings and Deligne complexes of Euclidean types other than $\tilde{A}_n$.

More precisely, let us consider a simplicial complex X that is

- either a Euclidean building of type $\widetilde{B}_n$, $\widetilde{C}_n$ or $\widetilde{D}_n$,
- or the Deligne complex of Euclidean type $\widetilde{C}_n$.

Note that the Coxeter groups of types $\widetilde{B}_n$ and $\widetilde{D}_n$ may be considered as subgroups of the Coxeter group of Euclidean type $\widetilde{C}_n$, spanned by reflections with respect to subarrangements of hyperplanes; see below for more details. As a consequence, if X is a Euclidean building of type $\widetilde{B}_n$ or $\widetilde{D}_n$, we will consider it as a (possibly nonthick) Euclidean building of type $\widetilde{C}_n$.

Let $W(\widetilde{C}_n)$ denote the Euclidean Coxeter group of type $\widetilde{C}_n$. Its Coxeter complex $\Sigma(\widetilde{C}_n)$ identifies with $\mathbb{R}^n$, and reflections of $W(\widetilde{C}_n)$ correspond to reflections with respect to hyperplanes

$$\{x_i = k \mid 1 \leqslant i \leqslant n, k \in \mathbb{Z}\} \quad \text{and} \quad \{x_i \pm x_j = 2k \mid 1 \leqslant i \neq j \leqslant n, k \in \mathbb{Z}\}.$$

Note that we can see the following subarrangement has type $\widetilde{B}_n$:

$$\{x_i = 2k \mid 1 \leqslant i \leqslant n, k \in \mathbb{Z}\} \quad \text{and} \quad \{x_i \pm x_j = 2k \mid 1 \leqslant i \neq j \leqslant n, k \in \mathbb{Z}\}.$$

Also the following subarrangement has type $\widetilde{D}_n$:

$$\{x_i \pm x_j = 2k \mid 1 \leqslant i \neq j \leqslant n, k \in \mathbb{Z}\}.$$

This justifies that, in the case of Euclidean buildings, we may assume that X is a (possibly nonthick) Euclidean building of type $\widetilde{C}_n$.

The vertex set of $\Sigma(\widetilde{C}_n)$ identifies with $\mathbb{Z}^n$, and a strict fundamental domain of the action of $W(\widetilde{C}_n)$ on $\Sigma(\widetilde{C}_n)$ is given by the standard orthoscheme simplex σ with vertices

$$v_0 = (0, \dots, 0), \quad v_1 = (1, 0, \dots, 0), \quad \dots, \quad v_n = (1, 1, \dots, 1).$$

Hence $v_0, v_1, \dots, v_n$ are representatives of the $n+1$ orbits of vertices of $\Sigma(\widetilde{C}_n)$ under the action of $W(\widetilde{C}_n)$. This enables us to define a type function τ on vertices of $\Sigma(\widetilde{C}_n)$:

$$\tau\colon \Sigma(\widetilde{C}_n)^{(0)} \to \{0, 1, \dots, n\}, \quad g \cdot v_i \mapsto i.$$

More generally, we can define a partial order on vertices of $\Sigma(\widetilde{C}_n)$: say that $v < v'$ if v, v' are adjacent vertices and $\tau(v) < \tau(v')$. We can therefore also view $\Sigma(\widetilde{C}_n)$ as the geometric realization of the poset of its vertices.

We will now see how to extend the type function and the partial order on vertices of X.

Proposition 7.1 *Assume that X is a Euclidean building of type $\widetilde{C}_n$ or the Deligne complex of Euclidean type $\widetilde{C}_n$. There exists a type function $\tau\colon X^{(0)} \to \{0, 1, \dots, n\}$ such that adjacent vertices of X have different types. Moreover, let us define a partial order on vertices of X by setting $v < v'$ if v and v' are adjacent in X and $\tau(v) < \tau(v')$. Then X is the geometric realization of the poset of its vertices.*

Proof Let us first consider the case where X is a Euclidean building of type $\widetilde{C}_n$. Given any vertex v of X, consider any apartment $A \subset X$ containing v, and let us define $\tau(v)$ as defined with respect to the apartment $A \simeq \Sigma(\widetilde{C}_n)$. Since two apartments containing v differ by an element of the Weyl group $W(\widetilde{C}_n)$, which preserves the type, we deduce that τ is well-defined on $X^{(0)}$. Similarly, given any two adjacent vertices v, v' in X, say that $v < v'$ if $\tau(v) < \tau(v')$.

We will see that this relation is actually transitive on vertices of X: assume that three vertices v_1, v_2, v_3 of X satisfy $v_1 < v_2$ and $v_2 < v_3$. Then the link of v_2 is isomorphic to the join of two spherical buildings of types $B_{\tau(v_2)}$ and $B_{n-\tau(v_2)}$. Hence we see that v_1 is adjacent to v_3 in X, so $v_1 < v_3$.

Let us now consider the case where X is the Deligne complex of Euclidean type $\widetilde{C}_n$. Let $s_0, s_1, \dots, s_{n-1}, s_n$ denote the standard generators of the Artin group $A(\widetilde{C}_n)$. For each $0 \leqslant i \leqslant n$, consider the maximal spherical-type standard parabolic subgroup

$$P_i = \langle s_0, s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n \rangle \simeq A(B_i) \times A(B_{n-i}).$$

For each vertex gP_i in X, where $g \in A(\widetilde{C}_n)$ and $0 \leqslant i \leqslant n$, let us define $\tau(gP_i) = i$. Given any two vertices gP_i, hP_j of X, say that $gP_i < hP_j$ if they are adjacent in X (ie $gP_i \cap hP_j \neq \varnothing$) and $i < j$. As in Lemma 6.5, one checks that it is a partial order. □

We will endow X with the piecewise ℓ^∞ orthoscheme metric d given by the geometric realization of its poset of vertices. We will now describe the local structure of X at any vertex, starting with a general statement about orthoscheme complexes of posets.

Lemma 7.2 *Let L denote a graded poset, with an element $v \in L$ comparable to every element of L. Let $L^+ = \{w \in L \mid w \geqslant v\}$ and $L^- = \{w \in L \mid w \leqslant v\}$. Then the ℓ^∞ orthoscheme realization of L is locally isometric at v to the ℓ^∞ product of the ℓ^∞ orthoscheme realizations of L^+ and L^-.*

Proof Since the ℓ^∞ orthoscheme realization of L is obtained as a union of orthoschemes, it is sufficient to prove the result when L is a chain.

In other words, consider the standard ℓ^∞ n–orthoscheme $C_n = \{x \in \mathbb{R}^n \mid 1 \leqslant x_1 \leqslant x_2 \geqslant \cdots \geqslant x_n \geqslant 0\}$, with vertices $v_i = (1,\dots,1,0,\dots,0)$ with i ones and $n-i$ zeros for $0 \leqslant i \leqslant n$. We fix a particular vertex $v = v_j$ for some $0 \leqslant j \leqslant n$, and we want to describe locally C around v. The orthoscheme C_n is locally isometric at v to the space

$$T = \{x \in \mathbb{R}^n \mid 1 \geqslant x_1 \geqslant x_2 \geqslant \cdots \geqslant x_j \text{ and } x_{j+1} \geqslant x_{j+2} \geqslant \cdots \geqslant x_n \geqslant 0\},$$

which is isometric to the ℓ^∞ product of $T^- = \{(x_1,\dots,x_j) \in \mathbb{R}^j \mid 1 \geqslant x_1 \geqslant x_2 \geqslant \cdots \geqslant x_j\}$ and $T^+ = \{(x_{j+1},\dots,x_n) \in \mathbb{R}^{n-j} \mid x_{j+1} \geqslant x_{j+2} \geqslant \cdots \geqslant x_n \geqslant 0\}$.

The space T^+ is locally isometric at v to the vertex $(1,1,\dots,1)$ in the j–orthoscheme C_j, and T^- is locally isometric at v to the vertex $(0,0,\dots,0)$ in the $(n-j)$–orthoscheme C_{n-j}.

In conclusion, if L is a chain $v_0 < v_1 < \cdots < v = v_j < v_n$, then the geometric realization of $|L|$ is locally isometric at v to the product of the geometric realizations of the chains $L^- = (v_0 < v_1 < \cdots < v_j = v)$ and $L^+ = (v = v_j < v_{j+1} < \cdots < v_n)$. ◻

Proposition 7.3 *Assume that X is a Euclidean building of type $\widetilde{C}_n$ or the Deligne complex of Euclidean type $\widetilde{C}_n$. Fix a vertex $v \in X$ of type $\tau(v) = i \in \{0,1,\dots,n\}$.*

There exist posets $L_{1,i}$ (resp. $L'_{0,n-i}$), which are either the poset of vertices of a spherical building of type B_i (resp. B_{n-i}) or the poset of maximal proper parabolic subgroups of the Artin group of spherical type B_i (resp. B_{n-i}). Let us define the posets $L_i = \{1\} \cup L_{1,i}$ and $L'_{n-i} = L'_{0,n-i} \cup \{0\}$, and let us consider the geometric realizations $|L_i|$ and $|L'_{n-i}|$, endowed with the piecewise ℓ^∞ orthoscheme metrics.

Then X is locally isometric at v to the ℓ^∞ direct product $|L_i| \times |L'_{n-i}|$.

Proof The space X is locally isometric at v to the ℓ^∞ orthoscheme realization of the poset L_v of vertices of X comparable to v. The poset L_v is the disjoint union of $L_{1,i} = \{w \in L_v \mid w < v\} \sqcup \{v\} \sqcup L'_{0,n-i} = \{w \in L_v \mid w > v\}$, such that $L_{1,i} < \{v\} < L'_{0,n-i}$. We may identify the poset $L_i = L_{1,i} \cup \{1\}$ with the interval $L_{1,i} \cup \{v\}$ in L_v. Similarly, we may identify the poset $L'_{n-i} = L'_{0,n-i} \cup \{0\}$ with the interval $L'_{0,n-i} \cup \{v\}$ in L_v. According to Lemma 7.2, the orthoscheme realization $|L_v|$ of L_v is locally isometric at v to the direct ℓ^∞ product $|L_i| \times |L'_{n-i}|$ of the orthoscheme realizations of L_i and L'_{n-i}.

In case X is a Euclidean building to type $\widetilde{C}_n$, note that $L_{1,i}$ is the vertex set of a spherical building of type B_i (with reversed order), and that $L'_{0,n-i}$ is the vertex set of a spherical building of type B_{n-i}.

In the case X is the Deligne complex of type $\widetilde{C}_n$, note that $L_{1,i}$ is the poset of proper parabolic subgroups of the Artin group of type B_i (with reversed order), and that $L'_{0,n-i}$ is the poset of proper parabolic subgroups of the Artin group of type B_{n-i}. □

We can now prove the following.

Theorem 7.4 *Any Euclidean building of type $\widetilde{B}_n$, $\widetilde{C}_n$ or $\widetilde{D}_n$, or the Deligne complex of type $\widetilde{C}_n$, endowed with the piecewise ℓ^∞ metric d, is injective.*

Proof According to Proposition 7.3, the space X is locally isometric to a product of orthoscheme complexes of the type $|L|$, where $L = L_0 \cup \{0\}$ is a poset, and L_0 is either the poset of vertices of a spherical building of type B_i or the poset of maximal proper parabolic subgroups of the Artin group of spherical type B_i for some $0 \leqslant i \leqslant n$.

According to Propositions 6.4 and 6.6, we know that in each case L is a graded meet-semilattice with minimum 0 of rank i such that any upper bounded subset has a join. According to Theorem 6.3, we deduce that $|L|$ is injective.

We know X is uniformly locally injective. According to Theorem 1.14, we deduce that X is injective. □

We can also deduce that the thickening is a Helly graph.

Corollary 7.5 *The thickening of the vertex set of any Euclidean building of type $\widetilde{B}_n$, $\widetilde{C}_n$ or $\widetilde{D}_n$, or the extended Deligne complex of type $\widetilde{C}_n$, is a Helly graph.*

Proof Let X denote either a Euclidean building of type $\widetilde{B}_n$, $\widetilde{C}_n$ or $\widetilde{D}_n$, or the extended Deligne complex of type $\widetilde{C}_n$. We will see that X satisfies the assumptions of Theorem 1.15.

Since simplices of X have a well-defined order, we see that X is a simplicial complex with ordered simplices. For each maximal simplex σ of X, the minimal and maximal vertices of σ form a maximal edge in X. Therefore, X has maximal edges as in Definition 1.9. According to Theorem 1.15, we deduce that the thickening of X is a Helly graph. □

One can also deduce another proof of the result of Huang and Osajda that FC-type Artin groups are Helly (see [Huang and Osajda 2021, Theorem 5.8]), with moreover explicit Helly and injective models.

Theorem 7.6 *Let $A = A(\Gamma)$ denote an FC-type Artin group, with standard generating set S, and let $\leqslant_L$ denote the standard prefix order on A with respect to S.*

- *Consider the graph Y, with vertex set A, with an edge between $g, h \in A$ if there exists $a \in A$ and a spherical subset $T \subset S$ such that $a \leqslant_L g, h \leqslant_L a\Delta_T$, where Δ_T denotes the standard Garside element of $A(T)$. Then Y is a Helly graph.*
- *Consider the simplicial complex X, with vertex set A, with a k–simplex for each chain $g_0 <_L g_1 <_L \cdots <_L g_k \leqslant_L g_0\Delta_T$ for some spherical subset $T \subset S$. Endow X with the standard ℓ^∞ orthoscheme metric. Then X is an injective metric space.*

Proof Let us start by proving that X is injective. Note that X can be described as the union of (possibly not connected) complexes X_T, for $T \subset S$ spherical, by restricting the spherical subsets to be contained in T. So X is locally isometric to the geometric realization of the poset $L = \bigcup_{T \subset S \text{ spherical}} [1, \Delta_T]$. This poset is graded, with minimum, with bounded rank, and is a meet-semilattice.

Moreover, the flag condition from the FC-type Artin group A translates into the flag condition for the poset L. Indeed, consider x_1, x_2, x_3 which are pairwise upper bounded. Let $T_1, T_2, T_3 \subset S$ denote the supports of x_1, x_2, x_3 respectively. By assumption, for each $i \neq j$, the subset $T_i \cup T_j$ is spherical. As a consequence, the subset $T = T_1 \cup T_2 \cup T_3$ is a complete subset of S. According to the FC-type condition, we deduce that T is spherical. Hence $x_1, x_2, x_3 \leqslant_L \Delta_T$. So L is a flag poset.

According to Theorem 6.3, we deduce that $|L|$ is injective. So X is uniformly locally injective and according to Theorem 1.14, we deduce that X is injective.

We now turn to the proof that Y is a Helly graph, as in the proof of Corollary 7.5. Since simplices of X have a well-defined order, we see that X is a simplicial complex with ordered simplices. For each maximal simplex σ of X, the minimal and maximal vertices of σ form a maximal edge in X. Therefore X has maximal edges. According to Theorem 1.15, we deduce that the thickening Y of X is a Helly graph. □

8 Bicombings on Deligne complexes in types $\tilde{A}_n$ and $\widetilde{C}_n$

We now see that, in the Deligne complex of spherical types A_n, B_n and Euclidean types $\tilde{A}_n$, $\widetilde{C}_n$, we may find a convex bicombing.

Theorem 8.1 *Let X denote the Deligne complex of the Artin group A of spherical type A_n, B_n or Euclidean type $\tilde{A}_n$, $\widetilde{C}_n$. There is a metric d_X on X that admits a convex, consistent, reversible geodesic bicombing σ. Moreover, d_X and σ are invariant under A.*

Proof The statement for types A_n and $\tilde{A}_n$ is Theorem 4.6.

Assume that X is the Deligne complex of type B_n. We have seen in the proof of Proposition 6.6 that X may be realized as the fixed-point subspace of the Deligne complex Y of type A_{2n-1} for an involution s. Note that, since s is order-reversing, s induces an isometry of Y. According to Theorem 4.6, there exists a unique reversible, convex, consistent, geodesic bicombing σ_Y on Y for a metric d_Y. Since σ_Y is unique, we deduce that s preserves σ_Y, and that the fixed-point subspace X is σ–stable. Let d_X denote the restriction of d_Y to X, and let σ_X denote the restriction of σ_Y to X. We deduce that σ_X is a convex, consistent, reversible geodesic bicombing on (X, d_X). Moreover, d_X and σ_X are invariant under A.

Assume that X is the Deligne complex of type $\widetilde{C}_n$. According to [Digne 2012, Theorem 5.2] (it is also a consequence of Corollary 8.5), we see that X may be realized as the fixed-point subspace of the Deligne complex Y of type $\tilde{A}_{2n-1}$ for an involution s. Following the same arguments as in the type B_n case, we conclude that there exists a metric d_X on X and a convex, consistent, reversible geodesic bicombing σ_X on (X, d_X) that are both invariant under A. □

We will describe several consequences we can derive from the fact that the Deligne complex has a consistent convex bicombing, which were usually known when the Deligne complex had a CAT(0) metric.

Corollary 8.2 (Okonek) *The Deligne complex X of Euclidean type $\tilde{A}_n$ or $\widetilde{C}_n$ is contractible. In particular, the $K(\pi, 1)$ conjecture holds in these cases.*

Proof Metric spaces with a convex bicombing are contractible. □

The proof of Morris-Wright [2021] for the intersection of parabolic subgroups in FC-type Artin groups adapts directly to our situation. It relies mainly on the result by Cumplido et al. [2019] that in a spherical-type Artin group, the intersection of parabolic subgroups is a parabolic subgroup.

Corollary 8.3 *Let A denote the Artin group of Euclidean type $\tilde{A}_n$ or $\widetilde{C}_n$. The intersection of any family of parabolic subgroups of A is a parabolic subgroup.*

Proof Note that any proper parabolic subgroup of an Artin group of Euclidean type has spherical type. Since the proof of [Morris-Wright 2021, Theorem 3.1] uses only the existence of an A–equivariant consistent geodesic bicombing on X, it adapts to these cases. □

The results by Godelle [2007] describing centralizers and normalizers of parabolic subgroups also adapt to our case.

Corollary 8.4 *Let A denote the Artin group of Euclidean type $\tilde{A}_n$ or $\widetilde{C}_n$. Then A satisfies properties $(\star)$, $(\star\star)$ and $(\star\star\star)$ from [Godelle 2007]; notably, for any subset $X \subset S$, we have*

$$\mathrm{Com}_A(A_X) = N_A(A_X) = A_X \cdot QZ_A(X),$$

where the quasicentralizer of X is $QZ_A(X) = \{g \in A \mid g \cdot X = X\}$.

Note that Godelle's property $(\star\star)$ stating that a parabolic subgroup P is contained in a parabolic subgroup Q, then P is a parabolic subgroup of Q has been proved by Blufstein and Paris [2023].

Proof The proof of [Godelle 2007, Theorem 3.1] uses only the following properties of the Deligne complex, which hold for the Deligne complex X:

- There exists an A–equivariant consistent geodesic bicombing σ_X on X.
- Closed cells of X are σ_X–stable.

These assumptions are satisfied. □

Note that Cumplido uses these properties of parabolic subgroups to solve the conjugacy stability problem for parabolic subgroups; see [Cumplido 2022, Theorem 14].

The results by Crisp [2000] about symmetrical subgroups of Artin groups also extend to our case.

Corollary 8.5 *Let A denote the Artin group of Euclidean type $\tilde{A}_n$ or $\widetilde{C}_n$. For any group G of symmetries of the Artin system, the fixed-point subgroup A^G is isomorphic to an Artin group.*

For the explicit description of the Artin group A^G, we refer the reader to [Crisp 2000].

Proof The proof of [Crisp 2000, Theorem 23] uses only the following properties of the Deligne complex X:

- There exists an A–equivariant consistent geodesic bicombing σ_X on X.
- Subsets of X which are σ_X–stable are contractible.

Let us check the last condition: if $C \subset X$ is a σ_X–stable subset, then σ_X restricts to a convex bicombing on C, which implies that C is contractible. □

One particular case of interest is when A is of Euclidean type $\tilde{A}_{2n-1}$ for some $n \geqslant 1$, and the group G of symmetries of the Artin system (the $2n$–cycle) is generated by a symmetry of the cycle defining A. We recover the result from [Digne 2012] that the fixed-point subgroup is isomorphic to the Artin group of type $\widetilde{C}_n$.

For any divisor $1 < k < n$ of n, one may also consider the group G of symmetries of the Artin system of type $\tilde{A}_{n-1}$ (the n–cycle) generated by a rotation of k. The fixed-point subgroup is then isomorphic to the Artin group of Euclidean type $\tilde{A}_{k-1}$. When we see $A(\tilde{A}_{n-1})$ as the group of braids with n strands on the annulus, we may think of the fixed-point subgroup as the subgroup of braids which are invariant by a rotation of the annulus of angle $\frac{2\pi k}{n}$.

References

[Alonso and Bridson 1995] **J M Alonso**, **M R Bridson**, *Semihyperbolic groups*, Proc. Lond. Math. Soc. 70 (1995) 56–114 MR Zbl

[Basso 2024] **G Basso**, *Extending and improving conical bicombings*, Enseign. Math. 70 (2024) 165–196 MR Zbl

[Bessis 2006] **D Bessis**, *A dual braid monoid for the free group*, J. Algebra 302 (2006) 55–69 MR Zbl

[Birman et al. 1998] **J Birman**, **K H Ko**, **S J Lee**, *A new approach to the word and conjugacy problems in the braid groups*, Adv. Math. 139 (1998) 322–353 MR Zbl

[Blufstein and Paris 2023] **M A Blufstein**, **L Paris**, *Parabolic subgroups inside parabolic subgroups of Artin groups*, Proc. Amer. Math. Soc. 151 (2023) 1519–1526 MR Zbl

[Bourbaki 2002] **N Bourbaki**, *Lie groups and Lie algebras, chapters IV–VI*, Springer (2002) MR Zbl

[Bowditch 2020] **B H Bowditch**, *Median and injective metric spaces*, Math. Proc. Cambridge Philos. Soc. 168 (2020) 43–55 MR Zbl

[Brady and McCammond 2010] **T Brady**, **J McCammond**, *Braids, posets and orthoschemes*, Algebr. Geom. Topol. 10 (2010) 2277–2314 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

[Bruhat and Tits 1972] **F Bruhat**, **J Tits**, *Groupes réductifs sur un corps local*, Inst. Hautes Études Sci. Publ. Math. 41 (1972) 5–251 MR Zbl

[Chalopin et al. 2020a] **J Chalopin**, **V Chepoi**, **A Genevois**, **H Hirai**, **D Osajda**, *Helly groups*, preprint (2020) arXiv 2002.06895

[Chalopin et al. 2020b] **J Chalopin**, **V Chepoi**, **H Hirai**, **D Osajda**, *Weakly modular graphs and nonpositive curvature*, Mem. Amer. Math. Soc. 1309, Amer. Math. Soc., Providence, RI (2020) MR

[Charney 1995] **R Charney**, *Geodesic automation and growth functions for Artin groups of finite type*, Math. Ann. 301 (1995) 307–324 MR Zbl

[Charney 2004] **R Charney**, *The Deligne complex for the four-strand braid group*, Trans. Amer. Math. Soc. 356 (2004) 3881–3897 MR Zbl

[Charney 2008] **R Charney**, *Problems related to Artin groups*, preprint, Amer. Inst. Math. (2008) Available at https://people.brandeis.edu/~charney/papers/Artin_probs.pdf

[Charney and Davis 1995a] **R Charney**, **M W Davis**, *Finite $K(\pi,1)$s for Artin groups*, from "Prospects in topology", Ann. of Math. Stud. 138, Princeton Univ. Press (1995) 110–124 MR Zbl

[Charney and Davis 1995b] **R Charney**, **M W Davis**, *The $K(\pi,1)$–problem for hyperplane complements associated to infinite reflection groups*, J. Amer. Math. Soc. 8 (1995) 597–627 MR Zbl

[Crisp 2000] **J Crisp**, *Symmetrical subgroups of Artin groups*, Adv. Math. 152 (2000) 159–177 MR Zbl

[Cumplido 2022] **M Cumplido**, *The conjugacy stability problem for parabolic subgroups in Artin groups*, Mediterr. J. Math. 19 (2022) art. id. 237 MR Zbl

[Cumplido et al. 2019] **M Cumplido**, **V Gebhardt**, **J González-Meneses**, **B Wiest**, *On parabolic subgroups of Artin–Tits groups of spherical type*, Adv. Math. 352 (2019) 572–610 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

[Descombes and Lang 2015] **D Descombes**, **U Lang**, *Convex geodesic bicombings and hyperbolicity*, Geom. Dedicata 177 (2015) 367–384 MR Zbl

[Descombes and Lang 2016] **D Descombes**, **U Lang**, *Flats in spaces with convex geodesic bicombings*, Anal. Geom. Metr. Spaces 4 (2016) 68–84 MR Zbl

[Digne 2012] **F Digne**, *A Garside presentation for Artin–Tits groups of type $\tilde{C}_n$*, Ann. Inst. Fourier (Grenoble) 62 (2012) 641–666 MR Zbl

[Dougherty et al. 2020] **M Dougherty**, **J McCammond**, **S Witzel**, *Boundary braids*, Algebr. Geom. Topol. 20 (2020) 3505–3560 MR Zbl

[Dress 1984] **A W M Dress**, *Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces*, Adv. Math. 53 (1984) 321–402 MR Zbl

[Godelle 2007] **E Godelle**, *Artin–Tits groups with* CAT(0) *Deligne complex*, J. Pure Appl. Algebra 208 (2007) 39–52 MR Zbl

[Godelle and Paris 2012] **E Godelle**, **L Paris**, *Basic questions on Artin–Tits groups*, from "Configuration spaces", CRM Series 14, Ed. Norm., Pisa (2012) 299–311 MR Zbl

[Haettel 2022a] **T Haettel**, *Injective metrics on buildings and symmetric spaces*, Bull. Lond. Math. Soc. 54 (2022) 2297–2313 MR Zbl

[Haettel 2022b] **T Haettel**, *A link condition for simplicial complexes, and CUB spaces*, preprint (2022) arXiv 2211.07857

[Haettel and Huang 2024] **T Haettel**, **J Huang**, *Lattices, Garside structures and weakly modular graphs*, J. Algebra 656 (2024) 226–258 MR Zbl

[Haettel and Osajda 2021] **T Haettel**, **D Osajda**, *Locally elliptic actions*, *torsion groups*, *and nonpositively curved spaces*, preprint (2021) arXiv 2110.12431

[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

[Haettel et al. 2023] **T Haettel**, **N Hoda**, **H Petyt**, *Coarse injectivity, hierarchical hyperbolicity and semihyperbolicity*, Geom. Topol. 27 (2023) 1587–1633 MR Zbl

[Hoda 2023] **N Hoda**, *Crystallographic Helly groups*, Bull. Lond. Math. Soc. 55 (2023) 2991–3011 MR Zbl

[Huang and Osajda 2021] **J Huang**, **D Osajda**, *Helly meets Garside and Artin*, Invent. Math. 225 (2021) 395–426 MR Zbl

[Jeong 2023] **S G Jeong**, *The seven-strand braid group is* CAT(0), Manuscripta Math. 171 (2023) 563–581 MR Zbl

[Kasprowski and Rüping 2017] **D Kasprowski**, **H Rüping**, *The Farrell–Jones conjecture for hyperbolic and* CAT(0)*–groups revisited*, J. Topol. Anal. 9 (2017) 551–569 MR

[Lang 2013] **U Lang**, *Injective hulls of certain discrete metric spaces and groups*, J. Topol. Anal. 5 (2013) 297–331 MR Zbl

[Leary and Minasyan 2021] **I J Leary**, **A Minasyan**, *Commensurating HNN extensions: nonpositive curvature and biautomaticity*, Geom. Topol. 25 (2021) 1819–1860 MR

[van der Lek 1983] **H van der Lek**, *The homotopy type of complex hyperplane complements*, PhD thesis, Radboud University (1983) Available at `https://repository.ubn.ru.nl/handle/2066/148301`

[McCammond 2006] **J McCammond**, *An introduction to Garside structures*, preprint (2006) Available at `https://web.math.ucsb.edu/~mccammon/current/artin/0intro/intro-garside.pdf`

[McCammond 2017] **J McCammond**, *The mysterious geometry of Artin groups*, Winter Braids Lect. Notes 4 (2017) art. id. 1 MR Zbl

[McCammond and Sulway 2017] **J McCammond**, **R Sulway**, *Artin groups of Euclidean type*, Invent. Math. 210 (2017) 231–282 MR Zbl

[Michel 1999] **J Michel**, *A note on words in braid monoids*, J. Algebra 215 (1999) 366–377 MR Zbl

[Miesch 2014] **B Miesch**, *Injective metrics on cube complexes*, preprint (2014) arXiv 1411.7234

[Miesch 2017] **B Miesch**, *The Cartan–Hadamard theorem for metric spaces with local geodesic bicombings*, Enseign. Math. 63 (2017) 233–247 MR Zbl

[Möller et al. 2024] **P Möller**, **L Paris**, **O Varghese**, *On parabolic subgroups of Artin groups*, Israel J. Math. 261 (2024) 809–840 MR Zbl

[Morris-Wright 2021] **R Morris-Wright**, *Parabolic subgroups in FC-type Artin groups*, J. Pure Appl. Algebra 225 (2021) art. id. 106468 MR Zbl

[Nachbin 1950] **L Nachbin**, *A theorem of the Hahn–Banach type for linear transformations*, Trans. Amer. Math. Soc. 68 (1950) 28–46 MR Zbl

[Okonek 1979] **C Okonek**, *Das $K(\pi, 1)$–Problem für die affinen Wurzelsysteme vom Typ A_n,C_n*, Math. Z. 168 (1979) 143–148 MR

[Osajda and Valiunas 2024] **D Osajda**, **M Valiunas**, *Helly groups, coarsely Helly groups, and relative hyperbolicity*, Trans. Amer. Math. Soc. 377 (2024) 1505–1542 MR Zbl

[Paolini and Salvetti 2021] **G Paolini**, **M Salvetti**, *Proof of the $K(\pi, 1)$ conjecture for affine Artin groups*, Invent. Math. 224 (2021) 487–572 MR Zbl

[Świątkowski 2006] **J Świątkowski**, *Regular path systems and (bi)automatic groups*, Geom. Dedicata 118 (2006) 23–48 MR Zbl

[Tits 1966] **J Tits**, *Normalisateurs de tores, I: Groupes de Coxeter étendus*, J. Algebra 4 (1966) 96–116 MR Zbl

[Tits 1974] **J Tits**, *Buildings of spherical type and finite BN-pairs*, Lecture Notes in Math. 386, Springer (1974) MR Zbl

Institut Montpelliérain Alexander Grothendieck, Université de Montpellier, CNRS
Montpellier, France

IRL 3457, CRM-CNRS, Université de Montréal
Montréal, Canada

thomas.haettel@umontpellier.fr

Received: 15 February 2023 Revised: 12 September 2023

Algebraic & Geometric Topology **24**:7 (2024) 4061–4084
DOI: 10.2140/agt.2024.24.4061
Published: 9 December 2024

The real-oriented cohomology of infinite stunted projective spaces

WILLIAM BALDERRAMA

Let $E\mathbb{R}$ be an even-periodic real Landweber exact C_2–spectrum, and ER be its spectrum of fixed points. We compute the ER–cohomology of the infinite stunted real projective spectra P_j. These cohomology groups combine to form the $RO(C_2)$–graded coefficient ring of the C_2–spectrum

$$b(ER) = F(EC_{2+}, i_* ER),$$

which we show is related to $E\mathbb{R}$ by a cofiber sequence $\Sigma^\sigma b(ER) \to b(ER) \to E\mathbb{R}$. We illustrate our description of $\pi_\star b(ER)$ with the computation of some ER–based Mahowald invariants.

55N20, 55N22, 55N91, 55Q51

1 Introduction

The spectrum MU of complex cobordism plays a central role in both our conceptual and computational understanding of stable homotopy theory. Landweber [1968] introduced what is now known as the C_2–equivariant spectrum $M\mathbb{R}$ of real bordism, with underlying spectrum MU and fixed points $MR = MU^{\mathrm{h}C_2}$ the homotopy fixed points for the action of C_2 on MU by complex conjugation. Work of Araki [1979], Hu and Kriz [2001], and others, has shown that essentially all of the theory of complex-oriented homotopy theory may be carried out in the C_2–equivariant setting with $M\mathbb{R}$ in place of MU, leading to the rich subject of real-oriented homotopy theory. This subject has seen extensive study over the past two decades, with a notable increase in interest following the use of $M\mathbb{R}$ by Hill, Hopkins and Ravenel [Hill et al. 2016] to resolve the Kervaire invariant one problem.

There are real analogues of most familiar complex-oriented cohomology theories. An important family of examples is given by the real Johnson–Wilson theories $E\mathbb{R}(n)$, refining the usual Johnson–Wilson theories $E(n)$. These theories are Landweber flat over $M\mathbb{R}$, in the sense that they are $M\mathbb{R}$–modules and satisfy

$$E\mathbb{R}(n)_\star X \cong E\mathbb{R}(n)_\star \otimes_{M\mathbb{R}_\star} M\mathbb{R}_\star X$$

for any C_2–spectrum X. The fixed points $ER(n) = E\mathbb{R}(n)^{C_2} = E(n)^{\mathrm{h}C_2}$ are nonequivariant cohomology theories that are interesting in their own right; for example, $ER(1) \simeq KO_{(2)}$, and $ER(2)$ is a variant of $\mathrm{TMF}_0(3)_{(2)}$. One may regard the descent from $E(n)$ to $ER(n)$ as encoding a portion of the $E(n)$–based Adams–Novikov spectral sequence, and accordingly each $ER(n)$ detects infinite families in $\pi_* S$.

There is in general a tradeoff between the richness of a homology theory and the ease with which it may be computed. Kitchloo, Lorman, and Wilson have carried out extensive computations with $ER(n)$–theory [Kitchloo and Wilson 2007b; 2015; Lorman 2016; Kitchloo et al. 2017; 2018a], and their program has shown that these theories strike a very pleasant balance between richness and computability. Computations of $ER(2)^*\mathbb{R}P^n$ in particular have been applied to the nonimmersion problem for real projective spaces, with computations for $n = 2k$ in [Kitchloo and Wilson 2008a], $n = 16k + 1$ in [Kitchloo and Wilson 2008b], and $n = 16k + 9$ by Banerjee [2010].

This paper contributes to the above story. Let $E\mathbb{R}$ be a real Landweber exact C_2–spectrum in the sense of Hill and Meier [2017, Section 3.2]; we take this to include the assumption that $E\mathbb{R}$ is strongly even. Write E for the underlying spectrum of $E\mathbb{R}$ and $ER = E\mathbb{R}^{C_2} = E^{\mathrm{h}C_2}$ for its fixed points. Suppose moreover that $E\mathbb{R}$ is even-periodic, in the sense that $\pi_{1+\sigma}E\mathbb{R}$ contains a unit. This is equivalent to asking that the $M\mathbb{R}$–orientation of $E\mathbb{R}$ extends to an $MP\mathbb{R}$–orientation, where

$$MP\mathbb{R} \simeq \bigoplus_{n\in\mathbb{Z}} \Sigma^{n(1+\sigma)} M\mathbb{R}$$

is the real analogue of 2–periodic complex cobordism.

The primary goal of this paper is to compute the ER–cohomology of the infinite stunted projective spectra P_j. When $j > 0$, these are the spaces

$$P_j = \mathbb{R}P^\infty / \mathbb{R}P^{j-1};$$

in general, P_j is the Thom spectrum of $j\sigma$, where σ is the sign representation of C_2 regarded as a vector bundle over $BC_2 = \mathbb{R}P^\infty$. The cohomology ER^*P_* is of interest for at least a few reasons: first, it is one long exact sequence away from the groups $ER^*\mathbb{R}P^j$, which have so far only been studied at heights ≤ 2; second, there are C_2–equivariant Hurewicz maps $\pi_{c+w\sigma}S_{C_2} \to ER^{-c}P_w$, which are at least as nontrivial as the nonequivariant Hurewicz maps for ER; third, there is an interesting interplay between the C_2 appearing in $ER \simeq E^{\mathrm{h}C_2}$ and the C_2 appearing in $ER^*(P_w) \simeq ER^*(S^{w\sigma}_{\mathrm{h}C_2})$ which sheds some light on the nature of the C_2–spectrum $E\mathbb{R}$.

We record the basic properties of ER in Section 3. In particular, $\pi_0 ER \cong \pi_0 E$, the torsion in $\pi_* ER$ is supported on a single class $x \in \pi_1 ER$, there is a cofiber sequence $\Sigma ER \xrightarrow{x} ER \to E$, and the x–Bockstein spectral sequence for ER–cohomology agrees with the homotopy fixed point spectral sequence (HFPSS) from the E_2–page on.

Write $b(ER) = F(EC_{2+}, i_* ER)$ for the Borel C_2–spectrum on ER with trivial C_2–action. This satisfies $\pi_{c+w\sigma} b(ER) = ER^{-c}P_w$, and we shall compute ER^*P_* using the x–Bockstein spectral sequence

$$\pi_\star b(E)[x] \Rightarrow \pi_\star b(ER).$$

This concludes an investigation we began in [Balderrama 2021]. There, we computed the HFPSS $H^*(C_2; \pi_\star b(KU_2^\wedge)) \Rightarrow \pi_\star b(KO_2^\wedge)$ as a step in our description of the C_2–equivariant $K(1)$–local sphere. At the time, we were able to put the E_2–page into a more general context by computing $H^*(C_2; \pi_\star b(E))$

for more general even-periodic Landweber exact spectra E, but had no information about possible higher differentials. In this paper, we carry out the rest of the computation for real-oriented E. The results are summarized in Section 2 below.

1.1 Remark The reader may observe that by restricting to even-periodic spectra, we have ruled out the real Johnson–Wilson theories $ER(n)$ for $n \geq 2$. However, any real Landweber exact C_2–spectrum $E\mathbb{R}$ is a summand of the even-periodic theory $\bigoplus_{n\in\mathbb{Z}} \Sigma^{n(1+\sigma)} E\mathbb{R}$, so no real information has been lost. A more subtle point is that implicit in the definition of real Landweber exactness is the assumption that $E\mathbb{R}$ is a ring up to homotopy, and it is not known whether $E\mathbb{R}(n)$ always satisfies this. However, the partial multiplicative structure given in [Kitchloo et al. 2018b] is sufficient for our computation to apply to 2–periodic $ER(n)$–theory.

Acknowledgements We thank Hood Chatham for an enlightening conversation highlighting the role of Borel completeness in Theorem 2.1. This work was supported by NSF RTG grant DMS-1839968.

2 Summary

We now describe our results. We start with the following, which serves as the linchpin for our computation of $\pi_\star b(ER)$. Write $\rho \in \pi_{-\sigma} S_{C_2}$ for the Euler class of the sign representation and $\tau^{-2} \in \pi_{2\sigma-2} b(E)$ for the Thom class of $2\sigma = \mathbb{C} \otimes \sigma$. These classes are sometimes denoted by a_σ and u_σ^{-2}, but we will reserve those symbols for $E\mathbb{R}$ and $C_{2+} \otimes i_* ER$. Write $u \in \pi_2 E$ for the chosen unit, and set

$$\xi = \rho\tau^{-2} u \in \pi_\sigma b(E).$$

2.1 Theorem (Section 4) *The class ξ is a permanent cycle in the x–Bockstein spectral sequence, detecting a lift of x. Moreover, there is a cofiber sequence*

$$\Sigma^\sigma b(ER) \xrightarrow{\xi} b(ER) \to E\mathbb{R} \tag{1}$$

of C_2–spectra.

This cofiber sequence is a twisted form of the standard cofiber sequence

$$\Sigma^{-\sigma} b(ER) \xrightarrow{\rho} b(ER) \to C_{2+} \otimes i_* ER. \tag{2}$$

2.2 Example When $E = KU$, one can identify $b(ER) = F(EC_{2+}, KO_{C_2})$ and $E\mathbb{R} = K\mathbb{R}$, and $\xi = \pm\eta_{C_2}$ is the C_2–equivariant Hopf map. In this case, Theorem 2.1 recovers the real Wood cofibering $KO_{C_2}/(\eta_{C_2}) \simeq K\mathbb{R}$ (cf [Guillou et al. 2020, Proposition 10.13]).

To show that ξ is a permanent cycle detecting a lift of x, we first reduce to the universal case $E = MP$, then show that this is the only possibility compatible with norms on $b(MPR)$. Given this, the cofiber sequence of (1) is a mostly formal consequence of (2) and the fact that ξ differs from ρ by a unit in $\pi_\star b(E)$.

We now describe $\pi_\star b(ER)$. We start by fixing some notation for $\pi_\star b(E)$. Write $[2](z) \in E_0[\![z]\!]$ for the 2–series of the formal group law of E, and write $u_n \in E_0$ for the elements corresponding to the usual $v_n \in \pi_{2(2^n-1)}E$ by $u_n = u^{-(2^n-1)}v_n$. We may find series $h_n(z) \in E_0[\![z]\!]$ for $n \geq 0$, of the form $h_n(z) = u_n + O(z)$ and satisfying

$$[2](z) = zh_0(z), \quad h_n(z) \equiv u_n + z^{2^n} h_{n+1}(z) \pmod{u_0, \dots, u_{n-1}}.$$

Note in particular

$$[2](z) \equiv z^{2^n} h_n(z) \pmod{u_0, \dots, u_{n-1}}.$$

We now specialize to $\pi_\star b(E)$. Set

$$z = \rho\xi = \rho^2\tau^{-2}u, \quad h_n = h_n(z), \quad w_n = \rho^{2^{n+1}} h_{n+1} \equiv \tau^{2^{n+1}} u^{-2^n}(h_n - u_n),$$

the last congruence being modulo $(u_0, \dots, u_{n-1})$. We abbreviate $h = h_0$. This is the transfer element in $\pi_0 b(E) = E^0BC_2$, and we have

$$\pi_0 b(E) = \frac{E_0[\![z]\!]}{([2](z))}, \quad \pi_\star b(E) = \frac{E_0[\rho, \tau^{\pm 2}, u^{\pm 1}]^\wedge_\rho}{(\rho \cdot h)};$$

see for instance [Balderrama 2021, Section 2.1].

2.3 Theorem (Section 5) *Define the subring $Z \subset \pi_\star b(E)$ by*

$$Z = E_0(\rho, \xi, \tau^{2^{n+2}l}u^{2^{n+1}k}u_n, \tau^{2^{n+1}(2l+1)}u^{2^{n+1}k}h_n : n \geq 0; k, l \in \mathbb{Z})^\wedge_{(\rho,\xi)} \subset \pi_\star b(E),$$

and let $B \subset Z[x]$ be the ideal generated by the elements

$$\tau^{2^{n+2}l}u^{2^{n+1}k}u_n \cdot x^{2^{n+1}-1}, \quad \tau^{2^{n+1}(2l+1)}u^{2^{n+1}k}h_n \cdot x^{2^{n+1}-1}, \quad \tau^{2^{n+2}l}u^{2^{n+1}k}w_n \cdot x^{2^{n+1}-1}$$

for $n \geq 0$ and $k, l \in \mathbb{Z}$. Then $Z[x]/B$ is the x–adic associated graded of $\pi_\star b(ER)$.

2.4 Remark In integer degrees, $\pi_* b(ER)$ is very simply described:

$$\pi_* b(ER) \cong ER_*[\![z]\!]/([2](z));$$

see Corollary 4.3. This does not require the full computation of $\pi_\star b(ER)$, and follows as soon as one knows that ξ is a permanent cycle. In particular, $\pi_0 b(ER) \cong E^0BC_2$. To get a feeling for $\pi_\star b(ER)$ outside integer degrees, the reader may wish to peruse Tables 1 and 2, described in Remark 5.5, which list $\pi_0 b(ER)$–module generators for the groups $\pi_{c+w\sigma} b(ER)$ in a range.

2.5 Remark Implicit in Theorem 2.3 is the fact that $\tau^{2^{n+2}l}u^{2^{n+1}k}w_n \in Z$ for $n \geq 0$ and $k, l \in \mathbb{Z}$. In particular,

$$\begin{aligned}
\tau^{2^{n+2}(2l+1)}u^{2^{n+2}k}w_n &= \rho^{2^{n+1}} \cdot \tau^{2^{n+2}(2l+1)}u^{2^{n+2}k}h_{n+1}, \\
\tau^{2^{n+3}l}u^{2^{n+1}(2k+1)}w_n &= \xi^{2^{n+1}} \cdot \tau^{2^{n+2}(2l+1)}u^{2^{n+2}k}h_{n+1}, \\
\tau^{2^{n+3}l}u^{2^{n+2}k}w_n &\equiv \rho^{2^{n+1}} \cdot \tau^{2^{n+3}l}u^{2^{n+2}k}u_{n+1} + \xi^{2^{n+1}} \cdot \tau^{2^{n+3}l}u^{2^{n+2}k}w_{n+1}, \\
\tau^{2^{n+2}(2l-1)}u^{2^{n+1}(2k+1)}w_n &\equiv \xi^{2^{n+1}} \cdot \tau^{2^{n+3}l}u^{2^{n+2}k}u_{n+1} + \rho^{2^{n+1}} \cdot \tau^{2^{n+3}(l-1)}u^{2^{n+2}(k+1)}w_{n+1},
\end{aligned}$$

where the last two formulas hold mod $u_0, \dots, u_n$.

The ring $Z[x]/B$ is the E_∞–page of the x–Bockstein spectral sequence for $\pi_\star b(ER)$, obtained after running differentials which are generated by

$$d_{2^{n+1}-1}(u^{2^n}) = u_n x^{2^{n+1}-1}, \quad d_{2^{n+1}-1}(\tau^{2^{n+1}}) = -w_n x^{2^{n+1}-1}.$$

The differentials on u^{2^n} appear in the x–Bockstein spectral sequence for $\pi_* ER$, and are consequences of the computation of $\pi_\star M\mathbb{R}$ by Hu and Kriz [2001], as we review in Section 3. The differentials on $\tau^{2^{n+1}}$ are the core of our computation. These differentials turn out to be forced by the permanent cycle $\xi = \rho\tau^{-2}u$, by a Leibniz rule argument based on $d_{2^{n+1}-1}(\xi^{2^n}) = 0$. This argument would not be possible if one tried to compute each $ER^* P_j$ individually, and illustrates the strength of using the C_2–spectrum $b(ER)$ as a tool for packaging information about the cohomology of all stunted projective spectra into one object.

One might also try to understand $\pi_\star b(ER)$ through the ρ–Bockstein or the ξ–Bockstein spectral sequences. Using the cofiber sequences (2) and (1), these are of signature

$$\pi_\star(C_{2+} \otimes i_* ER)[\rho] \Rightarrow \pi_\star b(ER), \quad \pi_\star E\mathbb{R}[\xi] \Rightarrow \pi_\star b(ER).$$

Here, $\pi_\star(C_{2+} \otimes i_* ER) \cong \pi_* ER[u_\sigma^{\pm 1}]$ with $|u_\sigma| = 1-\sigma$, and in degrees $* + w\sigma$ the ρ–Bockstein spectral sequence is exactly the Atiyah–Hirzebruch spectral sequence for $ER^* P_w$ based on the standard cell structure of P_w. By construction, the differentials in these spectral sequences are controlled by the boundary maps

$$\mathrm{tr}(u_\sigma^{-1} \cdot -)\colon \pi_{\star+1-\sigma}(C_{2+} \otimes i_* ER) \to \pi_\star b(ER), \quad \partial\colon \pi_{\star+1+\sigma} E\mathbb{R} \to \pi_\star b(ER)$$

for the cofiber sequences (2) and (1). This first boundary map is exactly the transfer for the C_2–spectrum $b(ER)$. Although we do not know whether it is feasible to compute either the ρ–Bockstein or ξ–Bockstein spectral sequence directly, we can use our computation of $\pi_\star b(ER)$ to deduce the following.

Write $\bar{u} \in \pi_{1+\sigma} E\mathbb{R}$ for the invertible element guaranteed by the $MP\mathbb{R}$–orientation of $E\mathbb{R}$.

2.6 Theorem (Section 6) *The above transfer and boundary maps satisfy*

$$\mathrm{tr}(u_\sigma^{-1} \cdot u_\sigma^{2^n(2k+1)}) = \rho^{2^n-1}\tau^{2^{n+1}k} h_n x^{2^n-1} + O(x^{2^n}),$$
$$\partial(\bar{u}^{2^n(2k+1)}) = \xi^{2^n-1}\tau^{-2^{n+1}k} u^{2^{n+1}k} h_n x^{2^n-1} + O(x^{2^n})$$

for $n \geq 0$ and $k \in \mathbb{Z}$.

The error terms here are necessary as the classes $\tau^{2^{n+1}k} h_n$ and $\tau^{-2^{n+1}k} u^{2^{n+1}k} h_n$ have only been defined mod x. It is amusing to observe that Theorem 2.6 produces elements of arbitrarily high x–adic filtration in the C_2–equivariant Hurewicz image of $b(MPR)$; as far as we know, such families have not yet been constructed in the nonequivariant Hurewicz image of MPR.

Theorem 2.3 does not quite describe the ring $\pi_\star b(ER)$, but only its x–adic associated graded $Z[x]/B$. The latter is a good approximation to the former, particularly when compared to the ρ–adic and ξ–adic associated graded rings, where the classes ρ and ξ appear as simple 2–torsion classes. Still, taking the

x–adic associated graded does kill some information, and it seems to be a subtle problem to completely reconstruct the ring $\pi_\star b(ER)$. Although we shall not completely resolve this, we do discuss where to find hidden ρ and ξ–extensions. The importance of ρ–extensions is clear: as

$$\pi_{c+w\sigma}(b(ER)/(\rho^m)) = ER^{-c}(P_{w-m}^{w-1}),$$

one must understand the action of ρ if one wishes to extract information about the ER–cohomology of finite projective spaces. The importance of ξ–extensions is clear from the perspective of C_2–equivariant homotopy theory: just as important classes in the Hurewicz image of ER are supported on x, important classes in the C_2–equivariant Hurewicz image of $b(ER)$ are supported on ξ, such as the equivariant Hopf fibrations η_{C_2}, ν_{C_2}, and σ_{C_2} detected in $\pi_\star b(ER)$ by $h_1\xi$, $h_2\xi^2x$, and $h_3\xi^4x^3$ respectively, and so the action of ξ gives information about the behavior of these elements. The cofiber sequences (1) and (2) give information about ρ and ξ–extensions, leading to the following.

2.7 Theorem (Section 7) *There are extensions*

$$\rho\cdot\tau^{2(2^{n+1}k-r)}u^{2^{n+1}(2l+1)}h = \big(\tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\xi^{2r-1}+O(\rho)\big)x^{2^{n+2}-2r}+O(x^{2^{n+2}-2r+1}),$$
$$\xi\cdot\tau^{2(2^{n+1}k+r)}u^{2(2^{n}(2l+1)-r)}h = \big(\tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\rho^{2r-1}+O(\xi)\big)x^{2^{n+2}-2r}+O(x^{2^{n+2}-2r+1})$$

for $k,l\in\mathbb{Z}$, $n\geq 0$, *and* $1\leq r\leq 2^{n+1}-1$.

As with Theorem 2.3, implicit in this theorem is the fact that the terms on the left and right do in fact live in $\pi_\star b(ER)$, for example $\tau^4h = 2\tau^4+\rho\xi\cdot\tau^4h_1$. The error terms are present to remind the reader that these are extensions and not products: to resolve them would require describing how to lift classes from Z to $\pi_\star b(ER)$, and we shall not pursue this. In particular, if k is even then the h_{n+1} terms on the right may be replaced with u_{n+1} without affecting the theorem statement.

This concludes our description of $\pi_\star b(ER)$. Although $\pi_\star b(ER)$ is complicated, it is not impossible to work with. We illustrate this in Section 8 by computing some MPR–based Mahowald invariants. Li, Shi, Wang and Xu [Li et al. 2019] have shown that real bordism detects the Hopf elements, Kervaire classes, and $\bar\kappa$ family. These are the elements in π_*S detected in the classical Adams spectral sequence by the Sq^0–families generated by h_0, h_0^2, and g_1. We compute the iterated MPR–based Mahowald invariants of 2, 4, and $\bar\kappa$, showing that they line up with these Sq^0–families exactly.

3 Even-periodic real Landweber exact spectra

We begin by recording some properties of $E\mathbb{R}$ and ER. The material of this section is essentially a translation to the even-periodic setting of familiar facts about the real Johnson–Wilson theories. We would like to avoid confusion between elements of $\pi_\star b(ER)$ and $\pi_\star E\mathbb{R}$, so in this section we write a_σ instead of ρ for the Euler class of the sign representation, and use the symbol u_σ for what would previously have been written τ. In particular, these symbols have degrees

$$|a_\sigma| = -\sigma,\quad |u_\sigma| = 1-\sigma.$$

Before considering $E\mathbb{R}$, we consider the C_2–HFPSS in general. There is a cofiber sequence

$$S^{-\sigma} \xrightarrow{a_\sigma} S^0 \to C_{2+}. \tag{3}$$

Let X be a C_2–spectrum. Then we may identify

$$\pi_\star(C_{2+} \otimes X) \cong \pi^e_* X[u_\sigma^{\pm 1}],$$

and so the a_σ–Bockstein spectral sequence for X is of signature

$$E_1 = \pi^e_* X[u_\sigma^{\pm 1}, a_\sigma] \Rightarrow \pi_\star X,$$

where $\pi^e_* X$ are the homotopy groups of the underlying spectrum of X. This spectral sequence converges conditionally to the homotopy groups of the a_σ–completion of X, which may be identified as its Borel completion $F(EC_{2+}, X)$. Moreover we have the following fact; see for instance [Hill and Meier 2017, Lemma 4.8].

3.1 Lemma *For any C_2–spectrum X, the a_σ–Bockstein spectral sequence for X agrees with the HFPSS for X from the E_2–page on.* □

The proof amounts to identifying the a_σ–Bockstein spectral sequence with the Borel cohomology spectral sequence induced by the standard cellular filtration of EC_{2+}. This identification leads to the following.

3.2 Lemma *Let X be a C_2–spectrum, and write ψ^{-1} for the involution on $\pi^e_* X$. Then the d_1–differential in the a_σ–Bockstein spectral sequence for X is given by*

$$d_1(\alpha u_\sigma^n a_\sigma^m) = (\alpha - (-1)^n \psi^{-1}(\alpha)) u_\sigma^{n-1} a_\sigma^{m+1}$$

for $\alpha \in \pi^e_* X$. In particular, if X carries a product, then the differentials satisfy the Leibniz rule

$$d_r(\alpha\beta) = d_r(\alpha)\beta + \psi^{-1}(\alpha) d_r(\beta)$$

for $r \geq 1$, where the ψ^{-1} may be omitted for $r \geq 2$. □

Now let $E\mathbb{R}$ be as in the introduction: a strongly even and even-periodic and real Landweber exact C_2–spectrum in the sense of [Hill and Meier 2017, Section 3.2], with underlying spectrum E. This set of assumptions means three things. First, $E\mathbb{R}$ is a homotopy commutative C_2–ring spectrum equipped with a multiplicative orientation $MP\mathbb{R} \to E\mathbb{R}$. In particular, there is an invertible element $\bar{u} \in \pi_{1+\sigma} E\mathbb{R}$ coming from the generator of the $n = 1$ summand of $MP\mathbb{R} \simeq \bigoplus_{n\in\mathbb{Z}} \Sigma^{n(1+\sigma)} M\mathbb{R}$. Second, $\pi_0 E\mathbb{R} \cong \pi_0 E$, and in general

$$E\mathbb{R}_\star X \cong E_0 \otimes_{MP_0} MP\mathbb{R}_\star X$$

for any C_2–spectrum X. Third, $\pi_{-1} E\mathbb{R} = 0$. Implicit in these is the fact that $MP\mathbb{R}$ itself satisfies these conditions. This is nontrivial, and follows from work of Hu and Kriz [2001] on real cobordism. We recall the key calculation.

3.3 Lemma *The C_2–spectrum $E\mathbb{R}$ is Borel complete, with a_σ–Bockstein spectral sequence*

$$E_1 = E_0[\bar{u}^{\pm 1}, u_\sigma^{\pm 1}, a_\sigma] \Rightarrow \pi_\star E\mathbb{R},$$

where

$$|\bar{u}| = 1+\sigma, \quad |u_\sigma| = 1-\sigma, \quad |a_\sigma| = -\sigma.$$

The differentials are $E_0[\bar{u}^{\pm 1}, a_\sigma]$–linear, and are generated by

$$d_{2^{n+1}-1}(u_\sigma^{2^n}) = u_n \bar{u}^{2^n-1} a_\sigma^{2^{n+1}-1},$$

where $u_n = u^{-(2^n-1)} v_n \in E_0$. In particular, $\pi_0 E\mathbb{R} = \pi_0 E$.

Proof We first verify the given description of the a_σ–Bockstein spectral sequence. The E_1–page of the a_σ–Bockstein spectral sequence is given by $E_*[u_\sigma^{\pm 1}, a_\sigma]$. To put this in the desired form, we set $\bar{u} = u_\sigma^{-1} u$ with $u \in \pi_2 E$ the unit; when $E = MP$, this generates the $n=1$ summand of $MP\mathbb{R} \simeq \bigoplus_{n\in\mathbb{Z}} \Sigma^{n(1+\sigma)} M\mathbb{R}$. As $E\mathbb{R}$ is Landweber exact over $MP\mathbb{R}$, the a_σ–Bockstein spectral sequence for $E\mathbb{R}$ is tensored down from the a_σ–Bockstein spectral sequence for $MP\mathbb{R}$, and here the computation is known by work of Hu and Kriz [2001].

The C_2–spectrum $M\mathbb{R}$ is shown to be Borel complete in [Hu and Kriz 2001, Theorem 4.1], and Landweber exactness extends the proof to $E\mathbb{R}$. By the Tate fracture square, $E\mathbb{R}$ is Borel complete if and only if the map $\Phi^{C_2} E\mathbb{R} \to E\mathbb{R}^{tC_2}$ is an equivalence, where Φ^{C_2} denotes the functor of geometric fixed points. Landweber exactness implies

$$\pi_* \Phi^{C_2} E\mathbb{R} \cong E_0 \otimes_{MP_0} \pi_* \Phi^{C_2} MP\mathbb{R} \cong E_0/(u_0, u_1, \ldots)[x^{\pm 1}],$$

where $x = a_\sigma \bar{u} \in \pi_1 E\mathbb{R}$, the last identification coming from the equivalence $\Phi^{C_2} M\mathbb{R} \simeq MO$. This is exactly what one obtains computing $\pi_* E\mathbb{R}^{tC_2}$ by the Tate spectral sequence, which may itself be obtained from the above description of the a_σ–Bockstein spectral sequence by inverting a_σ. Thus $E\mathbb{R}$ is Borel complete as claimed. □

We now pass to the nonequivariant spectrum $ER = E\mathbb{R}^{C_2} \simeq E^{hC_2}$. Note that $\pi_* ER$ is the portion of $\pi_\star E\mathbb{R}$ located in integer degrees, and write $x = a_\sigma \bar{u} \in \pi_1 ER$. We then have the following analogue of [Kitchloo and Wilson 2007a] and [Kitchloo and Wilson 2008a, Theorem 4.2].

3.4 Proposition *There is a cofiber sequence*

$$\Sigma ER \xrightarrow{x} ER \to E,$$

and thus for any spectrum X an x–Bockstein spectral sequence

$$E_1 = (E^* X)[x] \Rightarrow ER^* X,$$

and this agrees with the HFPSS

$$E_2 = H^*(C_2; E^* X) \Rightarrow ER^* X$$

from the E_2–page on. Write ψ^{-1} for the involution on E^*X. Then the d_1–differential in the x–Bockstein spectral sequence is given by

$$d_1(\alpha) = (\alpha - \psi^{-1}(\alpha))u^{-1}x$$

for $\alpha \in E^*X$, and the differentials satisfy the Leibniz rule

$$d_r(\alpha\beta) = d_r(\alpha)\beta + \psi^{-1}(\alpha)d_r(\beta)$$

for $r \geq 1$, where the ψ^{-1} may be omitted for $r \geq 2$. All x–Bockstein differentials are E_0–linear, and when $X = S^0$ they are generated by

$$d_{2^{n+1}-1}(u^{2^n}) = u_n x^{2^{n+1}-1}.$$

Proof As $x = a_\sigma \bar{u}$ and $\bar{u}$ is invertible in $E\mathbb{R}$, (3) implies that there is a cofiber sequence

$$\Sigma E\mathbb{R} \xrightarrow{x} E\mathbb{R} \to C_{2+} \otimes E\mathbb{R}$$

of C_2–spectra. Passing to fixed points yields the corresponding cofiber sequence for ER. The remaining facts follow from the previous lemmas. □

3.5 Remark Figure 1 depicts the E_∞ page of the x–Bockstein spectral sequence for ER.

The lines of slope 1 depict x–towers. Everything in filtration $\geq 2^n - 1$ is a module over $E_0/(u_0, \ldots, u_{n-1})$, and these regions are separated by dashed lines. The terms on the bottom describe the 0–line $(\pi_* ER)/(x)$. For example, $(\pi_{16} ER)/(x) \subset \pi_{16} E \cong E_0\{u^8\}$ is the E_0–submodule generated by $(2u^8, u_1 u^8, u_2 u^8)$; as $2u^8 \cdot x = 0$ and $u_1 u^8 \cdot x^3 = 0$, the x–tower out of this is supported on $u_1 u^8$ and $u_2 u^8$, and on just

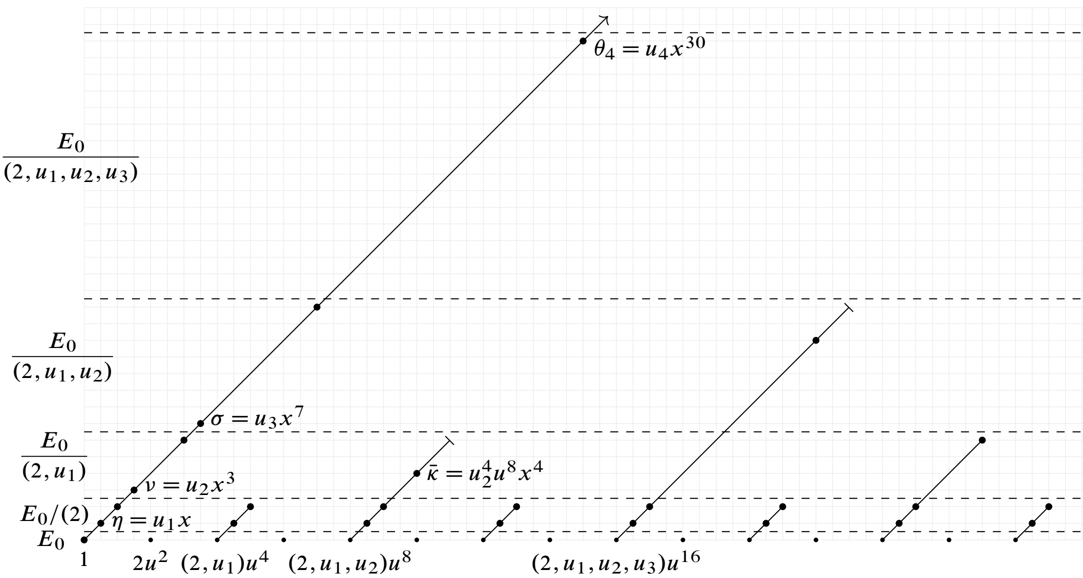


Figure 1

$u_2 u^8$ starting in filtration 3. The solid circles in positive filtration indicate degrees where $\pi_* ER$ has nontrivial Hurewicz image, with some notable elements labeled (see Section 8).

Similar charts appear in [Hahn and Shi 2020, Section 6].

4 Comparing $b(ER)$ and $E\mathbb{R}$

We are now in a position to consider Theorem 2.1. The first order of business is to identify $\xi = \rho\tau^{-2}u$ in $\pi_\sigma b(E)$ as a permanent cycle in the x–Bockstein spectral sequence for $\pi_\star b(ER)$. As

$$\rho\colon \pi_\sigma b(ER) \to \pi_0 b(ER) = ER^0\mathbb{R}P^\infty$$

is the inclusion of a summand, the fact that ξ is a permanent cycle in the x–Bockstein spectral sequence for $\pi_\star b(ER)$ is predicted by the computation of $ER(n)^*\mathbb{R}P^\infty$ by Kitchloo and Wilson [2008a, Theorem 1.2]; see also [Kitchloo et al. 2017]. However, we take a different approach that sheds light on additional aspects of ξ.

Because the x–Bockstein spectral sequence for $\pi_\star b(ER)$ agrees with the HFPSS

$$E_2 = H^*(C_2; \pi_\star b(E)) \Rightarrow \pi_\star b(ER)$$

from the E_2–page on, we can just as well work with the HFPSS in this section.

4.1 Lemma *We have*

$$\pi_\star b(E) = \frac{E_0[\rho, \tau^{\pm 2}, u^{\pm 1}]^\wedge_\rho}{(\rho\cdot h)},$$

and C_2 acts on $\pi_\star b(E)$ by the E_0–linear multiplicative involution ψ^{-1} satisfying

$$\psi^{-1}(\rho) = \rho, \quad \psi^{-1}(u) = -u, \quad \psi^{-1}(\tau^2) = \tau^2(h-1).$$

In particular, ξ is fixed under the action of ψ^{-1}.

Proof The structure of $\pi_\star b(E)$ is as described in [Balderrama 2021, Section 2.1]. That ψ^{-1} fixes ξ follows immediately. □

4.2 Proposition *The class ξ is a permanent cycle in the HFPSS for $\pi_\star b(ER)$, detecting a lift of x.*

Proof By assumption, $E\mathbb{R}$ is $MP\mathbb{R}$–oriented. As $\xi = \rho\tau^{-2}u$ lifts to $\pi_\sigma b(MP)$ and x lifts to $\pi_1 MPR$, it suffices to prove the proposition with ER replaced by MPR.

As MP is an $\mathbb{E}_\infty$ ring, and complex conjugation acts on MP by $\mathbb{E}_\infty$ automorphisms, there is a C_2–equivariant external squaring operation

$$\mathrm{Sq}\colon \pi_n MP \to \pi_{n(1+\sigma)} b(MP).$$

As Sq is additive modulo transfers and ρ annihilates the transfer ideal, the composite $\rho\cdot\mathrm{Sq}$ is additive and so induces a map

$$Q\colon H^1(C_2; \pi_{n+1} MP) \to H^1(C_2; \pi_{n(1+\sigma)+1} b(MP))$$

in group cohomology. By [Balderrama 2024, Theorem 1.0.1], if $a \in H^1(C_2;\pi_{n+1}MP)$ is a permanent cycle detecting $\alpha \in \pi_n MPR$, then $Q(a)$ is a permanent cycle weakly detecting $\mathrm{Sq}(\alpha) \in \pi_{n(1+\sigma)}b(MPR)$. Now recall that x represents the generator of $H^1(C_2;\mathbb{Z}\{u\}) \cong \mathbb{Z}/(2) \subset H^1(C_2;\pi_2 MP)$. The $\mathbb{E}_\infty$ structure on periodic cobordism is such that $\mathrm{Sq}(u) = \tau^{-2}u^2$; see for instance the paragraph after [Ando et al. 2004, Lemma 4.3], noting that u and $\tau^{-2}u^2$ are the periodic Thom classes for $\mathbb{C}$ and $\mathbb{C}[C_2]$ respectively. Thus $Q(x) = \xi x$, and it follows that ξx detects $\mathrm{Sq}(x)$. As $\mathrm{Sq}(x)$ lifts x^2, this is only possible if ξ is a permanent cycle detecting a lift of x as claimed. □

The following corollary is not needed for Theorem 2.1, but will be useful later on in understanding the structure of $\pi_\star b(ER)$. It is a direct analogue of [Kitchloo and Wilson 2008a, Theorem 1.2].

4.3 Corollary *In integer degrees, we have*

$$\pi_* b(ER) \cong ER_*[\![z]\!]/([2](z)),$$

where $z = \rho\xi$. In particular, $\pi_\star b(ER)$ is a module over $\pi_0 b(ER) \cong E^0 BC_2 \cong E_0[\![z]\!]/([2](z))$.

Proof The x–Bockstein spectral sequence for $\pi_* b(ER)$ takes the form

$$E_1 = \pi_* b(E)[x] \Rightarrow \pi_* b(ER).$$

Recall that

$$\pi_* b(E) \cong E_*[\![z]\!]/([2](z)), \quad z = \rho\xi.$$

As ρ and ξ are permanent cycles, so is z. Thus the differentials in the x–Bockstein spectral sequence for $\pi_* b(ER)$ are induced by those for $\pi_* ER$, leading to the given description of $\pi_* b(ER)$. □

We now relate $b(ER)$ and $E\mathbb{R}$. These live in the full subcategory $\mathrm{Sp}^{BC_2} \subset \mathrm{Sp}^{C_2}$ of Borel complete C_2–spectra, equivalent to the category of spectra with C_2–action. The functor

$$b\colon \mathrm{Sp} \to \mathrm{Sp}^{BC_2}, \quad b(X) = F(EC_{2+}, i_* X)$$

is the diagonal, endowing a spectrum with the trivial C_2–action. In particular, it is left adjoint to the functor of homotopy fixed points, and if $X \in \mathrm{Sp}^{BC_2}$ then the counit of this adjunction gives a canonical map

$$b(X^{\mathrm{h}C_2}) \to X.$$

Specializing to $X = E\mathbb{R}$, we have the following.

4.4 Theorem *The canonical map $b(ER) \to E\mathbb{R}$ fits into a cofiber sequence*

$$\Sigma^\sigma b(ER) \xrightarrow{\xi} b(ER) \to E\mathbb{R} \tag{4}$$

of C_2–spectra.

Proof As $E\mathbb{R}$ is strongly even, we have $\pi_\sigma E\mathbb{R} = 0$. As the maps in (4) are $b(ER)$–linear, their composite must be null. As $b(ER)$ and $E\mathbb{R}$ are x–complete, it then suffices to show that (4) is a cofiber sequence after coning off x. As $b(ER)/(x) \simeq b(E)$ and $E\mathbb{R}/(x) \simeq C_{2+} \otimes i_* E$, (4) with x coned off takes the form

$$\Sigma^\sigma b(E) \xrightarrow{\xi} b(E) \to C_{2+} \otimes i_* E,$$

which is a cofiber sequence as now $\xi = \rho \cdot \tau^{-2} u$ differs from ρ by a unit. □

Theorem 2.1 follows by combining Proposition 4.2 and Theorem 4.4.

5 The Bockstein spectral sequence

We now compute the x–Bockstein spectral sequence

$$\pi_\star b(E)[x] \Rightarrow \pi_\star b(ER).$$

We maintain notation from the introduction. In particular, recall that h_n and w_n are defined in terms of the 2–series of E by specializing

$$[2](z) = z h_0(z), \quad h_n(z) \equiv u_n + z^{2^n} h_{n+1}(z) \pmod{u_0, \dots, u_{n-1}},$$
$$w_n = \rho^{2^{n+1}} h_{n+1} \equiv \tau^{2^{n+1}} u^{-2^n}(h_n - u_n) \pmod{u_0, \dots, u_{n-1}}$$

to $z = \rho\xi = \rho^2\tau^{-2}u$. As with u_n, these classes are well defined modulo $(u_0, \dots, u_{n-1})$. We begin by describing what will be the cycles and boundaries of the x–Bockstein spectral sequence for $\pi_\star b(ER)$. Let $Z_{2^{n+1}-1} \subset \pi_\star b(E)$ be the subring

$$E_0\big(\rho, \xi, u^{\pm 2^{n+1}}, \tau^{2^{\pm n+2}}, \tau^{2^{i+2}l} u^{2^{i+1}k} u_i, \tau^{2^{i+1}(2l+1)} u^{2^{i+1}k} h_i : 0 \le i \le n; k, l \in \mathbb{Z}\big)^\wedge_{(\rho,\xi)},$$

and let $B_{2^{n+1}-1} \subset Z_{2^{n+1}-1}[x]$ be the ideal generated by

$$\tau^{2^{i+2}l} u^{2^{i+1}k} u_i \cdot x^{2^{i+1}-1}, \quad \tau^{2^{i+1}(2l+1)} u^{2^{i+1}k} h_i \cdot x^{2^{i+1}-1}, \quad \tau^{2^{i+2}l} u^{2^{i+1}k} w_i \cdot x^{2^{i+1}-1}$$

for $0 \le i \le n$ and $k, l \in \mathbb{Z}$. We also declare $Z_0 = \pi_\star b(E)$ and $B_0 = (0)$, and for $2^{n+1}-1 \le r < 2^{n+2}-1$ we write $Z_r = Z_{2^{n+1}-1}$ and $B_r = Z_{2^{n+1}-1}$. Thus there are inclusions

$$0 = B_0 \subset B_1 \subset B_2 \subset \cdots \subset Z_2[x] \subset Z_1[x] \subset Z_0[x] = \pi_\star b(E)[x].$$

5.1 Theorem *The x–Bockstein spectral sequence for $\pi_\star b(ER)$ supports differentials*

$$d_{2^{n+1}-1}(u^{2^n}) = u_n x^{2^{n+1}-1}, \quad d_{2^{n+1}-1}(\tau^{2^{n+1}}) = -w_n x^{2^{n+1}-1},$$

and we may identify $Z_r[x]$ and B_r as its r–cycles and r–boundaries.

Proof We proceed by induction, treating the inductive step first.

Let $n \ge 1$, and suppose we have verified $E_{2^n} \cong Z_{2^n-1}[x]/B_{2^n-1}$. In particular, E_{2^n} is generated by the permanent cycles ρ and ξ, the classes $\tau^{2^{i+2}l} u^{2^{i+1}k} u_i$ and $\tau^{2^{i+1}(2l+1)} u^{2^{i+1}k} h_i$ for $i < n$, and the classes $u^{\pm 2^n}$ and $\tau^{\pm 2^{n+1}}$. As the classes $\tau^{2^{i+2}l} u^{2^{i+1}k} u_i$ and $\tau^{2^{i+1}(2l+1)} u^{2^{i+1}k} h_i$ are x^{2^n-1}–torsion for $i < n$, having survived to the E_{2^n}–page they must be permanent cycles. It follows that the next differentials are determined by their effect on u^{2^n} and $\tau^{2^{n+1}}$.

The differential $d_{2^{n+1}-1}(u^{2^n}) = u_n x^{2^{n+1}-1}$ follows from Proposition 3.4. Now write

$$d_{2^{n+1}-1}(\tau^{2^{n+1}}) = \alpha \cdot x^{2^{n+1}-1}.$$

As ξ is a permanent cycle, the Leibniz rule implies

$$0 = d_{2^{n+1}-1}(\xi^{2^n}) = d_{2^{n+1}-1}(\rho^{2^n}\tau^{-2^{n+1}}u^{2^n}) = \rho^{2^n}(\tau^{-2^{n+2}}\alpha u^{2^n} + \tau^{-2^{n+1}}u_n)x^{2^{n+1}-1}.$$

This is on the $E_{2^{n+1}-1} = E_{2^n}$–page, and so combines with our inductive hypothesis to imply

$$z^{2^{n-1}}(u_n + \tau^{-2^{n+1}}u^{2^n}\alpha) \equiv 0 \pmod{u_0, \dots, u_{n-1}, z^{2^{n-1}}h_n}.$$

As $\alpha \equiv 0 \pmod{\rho}$, this is only possible if

$$u_n + \tau^{-2^{n+1}}u^{2^n}\alpha \equiv h_n \pmod{u_0, \dots, u_{n-1}},$$

and thus

$$d_{2^{n+1}-1}(\tau^{2^{n+1}}) = \alpha x^{2^{n+1}-1} = \tau^{2^{n+1}}u^{-2^n}(h_n - u_n)x^{2^{n+1}-1} = w_n x^{2^{n+1}-1}$$

as claimed.

To identify boundaries and cycles, observe that as a general property of the x–Bockstein spectral sequence, if we write $Z'_r[x]$ for its r–cycles then

$$Z'_r = \operatorname{Ker}(d_r : Z_{r-1} \to E_r) = \operatorname{Ker}(d_r : Z_{r-1} \to E_r[x^{-1}]),$$

ie to compute cycles it suffices to work in the x–inverted x–Bockstein spectral sequence, or equivalently the x–Bockstein spectral sequence with x set to 1. Our inductive hypothesis implies

$$\frac{E_{2^n}}{(x-1)} \cong \frac{(E_0/(u_0, \dots, u_{n-1}))[\rho, \xi, u^{\pm 2^n}, \tau^{\pm 2^{n+1}}]^{\wedge}_{(\rho,\xi)}}{(\xi^{2^n} - \rho^{2^n}\tau^{-2^{n+1}}u^{2^n}, \rho^{2^n}(u_n\tau^{2^{n+1}} + \rho^{2^{n+1}}h_{n+1}u^{2^n}))},$$

and we have just produced the differentials

$$d(u^{2^n}) = u_n, \quad d(\tau^{2^{n+1}}) = \rho^{2^{n+1}}h_{n+1}.$$

Thus $\operatorname{Ker}(d)$ is generated over $E_0(\rho, \xi, u^{\pm 2^{n+1}}, \tau^{\pm 2^{n+2}})^{\wedge}_{(\rho,\xi)}$ by $u_n\tau^{2^{n+1}} + \rho^{2^{n+1}}h_{n+1}u^{2^n} = \tau^{2^{n+1}}h_n$, and this leads to $Z'_{2^{n+1}-1} = Z_{2^{n+1}-1}$ as claimed. The identification of boundaries follows immediately.

The base case, concerning the d_1–differential and identification of the E_2–page, can be handled by considering $0 = d_1(\xi)$ just like the above, only taking into account the twist in the Leibniz rule for d_1 given in Proposition 3.4. Alternately, one may just use the formula $d_1(a) = (a - \psi^{-1}(a))x$ given there, where the action of ψ^{-1} is given in Lemma 4.1. □

The ring Z and ideal $B \subset Z[x]$ of the introduction may be identified as $Z = \bigcap_r Z_r$ and $B = \bigcup_r B_r$. Thus Theorem 2.3 follows from Theorem 5.1 by letting $r \to \infty$.

5.2 Remark Although we have relied on the known computation of $\pi_* MPR$ in computing $\pi_\star b(MPR)$, this was not actually necessary: the proof of Theorem 5.1 gives an independent computation, as we now explain.

Note that no computation was needed to produce $x \in \pi_1 MP\mathbb{R}$ or prove $MP\mathbb{R}/(x) \simeq C_{2+} \otimes i_* MP$, as $x = a_\sigma \bar{u}$ where $\bar{u}$ generates the $n = 1$ summand of $MP\mathbb{R} \simeq \bigoplus_{n\in\mathbb{Z}} \Sigma^{n(1+\sigma)} M\mathbb{R}$. Thus it suffices to describe the x–Bockstein spectral sequence $MP_0[u^{\pm 1}, x] \Rightarrow MPR_*$. This is MP_0–linear by [Hu and Kriz 2001, Proposition 2.27], which uses the theory of real orientations but not the computation of $\pi_\star M\mathbb{R}$. Thus it suffices to produce the differentials $d_{2^{n+1}-1}(u^{2^n}) = u_n x^{2^{n+1}-1}$. The differential $d_1(u) = 2x$ follows from the involution $\psi^{-1}(u) = -u$, so suppose inductively that we have computed $d_{2^{n+1}-1}(u^{2^n}) = u_n x^{2^{n+1}-1}$.

Next note that no computation was needed in Proposition 4.2 to prove ξ is a permanent cycle. The argument in Theorem 5.1 now applies to show $d_{2^{n+1}-1}(\tau^{2^{n+1}}) = -\rho^{2^{n+1}} h_{n+1} x^{2^{n+1}-1}$.

As in Section 4, there is a canonical map $q\colon b(MP^{hC_2}) \to F(EC_{2+}, MP\mathbb{R})$. Here, we write MP^{hC_2} and $F(EC_{2+}, MP\mathbb{R})$ instead of MPR and $MP\mathbb{R}$ as the proof that $MP\mathbb{R}$ is Borel complete relies on knowledge of its x–Bockstein spectral sequence. The map q fits into a diagram of cofiber sequences:

$$\begin{array}{ccccccc}
\Sigma b(MP^{hC_2}) & \xrightarrow{\quad x \quad} & b(MP^{hC_2}) & \longrightarrow & b(MP) & \xrightarrow{\quad \partial \quad} & \Sigma^2 b(MP^{hC_2}) \\
\downarrow q & & \downarrow q & & \downarrow p & & \downarrow q \\
\Sigma F(EC_{2+}, MP\mathbb{R}) & \xrightarrow{x} & F(EC_{2+}, MP\mathbb{R}) & \longrightarrow & C_{2+} \otimes i_* MP & \xrightarrow{\partial'} & \Sigma^2 F(EC_{2+}, MP\mathbb{R})
\end{array}$$

The x–Bockstein differential $d_{2^{n+1}-1}(\tau^{2^{n+1}}) = -\rho^{2^{n+1}} h_{n+1} x^{2^{n+1}-1}$ implies

$$\partial(\tau^{2^{n+1}}) = -\rho^{2^{n+1}} h_{n+1} x^{2^{n+1}-2}$$

mod higher filtration, and as $p(\tau^2) = u_\sigma^2$ and $q(\rho) = a_\sigma$ it follows that

$$\begin{aligned}
\partial'(u^{2^{n+1}}) = \partial'(\bar{u}^{2^{n+1}} u_\sigma^{2^{n+1}}) &= \bar{u}^{2^{n+1}} q(\partial(\tau^{2^{n+1}})) \\
&= \bar{u}^{2^{n+1}} q(-\rho^{2^{n+1}} h_{n+1} x^{2^{n+1}-2}) \\
&= \bar{u}^{2^{n+1}} a_\sigma^{2^{n+1}} u_{n+1} x^{2^{n+1}-2} = u_{n+1} x^{2^{n+2}-2}
\end{aligned}$$

mod higher filtration. This gives the next x–Bockstein differential

$$d_{2^{n+2}-1}(u^{2^{n+1}}) = u_{n+1} x^{2^{n+2}-1},$$

completing the induction.

We end this subsection with some observations about the structure of $\pi_\star b(ER)$.

5.3 Proposition *The C_2–spectrum $b(ER)$ has the gap*

$$\pi_{*\sigma-1} b(ER) = 0.$$

u_0, u_1 $\rho^8 x^8, h_2$	$\xi u_1, \xi h_2$ $\rho^7 x^8$	$\xi^2 u_1, \xi^2 h_2$ $\rho^6 x^8$	$\xi^3 u_1, \xi^3 h_2$ $\rho^5 x^8$	$u_0, \xi^4 h_2$ $\rho^4 x^8$	$\rho^3 u_1, \xi^5 h_2$ $\rho^3 x^8$	$\rho^2 u_1, \xi^6 h_2$ $\rho^2 x^8$	$\rho u_1, \xi^7 h_2$ ρx^8	u_0, u_1 x^8
$\rho^7 x^7$	$\rho^6 x^7$	$\rho^5 x^7$	$\rho^4 x^7$	$\rho^3 x^7$	$\rho^2 x^7$	ρx^7	x^7	ξx^7
h $\rho^6 x^6$	$\rho^5 x^6$	$h_1 x^2$ $\rho^4 x^6$	$\xi h_1 x^2$ $\rho^3 x^6$	h $\rho^2 x^6$	ρx^6	x^6	ξx^6	h $\xi^2 x^6$
$\rho^5 x^5$	$h_1 x$ $\rho^4 x^5$	$\xi h_1 x$ $\rho^3 x^5$	$\rho^2 x^5$	ρx^5	x^5	ξx^5	$\xi^2 x^5$	$\rho h_1 x$ $\xi^3 x^5$
u_0, h_1 $\rho^4 x^4$	ξh_1 $\rho^3 x^4$	$\xi^2 h_1$ $\rho^2 x^4$	$\xi^3 h_1$ ρx^4	u_0 x^4	$\rho^3 h_1$ ξx^4	$\rho^2 h_1$ $\xi^2 x^4$	ρh_1 $\xi^3 x^4$	u_0, h_1 $\xi^4 x^4$
$\rho^3 x^3$	$\rho^2 x^3$	ρx^3	x^3	ξx^3	$\xi^2 x^3$	$\xi^3 x^3$	$\xi^4 x^3$	$\xi^5 x^3$
h $\rho^2 x^2$	ρx^2	x^2	ξx^2	h $\xi^2 x^2$	$\xi^3 x^2$	$\xi^4 x^2$	$\xi^5 x^2$	h $\xi^6 x^2$
ρx	x	ξx	$\xi^2 x$	$\xi^3 x$	$\xi^4 x$	$\xi^5 x$	$\xi^6 x$	$\rho u_1 x$ $\xi^7 x$
1	ξ	ξ^2	ξ^3	u_0 ξ^4	$\rho^3 u_1$ ξ^5	$\rho^2 u_1$ ξ^6	ρu_1 ξ^7	u_0, u_1 ξ^8

Table 1

Proof Declare the *coweight* of a degree $c + w\sigma$ to be the quantity c, so that we are claiming $\pi_\star b(ER)$ vanishes in coweight -1. By Theorem 5.1, $\pi_\star b(ER)$ is generated over the coweight 0 classes $E_0(\rho,\xi)^\wedge_{(\rho,\xi)}$ by the class x in coweight 1 and the classes

$$\tau^{2^{n+2}l} u^{2^{n+1}k} u_n, \quad \tau^{2^{n+1}(2l+1)} u^{2^{n+1}k} h_n$$

in coweights of the form $2^{n+1}t$. These classes are killed by $x^{2^{n+1}-1}$, and therefore cannot support long enough x–towers to reach coweight -1. □

5.4 Proposition *If E is L_d–local, then $b(ER)$ is $u^{\pm 2^{d+1}}$ and $\tau^{\pm 2^{d+1}}$–periodic. Moreover,*

$$x^{2^{d+1}-1} = 0, \quad \rho^{2^d} x^{2^d-1} = 0, \quad \xi^{2^d} x^{2^d-1} = 0$$

in $Z[x]/B$.

Proof Recall that E is L_d–local provided u_d is invertible in $E_0/(u_0, \dots, u_{d-1})$, or equivalently if the ideal $(u_0, \dots, u_d) \subset E_0$ generates the entire ring. Thus as $u_i u^{\pm 2^{d+1}}$ is a permanent cycle for $i \le d$, it follows that $u^{\pm 2^{d+1}}$ is also a permanent cycle. Likewise, as $u_i x^{2^{d+1}-1} = 0$ for $i \le d$, it follows that $x^{2^{d+1}-1} = 0$.

$\xi u_1, \rho^7 h_2$ ξx^8	$\xi^2 u_1, \rho^6 h_2$ $\xi^2 x^7$	$\xi^3 u_1, \rho^5 h_2$ $\xi^3 x^7$	$u_0, \rho^4 h_2$ $\xi^4 x^8$	$\rho^3 u_1, \rho^3 h_2$ $\xi^5 x^7$	$\rho^2 u_1, \rho^2 h_2$ $\xi^6 x^8$	$\rho u_1, \rho h_2$ $\xi^7 x^8$	u_0, u_1 $\xi^8 x^8, h_2$
$\xi^2 x^7$	$\xi^3 x^7$	$\xi^4 x^7$	$\xi^5 x^7$	$\xi^6 x^7$	$\xi^7 x^7$	$\xi^8 x^7$	$\xi^9 x^7$
$\rho h_1 x^2$ $\xi^3 x^6$	$x^2 h_1$ $\xi^4 x^6$	$x^2 \xi h_1$ $\xi^5 x^6$	h $\xi^6 x^6$	$\xi^7 x^6$	$\xi^8 x^6$	$\xi^9 x^6$	h $\xi^{10} x^6$
$h_1 x$ $\xi^4 x^5$	$x \xi h_1$ $\xi^5 x^5$	$\xi^6 x^5$	$\xi^7 x^5$	$\xi^8 x^5$	$\xi^9 x^5$	$\xi^{10} x^5$	$x \rho h_1$ $\xi^{11} x^5$
ξh_1 $\xi^5 x^4$	$\xi^2 h_1$ $\xi^6 x^4$	$\xi^3 h_1$ $\xi^7 x^4$	u_0 $\xi^8 x^4$	$\rho^3 h_1$ $\xi^9 x^4$	$\rho^2 h_1$ $\xi^{10} x^4$	ρh_1 $\xi^{11} x^4$	u_0, h_1 $\xi^{12} x^4$
$\xi^6 x^3$	$\xi^7 x^3$	$\xi^8 x^3$	$\xi^9 x^3$	$\xi^{10} x^3$	$\xi^{11} x^3$	$\xi^{12} x^3$	$x^3 \rho^3 u_2$ $\xi^{13} x^3 k$
$\rho u_1 x^2$ $\xi^7 x^2$	$x^2 u_1$ $\xi^8 x^2$	$x^2 \xi u_1$ $\xi^9 x^2$	$x^2 \xi^2 u_1 / x^2 \rho^6 u_2$ $h, \xi^{10} x^2$	$x^2 \rho^5 u_2$ $\xi^{11} x^2$	$x^2 \rho^4 u_2$ $\xi^{12} x^2$	$x^2 \rho^3 u_2$ $\xi^{13} x^2$	$h, x^2 \rho^2 u_2$ $\xi^{14} x^2$
$u_1 x$ $\xi^8 x$	$x \xi u_1$ $\xi^9 x$	$x \xi^2 u_1 / x \rho^6 u_2$ $\xi^{10} x$	$x \rho^5 u_2$ $\xi^{11} x$	$x \rho^4 u_2$ $\xi^{12} x$	$x \rho^3 u_2$ $\xi^{13} x$	$x \rho^2 u_2$ $\xi^{14} x$	$x \rho u_1, x \rho u_2$ $\xi^{15} x$
$\xi u_1, \rho^7 u_2$ ξ^9	$\xi^2 u_1, \rho^6 u_2$ ξ^{10}	$\xi^3 u_1, \rho^5 u_2$ ξ^{11}	$u_0, \rho^4 u_2$ ξ^{12}	$\rho^3 u_1, \rho^3 u_2$ ξ^{13}	$\rho^2 u_1, \rho^2 u_2$ ξ^{14}	$\rho u_1, \rho u_2$ ξ^{15}	u_0, u_1 ξ^{16}, u_2

Table 2

Next, as $h_d(z) = u_d + O(z)$, it follows by Weierstrass preparation that

$$(u_0, \dots, u_{d-1}, h_d) \subset \pi_0 b(ER) \cong E^0 BC_2$$

(see Corollary 4.3) generates the entire ring. As $u_i \tau^{\pm 2^{d+1}}$ for $i < d$ and $h_d \tau^{\pm 2^{d+1}}$ are permanent cycles, it follows that $\tau^{\pm 2^{d+1}}$ is a permanent cycle. Next, note that

$$w_{d-1} = \rho^{2^d} h_d, \quad \tau^{2^{d+1}} u^{2^d} w_{d-1} = \xi^{2^d} h_d.$$

As $u_i x^{2^d - 1} = 0$ for $i < d$, the identities $w_{d-1} x^{2^d - 1} = 0$ and $\tau^{2^{d+1}} u^{2^d} w_{d-1} x^{2^d - 1} = 0$ then imply $\rho^{2^d} x^{2^d - 1} = 0$ and $\xi^{2^d} x^{2^d - 1} = 0$. □

5.5 Remark Tables 1 and 2 may be helpful in getting acquainted with the general shape of $\pi_\star b(ER)$, and especially for visualizing the arguments in Sections 6 and 7.

These describe generators of $\pi_{c+w\sigma} b(ER)$ as a module over $\pi_0 b(ER) \cong E^0 BC_2$ in coweight $0 \le c \le 8$ and stem $0 \le c + w \le 16$, the first table containing stems $0 \le c + w \le 8$ and second $9 \le c + w \le 16$. It is arranged by stem and coweight: the box at coordinate (s, c) contains a list of generators for $\pi_{c+(s-c)\sigma} b(ER)$. For space reasons, we have omitted any $\tau^{2i} u^j$ terms. These may recovered by comparing degrees: for example, the box in coordinate $(8,5)$ has entries $\rho h_1 x$ and $\xi^3 x^5$, and this means that $\pi_{5+3\sigma} b(ER)$ is generated over $\pi_0 b(ER)$ by $\rho \cdot \tau^{-4} u^4 h_1 \cdot x$ and $\xi^3 x^5$.

The entry $x\xi^2 u_1/x\rho^6 u_2$ indicates that either $x\xi^2 u_1$ or $x\rho^6 u_2$ may be chosen as a generator, and likewise for $x^2\xi^2 u_1/x^2\rho^6 u_2$. This sort of choice also appears on the 0–line: for example, in box $(5,0)$ one could replace $\rho^3 u_1$ with ξu_0.

These tables assume that E has sufficiently large height, say $E = MP$.

6 Transfers

Recall that there are cofiber sequences

$$\Sigma^{-\sigma} b(ER) \xrightarrow{\rho} b(ER) \to C_{2+} \otimes i_* ER, \quad \Sigma^{\sigma} b(ER) \xrightarrow{\xi} b(ER) \to E\mathbb{R} \tag{5}$$

of C_2–spectra. The first is a general cofiber sequence that exists for any C_2–spectrum, given that $C_{2+} \otimes b(ER) \simeq C_{2+} \otimes i_* ER$, and the second was shown in Theorem 4.4. Here,

$$\pi_\star(C_{2+} \otimes i_* ER) \cong \pi_* ER[u_\sigma^{\pm 1}], \quad |u_\sigma| = 1 - \sigma,$$

and $\pi_\star E\mathbb{R}$ was described in Section 3.

Associated to the cofiber sequences of (5) are boundary maps

$$\operatorname{tr}(u_\sigma^{-1} \cdot -)\colon \pi_{\star+1-\sigma}(C_{2+} \otimes i_* ER) \to \pi_\star b(ER), \quad \partial\colon \pi_{\star+1+\sigma} E\mathbb{R} \to \pi_\star b(ER).$$

The first of these is the transfer for the C_2–spectrum $b(ER)$. Both are $\pi_\star b(ER)$–linear.

6.1 Proposition *The above transfer and boundary maps satisfy*

$$\operatorname{tr}(u_\sigma^{-1} \cdot u_\sigma^{2^n(2k+1)}) = \rho^{2^n-1} \tau^{2^{n+1}k} h_n x^{2^n-1} + O(x^{2^n}),$$
$$\partial(\bar{u}^{2^n(2k+1)}) = \xi^{2^n-1} \tau^{-2^{n+1}k} u^{2^{n+1}k} h_n x^{2^n-1} + O(x^{2^n})$$

for $n \geq 0$ and $k \in \mathbb{Z}$.

Proof The error terms are present just because $\tau^{2^{n+1}k} h_n$ and $\tau^{-2^{n+1}k} u^{2^{n+1}k} h_n$ have only been defined mod x, so we omit them in the proof.

First consider the case $n = 0$. These claimed values are not hidden in the x–Bockstein spectral sequence, so it suffices to show that they hold after coning off x. After coning off x, the cofiber sequences (5) take the form

$$\Sigma^{-\sigma} b(E) \xrightarrow{\rho} b(E) \to C_{2+} \otimes i_* E, \quad \Sigma^{\sigma} b(E) \xrightarrow{\rho\tau^{-2}u} b(E) \to C_{2+} \otimes i_* E. \tag{6}$$

In particular, $\partial(\bar{u}\alpha) = \operatorname{tr}(\alpha)$. By [Hopkins et al. 2000, Remark 6.15], the transfer

$$\operatorname{tr}\colon \pi_0 E \to E^0 BC_2 \cong \pi_0 b(E)$$

satisfies $\mathrm{tr}(1) = h$. The proof is to observe that h is the unique class which satisfies

$$\rho \cdot h = 0, \quad h \equiv 2 \pmod{\rho}.$$

As tr and ∂ are $\pi_\star b(E)$–linear, we deduce

$$\mathrm{tr}(u_\sigma^{-1} \cdot u_\sigma^{2n+1}) = \tau^{2n}\,\mathrm{tr}(1) = \tau^{2n} h, \quad \partial(\bar{u}^{2n+1}) = \tau^{-2n} u^{2n} \partial(\bar{u}) = \tau^{-2n} u^{2n} h$$

as claimed. The argument is essentially the same for $n \geq 1$. Observe that

$$\rho^{2^n} \tau^{2^{n+1}k} h_n = \tau^{2^{n+1}k} w_{n-1}, \quad \xi^{2^n} \tau^{-2^{n+1}k} h_n = \tau^{-2^{n+1}(k+1)} u^{2^n} w_{n-1}.$$

In particular $\rho^{2^n-1}\tau^{2^{n+1}k} h_n x^{2^n-1}$ and $\xi^{2^n-1}\tau^{-2^{n+1}k} u^{2^{n+1}k} h_n x^{2^n-1}$ generate the kernels of ρ and ξ in their respective degrees. As the kernels of ρ and ξ are generated by the images of tr and ∂, this gives the claimed values of tr and ∂ up to multiplication by a unit, which may then be ruled out by working in the universal case $E = MP$. ◻

7 Hidden extensions

We now turn our attention to hidden extensions. We begin with a general discussion. Write $Z[x]/B$ for the x–adic associated graded of $\pi_\star b(ER)$, as computed in Section 5. In general, hidden extensions in the x–Bockstein spectral sequence arise from the failure of $\pi_\star b(ER)$ to be isomorphic to $Z[x]/B$, and especially for relations to fail to lift through the map

$$\pi_\star b(ER) \to (\pi_\star b(ER))/(x) \cong Z \subset \pi_\star b(E). \tag{7}$$

Recall that

$$\pi_\star b(E) = \frac{E_0[\rho, \tau^{\pm 2}, u^{\pm 1}]^\wedge_\rho}{(\rho \cdot h)}.$$

This indicates that the simple indecomposable hidden extensions will be those ρ and ξ–extensions lifting relations of the form

$$\rho \cdot \tau^{2i} u^j h = 0, \quad \xi \cdot \tau^{2i} u^j h = 0, \tag{8}$$

where i and j are such that $\tau^{2i} u^j h \in Z$.

If a relation of this sort lifts to $\pi_\star b(ER)$, then necessarily the corresponding $\tau^{2i} u^j h$ is in the image of the transfer or boundary studied in the previous section. These classes are generally not in the image of the transfer or boundary, and so one knows from the start that the relations in (8) generally lift to nontrivial hidden extensions in $\pi_\star b(ER)$.

One can use Proposition 6.1 to compute some of these directly:

$$\xi \cdot \tau^{2^{n+1}(2k+1)} h = \xi \cdot \mathrm{tr}(u_\sigma^{2^{n+1}(2k+1)}) = \mathrm{tr}(x u_\sigma^{-1} \cdot u_\sigma^{2^{n+1}(2k+1)}) = \rho^{2^{n+1}-1} \tau^{2^{n+1}k} h_{n+1} x^{2^{n+1}}$$

by Frobenius reciprocity, and likewise

$$\rho \cdot \tau^{-2^{n+1}(2k+1)} u^{2^{n+1}(2k+1)} h = \rho \cdot \partial(\bar{u}^{2^{n+1}(2k+1)+1}) = \xi^{2^{n+1}-1} \tau^{2^{n+2}k} u^{2^{n+2}k} h_{n+1} x^{2^{n+1}}.$$

In general however, a more indirect approach is necessary. Consider the cofiber sequences

(9) $$\Sigma^{-\sigma} b(ER) \xrightarrow{\rho} b(ER) \xrightarrow{p} C_{2+} \otimes i_* ER, \quad \Sigma^{\sigma} b(ER) \xrightarrow{\xi} b(ER) \xrightarrow{q} E\mathbb{R}.$$

The long exact sequences associated to these imply that the image of ρ is equal to the kernel of the forgetful map $p\colon \pi_\star b(ER) \to \pi_{|\star|} ER$, and that the image of ξ is equal to the kernel of the canonical map $q\colon \pi_\star b(ER) \to \pi_\star E\mathbb{R}$. To find elements of these kernels, one looks for elements in $\pi_\star b(ER)$ that lift the relations $u_n x^{2^{n+1}-1} = 0$. This relation already holds in $\pi_\star b(ER)$, so we need only consider lifts involving the filtration-shifting identities $p(\xi) = x$ and $q(\rho) = a_\sigma$. In this way we focus our attention on those classes of the form

(10) $$\tau^{2^{n+2}k} u^{2^{n+2}l} h_{n+1} \xi^r x^s, \quad \tau^{2^{n+2}k} u^{2^{n+2}l} h_{n+1} \rho^r x^s,$$

where $r + s \geq 2^{n+1} - 1$ and $r \geq 1$ and $s < 2^{n+2} - 1$. By the preceding discussion, the former must be in the image of ρ and the latter in the image of ξ, and when this is not the case in $Z[x]/B$ there must be a hidden extension making it so. If $r + s > 2^{n+2} - 1$, then the witness to the classes in (10) being in the image of ρ or ξ may be obtained by multiplying a smaller witness with some suitable power of ρ or ξ and x. Thus we are led to focus on the case where $r + s = 2^{n+2} - 1$. We will show that when s is even, the necessary hidden extensions are exactly those lifting the relations in (8). First, a couple observations.

7.1 Lemma *Fix positive integers $r + s = 2^{n+2} - 1$ with s even. Then the classes*

$$\tau^{2^{n+2}k} u^{2^{n+2}l} h_{n+1} \xi^r x^s, \quad \tau^{2^{n+2}k} u^{2^{n+2}l} h_{n+1} \rho^r x^s$$

are not in the image of ρ or ξ respectively in $Z[x]/B$, at least when $E = MP$.

Proof Consider the first case. Suppose towards contradiction that

$$\tau^{2^{n+2}k} u^{2^{n+2}l} h_{n+1} \xi^r x^s = \rho \alpha x^s$$

for some $\alpha \in Z$. As the x–Bockstein spectral sequence has only odd differentials and s is even, necessarily we can divide out by x to obtain

(11) $$(\tau^{2^{n+2}k} u^{2^{n+2}l} h_{n+1} \xi^r - \rho\alpha) x^{s-1} = 0.$$

This means that $\tau^{2^{n+2}k} u^{2^{n+2}l} h_{n+1} \xi^r - \rho\alpha$ detects some class $\theta \in \pi_\star b(MPR)$ satisfying $\theta \cdot x^{s-1} = 0$. Write $p\colon \pi_\star b(MPR) \to \pi_{|\star|} MPR$ for the restriction. As $p(\xi) = x$, necessarily $p(\theta)$ is detected by $u^{2^{n+2}l} u_{n+1} x^r$. Thus

$$0 = p(\theta \cdot x^{s-1}) \equiv u^{2^{n+2}l} u_{n+1} x^{r+s-1} \pmod{x^{r+s}}$$

in $\pi_* MPR$. As $r + s - 1 < 2^{n+2} - 1$, this is incompatible with the structure of the x–Bockstein spectral sequence for $\pi_* MPR$, a contradiction. The second case is identical, only instead using the map $b(MPR) \to MP\mathbb{R}$ in place of the restriction. □

7.2 Lemma *Suppose that i and j are such that $\tau^{2i} u^j h \in Z$. Then $\tau^{2i} u^j h$ generates the kernels of ρ and ξ in its degree of $Z[x]/B$ as a module over $\pi_0 b(ER)$.*

Proof The class $\tau^{2i}u^j h$ generates the kernels of ρ and ξ in Z, as this is the case in $\pi_\star b(E)$. Thus the lemma follows from the following observation: $Z[x]/B$ contains no x–divisible elements in the kernel of ρ or ξ in even degrees, that is in degrees of the form $c+w\sigma$ with both c and w even. Indeed, any x–divisible element in even degree and in a given filtration must be of the form $\alpha x^{2r}=0$ with $\alpha\in Z$ in even degree. As α is in even degree and B is generated by classes of the form $w\cdot x^?$ with w in even degree, relations $\rho\alpha x^{2r}=0$ or $\xi\alpha x^{2r}=0$ are only possible if $\alpha x^{2r}=0$ already, proving the lemma. □

We may now give the main theorem of this subsection.

7.3 Theorem *There are extensions*

$$\rho\cdot\tau^{2(2^{n+1}k-r)}u^{2^{n+1}(2l+1)}h=\big(\tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\xi^{2r-1}+O(\rho)\big)x^{2^{n+2}-2r}+O(x^{2^{n+2}-2r+1}),$$
$$\xi\cdot\tau^{2(2^{n+1}k+r)}u^{2(2^n(2l+1)-r)}h=\big(\tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\rho^{2r-1}+O(\xi)\big)x^{2^{n+2}-2r}+O(x^{2^{n+2}-2r+1})$$

for $k,l\in\mathbb{Z}$, $n\geq 0$, and $1\leq r\leq 2^{n+1}-1$.

Proof It suffices to produce these extensions in the universal case $E=MP$. This ensures that the terms on the right are nonzero, so that these are nontrivial extensions. As discussed above, the cofiber sequences of (9) show that the terms

$$\tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\xi^{2r-1}x^{2^{n+2}-2r},\quad \tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\rho^{2r-1}x^{2^{n+2}-2r}$$

must be in the image of ρ and ξ respectively. By Lemma 7.1, this is not the case in $Z[x]/B$, so there must be hidden extensions making it so. In other words, there must be hidden extensions of the form

$$\rho\cdot\alpha=\big(\tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\xi^{2r-1}+O(\rho)\big)x^{2^{n+2}-2r}+O(x^{2^{n+2}-2r+1}),$$
$$\xi\cdot\beta=\big(\tau^{2^{n+2}k}u^{2^{n+2}l}h_{n+1}\rho^{2r-1}+O(\xi)\big)x^{2^{n+2}-2r}+O(x^{2^{n+2}-2r+1}),$$

where α and β are detected by classes in $Z[x]/B$ killed by ρ and ξ respectively. The error terms ensure that we do not need to pin down α and β precisely, but only the $\pi_0b(MPR)$–submodule of $Z[x]/B$ that they generate. By Lemma 7.2, the extensions given in the theorem statement are the only possibilities in these degrees. □

7.4 Remark This leaves open the problem of finding witnesses to the classes of (10) being in the image of ρ or ξ in the case where $r+s=2^{n+2}-1$ and r is even. In some cases no hidden extension is necessary, for example

$$\rho^{2^{n+1}}h_{n+1}x^{2^{n+1}-1}=w_nx^{2^{n+1}-1}=0,$$
$$\xi^{2^{n+1}}h_{n+1}x^{2^{n+1}-1}=\tau^{-2^{n+2}}u^{2^{n+1}}w_nx^{2^{n+1}-1}=0.$$

However, the general situation seems to be rather subtle. For example, for $h_2\xi^2x^5$ to be in the image of ρ, the only possibility is that $\rho\tau^{-4}u^4h_1x$ detects a class satisfying

$$\rho\cdot\rho\tau^{-4}u^4h_1x=h_2\xi^2x^5+O(\rho).$$

On the other hand,

$$\rho^2\tau^{-4}u^4h_1 = \tau^{-4}u^4w_0, \quad \tau^{-4}u^4w_0\cdot x = 0$$

in $Z[x]/B$. This indicates the existence of a mixed extension along the lines of

$$\rho^2\tau^{-4}u^4h_1 = \tau^{-4}u^4w_0 + h_1\xi^2x^4 + O(\rho).$$

Note that if $\theta \in \pi_\star b(ER)$ is detected by $\tau^{-4}u^4w_0$, then so is $\theta + h_1\xi^2x^4$. Thus for such an extension to even be defined, one must specify some information about how one lifts elements from Z to $\pi_\star b(ER)$, and these considerations are outside the scope of our investigation.

8 Some Mahowald invariants

We end by giving some examples of computations within the ring $\pi_\star b(ER)$. Our examples will center around the following definition.

8.1 Definition Given a spectrum A, the *A–based Mahowald invariant* is a multivalued function

$$R_A\colon \pi_*A \rightharpoonup \pi_*A,$$

ie a relation on π_*A, defined as follows: given $y \in \pi_n A$ and $z \in \pi_{n+k}A$, we say $z \in R_A(y)$ if z lifts to a class $\zeta \in \pi_\star b(A)$ such that $\rho^N y = \rho^{N+k}\zeta$ for $N \gg 0$, and moreover k is as large as possible.

8.2 Remark There are natural maps $\pi_n A \to \pi_n A^{tC_2}$ and $\pi_{c+w\sigma}b(A) \to \pi_c A^{tC_2}$, and the condition $\rho^N y = \rho^{N+k}\zeta$ for $N \gg 0$ amounts to asking that $y = \zeta$ in $\pi_* A^{tC_2}$. When $A = S$, this construction recovers the classical Mahowald invariant, commonly called the root invariant. See [Mahowald and Ravenel 1993] for additional background, [Bruner and Greenlees 1995] for the relation to C_2–equivariant homotopy theory, which connects Definition 8.1 to other definitions, [Behrens 2007] for the state of the art in S–based Mahowald invariants at the prime 2, [Quigley 2022] for further discussion of A–based Mahowald invariants with $A \neq S$, and [Li et al. 2022] for more information about spectra related to ER^{tC_2}.

Li, Shi, Wang and Xu [Li et al. 2019] prove that the Hurewicz image of real bordism detects the Hopf elements, Kervaire classes, and $\bar{\kappa}$ family. These are the elements in π_*S detected on the E_2–page of the Adams spectral sequence by the classes h_i, h_j^2, and g_{k+1} respectively; note there is no claimed relation between h_i here and the elements h_i in $\pi_\star b(E)$. These classes arrange into Sq^0 families, ie

(12) $$\mathrm{Sq}^0(h_i) = h_{i+1}, \quad \mathrm{Sq}^0(h_j^2) = h_{j+1}^2, \quad \mathrm{Sq}^0(g_{k+1}) = g_{k+2}.$$

Informally, this means that they arise as iterated Mahowald invariants at the level of Ext. Of course this cannot lift to the level of homotopy, as not all of these classes are permanent cycles; still, it is known that $\eta \in R_S(2)$, $\nu \in R_S(\eta)$, and $\sigma \in R_S(\nu)$, and it is conjectured that $\theta_{j+1} \in R_S(\theta_j)$ for $j \geq 3$ provided θ_{j+1} exists, see [Mahowald and Ravenel 1993, Proposition 2.4].

We can compute the iterated MPR–based Mahowald invariants of the classes 2, $\theta_0 = 4$, and $\bar{\kappa}$, yielding an analogue of (12). Our computation works just as well for ER in a range depending on the height of E. First we need to know how $\bar{\kappa}$ sits inside $\pi_\star MPR$.

8.3 Lemma *The class $\bar{\kappa}$ is detected by MPR, with Hurewicz image $u_2^4 u^8 x^4$.*

Proof If $\bar{\kappa}$ is detected by MPR, then it is detected by MR. As $\pi_{20} MR = \mathbb{Z}/(2)\{u_2^4 u^8 x^4\} \subset \pi_{20} MPR$, it suffices just to show that $\bar{\kappa}$ is detected by MR, which was shown in [Li et al. 2019]. Alternately, as there is a ring map $MR \to \mathrm{TMF}_0(3)$ [Hill and Meier 2017], it suffices to show that $\bar{\kappa}$ is detected in the latter, and here one may appeal to [Mahowald and Rezk 2009]. □

We now abbreviate $R = R_{MPR}$.

8.4 Theorem *Define elements*

$$a_n = u_n x^{2^n - 1} \in \pi_{2^n - 1} MPR, \quad b_m = u_{m+1}^4 u^{2^{m+2}} x^{2^{m+2}-4} \in \pi_{4(3 \cdot 2^m - 1)} MPR$$

for $n \geq 0$ and $m \geq 1$, so that for example $a_0 = 2$ and $b_1 = \bar{\kappa}$. Then there are MPR–based Mahowald invariants

$$a_{n+1} \in R(a_n), \quad a_{n+1}^2 \in R(a_n^2), \quad b_{m+1} \in R(b_m).$$

Proof First consider a_n. As

$$h_n \equiv u_n + \rho^{2^n} \xi^{2^n} h_{n+1} \pmod{u_0, \dots, u_{n-1}},$$

the relation $\rho^{2^n} h_n \cdot x^{2^n - 1} = 0$ implies

$$\rho^{2^n} \cdot u_n x^{2^n - 1} = -\rho^{2^{n+1}} \cdot h_{n+1} \xi^{2^n} x^{2^n - 1}. \tag{13}$$

There are no further relations and $h_{n+1} \xi^{2^n} x^{2^n - 1}$ lifts $u_{n+1} x^{2^{n+1}-1} = a_{n+1}$, yielding $a_{n+1} \in R(a_n)$. The case of a_n^2 is identical, only we must apply (13) twice:

$$\rho^{2^n} \cdot u_n^2 x^{2(2^n - 1)} = -\rho^{2^{n+1}} \cdot h_{n+1} \xi^{2^n} u_n x^{2(2^n - 1)} = \rho^{3 \cdot 2^n} \cdot h_{n+1}^2 \xi^{2^{n+1}} x^{2(2^n - 1)}.$$

Now consider b_m. As $2^{m+2} - 4 \geq 2^{m+1} - 1$ for $m \geq 1$, we may apply (13) thrice to obtain

$$\rho^{2^{m+1}} \cdot u_{m+1}^4 u^{2^{m+2}} x^{2^{m+2}-4} = \rho^{2^{m+3}} \cdot u_{m+1} u^{2^{m+2}} \cdot \xi^{3 \cdot 2^{m+1}} h_{m+2}^3 x^{2^{m+2}-4}. \tag{14}$$

At this point additional care is needed: we cannot apply (13) again, as despite appearances $u_{m+1} u^{2^{m+2}}$ is indecomposable. Instead, the relation $\rho \cdot h = 0$ gives

$$0 \equiv u_{m+1} \rho^{2^{m+2}-1} \tau^{-2^{m+2}+2} u^{2^{m+1}-1} + h_{m+2} \rho^{2^{m+3}-1} \tau^{-2^{m+3}+2} u^{2^{m+2}-1} \pmod{u_0, \dots, u_m}$$

in $\pi_\star b(MP)$, and thus

$$\begin{aligned} u_{m+1} u^{2^{m+2}} \cdot \xi^{2^{m+2}-1} &= u_{m+1} u^{2^{m+2}} \cdot \rho^{2^{m+2}-1} \tau^{-2^{m+3}-2} u^{2^{m+2}-1} \\ &\equiv \tau^{-2^{m+3}} u^{2^{m+3}} h_{m+3} \cdot \rho^{3 \cdot 2^{m+1}} \xi^{2^{m+1}-1} \pmod{u_0, \dots, u_m}. \end{aligned}$$

Substituting this into (14) yields

$$\rho^{2^{m+1}} \cdot u_{m+1}^4 u^{2^{m+2}} x^{2^{m+2}-4} = \rho^{7\cdot 2^{m+1}} \cdot \tau^{-2^{m+3}} u^{2^{m+3}} h_{m+2}^4 \cdot \xi^{2^{m+2}} x^{2^{m+2}-4}.$$

We cannot pull this class back any further. Thus, as $\tau^{-2^{m+3}} u^{2^{m+3}} h_{m+2}^4 \cdot \xi^{2^{m+2}} x^{2^{m+2}-4}$ lifts

$$u_{m+2}^4 u^{2^{m+3}} x^{2^{m+3}-4} = b_{m+1},$$

we obtain $b_{m+1} \in R(b_m)$. □

References

[Ando et al. 2004] **M Ando**, **M J Hopkins**, **N P Strickland**, *The sigma orientation is an H_∞ map*, Amer. J. Math. 126 (2004) 247–334 MR Zbl

[Araki 1979] **S Araki**, *Orientations in τ–cohomology theories*, Japan. J. Math. 5 (1979) 403–430 MR Zbl

[Balderrama 2021] **W Balderrama**, *The C_2–equivariant $K(1)$–local sphere*, preprint (2021) arXiv 2103.13895

[Balderrama 2024] **W Balderrama**, *Total power operations in spectral sequences*, Trans. Amer. Math. Soc. 377 (2024) 4779–4823 MR Zbl

[Banerjee 2010] **R Banerjee**, *Real Johnson–Wilson theories and non-immersions of projective spaces*, PhD thesis, Johns Hopkins University (2010) Available at `https://www.proquest.com/docview/365981348`

[Behrens 2007] **M Behrens**, *Some root invariants at the prime 2*, from "Proceedings of the Nishida Fest", Geom. Topol. Monogr. 10, Geom. Topol. Publ., Coventry (2007) 1–40 MR Zbl

[Bruner and Greenlees 1995] **R Bruner**, **J Greenlees**, *The Bredon–Löffler conjecture*, Exp. Math. 4 (1995) 289–297 MR Zbl

[Guillou et al. 2020] **B J Guillou**, **M A Hill**, **D C Isaksen**, **D C Ravenel**, *The cohomology of C_2–equivariant $\mathcal{A}(1)$ and the homotopy of* ko_{C_2}, Tunis. J. Math. 2 (2020) 567–632 MR Zbl

[Hahn and Shi 2020] **J Hahn**, **X D Shi**, *Real orientations of Lubin–Tate spectra*, Invent. Math. 221 (2020) 731–776 MR Zbl

[Hill and Meier 2017] **M A Hill**, **L Meier**, *The C_2–spectrum* $\mathrm{Tmf}_1(3)$ *and its invertible modules*, Algebr. Geom. Topol. 17 (2017) 1953–2011 MR Zbl

[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

[Hopkins et al. 2000] **M J Hopkins**, **N J Kuhn**, **D C Ravenel**, *Generalized group characters and complex oriented cohomology theories*, J. Amer. Math. Soc. 13 (2000) 553–594 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

[Kitchloo and Wilson 2007a] **N Kitchloo**, **W S Wilson**, *On fibrations related to real spectra*, from "Proceedings of the Nishida Fest", Geom. Topol. Monogr. 10, Geom. Topol. Publ., Coventry (2007) 237–244 MR Zbl

[Kitchloo and Wilson 2007b] **N Kitchloo**, **W S Wilson**, *On the Hopf ring for $ER(n)$*, Topology Appl. 154 (2007) 1608–1640 MR Zbl

[Kitchloo and Wilson 2008a] **N Kitchloo**, **W S Wilson**, *The second real Johnson–Wilson theory and nonimmersions of RP^n*, Homology Homotopy Appl. 10 (2008) 223–268 MR Zbl

[Kitchloo and Wilson 2008b] **N Kitchloo**, **W S Wilson**, *The second real Johnson–Wilson theory and nonimmersions of RP^n, II*, Homology Homotopy Appl. 10 (2008) 269–290 MR Zbl

[Kitchloo and Wilson 2015] **N Kitchloo**, **W S Wilson**, *The $ER(n)$–cohomology of $BO(q)$ and real Johnson–Wilson orientations for vector bundles*, Bull. Lond. Math. Soc. 47 (2015) 835–847 MR Zbl

[Kitchloo et al. 2017] **N Kitchloo**, **V Lorman**, **W S Wilson**, *Landweber flat real pairs and $ER(n)$–cohomology*, Adv. Math. 322 (2017) 60–82 MR Zbl

[Kitchloo et al. 2018a] **N Kitchloo**, **V Lorman**, **W S Wilson**, *The $ER(2)$–cohomology of $B\mathbb{Z}/(2^q)$ and $\mathbb{CP}^n$*, Canad. J. Math. 70 (2018) 191–217 MR Zbl

[Kitchloo et al. 2018b] **N Kitchloo**, **V Lorman**, **W S Wilson**, *Multiplicative structure on real Johnson–Wilson theory*, from "New directions in homotopy theory", Contemp. Math. 707, Amer. Math. Soc., Providence, RI (2018) 31–44 MR Zbl

[Landweber 1968] **P S Landweber**, *Conjugations on complex manifolds and equivariant homotopy of MU*, Bull. Amer. Math. Soc. 74 (1968) 271–274 MR Zbl

[Li et al. 2019] **G Li**, **X D Shi**, **G Wang**, **Z Xu**, *Hurewicz images of real bordism theory and real Johnson–Wilson theories*, Adv. Math. 342 (2019) 67–115 MR Zbl

[Li et al. 2022] **G Li**, **V Lorman**, **J D Quigley**, *Tate blueshift and vanishing for real oriented cohomology theories*, Adv. Math. 411 (2022) art. id. 108780 MR Zbl

[Lorman 2016] **V Lorman**, *The real Johnson–Wilson cohomology of $\mathbb{CP}^\infty$*, Topology Appl. 209 (2016) 367–388 MR Zbl

[Mahowald and Ravenel 1993] **M E Mahowald**, **D C Ravenel**, *The root invariant in homotopy theory*, Topology 32 (1993) 865–898 MR Zbl

[Mahowald and Rezk 2009] **M Mahowald**, **C Rezk**, *Topological modular forms of level 3*, Pure Appl. Math. Q. 5 (2009) 853–872 MR Zbl

[Quigley 2022] **J D Quigley**, *tmf-based Mahowald invariants*, Algebr. Geom. Topol. 22 (2022) 1789–1839 MR Zbl

Department of Mathematics, University of Virginia
Charlottesville, VA, United States

williamb@math.uni-bonn.de

Received: 13 March 2023 Revised: 3 July 2023 msp

Algebraic & Geometric Topology **24**:7 (2024) 4085–4101
DOI: 10.2140/agt.2024.24.4085
Published: 9 December 2024

Fourier transforms and integer homology cobordism

MIKE MILLER EISMEIER

We explore the *Fourier transform* of the d–invariants, which is particularly well behaved with respect to connected sum. As corollaries, we show that lens spaces are cancellable in the monoid of 3–manifolds up to integer homology cobordism, and we recover a theorem of González-Acuña and Short on Alexander polynomials of knots with reducible surgeries.

57K31, 57R90

1 Introduction

The relation of homology cobordism between 3–manifolds has a long and interesting history. Fix a ring R. Let Y and Y' be closed oriented 3–manifolds, and suppose W is a compact oriented 4–manifold whose boundary ∂W is oriented diffeomorphic to $Y' - Y$. If both maps

$$i_*\colon H_*(Y;R) \to H_*(W;R), \quad i'_*\colon H_*(Y';R) \to H_*(W;R)$$

are isomorphisms, we say that W is an R–homology cobordism and that Y and Y' are R–homology cobordant.

This relation is most well studied when $H_*(Y;R) \cong H_*(S^3;R)$, in which case Y is called an R–homology sphere. The set of R–homology spheres modulo R–homology cobordism form a group Θ^3_R called the "R–homology cobordism group". The group operation is connected sum, the neutral element is $[S^3]$, and the inverse of $[Y]$ is $[-Y]$.

If one instead considers the set of all 3–manifolds modulo R–homology cobordism, the resulting object is a *monoid*, which we denote by $\widehat{\Theta}^3_R$; the R–homology cobordism group is the submonoid of invertible elements.[1] Some things are known about this monoid; for instance, every equivalence class contains an irreducible 3–manifold [Livingston 1981] or better yet a hyperbolic 3–manifold [Myers 1983]. In another direction, there are obstructions to finding a Seifert-fibered manifold in a given equivalence class [Cochran and Tanner 2014] or more generally to finding a graph manifold whose graph is a tree in a given equivalence class [Doig and Horn 2017].[2]

In this note, we will investigate integer homology cobordism between a class of 3–manifolds which are not integer homology spheres. In what follows, we suppress the ring $R = \mathbb{Z}$ from notation, and write integer homology and cohomology groups as $H_*(Y)$ and $H^*(Y)$.

Theorem 1 *Suppose L and L' are connected sums of lens spaces. If L and L' are integer homology cobordant by a cobordism W, then L is oriented diffeomorphic to L', and the induced map $W_*\colon H_1(L) \to H_1(L')$ respects the natural direct sum decompositions.*[3]

Further, if Y is any closed, oriented 3–manifold and $L \# Y$ is integer homology cobordant to $L' \# Y$, then L is oriented diffeomorphic to L'.

The first part of the result, that the oriented diffeomorphism type of L and L' is determined by their integer homology cobordism type, is not new. It follows from the more general results of [Greene 2013] on alternating links, and indeed Greene's results imply that double-branched covers of alternating links are determined by their homology cobordism type. Independent proofs of the more restrictive claim that the oriented diffeomorphism type of a lens space is determined by its d–invariants have also appeared [Doig and Wehrli 2015; Némethi 2005]. Even before then, the integer homology cobordism classification of lens spaces of odd order goes back to [Fintushel and Stern 1987].

Our argument is independent of [Greene 2013], depending only on the computation of Reidemeister torsion for lens spaces and its relationship to their d–invariants established in [Némethi 2005]. Theorem 1 is also stronger in two ways: first, it constraints the structure of the span $H_1(L) \leftarrow H_1(W) \to H_1(L')$ of any homology cobordism relating L and L'; second, it establishes that connected sums of lens spaces are *cancellable* in $\widehat{\Theta}^3_{\mathbb{Z}}$.

The proof of this theorem is presented in Section 4. The key point is that — provided a certain nonvanishing property holds — one can recover (a reduced version of) the d–invariants of a connected summand from those of a connected sum.

[1] If $H_1(Y;R)$ is nonzero, then Y is not invertible in $\widehat{\Theta}^3_R$; the purported inverse Y' should support an R–homology cobordism between $Y \# Y'$ and S^3, but $|H_1(Y \# Y';R)| \geq |H_1(Y;R)| > 1 = |H_1(S^3;R)|$.

[2] Though see the MR review of [Doig and Horn 2017] for some errata and [Suciu 2022, Proposition 9.2] for a simplified argument.

[3] This is meant in an unordered sense. Precisely, suppose $H_1(L) \cong \bigoplus_i H_1(L_i)$ and similarly for L'. Then W_* sends each $H_1(L_i)$ isomorphically onto some $H_1(L'_j)$.

To recover the d–invariants of the summands, we find it useful to pass to the *Fourier transform*. Given a function $f\colon A \to \mathbb{C}$ on a finite abelian group, its Fourier transform is instead a function on the *dual group* $A^\vee = \mathrm{Hom}(A, S^1)$, defined by $\hat{f}(\phi) = \frac{1}{|A|}\sum_{a\in A} f(a)\overline{\phi(a)}$. Here A will be $H^2(Y) \cong H_1(Y)$, and we will pick a base spin^c structure to consider the d–invariants as a function on $H_1(Y)$.

The use of Fourier transforms is well established in the theory of Reidemeister torsion: analytic interpretations of the Reidemeister torsion (eg [Fried 1987; Ray and Singer 1971]) interpret the torsion as a function of oriented flat bundles, and hence take as input a representation $\phi\colon H_1(Y;\mathbb{Z}) \to S^1$. It is also profitable to rephrase the surgery relation in terms of Fourier transforms; Nicolaescu [2004] used this to compare Reidemeister torsion to a Seiberg–Witten invariant. See also the discussion peppered throughout [Nicolaescu 2003].

Here we use a simple property of Fourier transforms. Given two groups A and A' and functions $f\colon A \to \mathbb{C}$ and $f'\colon A' \to \mathbb{C}$, the *direct sum* $(f \oplus f')(a, a') = f(a) + f(a')$ on $A \times A'$ has an especially simple Fourier transform; one can effectively read off the values of $\hat{f}(\phi)$ and $\hat{f}'(\psi)$ from the knowledge of $\widehat{f \oplus f'}$ whenever ϕ or ψ are nontrivial homomorphisms. See Proposition 12 for a precise statement.

Applying this purely algebraic observation to d–invariants, one can recover (a reduced version of) the d–invariants of summands from those of a connected sum. Provided they satisfy a certain nonvanishing property, this recovery process is well defined up to automorphism of $H^2(Y)\times H^2(Y')$, and thus preserved by integer homology cobordisms. Because the d–invariants of lens spaces satisfy this nonvanishing property, and these reduced d–invariants — equivalent to Reidemeister torsion for L–spaces — classify lens spaces up to oriented diffeomorphism, the main theorem follows.

In fact, the proof of Theorem 1 yields a stronger claim: there exist monoid homomorphisms $c_{p,q}\colon \widehat{\Theta}^3_{\mathbb{Z}} \to \mathbb{N}$ with $c_{p,q}(L(p,q)) = 1$ and $c_{p,q}(L(r,s)) = 0$ unless $L(r,s)$ is oriented diffeomorphic to $L(p,q)$. The purely algebraic part of this claim is the content of Corollary 19 in Section 2, while the relevant computation for lens spaces is given in Proposition 26 in Section 4.

If one considers the *Grothendieck group* $\mathrm{Gr}(\widehat{\Theta}^3_{\mathbb{Z}})$, the group whose elements are pairs $([Y],[Z])$ with $([Y],[Z]) = ([Y'],[Z'])$ if there is an integer homology cobordism $Y \# Z' \sim Y' \# Z$, the existence of these homomorphisms $c_{p,q}$ shows that lens spaces (up to oriented diffeomorphism) span a $\mathbb{Z}^\infty$ summand of the Grothendieck group.

As an aside, the perspective of Fourier transforms appears useful whenever one has an invariant which is additive in the sense above, including both d–invariants and Reidemeister torsion. To demonstrate this, in Section 3 we reprove [González-Acuña and Short 1986, Theorem 2.2]: if K is a knot with reducible surgery $S_n(K) \cong Y \# Y'$ with $H_1(Y) = \mathbb{Z}/p$ and $H_1(Y') = \mathbb{Z}/q$, the Alexander polynomial of K is divisible by that of the (p,q) torus knot; $\Delta(T_{p,q}) \mid \Delta(K)$. We hope that Fourier transforms can be a useful organizing tool in other contexts as well.

Acknowledgements The author would like to thank Danny Ruberman for a useful discussion during the preparation of this note, as well as Tye Lidman for comments on an early draft and suggesting that one might reprove [González-Acuña and Short 1986, Theorem 2.2] using the Fourier transform technique. The basic algebraic observation used here came about during conversations with Aiden Sagerman on a related question about products of punctured lens spaces.

2 Weighted torsors and Fourier transforms

In this section we cover the purely algebraic aspects of the main result: the context of weighted torsors, the definition of the Fourier transform, and the process of recovering summands of the direct sum of weighted torsors using Fourier transforms.

2.1 Weighted torsors

Here we rapidly define the relevant algebraic objects. Though the initial discussion about torsors is valid for any group, the relevant groups for us will be abelian, so we use additive notation.

Definition 2 A *torsor* is a pair (A, S), where A is a group and S carries a free and transitive (right) action by A.

A choice of element $s \in S$ gives rise to a bijection $m_s\colon A \cong S$, given by sending $a \mapsto s + a$; for a different choice of element $s' \in S$ with $s' = s + a_0$, the bijections differ by $(m_s^{-1} m_{s'})(a) = a_0 + a$.

This observation shows that one may think of a torsor as a group where one has forgotten which element is the identity, or as A "modulo translation".

Definition 3 An *isomorphism of torsors* is a pair $(f, g)\colon (A, S) \to (A', S')$, where $f\colon A \to A'$ is a group isomorphism and $g\colon S \to S'$ is a function satisfying

$$g(s + a) = g(s) + f(a).$$

By transitivity of the group actions, g is necessarily a bijection. If one chooses basepoints $s \in S$ and $s' \in S'$, and we have $g(s) = s' + a'$, then the map $m_{s'}^{-1} g m_s\colon A \to A$ is given by

$$(m_{s'}^{-1} g m_s)(a) = m_{s'}^{-1} g(s + a) = m_{s'}^{-1}(g(s) + f(a)) = m_{s'}^{-1}(s' + a' + f(a)) = a' + f(a).$$

Thinking of S as a group where we've forgotten the identity element (or as a sort of affine space), one should imagine g to be an affine function whose "linear part" is the homomorphism f; indeed, one can recover f from g.

Remark 4 The group of automorphisms of (A, S) is a group sometimes called the *holomorph* of A, and can be understood as the group of affine automorphisms of A.

Definition 5 A *weighted torsor* is a torsor (A, S) equipped with a function $d\colon S \to \mathbb{C}$. If $S = A$, we call (A, d) a *weighted group*.

If one chooses a basepoint $s \in S$, we write $d_s \colon A \to \mathbb{C}$ for the function $d_s(a) = (dm_s)(a) = d(s+a)$. For a different choice of basepoint $s' = s + a_0$, we have $d_{s'}(a) = d_s(a + a_0)$.

Definition 6 An *isomorphism of weighted torsors* $(A, S, d) \to (A', S', d')$ is a torsor isomorphism $(f, g) \colon (A, S) \to (A', S')$ which has $d'(g(s)) = d(s)$.

If one prefers to think entirely in terms of the group A (having chosen an arbitrary basepoint), a weighted torsor is a function $d \colon A \to \mathbb{C}$, with d considered equivalent to $d_b(a) = d(a+b)$ for any $b \in A$. In this perspective, an isomorphism of weighted torsors $(A, d) \cong (A', d')$ amounts to an *affine* isomorphism $f + a' \colon A \to A'$ which has

$$d'(f(a) + a') = d(a).$$

2.2 Fourier transforms and weighted duals

Given an abelian group A, its *Pontryagin dual* is the group $A^\vee = \operatorname{Hom}(A, S^1)$.

Convention For the rest of this note, abelian groups A and torsors (A, S) are assumed to be *finite*. This is true in all cases of interest to us, and simplifies discussions of Fourier transforms.

Given a function $d \colon A \to \mathbb{C}$, we can take its *Fourier transform* $\hat{d} \colon A^\vee \to \mathbb{C}$, defined as

$$\hat{d}(\phi) = \frac{1}{|A|} \sum_{a \in A} d(a) \overline{\phi(a)}.$$

Remark 7 This differs from Nicolaescu's definition [2003, Section 1.6] by a scalar factor of $1/|A|$. Our definition makes some important formulas later slightly simpler.

If $d' \colon A \to \mathbb{C}$ is defined by $d'(a) = d(a + a')$, then

$$\hat{d}'(\phi) = \frac{1}{|A|} \sum_{a \in A} d'(a) \overline{\phi(a)} = \frac{1}{|A|} \sum_{a \in A} d'(a - a') \overline{\phi(a - a')} = \frac{1}{|A|} \sum_{a \in A} d(a) \overline{\phi(a)} \phi(a') = \hat{d}(\phi)\phi(a').$$

This computation inspires the following definition, which we phrase intrinsically on the dual $B = A^\vee$; the statement below implicitly uses the isomorphism $(A^\vee)^\vee \cong A$ for finite A. The terminology follows [Nicolaescu 2003, Definition 3.22] (though notice that Nicolaescu allows for a sign ambiguity, and we do not).

Definition 8 If B is an abelian group equipped with weights $\hat{d}$ and $\hat{d}'$, and there exists some $\psi \in B^\vee$ such that $\hat{d}'(b) = \hat{d}(b)\psi(b)$ for all $b \in B$, we say that d and d' are *t–equivalent*. We say that weighted groups $(B, \hat{d})$ and $(B', \hat{d}')$ are *t–isomorphic* if there exists an isomorphism $f \colon B \to B'$ and element $\phi \in B^\vee$ such that

$$\hat{d}'(f(b)) = \hat{d}(b)\phi(b)$$

for all $b \in B$.

The discussion above shows that given a weighted torsor (A, S, d), choosing a basepoint $s \in S$ and taking the Fourier transform of d_s gives us a weighted group $(A^\vee, \hat{d}_s)$, well defined up to t–equivalence. Furthermore, it is clear that isomorphic weighted torsors give rise to t–isomorphic weighted groups.

Notice that $\hat{d}(1) = \sum_{a \in A} d(a)$, and if $\hat{d}$ is t–isomorphic to $\hat{d}'$, then $\hat{d}(1) = \hat{d}'(1)$. Later, this term will cause us some minor irritation, so we do what we can to remove it.

Definition 9 A weighted torsor (A, S, d) is *reduced* if $\sum_{s \in S} d(s) = 0$; equivalently, $\hat{d}(1) = 0$. Given any weighted torsor (A, S, d), its *reduced part* is given by (A, S, d^r), where

$$d^r(s) = d(s) - \frac{1}{|A|} \sum_{s' \in S} d(s');$$

that is, we subtract off the average value.

It is clear that d^r is reduced, and that reduction doesn't change the value of the Fourier transform at any $\phi \neq 1$. To see this, we need a small useful lemma.

Lemma 10 *If A is a finite abelian group and $\phi \colon A \to S^1$ is a homomorphism, then*

$$\sum_{a \in A} \phi(a) = \begin{cases} 0 & \text{if } \phi \neq 1, \\ |A| & \text{if } \phi = 1. \end{cases}$$

Proof This is a special case of the orthogonality relations for irreducible characters [Serre 1977, Theorem 2.3.3]. The proof is included for completeness.

If ϕ is trivial this is obvious. For ϕ nontrivial, write ζ for a generator of $\phi(A)$ so that ζ is a primitive m^{th} root of unity for some $m > 1$. We have

$$\sum_{a \in A} \phi(a) = \frac{|A|}{m} \sum_{k=0}^{m-1} \zeta^k.$$

But $\sum_{k=0}^{m-1} \zeta^k = (1 - \zeta^m)/(1 - \zeta) = 0$ for $\zeta \neq 1$ a nontrivial m^{th} root of unity. □

It follows that if two weighted torsors have $d'(a) = d(a) + c$ for all $a \in A$ and some constant c, then

$$\hat{d}'(\phi) = \frac{1}{|A|} \sum_{a \in A} d(a)\overline{\phi(a)} + \frac{1}{|A|} \sum_{a \in A} c\overline{\phi(a)} = \begin{cases} \hat{d}(1) + c & \text{if } \phi = 1, \\ \hat{d}(\phi) & \text{if } \phi \neq 1. \end{cases}$$

Corollary 11 *If (A, S, d) is a weighted torsor, its reduced part (A, S, d^r) satisfies*

$$\hat{d}^r(\phi) = \begin{cases} 0 & \text{if } \phi = 1, \\ \hat{d}(\phi) & \text{if } \phi \neq 1. \end{cases}$$

2.3 Direct sums of weighted torsors

Given two weighted torsors (A, S, d) and (A', S', d') we say their *direct sum* is the weighted torsor $(A \times A', S \times S', d \oplus d')$, where

$$(d \oplus d')(s, s') = d(s) + d(s').$$

Notice that the $(A \times A')^\vee$ is naturally isomorphic to $A^\vee \times (A')^\vee$; if $\phi\colon A \to S^1$ and $\psi\colon A' \to S^1$ are homomorphisms, these give rise to the homomorphism $\phi\psi\colon A \times A' \to S^1$ by pointwise multiplication, $(\phi\psi)(a,a') = \phi(a)\psi(a')$.

The observation which motivated the present note is the following calculation of the Fourier transform of a direct sum of weighted torsors.

Proposition 12 *If $d \oplus d'\colon A \times A' \to \mathbb{Q}$ is the direct sum of two weighted groups, the Fourier transform satisfies*

$$\widehat{(d \oplus d')}(\phi\psi) = \begin{cases} \hat{d}(\phi) & \text{if } \phi \neq 1 \text{ and } \psi = 1, \\ \hat{d}'(\psi) & \text{if } \phi = 1 \text{ and } \psi \neq 1, \\ \hat{d}(1) + \hat{d}'(1) & \text{if } \phi = \psi = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof We have

$$\begin{aligned} \widehat{(d \oplus d')}(\phi\psi) &= \frac{1}{|A||A'|} \sum_{a,a'} (d \oplus d')(a,a') \overline{\phi\psi(a,a')} \\ &= \frac{1}{|A||A'|} \sum_{a,a'} (d(a) + d'(a')) \overline{\phi(a)}\,\overline{\psi(a')} \\ &= \frac{1}{|A||A'|} \biggl(\sum_{a,a'} d(a) \overline{\phi(a)}\,\overline{\psi(a')} \biggr) + \frac{1}{|A||A'|} \biggl(\sum_{a,a'} d'(a') \overline{\psi(a')}\,\overline{\phi(a)} \biggr) \\ &= \biggl(\frac{1}{|A|} \sum_{a} d(a) \overline{\phi(a)} \biggr) \biggl(\frac{1}{|A'|} \sum_{a'} \bar{\psi}(a') \biggr) + \biggl(\frac{1}{|A'|} \sum_{a'} d'(a') \overline{\psi(a')} \biggr) \biggl(\frac{1}{|A|} \sum_{a} \bar{\phi}(a) \biggr). \end{aligned}$$

By Lemma 10, the first term vanishes when $\psi \neq 1$ and is $|A'|\hat{d}(\phi)$ when $\psi = 1$, while the second term vanishes when $\phi \neq 1$ and is $\hat{d}'(\psi)$ when $\phi = 1$. This gives the stated claim in all cases except $\phi, \psi = 1$; in that case, it gives $\hat{d}(1) + \hat{d}'(1)$. □

In particular, for *nontrivial* ϕ and ψ, one can read off the values of $\hat{d}(\phi)$ and $\hat{d}'(\psi)$ from the Fourier transform of the direct sum $\widehat{d \oplus d'}$. In the nonreduced case, the fact that $\hat{d}(1)$ and $\hat{d}'(1)$ are combined in $\widehat{(d \oplus d')}(1)$ means that we cannot recover this information from the Fourier transform of the direct sum. This is why we restrict attention to reduced weighted torsors.

2.4 Nonvanishing properties and recovering invariants of summands

We will now be more precise about the process of recovering the summands of a direct sum of weighted torsors in a way which is well defined up to isomorphism. To do so, we must make some further assumptions.

Definition 13 A weighted group $(B, \hat{d})$ has the *nonvanishing property* if $\hat{d}(b) \neq 0$ for all *nontrivial* elements $b \in B$.

Notice that this property is well defined up to t–isomorphism, because if $(B,\hat{d})$ and $(B',\hat{d}')$ are t–isomorphic, we have an isomorphism $f\colon B\to B'$, an element $\phi\in B^\vee$, and an equality $\hat{d}'(f(b))=\hat{d}(b)\phi(b)$. Because b is nontrivial if and only if $f(b)$ is, and $\phi(b)\in S^1$ is nonzero, $\hat{d}'$ has the nonvanishing property if and only if $\hat{d}$ does.

Definition 14 If $(B,\hat{d})$ is a weighted group, a *special subgroup* is a *nontrivial* subgroup $C\subset B$ such that $\hat{d}(c)\neq 0$ for all nontrivial $c\in C$. A *maximal special subgroup* is a special subgroup which is maximal among special subgroups.

Notice that special subgroups are well defined up to t–equivalence of weights, and that t–isomorphism preserves maximal special subgroups: if $f\colon(B,\hat{d})\to(B',\hat{d}')$ is a t–isomorphism and $C\subset B$ is a special subgroup, then $f(C)$ is too, and vice versa. Further, a t–isomorphism maps $(C,\hat{d}|_C)$ t–isomorphically onto $(f(C),\hat{d}'|_{f(C)})$. In particular, the maximal special subgroups (considered as weighted groups up to t–isomorphism) are t–isomorphism invariants of $(B,\hat{d})$.

It immediately follows from this that isomorphisms between direct sums of weighted torsors with the nonvanishing property are rather constrained.

Corollary 15 *Suppose $\{(A_i,d_i)\}_{i=1}^n$ is a collection of weighted groups whose Fourier transforms satisfy the nonvanishing property, and similarly with $\{(A'_j,d'_j)\}_{j=1}^m$. Write $(A,d)=\bigoplus_{i=1}^n(A_i,d_i)$ and similarly for (A',d'). If there is an affine isomorphism of weighted groups $\varphi+a'\colon(A,d)\cong(A',d')$, then $n=m$ and the map φ preserves the direct sum decompositions, in the sense that for all i we have $\varphi(A_i)=A'_j$ for some j.*

Proof By Proposition 12, the maximal special subgroups of $A^\vee=A_1^\vee\oplus\cdots\oplus A_n^\vee$ are precisely the coordinate axes $A_i^\vee$ (where all coordinates but the i^{th} are nonzero). Comparing the number of maximal special subgroups, we see that $m=n$. Because $\varphi^\vee$ maps maximal special subgroups bijectively to maximal special subgroups, for some permutation σ we have $\varphi^\vee((A'_{\sigma(i)})^\vee)=A_i^\vee$ for all i. That is, if $\psi'\colon A'\to S^1$ is any homomorphism, then ψ' factors through $\pi'_{\sigma(i)}\colon A'\to A'_{\sigma(i)}$ if and only if there exists some $\psi\colon A_i\to S^1$ with

$$\psi\pi_i=\psi'\pi'_{\sigma(i)}\varphi.$$

It will follow that $\varphi(A_i)\subset A'_{\sigma(i)}$, and then equality follows because these have the same cardinality (their duals do) and φ is injective. To see this first claim, pick $x_i\in A_i$, and consider $y_j=\pi_j\varphi(x_i)$. If y_j is nonzero, there is some homomorphism $\psi'_j\colon A'_j\to S^1$ with $\psi'_j(y_j)\neq 0$. By the discussion above,

$$\psi'_j(y_j)=\psi'_j\pi'_j\varphi(x_i)=\psi\pi_{\sigma^{-1}(j)}(x_i).$$

This can only be nonzero if $\sigma^{-1}(j)=i$ by assumption, so $\varphi(x_i)$ indeed lies in $A'_{\sigma(i)}$. □

We will prove the main theorem similarly, by counting maximal special subgroups (considered as weighted groups up to t–isomorphism); we introduce notation for this special concept.

Definition 16 Given a weighted group $(B, \hat{d})$, we associate the multiset

$$MS(B, \hat{d}) = \{[C, \hat{d}|_C] \mid C \subset B \text{ is a maximal special subgroup}\}$$

of special subgroups of B equipped with the restriction of $\hat{d}$, considered up to t–isomorphism.

Recall here that a multiset M is a set (by an abuse of notation written with the same name M) where each element $x \in M$ is equipped with a weight $w_M(x) \geq 1$ labeling how many times it occurs in the multiset.

Notice that if $(B, \hat{d})$ is t–isomorphic to $(B', \hat{d}')$, then the multisets $MS(B, \hat{d})$ and $MS(B', \hat{d}')$ are isomorphic (there is a weight-preserving bijection between them). If (A, S, d) is a weighted torsor, the multiset $MS(A^\vee, \hat{d}_s)$ is an invariant of (A, S, d); isomorphic weighted torsors give rise to isomorphic multisets. We write $MS(A, S, d)$ for $MS(A^\vee, \hat{d}_s)$ for some choice of $s \in S$.

If M and N are multisets, we write $M \cup N$ for the multiset whose underlying set is the union of the underlying sets of M and N, and whose weight is $w_{M \cup N}(x) = w_M(x) + w_N(x)$. (Here we write $w_N(x) = 0$ if x does not lie in N, and similarly with M.)

The crucial observation, almost immediate from Proposition 12, is that this multiset is additive.

Proposition 17 *If (A, S, d) and (A', S', d') are **reduced** weighted torsors, then*

$$MS(A \times A', S \times S', d \oplus d') = MS(A, S, d) \cup MS(A', S', d').$$

Proof For convenience, we write $d^\oplus = (d \oplus d')$, and $\hat{d}^\oplus$ for its Fourier transform.

As mentioned above, Proposition 12 implies that the maximal special subgroups of $(A \times A')^\vee \cong A^\vee \times (A')^\vee$ are precisely the maximal special subgroups of $A^\vee \times \{1\}$ and $\{1\} \times (A')^\vee$.

Given a maximal special subgroup $C \subset A^\vee$ (or similarly $C' \subset (A')^\vee$), what remains is to compare the restriction of $\hat{d}$ to C with the restriction of $\hat{d}^\oplus$ to $C \times \{1\}$, but

$$\hat{d}^\oplus|_{C \times \{1\}} = \hat{d}|_C$$

by the formula from Proposition 12 and the assumption that d and d' are *reduced* weighted torsors. □

We can now define monoid homomorphisms from the appropriate monoid to $\mathbb{N}$.

Definition 18 We write $\widehat{\Theta}_{\mathrm{WT}}$ for the monoid whose elements are weighted torsors up to isomorphism, and whose product operation is direct sum.

There is a corresponding monoid $\widehat{\Theta}_{\mathrm{RWT}}$ of reduced weighted torsors, and the map $d \mapsto (d^r, \operatorname{avg} d)$ defines a monoid isomorphism $\widehat{\Theta}_{\mathrm{WT}} \cong \widehat{\Theta}_{\mathrm{RWT}} \times \mathbb{C}$.

Write RWTN for the set of reduced weighted torsors whose Fourier transforms satisfy the nonvanishing property, considered up to isomorphism; because these are considered up to isomorphism, we may think of these as weighted groups up to affine isomorphism and drop the torsor S from notation.

For each $[A,d] \in \mathrm{RWTN}$, we define a map $c_{A,d} \colon \widehat{\Theta}_{\mathrm{WT}} \to \mathbb{N}$ by

$$c_{[A,d]}(B,T,f) = \#\text{ of occurrences of } [A^\vee, \hat{d}] \text{ in } MS(B,T,f^r).$$

That is, $c_{[A,d]}(B,T,f)$ is the weight $w_{MS(B,T,f^r)}([A^\vee,\hat{d}])$.

Corollary 19 *The functions $c_{[A,d]}$ are monoid homomorphisms. If (A',d') is another reduced weighted torsor whose Fourier transform has the nonvanishing property, then*

$$c_{[A,d]}(A',d') = \begin{cases} 1 & \text{if } (A,d) \text{ is isomorphic to } (A',d'), \\ 0 & \text{otherwise.} \end{cases}$$

Proof A maximal special subgroup is defined to be nontrivial, so for the trivial weighted torsor with underlying group 1 and zero weighting, $MS(1,0) = \varnothing$; so $c_{[A,d]}(1,0) = 0$ and thus c sends neutral element to neutral element. Additivity follows immediately from Proposition 17 and the fact that taking the reduced part $d \mapsto d^r$ commutes with direct sums. So $c_{[A,d]}$ is a monoid homomorphism.

Because the Fourier transform of (A',d') has the nonvanishing property, $MS(A,d) = \{[(A')^\vee, \hat{d}']\}$. If $[A^\vee,\hat{d}]$ appears in this singleton set, then in fact $((A')^\vee,\hat{d}')$ is t–isomorphic to $(A^\vee,\hat{d})$, and hence $(A,d) \cong (A',d')$. □

It follows that the functions c assemble into a surjective monoid homomorphism $c \colon \widehat{\Theta}_{\mathrm{WT}} \to \mathbb{N}^{\mathrm{RWTN}}$, which behaves particularly well on reduced weighted torsors whose Fourier transforms have the nonvanishing property: there is a map $\mathbb{N}^{\mathrm{RWTN}} \to \widehat{\Theta}_{\mathrm{WTN}}$ whose composition with c is the identity.

3 A theorem of González-Acuña and Short

Before moving on to the main theorem, we use this opportunity to give an alternative proof of [González-Acuña and Short 1986, Theorem 2.2], suggested to the author by Tye Lidman.

For context, if Σ is a homology sphere and $K = C_{p,q}(K')$ is the cable of another knot $K' \subset \Sigma$, then the pq–surgery satisfies $\Sigma_{pq}(K) \cong \Sigma_{p/q}(K') \# L(q,p)$. When $\Sigma = S^3$, the cabling conjecture [González-Acuña and Short 1986, Conjecture A] predicts that this construction gives the *only examples* of knots with reducible surgery. Among their evidence was the following theorem.

Theorem 20 *Let $K \subset \Sigma$ be a knot in an integer homology sphere with reducible surgery*

$$\Sigma_{n/m}(K) \cong Y_1 \# Y_2,$$

where $|H_1(Y_1)| = p > 1$ and $|H_1(Y_2)| = q > 1$. Then the polynomial

$$\Delta_{p,q} = \frac{(t^{pq}-1)(t-1)}{(t^p-1)(t^q-1)}$$

divides the Alexander polynomial Δ_K.

Note that the Alexander polynomial of a cable knot $K = C_{p,q}(K')$ satisfies $\Delta_K = \Delta_{p,q}\Delta_{K'}$.

Proof The proof makes use of the Reidemeister torsion of a 3–manifold, and we quickly recall some properties from [Nicolaescu 2003, Section 3.7]. When Y is a rational homology sphere, its Reidemeister torsion $T_Y\colon H_1(Y)\to\mathbb{Q}$ makes $H_1(Y)$ into a weighted torsor, well defined up to isomorphism and multiplication by ±1. When $H_1(Y)\cong\mathbb{Z}$, by contrast, $T_Y(t)$ should be understood as a rational function $T_Y\in\mathbb{Q}(t)$, well defined up to multiplication by $\pm t^k$. The Fourier transform $\widehat{T}_Y:\mathbb{C}\to\mathbb{C}$ is a meromorphic function defined by evaluating T_Y on a given complex number, and is well defined up to a variation on t–equivalence, $\widehat{T}'(z)\sim\pm z^k\widehat{T}(z)$.

In the case of knot complements, the Reidemeister torsion is related to the Alexander polynomial by the formula

$$T_{\Sigma\setminus K}(t)=\pm\frac{t^k\Delta_K(t)}{1-t};$$

this first appeared as [Milnor 1962, Theorem 4].

We will use the surgery formula for the Fourier-transformed Reidemeister torsion as stated in [Nicolaescu 2003, Theorem 3.23]: if ζ is a primitive n^{th} root of unity, then

$$\widehat{T}_{\Sigma_{n/m}(K)}(\zeta)=\frac{\widehat{T}_{\Sigma\setminus K}(\zeta)}{(1-\zeta)^{-1}}=\pm\frac{\zeta^k\Delta_K(\zeta)}{(1-\zeta^{-1})(1-\zeta)}.$$

In particular, the zeroes of $\widehat{T}_{\Sigma_{n/m}(K)}$ are identified with the n^{th} roots of unity ζ for which $\Delta_K(\zeta)=0$.

Here we use the canonical isomorphism $H_1(\Sigma_{n/m}(K))\cong\mathbb{Z}/n$, sending a meridian of the knot in $\Sigma\setminus K$ to 1 to identify $H_1^\vee$ with the group of n^{th} roots of unity.

Suppose K is a knot as in the statement of the theorem. The isomorphism $\mathbb{Z}/n\cong\mathbb{Z}/p\times\mathbb{Z}/q$ induced by the connected sum decomposition sends the elements (i,j) with i and j nontrivial to n^{th} roots of unity which are neither p^{th} nor q^{th} roots of unity. For rational homology spheres Y_1 and Y_2 we have $T_{Y_1\#Y_2}(i,j)=T_{Y_1}(i)+T_{Y_2}(j)$ [Turaev 2002, Theorem XII.1.2]. It follows from Proposition 12 that $\widehat{T}_{Y_1\#Y_2}(\zeta)=0$ for any n^{th} root of unity ζ which is neither a p^{th} nor q^{th} root of unity. Thus $\Delta_K(\zeta)=0$ for all such roots of unity, so Δ_K is divisible by

$$\prod_{\substack{\zeta=e^{2\pi ik/n}\\ 0\le k<n\\ p\nmid k\text{ and }q\nmid k}}(t-\zeta)=\frac{(t^{pq}-1)(t-1)}{(t^p-1)(t^q-1)}=\Delta_{p,q}$$

as claimed. □

4 d–invariants of 3–manifolds

If Y is a 3–manifold, there is a naturally associated torsor $(H^2(Y)_{\text{tors}};\text{Spin}^c(Y)_{\text{tors}})$, where the latter is the set of spin^c structures with torsion first Chern class. When Y is a rational homology sphere, every spin^c structure is torsion.

When we refer to a homology cobordism, we mean a pair (W, φ) of a compact oriented connected 4–manifold and a *chosen* orientation-preserving diffeomorphism $\varphi: \partial W \cong Y - Y'$ such that the corresponding maps $Y \to W$ and $Y' \to W$ induce isomorphisms on all integer cohomology groups.

Given a homology cobordism $W: Y \to Y'$ there is an induced isomorphism of torsors

$$(W_*, W_*^c): (H^2(Y)_{\mathrm{tors}}, \mathrm{Spin}^c_{\mathrm{tors}}(Y)) \to (H^2(Y')_{\mathrm{tors}}; \mathrm{Spin}^c_{\mathrm{tors}}(Y')).$$

We will make use of three weighted torsors; the first and third are associated to rational homology spheres, while the second is associated to an arbitrary 3–manifold:

- the *d–invariant* $d_Y: \mathrm{Spin}^c(Y) \to \mathbb{Q}$ [Ozsváth and Szabó 2003, Definition 4.1];
- the *twisted d–invariant* $\underline{d}_Y: \mathrm{Spin}^c_{\mathrm{tors}}(Y) \to \mathbb{Q}$ [Behrens and Golla 2018, Definition 3.1];
- the *Turaev–Reidemeister torsion* $T_Y: \mathrm{Spin}^c(Y) \to \mathbb{Q}$ [Turaev 2002, Chapter X].

Remark 21 The Turaev–Reidemeister torsion is often written as an $H^2(Y)$–equivariant map

$$\mathrm{Spin}^c(Y) \to \mathbb{Q}[H^2(Y)],$$

eg [Turaev 2002, Chapter I.4.1]. This gives rise to the function T_Y above by extracting the coefficient of $0 \in H^2(Y)$. This can be extended to an arbitrary 3–manifold, but the discussion is somewhat more intricate when $H^2(Y)$ is infinite: instead, the torsion defines an $H^2(Y)$–equivariant map to the fraction field $\mathbb{Q}(H^2(Y))$.

The twisted d–invariant is only used for a technical reason, to allow connected sums with arbitrary 3–manifolds instead of merely rational homology spheres. When Y is a rational homology sphere, we have the tautological equality $\underline{d}_Y(\mathfrak{s}) = d_Y(\mathfrak{s})$. The Turaev–Reidemeister torsion — and the relation to d–invariants — will be used exclusively for calculation.

First, we establish the relationship to the work from Section 2.

Lemma 22 *The assignment* $Y \mapsto (H^2(Y)_{\mathrm{tors}}, \mathrm{Spin}^c(Y)_{\mathrm{tors}}, \underline{d}_Y)$ *defines a monoid homomorphism* $\widehat{\Theta}_{\mathbb{Z}} \to \widehat{\Theta}_{\mathrm{WT}}$.

Proof This amounts to three claims: that $\underline{d}_Y(S^3) = 0$ (tautological), that the assignment $Y \mapsto \underline{d}_Y$ sends integer homology cobordisms to isomorphisms of weighted torsors (an immediate corollary of [Behrens and Golla 2018, Corollary 4.2]), and that $\underline{d}_Y$ is additive, in the sense that

$$\underline{d}_{Y \# Y'}(\mathfrak{s} \# \mathfrak{s}') = \underline{d}_Y(\mathfrak{s}) + \underline{d}_Y(\mathfrak{s}'),$$

which is [Behrens and Golla 2018, Proposition 3.7]. □

From this, we can immediately show that *provided the $\hat{d}$–invariants of the summands satisfy the nonvanishing property*, integer homology cobordisms between connected sums preserve the natural direct sum decomposition of their homology groups.

Corollary 23 *Suppose $Y = \#_{i=1}^n Y_i$ and $Y' = \#_{j=1}^m Y'_j$ are connected sums of 3–manifolds such that the $\hat{\underline{d}}$–invariants of each Y_i and Y'_j satisfy the nonvanishing property. If $W : Y \to Y'$ is a homology cobordism, then $n = m$ and the induced map $W_* \colon H_1(Y)_{\mathrm{tors}} \to H_1(Y')_{\mathrm{tors}}$ preserves the natural (unordered) direct sum decompositions.*

Proof As mentioned above, if $W\colon Y \to Y'$ is a homology cobordism, it induces an isomorphism of weighted torsors $(\mathrm{Spin}^c_{\mathrm{tors}}(Y), \underline{d}) \cong (\mathrm{Spin}^c_{\mathrm{tors}}(Y'), \underline{d}')$. The statement follows immediately from Corollary 15. □

The following corollary is simply an application of Corollary 19, applied to these particular weighted torsors.

Corollary 24 *Suppose Y_i is a collection of 3–manifolds indexed by some set S such that*

- *the groups $H^2(Y_i)$ are nontrivial;*
- *the Fourier transforms $\hat{\underline{d}}_{Y_i}$ satisfy the nonvanishing property;*
- *the weighted torsors $(H^2(Y_i), \mathrm{Spin}^c(Y_i)_{\mathrm{tors}}, \underline{d}^r)$ are pairwise nonisomorphic.*

Then there is a homomorphism $c\colon \widehat{\Theta}_{\mathbb{Z}} \to \mathbb{N}^S$ with $c_i(Y_i) = 1$ and $c_i(Y_j) = 0$ for $i \neq j$. In particular, the Y_i are linearly independent in $\widehat{\Theta}_{\mathbb{Z}}$ and span a $\mathbb{Z}^S$–summand of the Grothendieck group $\mathrm{Gr}(\widehat{\Theta}_{\mathbb{Z}})$.

Here $\underline{d}^r$ is the reduced part of $\underline{d}$ as in Definition 9.

To prove Theorem 1, we need to show that the lens spaces $L(p,q)$ — considered up to oriented diffeomorphism — satisfy the assumptions of the corollary. This is classical; the crucial observation is that the reduced d–invariant recovers the Turaev–Reidemeister torsion.

Lemma 25 *If Y is an L–space, then*

$$T(\mathfrak{s}) = \tfrac{1}{2}(\underline{d}^r(\mathfrak{s})).$$

This follows immediately from [Rustamov 2005, Theorems 5.3.3–4]. For lens spaces (and thus their connected sums, as both sides of this equality are additive) this was proven earlier in [Némethi 2005, Section 10.7], and indeed Némethi's result is used in Rustamov's argument.

The Turaev–Reidemeister torsion of lens spaces is classical, and the sign-refined version only slightly less so. For an appropriate choice of base spin^c structure and an appropriate isomorphism $H^2(L(p,q)) \cong \mathbb{Z}/p$, we have [Némethi and Nicolaescu 2002, Section 7.1] for any nontrivial p^{th} root of unity ζ

$$\hat{T}(\zeta) = \frac{1}{p(1-\zeta^{-1})(1-\zeta^{-q})}.$$

Notice that this formula has an extra factor of $1/p$ compared to Nicolaescu's, owing to the change of convention discussed in Remark 7.

We write $d^r_{p,q}\colon \mathrm{Spin}^c(L(p,q)) \to \mathbb{Q}$ for the reduced d–invariants of the lens space $L(p,q)$.

The lemma above establishes that (after choosing an appropriate base spin^c structure and isomorphism $L(p,q) \cong \mathbb{Z}/p$) we have for all nontrivial p^{th} roots of unity

$$(\hat{d}^r_{p,q})(\zeta) = \frac{1}{2p(1-\zeta^{-1})(1-\zeta^{-q})}.$$

Proposition 26 *The Fourier transforms of the weighted torsors $d^r_{p,q}$ have the nonvanishing property. Furthermore, if $d^r_{p,q} \cong d^r_{r,s}$, then $p = r$ and $s \equiv q^{\pm 1} \bmod p$, so $L(p,q)$ and $L(r,s)$ are oriented diffeomorphic.*

Proof That these have the nonvanishing property is obvious.

What remains is the essentially classical claim that signed Turaev–Reidemeister torsion classifies lens spaces up to oriented diffeomorphism; we include a short proof for completeness. We may as well assume $p > 2$.

If $d^r_{p,q} \cong d^r_{r,s}$ then $\mathbb{Z}/p \cong \mathbb{Z}/r$, so $p = r$. If $\hat{d}^r_{p,q}$ is t–isomorphic to $\hat{d}^r_{p,s}$, such an isomorphism induces a t–isomorphism $\hat{f}_{p,q} \cong \hat{f}_{p,s}$ between the simpler functions

$$\hat{f}_{p,q}(\zeta) = (1-\zeta)(1-\zeta^q) = 1 - \zeta - \zeta^q + \zeta^{q+1}.$$

If $q = 1$ this is the Fourier transform of the function $f_{p,1} = (1, -2, 1, 0, \dots, 0)$, where we list off the values $f_{p,q}(i)$ in order starting at 0. If $q = p-1$ we have instead $f_{p,p-1} = (2, -1, 0, \dots, 0, -1)$. For $1 < q < p-1$, we have $f_{p,q}(0) = f_{p,q}(q+1) = 1$, while $f_{p,q}(1) = f_{p,q}(q) = -1$, and all other values are zero.

It is transparent that there is no affine isomorphism of $\mathbb{Z}/p$ taking $f_{p,1}$ or $f_{p,p-1}$ to any of the other functions above, as the values are different; this reduces us to the case $1 < q < p-1$.

Now suppose there exists some integer k prime to p and some integer ℓ such that

$$f_{p,s}(ki + \ell) = f_{p,q}(i) \tag{1}$$

for all i. Since $f_{p,q}(0) = 1$, we have $f_{p,s}(\ell) = 0$ and thus either $\ell \equiv 0$ or $\ell \equiv s+1$. Because $i \mapsto ki + \ell$ is a bijection, in the former case we must have $k(q+1) \equiv s+1$ and in the latter case $k(q+1)+s+1 \equiv 0$. We handle these two cases separately.

(i) If $\ell \equiv 0$, then applying (1) to $i \equiv 1$, we have either $k \equiv 1$ (which gives $q+1 \equiv s+1$ and hence $q \equiv s$) or $k \equiv s$ (which gives $qs + s \equiv s+1$ so $qs \equiv 1$).

(ii) If $\ell \equiv s+1$, then applying (1) to $i \equiv 1$, we have either $k+s+1 \equiv 1$ (in which case $k \equiv -s$ so $-sq-s+s+1 \equiv 0$ and thus $qs \equiv 1$) or $k+s+1 \equiv s$ (in which case $k \equiv -1$ so $-q-1+s+1 \equiv 0$ and $q \equiv s$).

In any of the four possibilities for the values of k and ℓ modulo p, we see that the desired claim holds. □

The main theorem follows immediately from this proposition, as well as Corollaries 23 and 24.

5 Questions

We close with a handful of questions inspired by the results above.

Question 1 Which collections of 3–manifolds satisfy the hypotheses of Corollary 24? Does this class include spherical 3–manifolds, or double-branched covers of alternating links?

It would follow that manifolds in this class are integer homology cobordant if and only if they are diffeomorphic. For spherical 3–manifolds, an argument might proceed by an explicit computation of their d–invariants (or of their Reidemeister torsions); for alternating links an argument might follow through the lattice-theoretic techniques of [Greene 2013].

In another direction, recall that [Lisca 2007] provides a complete classification of connected sums of lens spaces up to rational homology cobordism. It would be interesting to determine the intermediate case of R–homology cobordism for, say, $R = \mathbb{Z}/p$.

Question 2 Determine the R–homology cobordism classification of connected sums of lens spaces for various rings R.

The author has made no attempt to investigate this. It would similarly be interesting to follow up on the classification of rational homology ribbon cobordisms between connected sum of lens spaces given in [Huber 2021, Theorem 1.3] and classify the R–homology ribbon cobordisms for various rings R.

It would be interesting to understand better the interaction between the integer homology *group* and the larger integer homology *monoid*. In the integer homology *group*, because all elements are invertible, all elements are also cancellative: if $[Y]+[Z]=[Z]$ we have $[Y]=0$. This is not true in an arbitrary monoid, so one might ask if integer homology spheres *remain* cancellative when we pass to the integer homology monoid.

This question can be phrased in terms of the Grothendieck group $\mathrm{Gr}(\widehat{\Theta}_{\mathbb{Z}})$ as follows.

Question 3 Does the map $\Theta_{\mathbb{Z}} \to \mathrm{Gr}(\widehat{\Theta}_{\mathbb{Z}})$ have nontrivial kernel? That is, can one find an integer homology 3–sphere Y and a closed oriented 3–manifold Z so that Y is not integer homology cobordant to S^3, but $Y \# Z$ is integer homology cobordant to Z?

One might imagine that the behavior of integer homology spheres under integer homology cobordism is somehow orthogonal to the behavior of rational homology spheres; very optimistically, one might believe that $j\colon \Theta_{\mathbb{Z}} \to \widehat{\Theta}_{\mathbb{Z}}$ *splits*, meaning that there is a homomorphism $\mu\colon \widehat{\Theta}_{\mathbb{Z}} \to \Theta_{\mathbb{Z}}$ with $\mu j = 1$, and one might then try to understand the structure of the Grothendieck group in terms of $\ker(\mu)$ and $\Theta_{\mathbb{Z}}$. If such a splitting exists at the level of monoids, we also have such a splitting at the level of groups when we pass to the Grothendieck group.

One way to guarantee this is *impossible*, while also giving an element in the kernel discussed above, is to find indivisible elements in $\Theta_{\mathbb{Z}}$ which become divisible in $\mathrm{Gr}(\widehat{\Theta}_{\mathbb{Z}})$. We pose the existence of such homology spheres as a question.

Question 4 Is there an integer homology sphere Y with the following properties?

(i) Y is indivisible in the integer homology cobordism group: if $n > 1$ and Y' is another integer homology sphere, there is no integer homology cobordism between Y and $\#^n Y'$.

(ii) There *does* exist an $n > 1$, an integer homology sphere Y', and a 3–manifold Z such that $Y \# Z$ is integer homology cobordant to $(\#^n Y') \# Z$.

Note that if this held, then $-Y \# nY'$ would give an infinite-order element in the kernel of $\Theta_{\mathbb{Z}} \to \mathrm{Gr}(\widehat{\Theta}_{\mathbb{Z}})$, answering Question 3 in the positive.

References

[Behrens and Golla 2018] **S Behrens**, **M Golla**, *Heegaard Floer correction terms, with a twist*, Quantum Topol. 9 (2018) 1–37 MR Zbl

[Cochran and Tanner 2014] **T D Cochran**, **D Tanner**, *Homology cobordism and Seifert fibered 3–manifolds*, Proc. Amer. Math. Soc. 142 (2014) 4015–4024 MR Zbl

[Doig and Horn 2017] **M I Doig**, **P D Horn**, *On the intersection ring of graph manifolds*, Trans. Amer. Math. Soc. 369 (2017) 1185–1203 MR Zbl

[Doig and Wehrli 2015] **M Doig**, **S Wehrli**, *A combinatorial proof of the homology cobordism classification of lens spaces*, preprint (2015) arXiv 1505.06970

[Fintushel and Stern 1987] **R Fintushel**, **R Stern**, *Rational homology cobordisms of spherical space forms*, Topology 26 (1987) 385–393 MR Zbl

[Fried 1987] **D Fried**, *Lefschetz formulas for flows*, from "The Lefschetz centennial conference, III" (A Verjovsky, editor), Contemp. Math. 58, Amer. Math. Soc., Providence, RI (1987) 19–69 MR Zbl

[González-Acuña and Short 1986] **F González-Acuña**, **H Short**, *Knot surgery and primeness*, Math. Proc. Cambridge Philos. Soc. 99 (1986) 89–102 MR Zbl

[Greene 2013] **J E Greene**, *Lattices, graphs, and Conway mutation*, Invent. Math. 192 (2013) 717–750 MR Zbl

[Huber 2021] **M Huber**, *Ribbon cobordisms between lens spaces*, Pacific J. Math. 315 (2021) 111–128 MR Zbl

[Lisca 2007] **P Lisca**, *Sums of lens spaces bounding rational balls*, Algebr. Geom. Topol. 7 (2007) 2141–2164 MR Zbl

[Livingston 1981] **C Livingston**, *Homology cobordisms of 3–manifolds, knot concordances, and prime knots*, Pacific J. Math. 94 (1981) 193–206 MR Zbl

[Milnor 1962] **J Milnor**, *A duality theorem for Reidemeister torsion*, Ann. of Math. 76 (1962) 137–147 MR Zbl

[Myers 1983] **R Myers**, *Homology cobordisms, link concordances, and hyperbolic 3–manifolds*, Trans. Amer. Math. Soc. 278 (1983) 271–288 MR Zbl

[Némethi 2005] **A Némethi**, *On the Ozsváth–Szabó invariant of negative definite plumbed 3–manifolds*, Geom. Topol. 9 (2005) 991–1042 MR Zbl

[Némethi and Nicolaescu 2002] **A Némethi**, **L I Nicolaescu**, *Seiberg–Witten invariants and surface singularities*, Geom. Topol. 6 (2002) 269–328 MR Zbl

[Nicolaescu 2003] **L I Nicolaescu**, *The Reidemeister torsion of 3–manifolds*, De Gruyter Studies in Mathematics 30, de Gruyter, Berlin (2003) MR Zbl

[Nicolaescu 2004] **L I Nicolaescu**, *Seiberg–Witten invariants of rational homology 3–spheres*, Commun. Contemp. Math. 6 (2004) 833–866 MR Zbl

[Ozsváth and Szabó 2003] **P Ozsváth**, **Z Szabó**, *Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary*, Adv. Math. 173 (2003) 179–261 MR Zbl

[Ray and Singer 1971] **D B Ray**, **I M Singer**, *R–torsion and the Laplacian on Riemannian manifolds*, Advances in Math. 7 (1971) 145–210 MR Zbl

[Rustamov 2005] **R Rustamov**, *On Heegaard Floer homology of three-manifolds*, PhD thesis, Princeton University (2005) MR Available at `https://www.proquest.com/docview/305385735`

[Serre 1977] **J-P Serre**, *Linear representations of finite groups*, Graduate Texts in Math. 42, Springer (1977) MR Zbl

[Suciu 2022] **A I Suciu**, *Cohomology jump loci of 3–manifolds*, Manuscripta Math. 167 (2022) 89–123 MR Zbl

[Turaev 2002] **V Turaev**, *Torsions of 3–dimensional manifolds*, Progr. Math. 208, Birkhäuser, Basel (2002) MR Zbl

Department of Mathematics and Statistics, University of Vermont
Burlington, VT, United States

`mike.miller-eismeier@uvm.edu`

Received: 5 June 2023 Revised: 14 August 2023

Algebraic & Geometric Topology **24**:7 (2024) 4103–4114
DOI: 10.2140/agt.2024.24.4103
Published: 9 December 2024

Profinite isomorphisms and fixed-point properties

MARTIN R BRIDSON

We describe a flexible construction that produces triples of finitely generated, residually finite groups $M \hookrightarrow P \hookrightarrow \Gamma$, where the maps induce isomorphisms of profinite completions $\widehat{M} \cong \widehat{P} \cong \widehat{\Gamma}$, but M and Γ have Serre's property FA while P does not. In this construction, P is finitely presented and Γ is of type F_∞. More generally, given any positive integer d, one can demand that M and Γ have a fixed point whenever they act by semisimple isometries on a complete CAT(0) space of dimension at most d, while P acts without a fixed point on a tree.

20E18, 20F67, 20J05; 20E08

1 Introduction

In the quest for profinite invariants of discrete groups, fixed-point properties have been a source of disappointment. For example, Aka [1] proved that the profinite completion of a finitely generated, residually finite group does not determine whether the group has property (T), ie whether the group can act without a global fixed point as a group of affine isometries of a Hilbert space. Cheetham-West, Lubotzky, Reid and Spitler [13] proved a similar theorem for actions on trees: they construct pairs of finitely presented, residually finite groups G_1 and G_2 such that $\widehat{G}_1 \cong \widehat{G}_2$ but G_1 has Serre's property FA whereas G_2 does not. (Here, $\widehat{G}_i$ denotes the profinite completion of G_i.)

In the present paper, we will improve upon this last result in two ways. First, we construct groups with these properties for which (G_2, G_1) is a Grothendieck pair, ie the isomorphism $\widehat{G}_1 \cong \widehat{G}_2$ is induced by a monomorphism of discrete groups $G_1 \hookrightarrow G_2$ (cf. [13, Question 4.1]). Secondly, we extend this result from actions on trees (the 1–dimensional case) to actions on d–dimensional CAT(0) spaces, with $d \geq 1$ arbitrary.

We say that a group G has *property* Fix_d if G fixes a point whenever it acts by semisimple isometries on a complete CAT(0) space of covering dimension at most d. Every isometry of a simplicial tree is semisimple, so Fix_1 implies Serre's property FA (and extends it to cover actions on complete $\mathbb{R}$–trees).

Theorem A *For every integer $d \geq 1$, there exist triples of residually finite groups $M \overset{i}{\hookrightarrow} P \overset{j}{\hookrightarrow} \Gamma$ such that*

1. *i and j induce isomorphisms $\widehat{M} \cong \widehat{P} \cong \widehat{\Gamma}$;*
2. *M is finitely generated, P is finitely presented, and Γ is of type F_∞;*

(3) *M and Γ have property* Fix_d*, but*

(4) *P is a nontrivial amalgamated free-product and therefore acts on a simplicial tree without a global fixed point.*

An artefact of our proof is that although M and Γ have Fix_d, they each contain a subgroup of finite index that can act on a tree without fixing a point (Remark 6.1).

The fixed-point properties required in the above theorem will be established using the following criterion, which is drawn from the circle of ideas developed in [6; 4].

Theorem B (Corollary 5.5) *If A is a finitely generated group with finite abelianisation and B is a finite group, then $A \wr B$ has* Fix_d *for $d = |B| - 1$.*

The first steps in our construction of the triples $M \hookrightarrow P \hookrightarrow \Gamma$ in Theorem A follow the template for constructing finitely presented Grothendieck pairs that originates in [10] and is explicit in Section 8 of [7]. We craft finitely presented groups Q that enjoy an array of properties relevant to our aims (Section 3.1); we use a suitably adapted form of the Rips construction (Proposition 3.1) to produce short exact sequences $1 \to N \to G \to Q \to 1$ with G finitely presented and residually finite, N finitely generated, and both N and G perfect; and we take a fibre product of several copies of $G \to Q$ to produce $N^d \hookrightarrow P_d \hookrightarrow G^d$ with P_d finitely presented. (A novel feature here is that we take the fibre product of several copies of $G \to Q$, not just two.) The triples $M \overset{i}{\hookrightarrow} P \overset{j}{\hookrightarrow} \Gamma$ we seek are obtained by taking finite extensions of N^d, P_d and G^d in a way that allows us to apply Theorem B.

There is a great deal of flexibility in this construction — see Section 7.

2 Preliminaries

In this section we gather the basic definitions and facts we need concerning profinite completions of groups and isometries of CAT(0) spaces.

2.1 Profinite completions

If $M_1 < M_2$ are normal subgroups of finite index in a group G, then there is a natural map $G/M_1 \to G/M_2$. Thus the finite quotients of G form a directed system. The *profinite completion* of G is the inverse limit of this system:

$$\widehat{G} := \varprojlim G/M.$$

The natural map $i \colon G \to \widehat{G}$ is injective if and only if G is residually finite. If G is finitely generated then, for every finite group Q, composition with i defines a bijection $\mathrm{Hom}(\widehat{G}, Q) \to \mathrm{Hom}(G, Q)$ that restricts to a bijection on the set of epimorphisms. In particular, G and $\widehat{G}$ have the same set of finite images, which we denote by $\mathfrak{F}(G)$. Thus $\widehat{G}_1 \cong \widehat{G}_2$ implies $\mathfrak{F}(G_1) = \mathfrak{F}(G_2)$. Less obviously, for finitely generated groups, $\mathfrak{F}(G_1) = \mathfrak{F}(G_2)$ implies $\widehat{G}_1 \cong \widehat{G}_2$ — see [17, pages 88–89]. (Note that $\widehat{G}_1 \cong \widehat{G}_2$ does not imply that there are any nontrivial homomorphisms $G_1 \to G_2$.)

A property $\mathfrak{P}$ of finitely generated, residually finite groups is said to be a *profinite invariant* if $\widehat{G}_1 \cong \widehat{G}_2$ implies that G_2 has $\mathfrak{P}$ whenever G_1 has $\mathfrak{P}$. Theorem A shows that Fix_d is not a profinite invariant.

A pair of finitely generated, residually finite groups $G_1 \overset{\iota}{\hookrightarrow} G_2$ is called a *Grothendieck pair* [15] if the induced map $\hat{\iota}\colon \widehat{G}_1 \to \widehat{G}_2$ is an isomorphism. For fixed G_2, there can be infinitely many nonisomorphic subgroups G_1 such that $G_1 \hookrightarrow G_2$ is a Grothendieck pair, even if one requires both G_1 and G_2 to be finitely presented [8].

2.2 Isometries of CAT(0) spaces

We refer the reader to [11] for basic facts about CAT(0) spaces. We write $\mathrm{Isom}(X)$ for the group of isometries of a CAT(0) space X and $\mathrm{Fix}(H)$ for the set of points in X fixed by each element of a subset $H \subset \mathrm{Isom}(X)$. Note that $\mathrm{Fix}(H)$ is closed and convex.

If X is complete, each closed, nonempty bounded subset is contained in a unique smallest ball; see [11, page 178]. If the bounded subset is an orbit of a subgroup $H < \mathrm{Isom}(X)$, then the centre of the ball will be fixed by H. This proves the following standard proposition.

Proposition 2.1 *If X is a complete* CAT(0) *space, then every finite subgroup of* $\mathrm{Isom}(X)$ *fixes a point in X.*

By combining the preceding bounded-orbit observation with the fact that $\mathrm{Fix}(H)$ is itself a CAT(0) space, one can prove the following standard fact — see [6, Corollary 2.5], for example.

Proposition 2.2 *Let X be a complete* CAT(0) *space. If the subgroups $H_1, \dots, H_n < \mathrm{Isom}(X)$ commute and $\mathrm{Fix}(H_i)$ is nonempty for $i = 1, \dots, n$, then $\bigcap_i \mathrm{Fix}(H_i)$ is nonempty.*

For an isometry $\gamma \in \mathrm{Isom}(X)$,

$$\mathrm{Min}(\gamma) := \{p \in X \mid d(p, \gamma.p) = |\gamma|\},$$

where $|\gamma| := \inf\{d(x, \gamma.x) \mid x \in X\}$. By definition, γ is *semisimple* if $\mathrm{Min}(\gamma)$ is nonempty. Every isometry of a complete $\mathbb{R}$–tree is semisimple. Semisimple isometries are divided into *hyperbolics* (also called loxodromics), for which $|\gamma| > 0$, and *elliptics*, which are the isometries with $\mathrm{Fix}(\gamma) \neq \varnothing$. If γ is hyperbolic then there exist γ–invariant isometrically embedded lines $\mathbb{R} \hookrightarrow X$ on which γ acts as a translation by $|\gamma|$; each such line is called an axis for γ. The union of these axes is $\mathrm{Min}(\gamma)$. The following extract from pages 229–231 of [11] summarises the properties of $\mathrm{Min}(\gamma)$ that we require.

Proposition 2.3 *Let X be a complete* CAT(0) *space and let $\gamma \in \mathrm{Isom}(X)$ be a hyperbolic isometry. Then:*

1. $\mathrm{Min}(\gamma)$ *splits isometrically* $\mathrm{Min}(\gamma) = Y \times \mathbb{R}$, *where $Y \times \{0\}$ is a closed, convex subspace of X.*
2. *γ acts trivially on Y and acts as translation by $|\gamma|$ on each of the lines $\{y\} \times \mathbb{R}$.*

(3) *The centraliser* $C(\gamma) < \mathrm{Isom}(X)$ *leaves* $\mathrm{Min}(\gamma)$ *and its splitting invariant, acting by translations of the second factor.*

(4) *If* $\delta \in C(\gamma)$ *is hyperbolic, then* $\mathrm{Min}(\gamma)$ *contains an axis for* δ.

3 A Rips construction and input groups

The purpose of this section is to produce the short exact sequences $1 \to N \to G \to Q \to 1$ described in the introduction.

3.1 The input groups Q

Our constructions require as input a group Q with the following properties:

- Q is of type F_3 (ie has a classifying space $K(Q,1)$ with finite 3–skeleton).
- $H_2(Q,\mathbb{Z}) = 0$.
- $\widehat{Q} = 1$.
- Q is a nontrivial amalgamated free product (and therefore does not have FA).

There are many ways to concoct groups Q with these properties. Indeed, *every* finitely presented group can be embedded (explicitly, with controlled geometry) into a finitely presented group that has no nontrivial finite quotients [3]; by replacing this enveloping group with its universal central extension one can force it to have trivial second homology; and by taking a free product of two copies of the resulting group one obtains a group Q satisfying all but the first of the above properties. If the group that one starts with is of type F_3, then so is Q. Likewise for type F (having a finite classifying space) and type F_∞ (having a classifying space with finite skeleta).

One can also find explicit groups of the desired form in the literature. For example, from [10] one could take

$$Q = \langle a, b, \alpha, \beta \mid ba^{-2}b^{-1}a^3,\ \beta\alpha^{-2}\beta^{-1}\alpha^3,\ [bab^{-1}, a]\beta^{-1},\ [\beta\alpha\beta^{-1}, \alpha]b^{-1} \rangle.$$

3.2 A convenient version of the Rips construction

There are many refinements of the Rips construction in the literature, with various properties imposed on the groups constructed. The following version suits our needs.

Proposition 3.1 *There exists an algorithm that, given a finite presentation* $\langle X \mid R \rangle$ *of a group* Q*, will construct a finite aspherical presentation* $\langle X \cup \{a_1, a_2\} \mid \widetilde{R} \cup V \rangle$ *for a group* G *so that*

(1) G *is hyperbolic, residually finite, of type* F*, and virtually special;*

(2) $N := \langle a_1, a_2 \rangle$ *is normal in* G;

(3) G/N *is isomorphic to* Q;

(4) *G is perfect if Q is perfect*;

(5) *if $\widehat{Q} = 1$ and $H_2(Q, \mathbb{Z}) = 0$, then N and G are both perfect.*

Proof With the exception of item (5), the proof is covered by Proposition 2.10 of [9]. The crucial property of residual finiteness is due to Wise [18; 19].

For item (5), we consider the 5–term exact sequence extracted from the corner of the LHS spectral sequence for $1 \to N \to G \to Q \to 1$:

$$H_2(Q, \mathbb{Z}) \to H_0(Q,\ H_1(N, \mathbb{Z})) \to H_1(G, \mathbb{Z}) \to H_1(Q, \mathbb{Z}) \to 0.$$

As all the other terms are zero, $H_0(Q,\ H_1(N, \mathbb{Z})) = 0$. By definition, $H_0(Q,\ H_1(N, \mathbb{Z}))$ is the group of coinvariants for the action of Q on $H_1(N, \mathbb{Z})$ that is induced by conjugation in G. As the abelian group $H_1(N, \mathbb{Z})$ is finitely generated, its automorphism group is residually finite. As Q has no nontrivial finite quotients, its action on $H_1(N, \mathbb{Z})$ must be trivial. Therefore $H_1(N, \mathbb{Z}) = H_0(Q,\ H_1(N, \mathbb{Z})) = 0$. □

4 Fibre products

Our proof of Theorem A relies on the various properties of fibre products that we establish in this section. These properties cover three topics: the finiteness properties of fibre products, their behaviour with respect to profinite completions, and their interaction with wreath products.

4.1 Fibre products and finiteness properties

For $i = 1, \dots, d$, let $p_i : G_i \to Q$ be an epimorphism of groups. The *fibre product* of this family of maps is

$$P_d = \{(g_1, \dots, g_d) \mid p_i(g_i) = p_j(g_j),\ i, j = 1, \dots, d\} < G_1 \times \cdots \times G_d.$$

The case $p_1 = \cdots = p_d$ will be of particular interest in this article.

P_d is the preimage of the diagonal subgroup

$$Q_d^{\Delta} := \{(q, \dots, q) \mid q \in Q\} < Q \times \cdots \times Q$$

and there is a short exact sequence

$$1 \to N^{(d)} \to P_d \to Q_d^{\Delta} \to 1, \tag{4-1}$$

where $N_i = \ker p_i$ and $N^{(d)} = N_1 \times \cdots \times N_d$.

We need a criterion to ensure that P_d is finitely presented; we will deduce this from the following *Asymmetric* 1-2-3 *Theorem* [12].

Theorem 4.1 [12] *For $i = 1, 2$, let $1 \to N_i \to G_i \xrightarrow{p_i} Q \to 1$ be a short exact sequence of groups. If G_1 and G_2 are finitely presented, Q is of type F_3, and at least one of the groups N_1, N_2 is finitely generated, then the fibre product $P < G_1 \times G_2$ is finitely presented.*

Corollary 4.2 *Suppose $d \geq 2$ and let $1 \to N_i \to G_i \xrightarrow{p_i} Q \to 1$ be a short exact sequence of groups, for $i = 1, \dots, d$. If the groups G_i are all finitely presented, the groups N_i are finitely generated, and Q is of type F_3, then the associated fibre product $P_d < G_1 \times \dots \times G_d$ is finitely presented.*

Proof We proceed by induction on d; the case $d = 2$ is covered by the theorem. Let $P_d < G_1 \times \dots \times G_d$ be the fibre product of $p_1, \dots, p_d$. For the inductive step, first note that

$$P_d \; < \; P_{d-1} \times G_d \; < \; G_1 \times \dots \times G_d$$

is the fibre product of the map $p_d \colon G_d \to Q$ and the composition $P_{d-1} \to Q^{\Delta}_{d-1} \to Q$, where $P_{d-1} \to Q^{\Delta}_{d-1}$ is the map from (4-1) and $Q^{\Delta}_{d-1} \to Q$ is the isomorphism $(q, \dots, q) \mapsto q$. To complete the proof, we apply the theorem, noting that P_{d-1} is finitely presented, by induction. □

We shall also need the following more elementary result.

Lemma 4.3 *For $i = 1, \dots, d$, let $p_i \colon G_i \twoheadrightarrow Q$ be an epimorphism of groups. If the groups G_i are finitely generated and Q is finitely presented, then the fibre product $P_d < G_1 \times \dots \times G_d$ is finitely generated.*

Proof As in the preceding proof, induction reduces us to the case $d = 2$. We fix a finite presentation $Q = \langle a_1, \dots, a_n \mid r_1, \dots, r_m \rangle$ and for $i = 1, 2$ choose $a_{ij} \in G_i$ such that $p_i(a_{ij}) = a_j$. We then add a finite set of elements $B_i \subset \ker p_i$ to obtain a finite generating set for G_i, and denote by ρ_{ik} the word obtained from r_k by replacing each a_j with a_{ij}. It is easy to check that the fibre product $P < G_1 \times G_2$ is generated by

$$\big\{(b_{1s}, 1),\ (1, b_{2s}),\ (a_{1j}, a_{2j}),\ (\rho_{1k}, 1) \mid j = 1, \dots, n;\ k = 1, \dots, m;\ b_{is} \in B_i\big\}. \qquad \square$$

4.2 Fibre products and Grothendieck pairs

The idea of constructing Grothendieck pairs using fibre products originates in the work of Platonov and Tavgen [16] and was extended in [2; 8; 10].

Lemma 4.4 [10, Lemma 2.2] *Let $1 \to N \to G \to Q \to 1$ be an exact sequence of finitely generated groups. If $\widehat{Q} = 1$ and $H_2(Q, \mathbb{Z}) = 0$, then $N \hookrightarrow G$ induces an isomorphism of profinite completions.*

The following variant of the Platonov–Tavgen argument will be useful.

Proposition 4.5 [8, Theorem 2.2] *Let $p_1 \colon G_1 \to Q$ and $p_2 : G_2 \to Q$ be epimorphisms with G_1 and G_2 finitely generated and Q finitely presented. Let $P < G_1 \times G_2$ be the associated fibre product. If $\widehat{Q} = 1$ and $H_2(Q, \mathbb{Z}) = 0$, then $P \hookrightarrow G_1 \times G_2$ induces an isomorphism of profinite completions.*

We need an extension to the case of $d \geq 2$ factors.

Theorem 4.6 *For $i = 1, \dots, d$, let $p_i \colon G_i \to Q$ be an epimorphism of finitely generated groups, and let $P_d < \Gamma := G_1 \times \dots \times G_d$ be the associated fibre product. If Q is finitely presented, $\widehat{Q} = 1$ and $H_2(Q, \mathbb{Z}) = 0$, then $P_d \hookrightarrow \Gamma$ induces an isomorphism of profinite completions.*

Proof We argue by induction on d, as in the proof of Corollary 4.2. In the inductive step, we appeal to Lemma 4.3 to ensure that P_{d-1} is finitely generated. We can then apply Proposition 4.5 to $p_d\colon G_d \to Q$ and $P_{d-1} \to Q^{\Delta}_{d-1} \cong Q$, noting that P_d is the fibre product of these maps. □

4.3 Fibre products and wreath products

Given groups A and B, with B finite, the *wreath product* $A \wr B$ is the semidirect product $A^B \rtimes B$, or more precisely $\left(\bigoplus_{b\in B} A_b\right) \rtimes B$, with fixed isomorphisms $\mu_b\colon A \to A_b$ so that $b \in B$ acts on $A_{b'}$ as $\mu_{bb'} \circ \mu_{b'}^{-1}$. We identify $A^{\Delta} < A \times \cdots \times A$ with its image under $(\mu_b)_{b\in B}$. The following trivial observation will play an important role in what follows.

Lemma 4.7 *$\langle A^{\Delta}, B\rangle < A \wr B$ is the direct product $A^{\Delta} \times B \cong A \times B$.*

Given B and a short exact sequence of groups $1 \to N \to G \to Q \to 1$, we take the direct product of $|B|$ copies of the sequence, indexed by the elements of B, and let B permute these copies by its left action on the indices. The resulting semidirect products give us a (nonexact) sequence of groups

$$N \wr B \hookrightarrow G \wr B \twoheadrightarrow Q \wr B.$$

The action of B preserves the fibre product $P_B < G^B = \bigoplus_{b\in B} G_b$ of the maps $G_b \to Q_b$, giving a semidirect product

$$P_B \rtimes B = \langle P_B, B\rangle < G \wr B$$

and a (nonexact) sequence of groups

$$N \wr B \hookrightarrow P_B \rtimes B \twoheadrightarrow \langle Q^{\Delta}, B\rangle < Q \wr B.$$

From Lemma 4.7 we deduce:

Lemma 4.8 *With the notation established above, there is surjection*

$$P_B \rtimes B \twoheadrightarrow Q^{\Delta} \cong Q.$$

5 Fixed point criteria

In this section we present criteria that guarantee fixed points for group actions on complete $\mathrm{CAT}(0)$ spaces of finite dimension. These criteria are extracted from the more general criteria explained in [4; 6].

The following result is a special case of [6, Corollary 3.6].

Proposition 5.1 *Let d be a positive integer and let X be a complete $\mathrm{CAT}(0)$ space of dimension less than d. Let $S_1, \dots, S_d \subset \mathrm{Isom}(X)$ be conjugates of a subset $S \subset \mathrm{Isom}(X)$ with $[s_i, s_j] = 1$ for all $s_i \in S_i$ and $s_j \in S_j$ with $i \neq j$. If every element of S (hence S_i) has a fixed point in X, then so does every finite subset of S (hence S_i).*

Corollary 5.2 *Let d be a positive integer, let X be a complete $\mathrm{CAT}(0)$ space of dimension less than d, and let $H_1, \dots, H_d < \mathrm{Isom}(X)$ be conjugate subgroups that pairwise commute. If each H_i is generated by a finite set of elliptic elements, then $D = \langle H_1, \dots, H_d \rangle$ has a fixed point in X.*

Proof Let $S = S_1$ be a finite set of elliptics generating H_1. We conjugate S to obtain a generating set S_i for each H_i. The proposition says that $\mathrm{Fix}(S_i) = \mathrm{Fix}(H_i)$ is nonempty, whence $\mathrm{Fix}(D)$ is nonempty, by Proposition 2.2. □

For $n \in \mathbb{N}$, an *n–flat* in a metric space X is an isometrically embedded copy of Euclidean space $\mathbb{E}^n \hookrightarrow X$.

Lemma 5.3 *If $K_0, \dots, K_d$ are groups with $\mathrm{Hom}(K_i, \mathbb{R}) = 0$ and X is a complete $\mathrm{CAT}(0)$ space that does not contain any $(d+1)$–flats, then there does not exist an action $\rho\colon K_0 \times \cdots \times K_d \to \mathrm{Isom}(X)$ such that each $\rho(K_i)$ contains a hyperbolic isometry.*

Proof We shall prove the lemma by induction, the case $d = 0$ being trivial. Assume that the lemma is true for $d' \le d-1$. The induction will be complete if we can derive a contradiction from the assumption that there are hyperbolic isometries $\gamma_i \in \rho(K_i)$ for $i = 0, \dots, d$. If this were the case, then, according to Proposition 2.3, the subspace $\mathrm{Min}(\gamma_0)$ would split isometrically as $Y \times \mathbb{R}$ and the centraliser $C(\gamma_0)$ of γ_0 in $\mathrm{Isom}(X)$ would preserve $\mathrm{Min}(\gamma_0)$ and its splitting, acting by translations on the second factor of $Y \times \mathbb{R}$. The group of translations is $\mathbb{R}$ and $\mathrm{Hom}(K_i, \mathbb{R}) = 0$, so $K_1 \times \cdots \times K_d$ must act trivially on the second factor. Thus we obtain an action of $K_1 \times \cdots \times K_d$ on $Y_0 = Y \times \{0\}$. Part (1) of Proposition 2.3 assures us that $Y_0 \subset X$ is closed and convex, hence a $\mathrm{CAT}(0)$ space, and part (4) tells us that $\gamma_i \in K_i$ acts as a hyperbolic isometry of Y_0, for $i = 1, \dots, d$. But $Y \times \mathbb{R} = \mathrm{Min}(\gamma_0)$ embeds isometrically in X, so Y_0 does not contain a d–flat. This contradicts our inductive hypothesis. □

Theorem 5.4 *Let G be a group and suppose that there is a subgroup $D = H_0 \times \cdots \times H_d < G$ with H_i conjugate to H_0 in G for $i = 1, \dots, d$. If H_0 is finitely generated and has finite abelianisation, then D has a fixed point whenever G acts by semisimple isometries on a complete $\mathrm{CAT}(0)$ space of dimension at most d.*

Proof The hypothesis $\dim(X) \le d$ is stronger than requiring that X contains no $(d+1)$–flat, so the preceding lemma tells us that there are no hyperbolic elements in the subgroups H_i. Because H_0 is finitely generated, Corollary 5.2 completes the proof. □

The following result was stated as Theorem B in the introduction.

Corollary 5.5 *If A is a finitely generated group with finite abelianisation and B is a finite group, then $A \wr B$ has Fix_d, where $d = |B| - 1$.*

Proof Let $G = A \wr B = \left(\bigoplus_{b \in B} A_b\right) \rtimes B$ and $D = \bigoplus_{b \in B} A_b$. Theorem 5.4 tells us that D has a fixed point whenever $A \wr B$ acts by semisimple isometries on a complete $\mathrm{CAT}(0)$ space X with $\dim(X) \le |B|-1$. Because $B < A \wr B$ normalises D, it leaves its set of fixed points $\mathrm{Fix}(D) \subset X$ invariant. $\mathrm{Fix}(D)$ is closed and convex, hence a complete CAT(0) space. Proposition 2.1 provides a point in $\mathrm{Fix}(D)$ that is fixed by B and hence by $A \wr B = \langle D, B \rangle$. □

6 Proof of Theorem A

Let Q be a group satisfying the conditions listed in Section 3.1. By Proposition 3.1, there is a short exact sequence

$$1 \to N \to G \to Q \to 1$$

with N finitely generated and perfect, and G hyperbolic (hence type F_∞), residually finite and perfect. Given $d \geq 2$, we fix a finite group B with $|B| = d+1$. Proceeding as in Section 4.3, we take the direct product of $|B|$ copies of this sequence, indexed by the elements of B, and take the fibre product of the maps $G_b \to Q$ to obtain

$$N^B \hookrightarrow P_B \hookrightarrow G^B.$$

The action of B permuting the direct factors of G^B leaves N^B and P_B invariant, so the above inclusions extend to

$$N \wr B \overset{i}{\hookrightarrow} P_B \rtimes B \overset{j}{\hookrightarrow} G \wr B.$$

We claim that this triple of groups has the properties required in Theorem A.

Towards showing that i induces an isomorphism of profinite completions, note first that Lemma 4.4 applies to $N^B \hookrightarrow P_B$, because N^B is normal in P_B with quotient Q. Likewise, Theorem 4.6 assures us that $P_B \hookrightarrow G^B$, the restriction of j, induces an isomorphism of profinite completions. Thus the maps $N^B \hookrightarrow P_B \hookrightarrow G^B$ induce isomorphisms $\widehat{N^B} \cong \widehat{P_B} \cong \widehat{G^B}$. The action of B permuting the factors of G^B extends to $\widehat{G^B}$, where it preserves the dense subgroups P_B and N^B. Since the operations of profinite completion and semidirect product with a finite group commute, we conclude that $\hat{\imath}$ and $\hat{\jmath}$ give isomorphisms $\widehat{N^B} \rtimes B \cong \widehat{P_B} \rtimes B \cong \widehat{G^B} \rtimes B$. This establishes Theorem A(1).

$N \wr B$ is finitely generated, since N is. Corollary 4.2 assures us that P_B is finitely presented, whence the finite extension $P_B \rtimes B$ is too. And since G is of type F_∞, so is $G \wr B$. (Indeed, Proposition 3.1 produces a group G that is of type F, ie has a finite classifying space, so G^B is of type F and $G \wr B$ is virtually of type F.) This establishes Theorem A(2).

N and G are finitely generated and perfect, so Corollary 5.5 tells us that $N \wr B$ and $G \wr B$ have Fix_d, since $|B| = d+1$. In contrast, $P_B \rtimes B$ maps onto Q, as in Lemma 4.8, and therefore it is a nontrivial amalgamated free product — in particular it does not have property FA or Fix_d. □

Remark 6.1 We remarked in the introduction that although $M = N \wr B$ and $\Gamma = G \wr B$ have Fix_d, where $d = |B| - 1$, they each have a subgroup of finite index that can act without a fixed point on a tree. For Γ this is obvious, since $G^B < \Gamma$ maps onto Q via projection to G. More indirectly, from Proposition 3.1

we know that G is virtually special and hence large, ie there is a finite-index subgroup $G_0 < G$ that maps onto a nonabelian free group, say $\mu\colon G_0 \twoheadrightarrow L$. Thus we can map G_0^B, which has finite index in Γ, onto L to produce fixed-point-free actions on trees.

For M, we consider $N_0 = N \cap G_0$, noting that as $\mu(N_0) < L$ is finitely generated and normal, $\mu(N_0)$ is either trivial or of finite index, since L is free. In fact, $\mu(N_0)$ must be the whole of L: as Q has no finite quotients, the restriction of $G \twoheadrightarrow Q$ to G_0 is onto and $G_0/N_0 \cong Q$, which means that G_0/N_0 cannot map onto $L/\mu(N_0)$ if the latter is a nontrivial free or finite group. It follows that N_0^B, which has finite index in M, maps onto L.

7 Flexibility and a decision problem

It is clear from the discussion in Section 3.1 that there is a great deal of flexibility in how one chooses the input groups Q. Consequently, one is free to impose various extra conditions on the Grothendieck pairs $P \rtimes B \hookrightarrow G \wr B$ that we have constructed. In particular, the range of pairs that one obtains is sufficient to accommodate many of the undecidability phenomena described in [5] and elsewhere. For example, by following the proof of [5, Theorem B] we obtain the following theorem. Similar results hold with condition Fix_d in place of FA.

Theorem 7.1 *There does not exist an algorithm that, given a finitely presented, residually finite group Γ that has property FA and a finitely presentable subgroup $u\colon P \hookrightarrow \Gamma$ with $\hat{u}\colon \hat{P} \to \hat{\Gamma}$ an isomorphism, can determine whether or not P has property FA.*

Proof As in [5], one can enhance the groups constructed in [14] to obtain a recursive sequence of finite presentations $\mathcal{Q}^{(m)} \equiv \langle S \mid R^{(m)} \rangle$ for groups $Q^{(m)}$, with S and $|R^{(m)}|$ fixed, so that (i) there is no algorithm to determine which of the groups are trivial, but (ii) if $Q^{(m)} \neq 1$ then it satisfies the properties listed in Section 3.1. We apply the algorithm of Proposition 3.1 to the presentations $\mathcal{Q}^{(m)}$ to obtain $G^{(m)} \twoheadrightarrow Q^{(m)}$, with an explicit presentation for $G^{(m)}$ and hence $G^{(m)} \times G^{(m)}$. The fibre product $P^{(m)} \hookrightarrow G^{(m)} \times G^{(m)}$ is given by the finite generating set described in Lemma 4.3, with B_i the given relators of $G^{(m)}$. Theorem 4.1 assures us that $P^{(m)}$ is finitely presentable. We pass from $P^{(m)} \hookrightarrow G^{(m)} \times G^{(m)}$ to $u_m\colon P^{(m)} \rtimes (\mathbb{Z}/2) \hookrightarrow G^{(m)} \wr (\mathbb{Z}/2)$ and then argue as in the proof of Theorem A to see that u_m induces an isomorphism of profinite completions, that $G^{(m)} \wr (\mathbb{Z}/2)$ has property FA (in fact Fix_1), and that $P^{(m)} \rtimes (\mathbb{Z}/2)$ maps onto $Q^{(m)}$.

Note that the groups $G^{(m)} \wr (\mathbb{Z}/2)$ are given by a recursive sequence of presentations and the maps u_m are given by a recursive sequence of generating sets for the subgroups $P^{(m)} \rtimes (\mathbb{Z}/2)$.

If $Q^{(m)} \neq 1$ then $P^{(m)} \rtimes (\mathbb{Z}/2)$ does not have FA, since it maps onto $Q^{(m)}$. But if $Q^{(m)} = 1$ then u_m is an isomorphism, so $P^{(m)}$ does have FA. And by construction, there is no algorithm to decide which of these alternatives holds. □

Acknowledgement

I thank the referee for their careful reading and helpful comments.

References

[1] **M Aka**, *Profinite completions and Kazhdan's property (T)*, Groups Geom. Dyn. 6 (2012) 221–229 MR Zbl

[2] **H Bass**, **A Lubotzky**, *Nonarithmetic superrigid groups: counterexamples to Platonov's conjecture*, Ann. of Math. 151 (2000) 1151–1173 MR Zbl

[3] **M R Bridson**, *Controlled embeddings into groups that have no non-trivial finite quotients*, from "The Epstein birthday schrift", Geom. Topol. Monogr. 1, Geom. Topol. Publ., Coventry (1998) 99–116 MR Zbl

[4] **M R Bridson**, *Helly's theorem*, CAT(0) *spaces, and actions of automorphism groups of free groups*, preprint, Oxford (2007)

[5] **M R Bridson**, *Decision problems and profinite completions of groups*, J. Algebra 326 (2011) 59–73 MR Zbl

[6] **M R Bridson**, *On the dimension of* CAT(0) *spaces where mapping class groups act*, J. Reine Angew. Math. 673 (2012) 55–68 MR Zbl

[7] **M R Bridson**, *Cube complexes, subgroups of mapping class groups, and nilpotent genus*, from "Geometric group theory", IAS/Park City Math. Ser. 21, Amer. Math. Soc., Providence, RI (2014) 379–399 MR Zbl

[8] **M R Bridson**, *The strong profinite genus of a finitely presented group can be infinite*, J. Eur. Math. Soc. 18 (2016) 1909–1918 MR Zbl

[9] **M R Bridson**, *The homology of groups, profinite completions, and echoes of Gilbert Baumslag*, from "Elementary theory of groups and group rings, and related topics", de Gruyter, Berlin (2020) 11–27 MR Zbl

[10] **M R Bridson**, **F J Grunewald**, *Grothendieck's problems concerning profinite completions and representations of groups*, Ann. of Math. 160 (2004) 359–373 MR Zbl

[11] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[12] **M R Bridson**, **J Howie**, **C F Miller III**, **H Short**, *On the finite presentation of subdirect products and the nature of residually free groups*, Amer. J. Math. 135 (2013) 891–933 MR Zbl

[13] **T Cheetham-West**, **A Lubotzky**, **A W Reid**, **R Spitler**, *Property FA is not a profinite property*, Groups Geom. Dyn. (online publication October 2024)

[14] **D J Collins**, **C F Miller III**, *The word problem in groups of cohomological dimension* 2, from "Groups St Andrews 1997 in Bath, I" (C M Campbell, E F Robertson, N Ruskuc, G C Smith, editors), Lond. Math. Soc. Lect. Note Ser. 260, Cambridge Univ. Press (1999) 211–218 MR Zbl

[15] **A Grothendieck**, *Représentations linéaires et compactification profinie des groupes discrets*, Manuscripta Math. 2 (1970) 375–396 MR Zbl

[16] **V P Platonov**, **O I Tavgen**, *Grothendieck's problem on profinite completions and representations of groups*, K–Theory 4 (1990) 89–101 MR Zbl

[17] **L Ribes**, **P Zalesskii**, *Profinite groups*, Ergebnisse der Math. 40, Springer (2000) MR Zbl

[18] **D T Wise**, *A residually finite version of Rips's construction*, Bull. Lond. Math. Soc. 35 (2003) 23–29 MR Zbl

[19] **D T Wise**, *Cubulating small cancellation groups*, Geom. Funct. Anal. 14 (2004) 150–214 MR Zbl

Mathematical Institute, University of Oxford
Oxford, United Kingdom

bridson@maths.ox.ac.uk

Received: 22 June 2023 Revised: 9 September 2023

Algebraic & Geometric Topology **24**:7 (2024) 4115–4125
DOI: 10.2140/agt.2024.24.4115
Published: 9 December 2024

Slice genus bound in DTS^2 from s–invariant

QIUYU REN

We prove a recent conjecture of Manolescu and Willis which states that the s–invariant of a knot in $\mathbb{RP}^3$ (as defined by them) gives a lower bound on its null-homologous slice genus in the unit disk bundle of TS^2. We also conjecture a lower bound in the more general case where the slice surface is not necessarily null-homologous, and give its proof in some special cases.

57K18; 57K10, 57K40

1 Introduction

Rasmussen [7] famously defined the s–invariant for knots in S^3 using Khovanov homology theory [3], and proved that for a knot K in S^3,

$$2g_4(K) \geq |s(K)|, \tag{1}$$

where $g_4(K)$ is the slice genus of K, which can be defined as the minimal genus of an orientable cobordism (in $S^3 \times [0,1]$) from K to the unknot.

Analogously, Manolescu, Marengon, Sarkar and Willis [5] and Manolescu and Willis [6] defined s–invariants ($\mathbb{Z}$–valued, like the usual s–invariant) for null-homologous knots in $S^1 \times S^2$ and for all knots in $\mathbb{RP}^3$, respectively, and proved the same inequality (1) in these settings. Here the slice genus $g_4(K)$ for K is still defined as the minimal genus of an orientable cobordism from K to an unknot depending on the homology class of K (there are two unknots in $\mathbb{RP}^3$, one null-homologous and one not).

For an integer d, let $D(d)$ denote the D^2–bundle over S^2 with Euler number d. Thus $D(0) = D^2 \times S^2$ with boundary $S^1 \times S^2$; $D(1) = \mathbb{CP}^2 \backslash B^4$ with boundary S^3; $D(2) = DTS^2$, the unit disk bundle of the tangent bundle of S^2, with boundary $\mathbb{RP}^3$. For a null-homologous properly embedded orientable connected surface $\Sigma \subset D(d)$ with boundary a knot $K \subset \partial D(d)$, for $d = 0, 1, 2$, the genus bound

$$2g(\Sigma) \geq -s(K) \tag{2}$$

was proved for $d = 0, 1$ [5, Theorem 1.15, Corollary 1.9] and conjectured for $d = 2$ [6, Conjecture 6.9]. We prove this conjecture.

Theorem *If* $(\Sigma, K) \subset (DTS^2, \mathbb{RP}^3)$ *is a null-homologous properly embedded orientable connected surface that bounds a knot* $K \subset \mathbb{RP}^3$*, then*

$$2g(\Sigma) \geq -s(K).$$

Remark 1 By reversing the orientation of $D(d)$, (2) for $d = 0, 1, 2$ implies

$$2g(\Sigma) \geq s(K)$$

for $d = 0, -1, -2$. Since a cobordism in $\partial D(d) \times [0, 1]$ from a null-homologous $K \subset \partial D(d)$ to an unknot can be capped off to become a slice surface in $D(\pm d)$, (2) for $d = 0, 1, 2$ can be considered as refinements of (1) for null-homologous knots K in $S^1 \times S^2, S^3, \mathbb{RP}^3$, respectively. In the only other case, namely when $K \subset \mathbb{RP}^3$ is not null-homologous, (1) is refined by (4) below with $[\Sigma] = \pm 1$.

When $d = 0$, $\Sigma \subset D^2 \times S^2$ being null-homologous is equivalent to $\partial\Sigma \subset S^1 \times S^2$ being null-homologous. Since the s–invariant is only defined for null-homologous knots in $S^1 \times S^2$, the null-homologous condition on Σ puts no restriction. When $d = 1, 2$, however, we could ask whether $s(K)$ gives genus bounds for slice surfaces of K in $D(d)$ that are not necessarily null-homologous. For $d = 1$ this is conjectured in [5] and proved by Ren [8, Corollary 1.5]. Explicitly, for any $(\Sigma, K) \subset (\mathbb{CP}^2 \backslash B^4, S^3)$ we have

(3) $$2g(\Sigma) \geq -s(K) - [\Sigma]^2 + |[\Sigma]|,$$

where $|\cdot|$ is the L^1–norm (equivalently, absolute value) on $H_2(\mathbb{CP}^2 \backslash B^4, S^3) \cong \mathbb{Z}$. We pose the following conjecture for the case $d = 2$.

Conjecture *If* $(\Sigma, K) \subset (DTS^2, \mathbb{RP}^3)$ *is a properly embedded orientable connected surface that bounds a knot* $K \subset \mathbb{RP}^3$*, then*

(4) $$2g(\Sigma) \geq -s(K) - \tfrac{1}{2}[\Sigma]^2.$$

The main theorem shows the conjecture when $[\Sigma] = 0$. In fact, the same proof applies to show it in a couple more cases.

Proposition 2 *Inequality* (4) *holds if* $[\Sigma] = \pm 1, \pm 2, \pm 3$ *in* $H_2(DTS^2, \mathbb{RP}^3) \cong \mathbb{Z}$.

In fact, in the various settings above, the s–invariants are defined for links as well as knots [1; 5; 6]. As remarked in [5; 6], (2) for $d = 0, 1, 2$ and (3) reduce to computing the s–invariants of some special family of links; similarly, (4) also reduces to computing the s–invariant of a family of links. We will explain these reductions in Section 2. In Section 3, we calculate the s–invariants that enable us to conclude the main theorem and Proposition 2. We also pose a technical question whose positive answer implies the conjecture above.

Acknowledgements

We thank Ciprian Manolescu for suggesting this problem and for helpful discussions. We also thank the referee for a careful reading of the paper.

2 Reduce to s–invariant calculations

In this section, we define links $T(d; p,q) \subset \partial D(d)$ for $p,q \geq 0$, such that (2) reduces to the calculation of the s–invariants of $T(d; p, p)$ for $d = 0, 1, 2$, and (3) and (4) reduce to that of $T(d; p,q)$ for $d = 1, 2$. The strategy is essentially due to [5; 6], but we carry it out explicitly for completeness.

Let $S \subset D(d)$ denote the core 2–sphere of the D^2–bundle $D(d) \to S^2$. Any properly embedded oriented surface $\Sigma \subset D(d)$ can be perturbed so that it intersects S transversely at some p points positively and some q points negatively. Removing a tubular neighborhood of S, we obtain a properly embedded surface $\Sigma_0 \subset \partial D(d) \times [0, 1]$, whose boundary on $\partial D(d) \times \{1\}$ is the original boundary of Σ and whose boundary on $\partial D(d) \times \{0\}$ is an oriented link in $\partial D(d)$ consisting of $p+q$ fibers of the S^1–bundle $\partial D(d) \to S^2$, p of which are oriented positively and q of which negatively. We denote the mirror of this link by $T(d; p,q)$; thus, Σ_0 is a cobordism from $\overline{T(d; p,q)}$ to $\partial\Sigma$.

Example 3 $T(0; p,q)$ is the disjoint union of $p+q$ knots of the form $* \times S^2 \subset S^1 \times S^2$, p of which are oriented upwards and q of which downwards. It is denoted by $F_{p,q}$ in [5].

Example 4 $\overline{\partial D(1)} \to S^2$ is the Hopf fibration, hence its fibers have pairwise linking number 1. Thus $T(1; p,q)$ is the torus link $T(p+q, p+q)$ in which p of the strands are oriented against the other q strands. It is denoted by $F_{p,q}(1)$ in [5] and $T(p+q, p+q)_{p,q}$ in [8].

Example 5 Think of $\mathbb{RP}^3$ as the 3–ball B^3 with antipodal points on the boundary identified. Then $T(2; p,q)$ can be obtained by standardly embedding a half-twist on $p+q$ strands, p of which are oriented against the other q, into B^3 such that the endpoints land on the boundary. This can be seen by realizing $T(2; p,q) \subset \mathbb{RP}^3$ as the quotient of $T(1; p,q) \subset S^3$ by the standard involution on S^3 which gives the quotient $\mathbb{RP}^3$. $T(2; p, p)$ is denoted by H_p in [6].

By [1; 5; 6], if Σ is an oriented cobordism in $Y \times [0, 1]$ between two (null-homologous if $Y = S^1 \times S^2$) oriented links L_0 and L_1 in Y for $Y = S^3, S^1 \times S^2, \mathbb{RP}^3$, such that every component of Σ has a boundary in L_0, then

$$s(L_1) - s(L_0) \geq \chi(\Sigma). \tag{5}$$

By construction, if $(\Sigma, L) \subset (D(d), \partial D(d))$ is a properly embedded oriented connected surface without closed components, by deleting a tubular neighborhood of the core $S \subset D(d)$, we obtain a cobordism Σ_0 from some $\overline{T(d; p,q)}$ to L, each of whose components has a boundary in L. Turning the cobordism upside down and applying (5) gives

$$s(T(d; p,q)) - s(\bar{L}) \geq \chi(\Sigma_0) = \chi(\Sigma) - p - q, \tag{6}$$

where the last inequality holds because topologically Σ_0 is Σ with $p+q$ disks removed from its interior.

The number $p-q$ equals to the homology class $[\Sigma]\in H_2(D(d),\partial D(d))$ upon an identification

$$H_2(D(d),\partial D(d))\cong H^2(D(d))\cong \mathbb{Z}.$$

Thus, for $[\Sigma]=p-q$ a fixed number, if $s(T(d;p,q))+p+q$ is independent of specific p,q, then (6) can be rewritten in terms of $s(\bar{L})$, $\chi(\Sigma)$, and $[\Sigma]$. This is the case for $d=0$ and $[\Sigma]=0$, as well as for $d=1$ with any $[\Sigma]$. Explicitly, by [5, Theorem 1.6; 8, Theorem 1.1] we have

$$s(T(0;p,p))=-2p+1,\quad s(T(1;p,q))=(p-q)^2-2\max(p,q)+1. \tag{7}$$

We conjecture this is also true for $d=2$ and any $[\Sigma]$, with

$$s(T(2;p,q))=\lfloor \tfrac{1}{2}(p-q)^2\rfloor-p-q+1, \tag{8}$$

and give a proof of it for $[\Sigma]=0,\pm1,\pm2,\pm3$.

In the special case when $L=K$ is a knot, we have $s(\bar{L})=s(\bar{K})=-s(K)$ by [5, Proposition 8.8(1); 6, Proposition 4.10; 7, Proposition 3.10]. In this case, plugging (7) into (6) gives (2) for $d=0,1$, and (3). Plugging (8) into (6) would give the conjectural inequality (4), although we are only able to prove it for $|p-q|\le 3$.

Remark 6 (1) It is easy to prove (8) for $pq=0$, since in this case $T(2;p,q)$ is a positive link and one can apply [6, Remark 6.3]. However this does not help in establishing (4).

(2) If (8) were true in general, one can proceed as in [8, Section 4] to determine the entire quantum filtration structure of the Lee homology (as defined in [6]) of $s(T(2;p,q))$.

3 s–invariants of $T(2;p,q)$

As explained in Section 2, the main theorem and Proposition 2 reduce to the following proposition.

Proposition 7 *For $p,q\ge 0$ with $|p-q|\le 3$, the s–invariant (as defined in [6]) of the link $T(2;p,q)\subset\mathbb{RP}^3$ defined in Section 2 is given by*

$$s(T(2;p,q))=\lfloor \tfrac{1}{2}(p-q)^2\rfloor-p-q+1.$$

Stošić [9, Theorem 3] calculated the Khovanov homology groups of the positive torus links $T(2n,2n)$ in their highest nontrivial homological grading $h=2n^2$. For dimension reasons, the Lee spectral sequence from $\mathrm{Kh}(T(2n,2n))\otimes\mathbb{Q}$ to the Lee homology $\mathrm{Kh}_{\mathrm{Lee}}(T(2n,2n))$ collapses immediately in this homological degree. This can be used to give an alternative proof of $s(T(1;p,p))=1-2p$, a fact that is reproved in [5, Theorem 1.7]. We prove Proposition 7 by adapting the argument of Stošić. It is worth remarking that the calculation of the more general $s(T(1;p,q))$ was done in [8] by pushing Stošić's argument slightly further. However, we were not able to achieve the same here to prove (8) in its full generality (see Remark 9).

3.1 Review of Khovanov homology in $\mathbb{RP}^3$

We first briefly review some properties of the Khovanov/Lee homology and the s–invariant of links in $\mathbb{RP}^3$, following [6]. We only give definitions that will be relevant to us. We assume the reader is familiar with the usual theory in S^3, in particular [3; 4; 7].

Think of $\mathbb{RP}^3 \backslash *$ as the twisted I–bundle over $\mathbb{RP}^2$; we see that links in $\mathbb{RP}^3$ can be represented by link diagrams in $\mathbb{RP}^2$, and two different diagrams of the same link are related by the usual three Reidemeister moves. Although the over/under strands are not well defined at a crossing, it is unambiguous to distinguish positive/negative crossings (if the link is oriented or has only one component) or to define $0/1$–resolutions at a crossing in such a link diagram.

Let $L \subset \mathbb{RP}^3$ be an oriented link with an oriented link diagram D. Let $2^n = (0 \to 1)^n$ denote the hypercube of complete resolutions seen as a directed graph, where n is the number of crossings in D. Every vertex v corresponds to a complete resolution D_v, which is assigned an abelian group $C(D_v)$, bigraded by two parameters q and k. Every edge e from a vertex v to a vertex w corresponds to a saddle from D_v to D_w, which is assigned four maps $\partial_0^e, \partial_-^e, \Phi_0^e, \Phi_+^e : C(D_v) \to C(D_w)$ of bidegree $(0,0)$, $(0,-2)$, $(4,0)$, and $(4,2)$, respectively. The *Khovanov complex* of D is

$$C(D) := \bigoplus_v C(D_v)[-n_- + |v|]\{n_+ - 2n_-\}$$

equipped with the differential $\partial := \sum_e \partial_0^e$. Here $[\,\cdot\,]$ denotes the homological grading shift, $\{\,\cdot\,\}$ denotes the shift in the first grading (called quantum grading) of $C(D_v)$, $n_\pm$ denotes the number of positive/negative crossings in D and $|v|$ denotes the number of 1's in v. The *Lee complex* is $C_{\mathrm{Lee}}(D) := C(D) \otimes \mathbb{Q}$ equipped with the differential $\partial_{\mathrm{Lee}} := \sum_e (\partial_0^e + \partial_-^e + \Phi_+^e + \Phi_0^e) \otimes \mathbb{Q}$. For our purpose, we also consider a *deformed Khovanov complex*, defined as $C'(D) := C(D)$ equipped with the differential $\partial' := \sum_e (\partial_0^e + \partial_-^e)$. The cohomologies of these three complexes are denoted by $\mathrm{Kh}(L)$, $\mathrm{Kh}_{\mathrm{Lee}}(L)$, and $\mathrm{Kh}'(L)$, respectively, which do not depend on the choice of the link diagram D.

The group $\mathrm{Kh}(L)$ is trigraded by h (homological grading), q (quantum grading), and k;[1] $\mathrm{Kh}'(L)$ is bigraded by h and q; $\mathrm{Kh}_{\mathrm{Lee}}(L)$ is graded by h and filtered by q. As a vector space, $\mathrm{Kh}_{\mathrm{Lee}}(L) \cong \mathbb{Q}^{2^{|L|}}$ is spanned by some generators $[s_{\mathfrak{o}}]$, where $\mathfrak{o}$ runs over the all possible orientations of L as an unoriented link.[2] When $\mathfrak{o}$ is the given orientation on L, $[s_{\mathfrak{o}}]$ sits in homological degree 0, and its quantum filtration degree plus 1 is defined as the *s–invariant* of L. As a filtered complex, the associated graded complex of $C_{\mathrm{Lee}}(L)$ is exactly $C'(L) \otimes \mathbb{Q}$. Thus, there is a spectral sequence with E_1–page $\mathrm{Kh}'(L) \otimes \mathbb{Q}$ that converges to $\mathrm{Kh}_{\mathrm{Lee}}(L)$, whose r^{th} differential has bidegree $(1, 4r)$.

The orientation on L plays only a minor role on the group $\mathrm{Kh}^\bullet(L)$, where $\bullet$ denotes one of the three favors we are considering. Explicitly, negating the orientation on a sublink $L' \subset L$ shifts its grading by

[1] The grading k takes values in $\{0, \pm 1\}$, and is related to the homology class of L. For our purpose we will not need to consider this grading in what follows.

[2] In fact, the definition of $[s_{\mathfrak{o}}]$ depends on some auxiliary choices, which we may ignore here.

$[2\ell]\{6\ell\}$, where ℓ is the linking number between L' and $L\backslash L'$ with the new orientations. Here the linking number of two disjoint oriented links in $\mathbb{RP}^3$ takes half integer values, and can be defined to be a half of the linking number between their lifts in S^3.

Every cobordism $\Sigma\colon L_0 \to L_1$ between two oriented links with diagrams D_0 and D_1 induces a chain map $C^\bullet(\Sigma)\colon C^\bullet(D_0) \to C^\bullet(D_1)$ with some grading shifts. By design, if $D_{0/1}$ are the $0/1$–resolutions at a crossing of a link diagram D of some link L, and Σ is the obvious saddle cobordism, then $C^\bullet(D)$ is isomorphic to the mapping cone of $C^\bullet(\Sigma)$ up to grading shifts. More explicitly, for our convenience we record that if $D_{0/1}$ are the $0/1$–resolutions of D at a positive crossing, and L_0 is assigned the induced orientation from L while L_1 is assigned any orientation, then

$$C'(D) \cong \mathrm{Cone}(C'(D_0) \to C'(D_1)[c]\{3c+1\})[1]\{1\},$$

where $c = n_-(D_1) - n_-(D)$. Thus we have the following exact triangle of deformed Khovanov homology groups:

$$\begin{array}{ccc} \mathrm{Kh}'(L_1)[c+1]\{3c+2\} & \longrightarrow & \mathrm{Kh}'(L) \\ & {\scriptstyle [1]}\nwarrow \quad \swarrow & \\ & \mathrm{Kh}'(L_0)\{1\} & \end{array} \tag{9}$$

If $\Sigma\colon L_0 \to L_1$ is an oriented cobordism, the induced map $C^\bullet(\Sigma)$ preserves the homological grading h and changes the quantum grading q by $\chi(\Sigma)$ (or is of q–filtered degree $\chi(\Sigma)$ in the case of C_{Lee}). Moreover, the induced map $\mathrm{Kh}_{\mathrm{Lee}}(\Sigma)$ sends a generator $[s_{\mathfrak{o}}] \in \mathrm{Kh}_{\mathrm{Lee}}(L_0)$ to some $\sum_{\mathfrak{o}'} \lambda_{\mathfrak{o}'}[s_{\mathfrak{o}'}] \in \mathrm{Kh}_{\mathrm{Lee}}(L_1)$, where $\mathfrak{o}'$ runs over orientations of L_1 such that there is an orientation on Σ making it an oriented cobordism $(L_0, \mathfrak{o}) \to (L_1, \mathfrak{o}')$, and $\lambda_{\mathfrak{o}'} \neq 0$ provided Σ has no closed components. In particular, this implies (5) because in that case there is exactly one choice of $\mathfrak{o}'$.

Finally, we remark that there are by definition two unknots in $\mathbb{RP}^3$. The *class*-0 *unknot* U_0 is an unknot in a small ball contained in $\mathbb{RP}^3$; the *class*-1 *unknot* U_1 is a copy of the standardly embedded $\mathbb{RP}^1 \subset \mathbb{RP}^3$. Both these unknots have rank-2 deformed Khovanov homology given by $\mathrm{Kh}'^{0,\pm 1}(U_i) = \mathbb{Z}$. The deformed Khovanov homology behaves as expected under the disjoint union of two links, one in $\mathbb{RP}^3$ and one in S^3. In particular, regarding U_0 as a knot in S^3, we have $\mathrm{Kh}'(L \sqcup U_0) = \mathrm{Kh}'(L)\{1\} \oplus \mathrm{Kh}'(L)\{-1\}$ for any link $L \subset \mathbb{RP}^3$.

3.2 Calculation of s

Now we are ready to prove Proposition 7. We first define two auxiliary families of links, T_n^i and S_n^i for $0 \le i \le n-1$. These should be compared with $D_{n,n-1}^i$ and $D_{n,n}^i$ in [8, Section 5].

Think of $\mathbb{RP}^2$ as D^2 with antipodal points on the boundary identified. A braid diagram can be placed into D^2 with its endpoints on ∂D^2, identified pairwise to give a link diagram in $\mathbb{RP}^2$; this is called by [6] the projective closure of the given braid. Let T_n^i be the link represented by the projective closure of $\sigma_{n-1}(\sigma_{n-2}\sigma_{n-1})\cdots(\sigma_2\cdots\sigma_{n-1})(\sigma_1\cdots\sigma_i)$, and S_n^i be the link presented by the projective closure

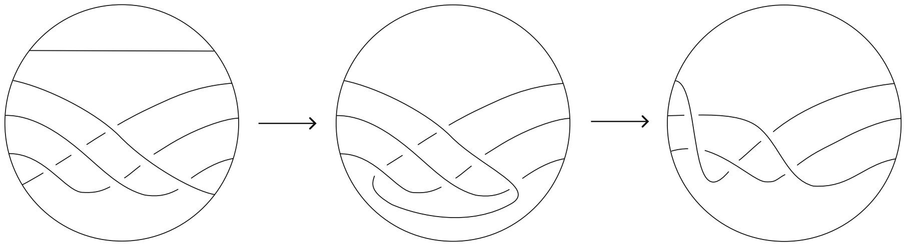


Figure 1: Check T_5^0 and S_3^2 are isotopic via Reidemeister moves.

of $\sigma_{n-1}(\sigma_{n-2}\sigma_{n-1})\cdots(\sigma_1\cdots\sigma_{n-1})(\sigma_{n-1}\cdots\sigma_{n-i})$. We equip T_n^i and S_n^i with the orientation where all strands are oriented upwards. By the description in Example 5, $T(2;n,0)$ is exactly the link T_n^{n-1}, and all $T(2;p,n-p)$ are T_n^{n-1} with a possibly different orientation. Also, by definition $S_n^0=T_n^{n-1}$, and it is easy to check by Reidemeister moves[3] that $T_n^0=S_{n-2}^{n-3}$ (see Figure 1 for the case $n=5$).

For $n\geq 2$ and $i>0$, resolving the crossing in the standard diagram of T_n^i that corresponds to the last word σ_i gives T_n^{i-1} as the 0–resolution and some other link $(T_n^i)_1$ as the 1–resolution. Similarly, resolving the crossing of S_n^i that corresponds to the last σ_{n-i} gives 0–resolution S_n^{i-1} and 1–resolution $(S_n^i)_1$. By Reidemeister moves, one may check that in fact (as unoriented links)

$$(T_n^i)_1=\begin{cases}T_{n-2}^{n-3}\sqcup U_0 & \text{if } i=n-1,\\ T_{n-2}^{i-1} & \text{if } i<n-1,\end{cases}\qquad (S_n^i)_1=\begin{cases}S_{n-2}^{i-2} & \text{if } i>1,\\ S_{n-2}^{0}\sqcup U_0 & \text{if } i=1.\end{cases}\tag{10}$$

Here U_0 is the class-0 unknot, and as a convention we define $T_0^{-1}=S_0^0=\varnothing$ to be the empty link.

Give $(T_n^i)_1$ and $(S_n^i)_1$ the orientations of the right hand sides in the identification (10); the skein exact triangle (9) gives exact triangles

$$\mathrm{Kh}'((T_n^i)_1)[n-1]\{3n-4\}\to\mathrm{Kh}'(T_n^i)\to\mathrm{Kh}'(T_n^{i-1})\{1\}\xrightarrow{[1]},\tag{11}$$

$$\mathrm{Kh}'((S_n^i)_1)[n]\{3n-1\}\to\mathrm{Kh}'(S_n^i)\to\mathrm{Kh}'(S_n^{i-1})\{1\}\xrightarrow{[1]}.\tag{12}$$

We prove a lemma that gives a "graphical lower bound" of the deformed Khovanov homology groups of T_n^i and S_n^i, in the spirit of [8, Theorem 2.1]. In fact, most parts of the statement won't be relevant for our purpose. But since the general statement is not much more complicated to state and to prove, we include it fully here.

Lemma 8 (1) $\mathrm{Kh}'^{h,q}(T_n^i)=0$ *for* $h<0$ *or* $h>\lfloor n^2/4\rfloor$ *or* $q-h<\lfloor n^2/2\rfloor-2n+1+i$.

(2) $\mathrm{Kh}'^{h,q}(S_n^i)=0$ *for* $h<0$ *or* $h>\lfloor n^2/4\rfloor+\lceil i/2\rceil$ *or* $q-h<\lfloor n^2/2\rfloor-n+i$ *or* $q-2h<\lfloor n^2/4\rfloor-n+i$.

Remark 9 The s–invariant of all $T(1;p,q)\subset S^3$ was deduced in [8, Theorem 2.1]. The difficulty that prevents us to similarly deduce (8) from Lemma 8 is that we were not able to establish an injectivity result like the addendum in [8, Theorem 2.1(1)] (there we actually have an isomorphism; however, only injectivity is needed for the proof, and only injectivity is expected in our case). See also Section 3.3.

[3] If one wishes to think the link diagrams as sitting in D^2, there will be two additional Reidemeister moves when one crosses the boundary, as illustrated in [2, Figure 1]. Note however the picture (e) there was incorrectly drawn.

Proof We induct on n and i. For $n=1$ this is immediate, since $S_1^0=T_1^0=U_1$ is the class-1 unknot. For $n=2$, $T_2^0=U_0$ satisfies (1). By (10) and (11), nonzero homology groups of T_2^1 are exactly

$$\mathrm{Kh}'^{h,q}(T_2^1)=\mathbb{Z},\quad (h,q)=(0,0),(0,2),(1,1),(1,3);$$

thus T_2^1 satisfies (1) and $S_2^0=T_2^1$ satisfies (2). By (10)(12), S_2^1 has

$$\mathrm{Kh}'^{h,q}(S_2^1)=\mathbb{Z},\quad (h,q)=(0,1),(0,3),(1,2),(2,6),$$

and zero elsewhere possibly except when $(h,q)=(1,4),(2,4)$; thus it satisfies (2). Now, by the induction hypothesis and (10), (11), and (12), (1) and (2) are inductively proved for $n>2$ by checking the elementary statements that

- the vanishing region (in the hq–coordinate plane) of $\mathrm{Kh}'(T_n^0)$ described in (1) is contained in that of $\mathrm{Kh}'(S_{n-2}^{n-3})$ described in (2);
- for $i>0$, the vanishing region of $\mathrm{Kh}'(T_n^i)$ is contained in that of both $\mathrm{Kh}'((T_n^i)_1)[n-1]\{3n-4\}$ and $\mathrm{Kh}'(T_n^{i-1})\{1\}$;
- the vanishing region of $\mathrm{Kh}'(S_n^0)$ is identical to that of $\mathrm{Kh}'(T_n^{n-1})$;
- for $i>0$, the vanishing region of $\mathrm{Kh}'(S_n^i)$ is contained in that of both $\mathrm{Kh}'((S_n^i)_1)[n]\{3n-1\}$ and $\mathrm{Kh}'(S_n^{i-1})\{1\}$. □

Proof of Proposition 7 We divide into four cases according to the value of $|p-q|$. We give the proof carefully for the case $|p-q|=0$, and more casually for the rest cases, as they will be similar.

Case 1 $|p-q|=0$.

Write $n=p+q=2m$. By Lemma 8, $\mathrm{Kh}'(T_n^i)$ vanishes in homological degrees $h>m^2$. We induct on m to show that

$$\operatorname{rank}\mathrm{Kh}'^{m^2,*}(T_n^i)=2\binom{i}{m}, \tag{13}$$

$$\inf\{q\mid \mathrm{Kh}'^{m^2,q}(T_n^{n-1})\neq 0\}=3m^2-2m. \tag{14}$$

When $m=1$, we have $T_2^0=U_0$ satisfies (13), and T_2^1 satisfies (13) and (14) by the description of $\mathrm{Kh}'(T_2^1)$ in the proof of Lemma 8.

Assume now $m>1$. We have $\mathrm{Kh}'^{m^2,*}(T_n^0)=0$ by Lemma 8(2) applied to $S_{n-2}^{n-3}=T_n^0$; thus T_n^0 satisfies (13). For $i>0$, (11) and (10) give

$$\operatorname{rank}\mathrm{Kh}'^{m^2,*}(T_n^i)\le\operatorname{rank}\mathrm{Kh}'^{m^2,*}(T_n^{i-1})+\operatorname{rank}\mathrm{Kh}'^{(m-1)^2,*}(T_{n-2}^{i-1}),\quad i<n-1, \tag{15}$$

$$\operatorname{rank}\mathrm{Kh}'^{m^2,*}(T_n^{n-1})\le\operatorname{rank}\mathrm{Kh}'^{m^2,*}(T_n^{n-2})+2\operatorname{rank}\mathrm{Kh}'^{(m-1)^2,*}(T_{n-2}^{n-3}). \tag{16}$$

Using (15) iteratively and (16), as well as the induction hypothesis, we obtain

$$\operatorname{rank}\mathrm{Kh}'^{m^2,*}(T_n^i)\le 2\binom{i}{m}.$$

On the other hand, due to the existence of the Lee spectral sequence from $\mathrm{Kh}' \otimes \mathbb{Q}$ to $\mathrm{Kh}_{\mathrm{Lee}}$, we have that $\operatorname{rank} \mathrm{Kh}'^{m^2,*}(T_n^{n-1})$ is bounded below by $\dim \mathrm{Kh}^{m^2}_{\mathrm{Lee}}(T_n^{n-1})$, which equals $\binom{2m}{m} = 2\binom{n-1}{m}$ as $\mathrm{Kh}^{m^2}_{\mathrm{Lee}}(T_n^{n-1})$ is generated by those generators $[s_{\mathfrak{o}}]$ for which $\mathfrak{o}$ is an orientation of T_n^{n-1} realizing $T(2;m,m)$ (note components in T_n^{n-1} have pairwise linking number $1/2$). We conclude that T_n^{n-1} satisfies (13), and so do all T_n^i, because the sharpness of the estimate above shows (15) and (16) are in fact equalities.

The sharpness of estimate also implies that the map $\mathrm{Kh}'(T_{n-2}^{n-3} \sqcup U_0) \to \mathrm{Kh}'(T_n^{n-1})$ in the exact triangle (11) is injective upon tensoring $\mathbb{Q}$. It follows that

$$\inf\{q \mid \mathrm{Kh}'^{m^2,q}(T_n^{n-1}) \neq 0\} \leq \inf\{q \mid \mathrm{Kh}'^{(m-1)^2,q}(T_{n-2}^{n-3}) \neq 0\} - 1 + 3n - 4 = 3m^2 - 2m.$$

Lemma 8(1) gives the reverse inequality, so (14) is also proved.

We return to the calculation of the s–invariant. The sharpness of the estimate of $\operatorname{rank} \mathrm{Kh}'(T_n^{n-1})$ also implies that the Lee spectral sequence from $\mathrm{Kh}'(T_n^{n-1}) \otimes \mathbb{Q}$ to $\mathrm{Kh}_{\mathrm{Lee}}(T_n^{n-1})$ collapses immediately at homological degree $h = m^2$. It follows that the lowest quantum filtration level of $\mathrm{Kh}^{m^2}_{\mathrm{Lee}}(T_n^{n-1})$ is at $q = 3m^2 - 2m$.

Taking into account the bidegree shift $[m^2]\{3m^2\}$, the s–invariant of $T(2;m,m)$ is equal to the quantum filtration degree of $[s_{\mathfrak{o}}] \in \mathrm{Kh}^{m^2}_{\mathrm{Lee}}(T_n^{n-1})$ minus $3m^2 - 1$, where $\mathfrak{o}$ is any orientation of T_n^{n-1} that realizes $T(2;m,m)$ (by symmetry all these $[s_{\mathfrak{o}}]$'s have the same filtration degree). Since $\mathrm{Kh}^{m^2}_{\mathrm{Lee}}(T_n^{n-1})$ is spanned by all such $[s_{\mathfrak{o}}]$, any element in $\mathrm{Kh}^{m^2}_{\mathrm{Lee}}(T_n^{n-1})$ has filtration degree no less than those of $[s_{\mathfrak{o}}]$; in other words, every $[s_{\mathfrak{o}}]$ sits in the lowest filtration level. It follows that

$$s(T(2;m,m)) = (3m^2 - 2m) - (3m^2 - 1) = -2m + 1,$$

proving Proposition 7 for $|p-q| = 0$.

Case 2 $|p-q| = 2$.

Write $n = p + q = 2m$. By an induction on m one can show that

$$\operatorname{rank} \mathrm{Kh}'^{m^2-1,*}(T_n^i) = 2\binom{i}{m+1} + 2\binom{i}{m-1},$$
$$\inf\{q \mid \mathrm{Kh}'^{m^2-1,q}(T_n^{n-1}) \neq 0\} = 3m^2 - 2m - 1.$$

Moreover, $\operatorname{rank} \mathrm{Kh}'^{m^2-1,*}(T_n^{n-1}) = \dim \mathrm{Kh}_{\mathrm{Lee}}(T_n^{n-1})$, which implies the collapsing of the Lee spectral sequence at $h = m^2 - 1$. After a bidegree shift $[m^2 - 1]\{3m^2 - 3\}$, we calculate that

$$s(T(2;m+1,m-1)) = (3m^2 - 2m - 1) - (3m^2 - 3 - 1) = -2m + 3,$$

proving the case $|p-q| = 2$.

Case 3 $|p-q| = 1$.

Write $n = p + q = 2m + 1$. By an induction on m (with base case $m = 0$) one can show that

$$\operatorname{rank} \mathrm{Kh}'^{m^2+m,*}(T_n^i) = 2\binom{i+1}{m+1},$$
$$\inf\{q \mid \mathrm{Kh}'^{m^2+m,q}(T_n^{n-1}) \neq 0\} = 3m^2 + m - 1.$$

Moreover, we also conclude the immediate collapsing of the Lee spectral sequence by a dimension count, and calculate that $s(T(2; m+1, m)) = (3m^2 + m - 1) - (3m^2 + 3m - 1) = -2m$.

Case 4 $|p - q| = 3$.

Write $n = p + q = 2m + 1$. From

$$\operatorname{rank} \mathrm{Kh}'^{m^2+m-2,*}(T_n^i) = 2\binom{i}{m+2} + 2\binom{i}{m-1},$$
$$\inf\{q \mid \mathrm{Kh}'^{m^2+m-2,q}(T_n^{n-1}) \neq 0\} = 3m^2 + m - 3 \quad (m > 0),$$

and a dimension count, we conclude as above that $s(T(2; m+2, m-1)) = -2m + 4$. We remark that in this case one need to take both $m = 0, 1$ as base cases for induction, where $T_3^0 = S_1^0 = T_1^0 = U_0$ and all $\mathrm{Kh}'(T_3^i)$ can be completely determined from (11). □

3.3 A question

As an analogue to [8, Question 6.1], we pose the following question, whose truth is verified in small examples ($n \leq 5$).

Question 10 *Is it true that the saddle cobordism $T_n^i \to T_n^{i-1}$ always induces a surjection on $\mathrm{Kh}' \otimes \mathbb{Q}$?*

A positive answer to Question 10 in the case $i = n - 1$ implies the saddle cobordism

$$T(2; n-2, 0) \sqcup U_0 \to T(2; n, 0)$$

is injective in $\mathrm{Kh}' \otimes \mathbb{Q}$. By the same argument as the proof of Theorem 1.1 ($m = n$) in [8], this implies (8); thus the conjectural genus bound (4). Of course, (8) is a much weaker statement than Question 10 and would follow from the surjectivity of

$$\mathrm{Kh}'^{pq, pq+\lfloor n^2/2 \rfloor - n}(T_n^{n-1}) \otimes \mathbb{Q} \to \mathrm{Kh}'^{pq, pq+\lfloor n^2/2 \rfloor - n - 1}(T_n^{n-2}) \otimes \mathbb{Q}$$

for all $p + q = n$ (see Remark 9).

References

[1] **A Beliakova**, **S Wehrli**, *Categorification of the colored Jones polynomial and Rasmussen invariant of links*, Canad. J. Math. 60 (2008) 1240–1266 MR Zbl

[2] **B Gabrovšek**, *The categorification of the Kauffman bracket Skein module of* $\mathbb{R}\mathrm{P}^3$, Bull. Aust. Math. Soc. 88 (2013) 407–422 MR Zbl

[3] **M Khovanov**, *A categorification of the Jones polynomial*, Duke Math. J. 101 (2000) 359–426 MR Zbl

[4] **E S Lee**, *An endomorphism of the Khovanov invariant*, Adv. Math. 197 (2005) 554–586 MR Zbl

[5] **C Manolescu**, **M Marengon**, **S Sarkar**, **M Willis**, *A generalization of Rasmussen's invariant, with applications to surfaces in some four-manifolds*, Duke Math. J. 172 (2023) 231–311 MR Zbl

[6] **C Manolescu**, **M Willis**, *A Rasmussen invariant for links in* $\mathbb{RP}^3$, preprint (2023) arXiv 2301.09764

[7] **J Rasmussen**, *Khovanov homology and the slice genus*, Invent. Math. 182 (2010) 419–447 MR Zbl

[8] **Q Ren**, *Lee filtration structure of torus links*, preprint (2023) arXiv 2305.16089

[9] **M Stošić**, *Khovanov homology of torus links*, Topology Appl. 156 (2009) 533–541 MR Zbl

Department of Mathematics, University of California, Berkeley
Berkeley, CA, United States

qiuyu_ren@berkeley.edu

Received: 3 July 2023 Revised: 9 October 2023

Algebraic & Geometric Topology **24**:7 (2024) 4127–4137
DOI: 10.2140/agt.2024.24.4127
Published: 9 December 2024

Relatively geometric actions of Kähler groups on CAT(0) cube complexes

COREY BREGMAN
DANIEL GROVES
KEJIA ZHU

We prove that for $n \geq 2$, a nonuniform lattice in $\mathrm{PU}(n,1)$ does not admit a relatively geometric action on a CAT(0) cube complex. As a consequence, if Γ is a nonuniform lattice in a noncompact semisimple Lie group G without compact factors that admits a relatively geometric action on a CAT(0) cube complex, then G is commensurable with $\mathrm{SO}(n,1)$. We also prove that if a Kähler group is hyperbolic relative to residually finite parabolic subgroups, and acts relatively geometrically on a CAT(0) cube complex, then it is virtually a surface group.

20F65, 22E40; 32J05, 32J27, 57N65

1 Introduction

A finitely generated group is called *cubulated* if it acts properly cocompactly on a CAT(0) cube complex. Agol [1], building on the work of Wise [29] and many others, proved that cubulated hyperbolic groups enjoy many important properties, and used this to solve several open conjectures in 3–manifold topology, in particular the virtual Haken and virtual fibering conjectures. Wise [29, Section 17] proved the virtual fibering conjecture in the noncompact finite-volume setting, using the relatively hyperbolic structure of the fundamental group.

Einstein and Groves define the notion of a *relatively geometric* action of a group pair $(\Gamma, \mathcal{P})$ on a CAT(0) cube complex [10]. For such an action, elements of $\mathcal{P}$ act elliptically. This allows the possibility that even though the elements of $\mathcal{P}$ might not act properly on any CAT(0) cube complex, there still may be a relatively geometric action. Relatively geometric actions are a natural generalization of proper actions and share many of the same features as in the proper case, especially when Γ is hyperbolic relative to $\mathcal{P}$.

Uniform lattices in $\mathrm{SO}(3,1)$ always act geometrically, and thus relatively geometrically, on CAT(0) cube complexes; see Bergeron and Wise [5]. Bergeron, Haglund and Wise [4] prove that in higher dimensions, lattices in $\mathrm{SO}(n,1)$ which are arithmetic of simplest type are cubulated. It also follows from this and Wise's quasiconvex hierarchy theorem [29] that many "hybrid" hyperbolic n–manifolds have cubulated fundamental groups. In the relatively geometric setting, using the work of Cooper and Futer [7], Einstein and Groves prove that nonuniform lattices in $\mathrm{SO}(3,1)$ also admit relatively geometric actions, relative to

Zhu is the corresponding author. All the authors are the joint first author.

their cusp subgroups [10]. In fact, they prove that if $(G, \mathcal{P})$ is hyperbolic relative to free abelian subgroups and the Bowditch boundary $\partial(G, \mathcal{P})$ is homeomorphic to S^2, then G is isomorphic to a nonuniform lattice in $\mathrm{SO}(3, 1)$ if and only if $(G, \mathcal{P})$ admits a relatively geometric action on a CAT(0) cube complex. This is a relative version of the work of Markovic [24] and Haïssinsky [21] in the convex–cocompact setting, giving an equivalent formulation of the Cannon conjecture in terms of actions on hyperbolic CAT(0) cube complexes. It is not known whether the above results extend to all lattices in $\mathrm{SO}(n, 1)$ for $n \geq 4$.

On the other hand, the work of Delzant and Gromov implies that uniform lattices in $\mathrm{PU}(n, 1)$ are not cubulated [8, Corollary, page 52]. Recall that a group Γ is *Kähler* if $\Gamma \cong \pi_1(X)$ for some compact Kähler manifold X. If $\Gamma \leq \mathrm{PU}(n, 1)$ is a torsion-free uniform lattice, then Γ acts freely properly discontinuously cocompactly on complex hyperbolic n–space $\mathbb{H}^n_{\mathbb{C}}$. The quotient $M = \Gamma \backslash \mathbb{H}^n_{\mathbb{C}}$ is a closed negatively curved Kähler manifold, and in particular Γ is a hyperbolic Kähler group. In this context, Delzant and Gromov showed that any infinite Kähler group that is hyperbolic and cubulated is commensurable to a surface group of genus $g \geq 2$ [8, Corollary, page 52]. Thus Γ is not cubulated for $n \geq 2$. Since every uniform lattice in $\mathrm{PU}(n, 1)$ is virtually torsion free, it follows that uniform lattices in $\mathrm{PU}(n, 1)$ are not cubulated if $n \geq 2$.

On the other hand, uniform lattices in $\mathrm{PU}(1, 1) = \mathrm{SO}(2, 1)$, are finite extensions of hyperbolic surface groups, and hence are hyperbolic and cubulated. Similarly, nonuniform lattices in $\mathrm{PU}(1, 1)$ are the orbifold fundamental groups of surfaces with finitely many cusps, and hence virtually free. Such lattices admit both proper cocompact and relatively geometric actions on CAT(0) cube complexes. Since the cusp subgroups of a nonuniform lattice in $\mathrm{PU}(n, 1)$ for $n \geq 2$ are virtually nilpotent but not virtually abelian, it follows from a result of Haglund [20] that such a lattice does not admit a proper action on a CAT(0) cube complex (see Proposition 4.2).

However, the parabolic subgroups do not yield such an obstruction to the existence of a relatively geometric action. Thus this leaves open the question of whether nonuniform lattices in $\mathrm{PU}(n, 1)$ admit relatively geometric actions on CAT(0) cube complexes for $n \geq 2$. Our first result answers this in the negative:

Theorem 1.1 *Let $\Gamma \leq \mathrm{PU}(n, 1)$ be a nonuniform lattice with $n \geq 2$, and let $\mathcal{P}$ be the collection of cusp subgroups of Γ. Then $(\Gamma, \mathcal{P})$ does not admit a relatively geometric action on a* CAT(0) *cube complex.*

Corollary 1.2 *Let Γ be a lattice in a noncompact semisimple Lie group G without compact factors. If either*

- (i) *Γ is uniform and cubulated hyperbolic, or*
- (ii) *Γ is nonuniform, hyperbolic relative to its cusp subgroups $\mathcal{P}$, and $(\Gamma, \mathcal{P})$ admits a relatively geometric action on a* CAT(0) *cube complex,*

then G is commensurable to $\mathrm{SO}(n, 1)$ *for some $n \geq 1$.*

We say that a relatively hyperbolic group pair $(\Gamma, \mathcal{P})$ is *properly* relatively hyperbolic if $\mathcal{P} \neq \{\Gamma\}$. The following result considers general relatively geometric actions of Kähler relatively hyperbolic groups on CAT(0) cube complexes (when the peripheral subgroups are residually finite):

Theorem 1.3 *Let* $(\Gamma, \mathcal{P})$ *be a properly relatively hyperbolic pair such that each element of* $\mathcal{P}$ *is residually finite. If* Γ *is Kähler and acts relatively geometrically on a* CAT(0) *cube complex, then* Γ *is virtually a hyperbolic surface group.*

We will deduce Theorem 1.1 from Theorem 1.3 in Section 4. In fact, nonuniform lattices in $\mathrm{PU}(n,1)$ are Kähler for $n \geq 3$; see Toledo [28, Theorem 2]. Hence Theorem 1.1 follows immediately from Theorem 1.3 in this range. However, our proof of Theorem 1.1 will work for all $n \geq 2$, and will not use this fact. In [9], Delzant and Py considered actions of Kähler groups on locally finite finite-dimensional CAT(0) cube complexes that are more general than geometric ones (see Theorem A for precise hypotheses), and showed that every such action virtually factors through a surface group. We remark that the cube complexes appearing in relatively geometric actions will in general not be locally finite.

We conclude the introduction with a sample application of Theorem 1.3:

Corollary 1.4 *Suppose that* A *and* B *are infinite residually finite groups which are not virtually free. No* $C'\left(\frac{1}{6}\right)$*–small cancellation quotient of* $A * B$ *is Kähler.*

Proof Let Γ be such a small cancellation quotient of $A * B$. By Einstein and Ng [13, Theorem 1.1], Γ is residually finite and admits a relatively geometric action on a CAT(0) cube complex. If Γ were Kähler, it would be a virtual surface group by Theorem 1.3. However, A embeds in Γ as an infinite-index subgroup, and the only infinite-index subgroups of virtual surface groups are virtually free. □

Outline In Section 2 we review the definition of a relatively geometric action of a group pair on a CAT(0) cube complex and the notion of group-theoretic Dehn fillings, and then collect some known results about these. In Section 3 we prove Theorem 1.3. In Section 4, after reviewing the Borel–Serre and toroidal compactifications of nonuniform quotients of complex hyperbolic space, we prove Theorem 1.1.

Acknowledgments Bregman was supported by NSF grant DMS-2052801. Groves was supported by NSF grants DMS-1904913 and DMS-2203343. Zhu would like to thank his advisor, Daniel Groves, for introducing him to the subject and answering his questions. He would like to thank his coadvisor, Anatoly Libgober, for his constant support and warm encouragement. We are also grateful to the referee for many helpful comments that improved the paper.

2 Actions on CAT(0) cube complexes

In this section we review the notion of a relatively geometric action of a group pair $(\Gamma, \mathcal{P})$ on a CAT(0) cube complex, defined by Einstein and Groves in [10]. We then introduce Dehn fillings of group pairs and recall some useful results from [11].

Definition 2.1 Let Γ be a group and $\mathcal{P}$ a collection of subgroups of Γ. A (cellular) action of Γ on a cell complex X is *relatively geometric with respect to $\mathcal{P}$* if

(i) $\Gamma \setminus X$ is compact,

(ii) each element of $\mathcal{P}$ acts elliptically on X, and

(iii) each cell stabilizer in X is either finite or else conjugate to a finite-index subgroup of an element of $\mathcal{P}$.

Recall that if $(\Gamma, \mathcal{P})$ is a relatively hyperbolic group pair and $\Gamma_0 \leq \Gamma$ has finite index then $(\Gamma_0, \mathcal{P}_0)$ is also a relatively hyperbolic group pair, where $\mathcal{P}_0$ is the set of representatives of the Γ_0–conjugacy classes of

$$\{P^g \cap \Gamma_0 \mid g \in \Gamma,\ P \in \mathcal{P}\} \tag{1}$$

Since $[\Gamma : \Gamma_0]$ is finite, $\mathcal{P}_0$ is still a finite collection of subgroups. It follows that if Γ admits a relatively geometric action on a cell complex X, then $(\Gamma_0, \mathcal{P}_0)$ also admits a relatively geometric action on X by restriction. Thus we have the following result:

Lemma 2.2 *Let $\Gamma_0 \leq \Gamma$ be a finite-index subgroup. If $(\Gamma, \mathcal{P})$ has a relatively geometric action on a cell complex X, then the restriction of this action to $(\Gamma_0, \mathcal{P}_0)$ is also relatively geometric, where $\mathcal{P}_0$ is defined as in (1).*

2.1 Dehn fillings

Dehn fillings first appeared in the context of 3–manifold topology and were subsequently generalized to the group-theoretic setting by Osin [26] and Groves and Manning [17]. We now recall the notion of a Dehn filling of a group pair $(G, \mathcal{P})$:

Definition 2.3 (Dehn filling) Given a group pair $(G, \mathcal{P})$, where $\mathcal{P} = \{P_1, \dots, P_m\}$, and a choice of normal subgroups of peripheral groups $\mathcal{N} = \{N_i \trianglelefteq P_i\}$, the *Dehn filling* of $(G, \mathcal{P})$ with respect to $\mathcal{N}$ is the pair $(\overline{G}, \overline{\mathcal{P}})$, where $\overline{G} = G/K$ and $K = \langle\langle \bigcup N_i \rangle\rangle$ is the normal closure in G of the group generated by $\{\bigcup_i N_i\}$ and $\overline{\mathcal{P}}$ is the set of images of $\mathcal{P}$ under this quotient. The N_i are called the *filling kernels*. When we want to specify the filling kernels we write $G(N_1, \dots, N_m)$ for the quotient $\overline{G}$.

Definition 2.4 (peripherally finite) If each normal subgroup N_i has finite index in P_i, the filling is said to be *peripherally finite*.

Definition 2.5 (sufficiently long) We say that a property $\mathcal{X}$ holds for all sufficiently long Dehn fillings of $(G, \mathcal{P})$ if there is a finite subset $B \subset G \setminus \{1\}$ such that whenever $N_i \cap B = \varnothing$ for all i, the corresponding Dehn filling $G(N_1, \dots, N_n)$ has property $\mathcal{X}$.

The proof of the next theorem relies on the notion of a $\mathcal{Q}$–filling of a collection of subgroups $\mathcal{Q}$ of G. Recall from [19] that given a subgroup $Q < G$, the quotient $G(N_1, \dots, N_m)$ is a *Q–filling* if, for all $g \in G$ and $P_i \in \mathcal{P}$, $|Q \cap P_i^g| = \infty$ implies $N_i^g \subseteq Q$. If $\mathcal{Q} = \{Q_1, \dots, Q_l\}$ is a family of subgroups, then $G(N_1, \dots, N_m)$ is a *$\mathcal{Q}$–filling* if it is a Q–filling for every $Q \in \mathcal{Q}$.

Let $\mathcal{Q}$ be a collection of finite-index subgroups of elements of $\mathcal{P}$ such that any infinite cell stabilizer contains a conjugate of an element of $\mathcal{Q}$. The following is proved in [11]:

Theorem 2.6 [11, Proposition 4.1 and Corollary 4.2] *Let $(\Gamma, \mathcal{P})$ be a relatively hyperbolic pair such that the elements of $\mathcal{P}$ are residually finite. If $(\Gamma, \mathcal{P})$ admits a relatively geometric action on a* CAT(0) *cube complex X, then*

(i) *for sufficiently long $\mathcal{Q}$–fillings $\Gamma \to \overline{\Gamma} = \Gamma / K$, the quotient $\overline{X} = K \backslash X$ is a* CAT(0) *cube complex, and*

(ii) *any sufficiently long peripherally finite $\mathcal{Q}$–filling of Γ is hyperbolic and virtually special.*

The following result is implicit in [11]. For completeness, we provide a proof.

Lemma 2.7 *In the context of Theorem 2.6(i), the action of $\overline{\Gamma}$ on $\overline{X}$ is relatively geometric.*

Proof Since $\overline{\Gamma} \backslash \overline{X} = \Gamma \backslash X$ the action is cocompact. Let $\overline{\mathcal{P}}$ be the induced peripheral structure on $\overline{\Gamma}$ (the image of $\mathcal{P}$). The fact that elements of $\overline{\mathcal{P}}$ act elliptically on $\overline{X}$ follows from the fact that elements of $\mathcal{P}$ act elliptically on X. Because each cell stabilizer of $\Gamma \curvearrowright X$ is either finite or conjugate to a finite index of subgroup of some $P_i \in \mathcal{P}$, this implies that the cell stabilizers of $\overline{\Gamma} \curvearrowright \overline{X}$ are conjugate to finite-index subgroups of $P_i/(K \cap P_i)$ (the elements of $\overline{\mathcal{P}}$). Thus the action of $\overline{\Gamma}$ on $\overline{X}$ is relatively geometric. □

3 Relatively geometric actions: the Kähler case

In this section we apply Theorem 2.6 to prove Theorem 1.3. The main idea is to use Dehn filling to produce a minimal action of a finite-index subgroup of Γ on a tree with finite kernel. A deep result of Gromov and Schoen implies that any Kähler group admitting a minimal action on a tree with finite kernel must be virtually a hyperbolic surface group [16] (see also [27, Theorem 6.1] for a detailed discussion and explanation).

Proof of Theorem 1.3 Suppose that $(\Gamma, \mathcal{P})$ acts relatively geometrically on a CAT(0) cube complex. Since the elements of $\mathcal{P}$ are residually finite, there exists a finite index $\Gamma_0 \leq \Gamma$ such that Γ_0 is torsion free and $\Gamma_0 \backslash X$ is special, by [11, Theorem 1.4].

Cutting along an embedded two-sided hyperplane H in $\Gamma_0 \backslash X$ yields a splitting of Γ_0 according to the complex of groups version of van Kampen's theorem [6, III.$\mathcal{C}$.3.11(5), III.$\mathcal{C}$.3.12, page 552].[1] The edge group of such a splitting is a hyperplane stabilizer for the Γ_0–action on X, which is relatively quasiconvex by [12, Corollary 4.11]. We may choose a hyperplane whose stabilizer is infinite-index in Γ_0 (if the first such hyperplane does not satisfy this requirement, then replace X by the hyperplane, and continue until we find such a hyperplane, possibly replacing Γ_0 by a further finite-index subgroup

[1] One can also see this tree directly by considering the dual tree to the collection of hyperplanes of X which project to H. See [19, Remark 1.1] for more details.

along the way). The action of Γ_0 on the Bass–Serre tree T associated to this splitting has finite kernel, since any normal subgroup contained in an infinite-index relatively quasiconvex subgroup is finite (since an infinite normal subgroup has full limit set in the Bowditch boundary, but an infinite-index relatively quasiconvex subgroup does not). Let F denote the kernel of the action of Γ_0 on T.

By a result of Gromov and Schoen in [16] (see [27, Theorem 6.1] for details), the induced action of Γ_0 on T factors through a surjective homomorphism $\varphi\colon \Gamma_0 \to \Delta_0$, where $\Delta_0 \leq \mathrm{PSL}_2(\mathbb{R})$ is a cocompact lattice. The kernel of φ is contained in F, and hence finite, so Γ_0 is commensurable up to finite kernels with Δ_0, which is itself virtually a hyperbolic surface group. Since any group commensurable up to finite kernels with a hyperbolic surface group is virtually a hyperbolic surface group, this means that Γ_0, and hence Γ, is virtually a hyperbolic surface group, as desired. □

4 Relatively geometric actions: lattices in PU(n, 1)

Let Γ be a nonuniform lattice in $\mathrm{PU}(n,1)$. Then Γ acts properly discontinuously on complex hyperbolic space $\mathbb{H}^n_{\mathbb{C}}$ and the quotient, which we henceforth denote by $M = \Gamma \backslash \mathbb{H}^n_{\mathbb{C}}$, is a noncompact orbifold of finite volume with finitely many cusps. Each cusp corresponds to a conjugacy class of subgroups stabilizing a parabolic fixed point in $\partial_\infty \mathbb{H}^n_{\mathbb{C}}$. Farb [14] proved that Γ is hyperbolic relative to the collection of these cusp subgroups, which we denote by $\mathcal{P}$. In this section, we prove Theorem 1.1, namely that $(\Gamma, \mathcal{P})$ does not admit a relatively geometric action on a CAT(0) cube complex.

Throughout the course of the proof, we pass freely to finite-index subgroups by invoking Lemma 2.2. In order to streamline the exposition, we do not refer to Lemma 2.2 each time. It is well known that Γ has a torsion-free subgroup of finite index, so (passing to this finite-index subgroup if necessary) for the remainder of this section we assume that $\Gamma \leq \mathrm{PU}(n,1)$ is torsion free.

4.1 The structure of cusps

We now briefly review the geometric structure of cusps in M. For more details see [15]. Recall that up to scaling, each horosphere in $\mathbb{H}^n_{\mathbb{C}}$ is isometric to $\mathcal{H}_{2n-1}(\mathbb{R})$, the $(2n-1)$–dimensional real Heisenberg group, equipped with a left-invariant metric. The Heisenberg group is a central extension

$$1 \to \mathbb{R} \to \mathcal{H}_{2n-1}(\mathbb{R}) \to \mathbb{R}^{2n-2} \to 1 \tag{2}$$

with extension 2–cocycle equal to the standard symplectic form

$$\omega = 2\sum_{i=1}^{n-1} dx_i \wedge dy_i,$$

where $(x_1, y_1, \ldots, x_{n-1}, y_{n-1})$ are coordinates on $\mathbb{R}^{2n-2}$. The Lie algebra $\mathfrak{h}_{2n-1}$ is 2–step nilpotent with basis $\{X_1,\ Y_1, \ldots, X_n,\ Y_n,\ Z\}$, where

$$[X_i, Y_i] = Z,$$

and all other brackets vanish. Thus Z generates the center of $\mathfrak{h}_{2n-1}$ representing the kernel $\mathbb{R}$ in (2), while the remaining coordinates project to the generators of $\mathbb{R}^{2n-2}$. Choosing the identity matrix I_{2n-1} as the inner product on $\mathfrak{h}_{2n-1}$, we see that the isometry group of $\mathcal{H}_{2n-1}(\mathbb{R})$ is isomorphic to $\mathcal{H}_{2n-1}(\mathbb{R}) \rtimes U(n-1)$, where the $\mathcal{H}_{2n-1}(\mathbb{R})$ factor is the action of $\mathcal{H}_{2n-1}(\mathbb{R})$ on itself by left translation and the unitary group $U(n-1)$ is the stabilizer of the identity. Indeed, any isometry which fixes $1 \in \mathcal{H}_{2n-1}(\mathbb{R})$ must also be a Lie algebra isomorphism; it therefore preserves the center $\langle Z \rangle$ and induces an isometry of $\mathbb{R}^{2n-2} \cong \langle X_1, Y_1, \cdots, X_{n-1}, Y_{n-1} \rangle$ preserving ω. We conclude that such an isometry lies in $U(n-1) = O_{2n-2}(\mathbb{R}) \cap \mathrm{Sp}_{2n-2}(\mathbb{R})$.

Definition 4.1 Let $\pi : \mathcal{H}_{2n-1}(\mathbb{R}) \rtimes U(n-1) \to U(n-1)$ be the projection. For $g \in \mathcal{H}_{2n-1}(\mathbb{R}) \rtimes U(n-1)$, we call $\pi(g)$ the *rotational part* of g.

The center of $\mathcal{H}_{2n-1}(\mathbb{R})$ is normal in $\mathcal{H}_{2n-1}(\mathbb{R})$ and invariant under any isometry in $U(n-1)$. Therefore we have a short exact sequence

$$1 \to \mathbb{R} = Z(\mathcal{H}_{2n-1}(\mathbb{R})) \to \mathcal{H}_{2n-1}(\mathbb{R}) \rtimes U(n-1) \to \mathbb{R}^{2n-2} \rtimes U(n-1) \to 1. \tag{3}$$

Since Γ is torsion free, each cusp subgroup $P \leq \Gamma$ is isomorphic to a discrete torsion-free cocompact subgroup of $\mathrm{Isom}(\mathcal{H}_{2n-1}(\mathbb{R}))$. In particular, $P_0 = P \cap \mathcal{H}_{2n-1}(\mathbb{R})$ is a discrete cocompact subgroup and $P \cap Z(\mathcal{H}_{2n-1}(\mathbb{R})) \cong \mathbb{Z}$. By (3), P fits into a short exact sequence

$$1 \to \mathbb{Z} = P \cap Z(\mathcal{H}_{2n-1}(\mathbb{R})) \to P \to \Lambda \to 1, \tag{4}$$

where Λ is a discrete cocompact subgroup of $\mathbb{R}^{2n-2} \rtimes U(n-1)$. It follows that Λ has a finite-index subgroup Λ_0 isomorphic to $\mathbb{Z}^{2n-2}$, which is the image of P_0.

On the level of quotient spaces, the sequence in (4) has the following translation: the quotient space $\mathcal{O} = \Lambda \backslash \mathbb{R}^{2n-2}$ is a Euclidean orbifold finitely covered by the $(2n-2)$–dimensional torus $T = \Lambda_0 \backslash \mathbb{C}^{n-1}$, and $\Sigma = P \backslash \mathcal{H}_{2n-1}(\mathbb{R})$ is the total space of an S^1–bundle over $\mathcal{O}$, ie there is a fiber sequence

$$S^1 \hookrightarrow \Sigma \to \mathcal{O}. \tag{5}$$

Since $\mathcal{O}$ need not be smooth, this is not generally a locally trivial fibration. However, as P is torsion free, Σ is smooth. Passing to the torus cover, we obtain an actual fiber bundle

$$S^1 \hookrightarrow \widehat{\Sigma} \to T.$$

Choosing $\widehat{\Sigma}$ to be a regular cover with fundamental group P_0, we see that the finite group $F = P/P_0$ acts on $\widehat{\Sigma}$ preserving the fibration, and hence defines a finite group of isometries of T. Thus the stabilizer of a point in T acts freely on the S^1 fiber. Since the action of F on $\widehat{\Sigma}$ is free, it follows that point stabilizers in T must be cyclic of finite order, and act by rotations on the fiber. Since $F \leq U(n-1)$, any abelian subgroup is diagonalizable. Thus, locally each point in T has a neighborhood of the form $(S^1 \times \mathbb{D}^{n-1})/(\mathbb{Z}/m\mathbb{Z})$ where $\mathbb{D} \subset \mathbb{C}$ is the open unit disk, and $\mathbb{Z}/m\mathbb{Z}$ acts on S^1 by rotation by $2\pi/m$ and on the polydisk $\mathbb{D}^{n-1}$ by a diagonal unitary matrix $\Delta = \mathrm{diag}(e^{2\pi k_1/m}, \ldots, e^{2\pi k_{n-1}/m})$ where at least one k_i is coprime to m. See Figure 1 for a schematic. Since F acts by rotation on each fiber, Σ is the boundary of a disk bundle over $\mathcal{O}$, which we denote by $E_{\mathcal{O}}$.

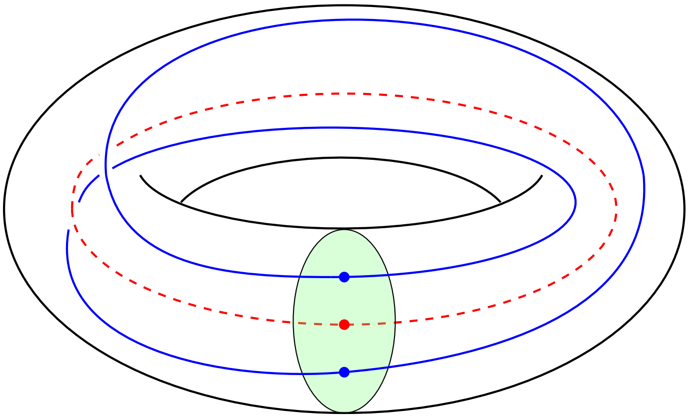


Figure 1: The local picture of the fibration in (5) near a singular point of $\mathcal{O}$. A nonsingular fiber (not dashed) winds $m = 2$ times around the singular fiber (dashed).

Recall that the center of $\mathcal{H}_{2n-1}(\mathbb{R})$ is quadratically distorted. It follows that the center of P is quadratically distorted as well. By [20, Theorem 1.5], there is no proper action of P on a CAT(0) cube complex. Therefore we have:

Proposition 4.2 *Let $\Gamma \leq \mathrm{PU}(n,1)$ be a nonuniform lattice, and suppose Γ acts on a* CAT(0) *cube complex X. The action of each cusp subgroup of Γ is not proper. In particular, Γ is not cubulated.*

4.2 The toroidal compactification of M

The toroidal compactification is a natural compactification of M which fills in the cusps with the Euclidean orbifolds described in Section 4.1. Let $\mathcal{O}_i$ be the Euclidean orbifold quotient of Σ_i, with corresponding disk bundle E_i. Thus we can identify $E_i \setminus \mathcal{O}_i$ with the cusp $\mathcal{C}_i$, then compactify M by adding $\bigsqcup_i \mathcal{O}_i$ at infinity. The result is a Kähler orbifold $\mathcal{T}(M)$ with boundary divisor $D = \bigsqcup_i \mathcal{O}_i$. The pair $(\mathcal{T}(M), D)$ is called the *toroidal compactification of M*. See [23; 2] for more details.

If the parabolic elements in Γ have trivial rotational part, then each $\mathcal{O}_i$ is a $(2n-2)$–dimensional torus, $\mathcal{T}(M)$ is a smooth Kähler manifold and D is a smooth divisor in $\mathcal{T}(M)$. Moreover, Hummel and Schroeder show that $\mathcal{T}(M)$ admits a nonpositively curved Riemannian metric [23]. In particular, $\mathcal{T}(M)$ is aspherical; if $\Delta = \pi_1(\mathcal{T}(M))$ then $\mathcal{T}(M)$ is a $K(\Delta, 1)$. It is clear from the construction that $\pi_1(\mathcal{T}(M))$ is the quotient of $\pi_1(M)$ by all the centers of the peripheral subgroups. The following lemma ensures that we can always find a finite cover of M whose toroidal compactification is smooth:

Lemma 4.3 *Let $\Gamma \leq \mathrm{PU}(n,1)$ be torsion free and let $M = \Gamma \setminus \mathbb{H}^n_{\mathbb{C}}$ be the quotient. There exists a finite cover $M' \to M$ such that the toroidal compactification of M' is smooth.*

Proof By the main theorem of [22, page 2453] , there exists a finite subset $F \subset \Gamma$ of parabolic isometries such that if $N \trianglelefteq \Gamma$ is a normal subgroup satisfying $F \cap N = \varnothing$, then any parabolic isometry in N

has no rotational part. Since Γ is residually finite and F is finite, we can find a finite-index normal subgroup $\Gamma' \trianglelefteq \Gamma$ such that $\Gamma' \cap F = \varnothing$. Therefore the finite cover $M' := \Gamma' \backslash \mathbb{H}^n_{\mathbb{C}}$ of M admits a toroidal compactification which is smooth. □

4.3 Proof of Theorem 1.1

Proof By Lemma 4.3 we may assume that $\Gamma \le \mathrm{SU}(n,1)$ is torsion free and the toroidal compactification $\mathcal{T}(M)$ is smooth. In particular, Γ and all of its peripheral subgroups are residually finite.

Suppose $(\Gamma, \mathcal{P})$ admits a relatively geometric action on a CAT(0) cube complex X. Given a finite-index subgroup $\Gamma_0 \le \Gamma$, let $\mathcal{P}_0$ be the induced peripheral structure on Γ_0, and let Δ_0 be $\pi_1(\mathcal{T}(M_0))$, where $M_0 = \Gamma_0 \backslash \mathbb{H}^n_{\mathbb{C}}$. Since the kernel of the quotient map $\Gamma_0 \to \Delta_0$ is normally generated by subgroups in $\mathcal{P}_0$, we get an induced peripheral structure $(\Delta_0, \mathcal{A}_0)$, where $\mathcal{A}_0$ is the collection of images of elements of $\mathcal{P}_0$. Our strategy is to show that there exists a finite-index subgroup $\Gamma_0 \le \Gamma$ such that the pair $(\Delta_0, \mathcal{A}_0)$ is relatively hyperbolic and admits a relatively geometric action on a CAT(0) cube complex. Since $\mathcal{T}(M_0)$ is smooth (since $\mathcal{T}(M)$ is), Δ_0 is Kähler. Thus, as $n \ge 2$, we will get a contradiction by Theorem 1.3.

Let $\mathcal{P} = \{P_1, \dots, P_k\}$ be the induced peripheral structure on Γ. Now let $Z(P_i)$ be the center of P_i. As in Section 2.1 let $\mathcal{Q}$ be a set of finite-index subgroups of elements of $\mathcal{P}$ such that any infinite stabilizer for the Γ–action on X contains a conjugate of an element of $\mathcal{Q}$. We apply Theorem 2.6(1) to a sufficiently long $\mathcal{Q}$–filling $\mathcal{Z} = \{Z_1, \dots, Z_k\}$ where $Z_i \le Z(P_i)$ is a finite-index subgroup. Residual finiteness of $Z(P_i)$ implies the existence of such sufficiently long $\mathcal{Q}$–fillings. We thus obtain a Dehn filling $\psi : \Gamma \to \Delta = \Gamma/K$ determined by the Z_i such that $Y = K \backslash X$ is a CAT(0) cube complex.

Let $(\Delta, \mathcal{A})$ be the induced peripheral structure on Δ. By Theorem 2.6, we know that $(\Delta, \mathcal{A})$ is relatively hyperbolic. Lemma 2.7 implies that the action of Δ on Y is relatively geometric.

Finally, we claim that there exists a finite-index subgroup $\Delta_0 \le \Delta$ that is torsion-free. Since the elements of $\mathcal{A}$ are virtually abelian, and hence residually finite, Δ is also residually finite by [11, Corollary 1.7]. Since Γ is torsion free, by [18, Theorem 4.1] so long as the filling $\Gamma \to \Delta$ is long enough (which we may assume without loss of generality), any element of finite order in Δ is conjugate into some element of $\mathcal{A}$. As there are finitely many elements of $\mathcal{A}$, each of which has only finitely many conjugacy classes of finite-order elements, we can find a finite-index subgroup $\Delta_0 \le \Delta$ which avoids each of these conjugacy classes, and hence is torsion free.

The induced peripheral structure $(\Delta_0, \mathcal{A}_0)$ is relatively hyperbolic and $\Delta_0 \curvearrowright Y$ is relatively geometric by Lemma 2.2. Let $\Gamma_0 = \psi^{-1}(\Delta_0)$ and let $\mathcal{P}_0 = \{P_{0,1}, \dots, P_{0,r}\}$ be the induced peripheral structure on Γ_0. Then $K \le \Gamma_0$, and since Δ_0 is torsion free, this implies $K \cap P_{0,i} = Z(P_{0,i})$ for each i. As $\mathcal{P}_0$ is the collection of cusp subgroups of $M_0 = \Gamma_0 \backslash \mathbb{H}^n_{\mathbb{C}}$, we conclude that $\Delta_0 = \pi_1(\mathcal{T}(M))$. Thus Δ_0 is Kähler and acts relatively geometrically on Y. By Theorem 1.3, we conclude that Δ_0 is virtually a hyperbolic surface group, which is impossible because Δ_0 contains a subgroup isomorphic to $\mathbb{Z}^{2n-2}$ and $n \ge 2$. □

4.4 Proof of Corollary 1.2

Proof A uniform lattice (resp. nonuniform lattice) Γ in a semisimple Lie group G is hyperbolic (resp. hyperbolic relative to its cusp subgroups $\mathcal{P}$) if and only if G has rank 1. One direction of this is proved by Farb [14, Theorem 4.11], and the other by Behrstock, Druțu and Mosher [3, Theorem 1.2]. Any rank-1 noncompact semisimple Lie group is commensurable with one of $\mathrm{SO}(n,1)$, $\mathrm{PU}(n,1)$, $\mathrm{Sp}(n,1)$ for $n \geq 2$, or the isometry group of the octonionic hyperbolic plane $\mathbb{H}^2_{\mathbb{O}}$. The last and $\mathrm{Sp}(n,1)$ have property (T), while $\mathrm{SO}(n,1)$ and $\mathrm{PU}(n,1)$ do not. Hence if Γ is commensurable with a lattice in $\mathrm{Sp}(n,1)$ or $\mathrm{Isom}(\mathbb{H}^2_{\mathbb{O}})$, then Γ has (T).

By a result of Niblo and Reeves [25, Theorem B], any action of a group with property (T) on a finite-dimensional CAT(0) cube complex has a global fixed point, so lattices in $\mathrm{Sp}(n,1)$ and $\mathrm{Isom}(\mathbb{H}^2_{\mathbb{O}})$ admit neither geometric nor relatively geometric actions on CAT(0) cube complexes. Hence if Γ is as in the statement of the result, it must be commensurable to a lattice in either $\mathrm{PU}(n,1)$ or $\mathrm{SO}(n,1)$. For $n \geq 2$, the uniform case of $\Gamma \leq \mathrm{PU}(n,1)$ is eliminated by the work of Delzant and Gromov [8, Corollary, page 52]. The corollary now follows from Theorem 1.1. □

References

[1] **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 MR Zbl

[2] **A Ash**, **D Mumford**, **M Rapoport**, **Y-S Tai**, *Smooth compactifications of locally symmetric varieties*, 2nd edition, Cambridge Univ. Press (2010) MR Zbl

[3] **J Behrstock**, **C Druțu**, **L Mosher**, *Thick metric spaces, relative hyperbolicity, and quasi-isometric rigidity*, Math. Ann. 344 (2009) 543–595 MR Zbl

[4] **N Bergeron**, **F Haglund**, **D T Wise**, *Hyperplane sections in arithmetic hyperbolic manifolds*, J. Lond. Math. Soc. 83 (2011) 431–448 MR Zbl

[5] **N Bergeron**, **D T Wise**, *A boundary criterion for cubulation*, Amer. J. Math. 134 (2012) 843–859 MR Zbl

[6] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[7] **D Cooper**, **D Futer**, *Ubiquitous quasi-Fuchsian surfaces in cusped hyperbolic 3–manifolds*, Geom. Topol. 23 (2019) 241–298 MR Zbl

[8] **T Delzant**, **M Gromov**, *Cuts in Kähler groups*, from "Infinite groups: geometric, combinatorial and dynamical aspects" (L Bartholdi, T Ceccherini-Silberstein, T Smirnova-Nagnibeda, A Zuk, editors), Progr. Math. 248, Birkhäuser, Basel (2005) 31–55 MR Zbl

[9] **T Delzant**, **P Py**, *Cubulable Kähler groups*, Geom. Topol. 23 (2019) 2125–2164 MR Zbl

[10] **E Einstein**, **D Groves**, *Relative cubulations and groups with a 2–sphere boundary*, Compos. Math. 156 (2020) 862–867 MR Zbl

[11] **E Einstein**, **D Groves**, *Relatively geometric actions on* CAT(0) *cube complexes*, J. Lond. Math. Soc. 105 (2022) 691–708 MR Zbl

[12] **E Einstein**, **D Groves**, **T Ng**, *Separation and relative quasiconvexity criteria for relatively geometric actions*, Groups Geom. Dyn. 18 (2024) 649–676 MR Zbl

[13] **E Einstein**, **T Ng**, *Relative cubulation of small cancellation free products*, preprint (2021) arXiv 2111.03008

[14] **B Farb**, *Relatively hyperbolic groups*, Geom. Funct. Anal. 8 (1998) 810–840 MR Zbl

[15] **W M Goldman**, *Complex hyperbolic geometry*, Clarendon, New York (1999) MR Zbl

[16] **M Gromov**, **R Schoen**, *Harmonic maps into singular spaces and p–adic superrigidity for lattices in groups of rank one*, Inst. Hautes Études Sci. Publ. Math. 76 (1992) 165–246 MR Zbl

[17] **D Groves**, **J F Manning**, *Dehn filling in relatively hyperbolic groups*, Israel J. Math. 168 (2008) 317–429 MR Zbl

[18] **D Groves**, **J F Manning**, *Dehn fillings and elementary splittings*, Trans. Amer. Math. Soc. 370 (2018) 3017–3051 MR Zbl

[19] **D Groves**, **J F Manning**, *Hyperbolic groups acting improperly*, Geom. Topol. 27 (2023) 3387–3460 MR Zbl

[20] **F Haglund**, *Isometries of* CAT(0) *cube complexes are semi-simple*, Ann. Math. Qué. 47 (2023) 249–261 MR Zbl

[21] **P Haïssinsky**, *Hyperbolic groups with planar boundaries*, Invent. Math. 201 (2015) 239–307 MR Zbl

[22] **C Hummel**, *Rank one lattices whose parabolic isometries have no rotational part*, Proc. Amer. Math. Soc. 126 (1998) 2453–2458 MR Zbl

[23] **C Hummel**, **V Schroeder**, *Cusp closing in rank one symmetric spaces*, Invent. Math. 123 (1996) 283–307 MR Zbl

[24] **V Markovic**, *Criterion for Cannon's conjecture*, Geom. Funct. Anal. 23 (2013) 1035–1061 MR Zbl

[25] **G Niblo**, **L Reeves**, *Groups acting on* CAT(0) *cube complexes*, Geom. Topol. 1 (1997) 1–7 MR Zbl

[26] **D V Osin**, *Peripheral fillings of relatively hyperbolic groups*, Invent. Math. 167 (2007) 295–326 MR Zbl

[27] **P Py**, *Lectures on Kähler groups*, preprint (2024) `https://www-fourier.ujf-grenoble.fr/~py/Documents/Livre-Kahler/lectures.pdf`

[28] **D Toledo**, *Examples of fundamental groups of compact Kähler manifolds*, Bull. London Math. Soc. 22 (1990) 339–343 MR Zbl

[29] **D T Wise**, *The structure of groups with a quasiconvex hierarchy*, Annals of Mathematics Studies 209, Princeton Univ. Press (2021) MR Zbl

Department of Mathematics, Tufts University
Medford, MA, United States

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago
Chicago, IL, United States

Department of Mathematics, University of California at Riverside
Riverside, CA, United States

corey.bregman@tufts.edu, dgroves@uic.edu, kejia.zhu@ucr.edu

https://sites.google.com/view/cbregman, http://homepages.math.uic.edu/~groves/, https://sites.google.com/view/kejiazhu

Received: 21 August 2023 Revised: 5 January 2024

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.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 24 Issue 7 (pages 3571–4137) 2024
