Geometry & Topology

Volume 27 (2023)

Issue 9 (pages 3387–3831)

msp

GEOMETRY & TOPOLOGY
msp.org/gt

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

The subscription price for 2023 is US $740/year for the electronic version, and $1030/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. Geometry & Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

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

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

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

Geometry & Topology **27**:9 (2023) 3387–3460
DOI: 10.2140/gt.2023.27.3387
Published: 5 December 2023

Hyperbolic groups acting improperly

DANIEL GROVES
JASON FOX MANNING

We study hyperbolic groups acting on CAT(0) cube complexes. The first main result is a structural result about the Sageev construction, in which we relate quasiconvexity of hyperplane stabilizers with quasiconvexity of cell stabilizers. The second main result generalizes both Agol's Theorem on cubulated hyperbolic groups and Wise's Quasiconvex Hierarchy Theorem.

20F65, 57M05

1 Introduction

In recent years, CAT(0) cube complexes have played a central role in many spectacular advances, most notably in Agol's proof of the Virtual Haken and Virtual Fibering Theorems in [1]. The main result of [1] is that a hyperbolic group which acts properly and cocompactly on a CAT(0) cube complex is virtually special. A key ingredient in Agol's proof was the work of Wise from [36], particularly Wise's Quasiconvex Hierarchy Theorem [36, Theorem 13.3]. One of the two main results of the current paper is Theorem D, which provides a simultaneous generalization of Agol's Theorem and Wise's Theorem. So far this generalization has been applied by Duong [12] and Einstein and Groves [13]. At the end of the introduction in Section 1.2, we explain how Theorem D (together with Theorem A) simplifies the proof of the Virtual Haken and Virtual Fibering Theorems for hyperbolic 3–manifolds, requiring only a single immersed quasi-Fuchsian surface instead of a ubiquitous family.

Cube complexes in group theory arise via the construction of Sageev [32] which takes as input a group G and a collection of *codimension-one* subgroups of G and produces a CAT(0) cube complex X, equipped with an isometric G–action on X with no global fixed point. The other main result of the current paper is Theorem A, which establishes some fundamental properties about the Sageev construction.

Sageev's construction works in great generality. However, in order to get more information from the G–action on X, it is useful to add geometric hypotheses. For example, if G is a hyperbolic group and the codimension-one subgroups are quasiconvex, Sageev proved that the associated cube complex is G–cocompact [33, Theorem 3.1]. Achieving a *proper* action is harder (see Bergeron and Wise [5] and Hruska and Wise [24] for conditions which ensure properness).

Even an improper action $G \curvearrowright X$ gives a description of G as the fundamental group of a *complex of groups* in the sense of Bridson and Haefliger (see [8, III.$\mathcal{C}$] or Section 2 below). In this description, the underlying space is the quotient $G\backslash X$ and the local groups can be identified with cell stabilizers for the action.

Our first main result links the geometry of the hyperplane stabilizers with that of the cell stabilizers.

Theorem A *Let G be hyperbolic. The following conditions on a cocompact G–action on a CAT(0) cube complex are equivalent*:

1. *All hyperplane stabilizers are quasiconvex.*
2. *All vertex stabilizers are quasiconvex.*
3. *All cell stabilizers are quasiconvex.*

Intersections of quasiconvex subgroups are quasiconvex, and cell stabilizers are intersections of vertex stabilizers. Therefore, the equivalence of (2) and (3) is trivial. We prove the equivalence of (1) and (2).

We remark that we actually prove the direction (1) $\Longrightarrow$ (2) in the more general setting of arbitrary finitely generated groups where we assume the relevant subgroups are *strongly* *quasiconvex* in the sense of Tran [35]. Note that in this more general setting, (2) and (3) are still equivalent. See Section 3 for more details. In Section 3.6 we explain how Theorem A implies the following result.

Corollary B *Suppose that G is a hyperbolic group acting cocompactly on a CAT(0) cube complex X with quasiconvex hyperplane stabilizers. Then*

(1) *X is δ–hyperbolic for some δ*;

(2) *there exists a $k \geq 0$ such that the fixed-point set of any infinite subgroup of G intersects at most k distinct cells*; *and*

(3) *the action of G on X is acylindrical* (in the sense of Bowditch [6, page 284]).

Anthony Genevois explained to us how conclusion (2) implies acylindricity for actions on hyperbolic CAT(0) cube complexes (see Section 3.6). The condition in (2) is not implied by acylindricity since X is not assumed to be locally compact.

Without the conclusion of δ–hyperbolicity, a more general version of Corollary B holds just as for Theorem A. See Remark 3.31 for more details.

In Sageev's construction, the stabilizers in G of hyperplanes in the resulting cube complex are commensurable with the chosen codimension-one subgroups of G. Therefore, we have the following result.

Corollary C *Let G be a hyperbolic group and let $\mathcal{H} = \{H_1, \dots, H_k\}$ be a collection of quasiconvex codimension-one subgroups. Let X be a CAT(0) cube complex obtained by applying the Sageev construction to $\mathcal{H}$.*

(1) *The stabilizers of cells in X are quasiconvex in G. In particular, they are finitely presented.*

(2) *X is δ–hyperbolic for some δ.*

(3) *There exists a $k \geq 0$ such that the fixed-point set of any infinite subgroup of G intersects at most k distinct cells.*

(4) *The action of G on X is acylindrical.*

As far as we are aware, even the corollary of item (1) that the cell stabilizers are finitely generated in the above result is new. We remark that the fact that cell stabilizers are finitely presented implies that the description of G as the fundamental group of the complex of groups associated to $G\backslash X$ is a finite description.

Some of the most dramatic uses of CAT(0) cube complexes have come from Haglund and Wise's theory of *special* cube complexes [20]. A cube complex is *special* if it admits a locally isometric immersion into the Salvetti complex of a right-angled Artin group. A group G is *virtually special* if there is a finite-index subgroup $G_0 \leq G$ and a CAT(0) cube complex X such that G_0 acts freely and cubically on X and $G_0 \backslash X$ is a compact special cube complex. (For some authors the quotient is allowed to be noncompact but have finitely many hyperplanes.)

As shown in [20], virtually special hyperbolic groups have many remarkable properties, such as being residually finite, linear over $\mathbb{Z}$ and possessing very strong subgroup separability properties.

Agol [1] proved that if a hyperbolic group G acts properly and cocompactly on a CAT(0) cube complex then G is virtually special. It is this result that implies the virtual Haken conjecture, as well as the virtual fibering conjecture (in the compact case), and many other results.

One of the key ingredients of the proof of Agol's Theorem, and another of the most important theorems in the area is Wise's Quasiconvex Hierarchy Theorem [36, Theorem 13.3] — see also [3, Theorem 10.2] — which states that if a hyperbolic group G can be expressed as $A*_C$ (resp. $A *_C B$), where C is quasiconvex in G and A is (resp. A and B are) virtually special then G is virtually special. This theorem can be rephrased as saying that if a hyperbolic group acts cocompactly on a *one-dimensional CAT*(0) *cube complex* (otherwise known as a "tree") with virtually special and quasiconvex cell stabilizers, then G is virtually special.

Our second main result is a common generalization of Agol's Theorem and Wise's Quasiconvex Hierarchy Theorem.

Theorem D *Suppose that G is a hyperbolic group acting cocompactly on a CAT*(0) *cube complex X with quasiconvex and virtually special cell stabilizers. Then G is virtually special.*

By Corollary C, Theorem D has the following immediate consequence.

Corollary E *Suppose that G is a hyperbolic group and that $\mathcal{H} = \{H_1, \dots, H_k\}$ is a collection of quasiconvex codimension-one subgroups. If the vertex stabilizers of the G–action on a cube complex obtained by applying the Sageev construction to $\mathcal{H}$ are virtually special, then G is virtually special.*

Since finding proper actions of hyperbolic groups on CAT(0) cube complexes is much harder than finding cocompact actions, Theorem D is expected to be a powerful new tool for proving that hyperbolic groups are virtually special. As mentioned above, Theorem D was used in [12] to show that random groups in the square model at density $< \frac{1}{3}$ are virtually special. Theorem D (as well as Corollary 6.6 below) are also applied in [13] to provide a characterization of relatively hyperbolic groups with 2–sphere boundary in terms of actions on cube complexes.

Theorem A is one of the key ingredients of the proof of Theorem D. We now explain how Theorem D is a consequence of the above-mentioned results of Agol and Wise, along with Theorem A and the following result (proved in Section 6).

Theorem F *Suppose that the hyperbolic group G acts cocompactly on a CAT(0) cube complex X and that cell stabilizers are virtually special and quasiconvex. There exists a quotient $\overline{G} = G/K$ such that*

(1) *the quotient $K\backslash X$ is a CAT(0) cube complex*;

(2) *the group $\overline{G}$ is hyperbolic*; *and*

(3) *the action of $\overline{G}$ on $K\backslash X$ is proper* (*and cocompact*).

Proof of Theorem D Consider the hyperbolic group G, acting on a CAT(0) cube complex X as in the statement of Theorem D. By Theorem F there exists a hyperbolic quotient $\overline{G} = G/K$ of G such that $K\backslash X$ is a CAT(0) cube complex, and the $\overline{G}$–action on $K\backslash X$ is proper and cocompact. Let $Z = K\backslash X$.

By Agol's Theorem [1, Theorem 1.1], there is a finite-index subgroup $\overline{G}_0$ of $\overline{G}$ such that $\overline{G}_0\backslash Z$ is special. Let G_0 be the preimage in G of $\overline{G}_0$. Clearly, the underlying space of $G_0\backslash X$ is the same as that of $\overline{G}_0\backslash Z$, and in particular all of the hyperplanes are two-sided and embedded.

We cut successively along these hyperplanes, applying the complex of groups version of the Seifert–Van Kampen Theorem [8, Example III.$\mathcal{C}$.3.11(5) and Exercise III.$\mathcal{C}$.3.12]. In this way, we obtain a hierarchy of G_0 with the following properties:

(1) The edge groups are quasiconvex (since they are stabilizers of hyperplanes, which are quasiconvex by Theorem A).

(2) The terminal groups are virtually special (since they are finite-index subgroups of the vertex stabilizers in G).

Therefore, G_0 admits a quasiconvex hierarchy terminating in virtually special groups, so G_0 is virtually special by Wise's Quasiconvex Hierarchy Theorem [36, Theorem 13.3] (see [3, Theorem 10.3] for a somewhat different account). Since G_0 is finite index in G, the group G is virtually special, as required. □

Remark 1.1 We thank one of the referees for pointing out that one can replace the use of the complex of groups Seifert–Van Kampen Theorem in the previous proof with the following argument: Once all of the hyperplanes in $G_0\backslash X$ are two-sided and embedded, lift to X and consider the trees dual to the hyperplanes. This gives a collection of G_0–trees. Order them in some way. If V stabilizes a vertex in the first tree, consider the V–action on the second tree. The stabilizers in V for this second action act on the third tree, and so on. In this way, a quasiconvex hierarchy for G_0 is obtained, and the proof finishes as above.

We now briefly outline the contents of this paper. In Section 2 we recall those parts of the theory of complexes of groups from [8] which we need. In Section 3, we prove Theorem A and Corollary B. The proof of Theorem A depends on a quasiconvexity criterion (Theorem A.3) which is proved separately in the appendix. We separate out Theorem A.3 and its proof both because it may be of independent interest and because the methods, unlike in the rest of the paper, are pure δ–hyperbolic geometry. In Section 4 we investigate conditions on a group G acting on a CAT(0) cube complex X and a normal subgroup $K \trianglelefteq G$ which imply that the quotient $K\backslash X$ is a CAT(0) cube complex. In Section 5 we translate these conditions into group-theoretic statements. In Section 6 we prove various results about Dehn filling (in particular, Theorem 6.5 and Corollary 6.6, which may be of independent interest) to see that the conditions from Section 5 are satisfied for certain subgroups K which arise as kernels of long Dehn filling maps. We use this to deduce Theorem F.

1.1 Notation and conventions

The notation $A \dot{<} B$ indicates that A is a finite-index subgroup of B; similarly, $A \dot{\lhd} B$ indicates A is a finite-index normal subgroup. We write conjugation as $a^x = xax^{-1}$, or sometimes as $\mathrm{Ad}(x)(a)$. For p an element of a G–set, we denote the G–orbit by $[\![p]\!]$.

1.2 Virtual Haken and fibering with a single surface

Let M be a closed hyperbolic 3–manifold, and let $\Gamma = \pi_1(M)$. Agol's proof that M is virtually Haken and virtually fibered in [1] proceeds via proving that Γ is virtually special. This relies on Bergeron and Wise's Theorem that Γ acts properly and cocompactly on a CAT(0) cube complex [5]. In turn, Bergeron and Wise rely on work of Kahn and Markovic [26], which provides a "ubiquitous"[1] family of immersed

[1] This terminology is from Cooper and Futer [10].

quasi-Fuchsian surfaces in M. That there is such an abundance of surfaces follows from the proofs in [26], but is not explicitly stated there.

Here we point out that the results in this paper show that the fact that Γ is virtually special follows from the existence of a *single* immersed quasi-Fuchsian surface in M. It is explained in [36] how virtual Haken and virtual fibering follow.

Theorem 1.2 *Suppose that M is a closed hyperbolic 3–manifold and that M contains an immersed quasi-Fuchsian surface. Then $\pi_1(M)$ is virtually special.*

Proof If M is nonorientable, we replace it by its orientation double cover. Let $\Gamma \cong \pi_1 M$ be a lattice in $\mathrm{Isom}^+(\mathbb{H}^3)$, so that $M \cong \Gamma\backslash\mathbb{H}^3$. We note that in this setting a subgroup $W < \Gamma$ is geometrically finite as a Kleinian group if and only if it is quasiconvex in Γ; see [27, Theorem 2] or [34, Theorem 1.1 and Proposition 1.3].

Let $H < \Gamma$ be the subgroup corresponding to the immersed quasi-Fuchsian surface. Since H is quasiconvex and codimension-one in Γ, we can apply the Sageev construction to obtain a cocompact action of Γ on a CAT(0) cube complex X with no global fixed point, and with hyperplane stabilizers conjugate to H. Theorem A implies that the vertex stabilizers for this action are quasiconvex in Γ. To apply Theorem D, we will show that the vertex stabilizers admit quasiconvex hierarchies and hence are virtually special.

Let $V < \Gamma$ be a vertex stabilizer. Since V is quasiconvex in Γ it is a geometrically finite subgroup of $\mathrm{Isom}^+(\mathbb{H}^3)$. As V has infinite index in Γ, it acts with infinite covolume on $\mathbb{H}^3$. An argument of Thurston shows that every finitely generated subgroup of V is also geometrically finite [30, Proposition 7.1].

Since Γ contains no parabolics, neither does V. Thus a small closed neighborhood N of the convex core of $H\backslash\mathbb{H}^3$ is a compact 3–manifold with nonempty boundary, and hence is irreducible in the sense that every embedded 2–sphere bounds a ball [29, Propositions 2.36 and 3.1]. A compact irreducible 3–manifold with nonempty boundary is Haken; see [22, Chapter 6; 25, Chapter III]. In particular it has a Haken hierarchy [25, IV.12]. This topological hierarchy of N gives a group-theoretic hierarchy of V. The edge groups in the hierarchy are finitely generated. The previously mentioned argument of Thurston then implies that the edge groups are geometrically finite and hence quasiconvex in Γ. In particular, this is a quasiconvex hierarchy, and we may apply Wise's Quasiconvex Hierarchy Theorem to conclude that V is virtually special.

Since all vertex stabilizers of the action $\Gamma \curvearrowright X$ are quasiconvex and virtually special, we may apply Theorem D to conclude that Γ is itself virtually special. □

Acknowledgements

We thank Richard Webb for suggesting that the direction (1) $\Longrightarrow$ (2) of Theorem A might hold in a more general setting than that of hyperbolic groups.

Thanks to Anthony Genevois for pointing out that his work allows us to deduce acylindricity (Corollary B(3)) from Corollary B(2).

We also thank Alessandro Sisto for pointing out the applicability of a "Greendlinger Lemma" type result in our joint work; see [19, Lemma 2.26]. This lemma inspired Theorem 6.7 in the current paper, which simplified the proof of Theorem 6.5.

Finally, we thank the referees whose careful readings and comments have led to substantial improvements in the exposition of this paper.

Groves was partially supported by the Simons Foundation, grant 342049, and the National Science Foundation, grants DMS-1507067 and DMS-1904913. Manning was also partially supported by the Simons Foundation, grant 524176, and the National Science Foundation, grant DMS-1462263.

2 Complexes of groups

In this section we give a brief account of those parts of the theory of complexes of groups which we need. Much more detail can be found in Bridson and Haefliger [8, III.$\mathcal{C}$].

2.1 Paths and homotopies in a category

The definitions here are mainly taken from [8, III.$\mathcal{C}$.A], though our notation is slightly different.

Let $\mathcal{C}$ be a category. For an arrow a of $\mathcal{C}$, we denote its source by $i(a)$ and its target by $t(a)$. An *oriented edge* of $\mathcal{C}$ is a symbol a^+ or a^-, where a is an arrow of $\mathcal{C}$. The source and target of an oriented edge are defined by

$$i(a^-) = i(a), \quad t(a^-) = t(a) \qquad \text{and} \qquad i(a^+) = t(a), \quad t(a^+) = i(a).$$

(We caution readers that this may be the opposite of what they expect. The signs are chosen so that concatenation of $+$ edges is homotopic to composition of the corresponding arrows; see Definition 2.1.)

We now define $\mathcal{C}$–paths. A *$\mathcal{C}$–path p of length* 0 is an object v of $\mathcal{C}$ with $i(p) = t(p) = v$. For $j > 0$, a *$\mathcal{C}$–path of length j* is a list $p = e_1 \cdot e_2 \cdots e_j$ where for each i we have $t(e_i) = i(e_{i+1})$. We have $i(p) = i(e_1)$ and $t(p) = t(e_j)$.

If p is a $\mathcal{C}$–path of length j, q is a $\mathcal{C}$–path of length k, and $t(p) = i(q)$, then the concatenation $p \cdot q$ is a $\mathcal{C}$–path of length $j + k$ with $i(p \cdot q) = i(p)$ and $t(p \cdot q) = t(q)$.[2]

The category $\mathcal{C}$ is *connected* if for any two objects v_0, v_1 in $\mathcal{C}$ there is a $\mathcal{C}$–path p with $i(p) = v_0$ and $t(p) = v_1$.

If p is a $\mathcal{C}$–path, then p is *nonbacktracking* if it contains no subpath of the form $a^+ \cdot a^-$ or $a^- \cdot a^+$.

Definition 2.1 *Homotopies* of $\mathcal{C}$–paths (see [8, III.$\mathcal{C}$.A.11]) are generated by the following *elementary homotopies*, valid whenever both sides are paths:

(1) $p \cdot a^+ \cdot a^- \cdot q \simeq p \cdot q$ or $p \cdot a^- \cdot a^+ \cdot q \simeq p \cdot q$;

(2) $p \cdot a^+ \cdot b^+ \cdot q \simeq p \cdot (ab)^+ \cdot q$ or $p \cdot b^- \cdot a^- \cdot q \simeq p \cdot (ab)^- \cdot q$ (here and below we write ab for the composition $a \circ b$); and

(3) $p \cdot \mathbb{1}_v^{\pm} \cdot q \simeq p \cdot q$ (where $\mathbb{1}_v$ is an identity arrow).

Any category has a *nerve* which is a simplicial complex whose 0–cells are the objects of $\mathcal{C}$, with 1–cells corresponding to arrows, 2–cells to composable pairs of arrows, and so on. The $\mathcal{C}$–paths we have just defined give edge-paths and the elementary homotopies correspond to simplicial homotopies in this complex.

2.2 Small categories without loops (scwols)

By a *scwol* (small category without loops) we mean a small category in which for every object v, the set of arrows from v to itself contains only the unit $\mathbb{1}_v$, and this unit $\mathbb{1}_v$ cannot be written as a composition of other arrows. An arrow is *trivial* if it is equal to $\mathbb{1}_v$ for some object v. We sometimes conflate v and $\mathbb{1}_v$. A (*nondegenerate*) *morphism* of scwols $f : \mathcal{A} \to \mathcal{B}$ is a functor which induces, for each object v of $\mathcal{A}$, a bijection between the arrows $\{a \mid i(a) = v\}$ and the arrows $\{a \mid i(a) = f(v)\}$.

[2] For purposes of concatenation, a $\mathcal{C}$–path of length 0 is regarded as an empty list.

Definition 2.2 (simple scwol, scwolification) A scwol in which there is at most one arrow with a given source and target will be called a *simple scwol.* Any small category $\mathcal{C}$ has a canonical quotient category $\mathrm{scwol}(\mathcal{C})$, which is a simple scwol. The objects of $\mathrm{scwol}(\mathcal{C})$ are equivalence classes of objects of $\mathcal{C}$, where $v \sim w$ if there are arrows a and b such that $i(a) = t(b) = v$ and $i(b) = t(a) = w$. Similarly, the arrows of $\mathrm{scwol}(\mathcal{C})$ are equivalence classes of arrows of $\mathcal{C}$, where $a \sim b$ whenever $i(a) \sim i(b)$ and $t(a) \sim t(b)$. We may refer to $\mathrm{scwol}(\mathcal{C})$ as the *scwolification of* $\mathcal{C}$. The map $\mathcal{C} \to \mathrm{scwol}(\mathcal{C})$ taking each object and arrow to its equivalence class will be called the *scwolification functor*.

Remark 2.3 The procedure of scwolification is natural. In particular, a group action on a small category $\mathcal{C}$ descends to an action on $\mathrm{scwol}(\mathcal{C})$.

A key example of a scwol is the (opposite) poset of cells of a simplicial or cubical complex, with arrows from each cell to all its faces. If two faces of some cell are glued together one obtains a nonsimple scwol.

Definition 2.4 [8, III.$\mathcal{C}$.1.3] A scwol $\mathcal{A}$ has a (*geometric*) *realization* which is a simplicial complex whose 0–cells are the objects of $\mathcal{A}$, with 1–cells corresponding to nontrivial arrows, 2–cells to composable pairs of such arrows, and so on.

The realization of $\mathcal{A}$ is naturally a subcomplex of the nerve of $\mathcal{A}$. Although the nerve is necessarily infinite-dimensional, the realization of a scwol has dimension equal to the length of the longest chain of nontrivial composable arrows. Every scwol which appears in the current paper has finite-dimensional realization.

Definition 2.5 If A is the realization of a scwol $\mathcal{A}$, then there is a canonical correspondence between combinatorial paths in the 1–skeleton of A and $\mathcal{A}$–paths without trivial arrows. If p is a combinatorial path in $A^{(1)}$, and q the corresponding $\mathcal{A}$–path, we say that p is the *realization* of q, and q is the *idealization* of p.

2.3 Complexes of groups

Definition 2.6 [8, III.$\mathcal{C}$.2.1] Let $\mathcal{A}$ be a scwol. A *complex of groups* $H(\mathcal{A})$ consists of

(1) for each object σ of $\mathcal{A}$, a *local group* (also called a *cell group*) H_σ;

(2) for each arrow a of $\mathcal{A}$, an injective group homomorphism $\psi_a \colon H_{i(a)} \to H_{t(a)}$ (if a is a trivial arrow, we require ψ_a to be the identity map);

(3) for each pair of composable arrows a and b with composition ab, a *twisting element* $z(a,b) \in H_{t(a)}$ (if either a or b is trivial, $z(a,b) = 1$).[3]

These data satisfy the following conditions (continuing to write ab for $a \circ b$) whenever all written compositions of arrows are defined:

(1) **Compatibility** $\operatorname{Ad}(z(a,b))\psi_{ab} = \psi_a\psi_b$.[4]

(2) **Cocycle** $\psi_a(z(b,c))z(a,bc) = z(a,b)z(ab,c)$.

The cocycle condition above applies to any arrangement of arrows of the form

$$\begin{tikzcd}
& \cdot \arrow[rrd, "ab"] \arrow[ddr, "b"] & & \\
\cdot \arrow[ur, "c"] \arrow[drr, "bc"'] & & & \cdot \\
& & \cdot \arrow[ur, "a"'] &
\end{tikzcd}$$

Definition 2.7 (the complex of groups coming from an action) Suppose G acts on a scwol $\mathcal{X}$ in such a way that any $g \in G$ fixing an object fixes every arrow from that object. Suppose further that $\mathcal{Y} = G\backslash\mathcal{X}$ is a scwol. We obtain a complex of groups $G(\mathcal{Y})$ once we have [8, III.$\mathcal{C}$.2.9]:

(1) For each object v of $\mathcal{Y}$, a choice of a lift $\tilde{v}$ to $\mathcal{X}$; this lift also determines lifts $\tilde{a}$ of all arrows a with $i(a) = v$.

(2) For each arrow a, a choice of $h_a \in G$ such that $t(h_a(\tilde{a})) = \widetilde{t(a)}$. (When a is a trivial arrow, we always take $h_a = 1$.)

Given these choices, one defines

(1) G_v as the stabilizer of $\tilde{v}$,

(2) $\psi_a = \operatorname{Ad}(h_a)|_{G_{i(a)}}$,

(3) $z(a,b) = h_a h_b h_{ab}^{-1}$.

The complex of groups $G(\mathcal{Y})$ can be used to recover the group G. There are two different ways of doing this. The first is explained in [8, III.$\mathcal{C}$.3.7], and involves $G(\mathcal{Y})$–*paths*. The second way is from [8, III.$\mathcal{C}$.A], and is the way that we proceed. The advantage to this second way, which uses categories and coverings of categories, is that lifting paths to covers is a canonical procedure (as with usual covering theory).

[3] In [8] the notation $g_{a,b}$ is used instead of $z(a,b)$.

[4] Recall $\operatorname{Ad}(z)(x) = zxz^{-1}$.

2.4 Fundamental groups and coverings of categories

In Definition 2.1 we defined homotopy of $\mathcal{C}$–paths, where $\mathcal{C}$ is a category.

Definition 2.8 Given a category $\mathcal{C}$ and an object v_0 of $\mathcal{C}$, the *fundamental group of $\mathcal{C}$ based at* v_0, denoted by $\pi_1(\mathcal{C}, v_0)$, is the set of homotopy classes of $\mathcal{C}$–loops based at v_0, with operation induced by concatenation of $\mathcal{C}$–paths.

Definition 2.9 [8, III.$\mathcal{C}$.A.15] Let $\mathcal{C}$ be a connected category. A functor $f\colon \mathcal{C}' \to \mathcal{C}$ is a *covering* if for each object σ' of $\mathcal{C}'$, the restriction of f to the collection of arrows that have σ' as their initial (resp. terminal) object is a bijection onto the set of arrows which have $f(\sigma')$ as their initial (resp. terminal) object.

The *universal cover* $\widetilde{\mathcal{C}}$ of a connected category $\mathcal{C}$ is described in [8, III.$\mathcal{C}$.A.19]: Fix a base vertex v_0 of $\mathcal{C}$, and define $\mathrm{Obj}(\widetilde{\mathcal{C}})$ to be the set of homotopy classes of $\mathcal{C}$–paths starting at v_0. If $[c]$ is a homotopy class of path, and α is an arrow from $t(c)$, then there is an arrow $\tilde{\alpha}$ of $\widetilde{\mathcal{C}}$ from $[c]$ to $[c \cdot \alpha^-]$. The projection $\pi\colon \widetilde{\mathcal{C}} \to \mathcal{C}$ sets $\pi([p]) = t(p)$ and if $\tilde{\alpha}$ is the arrow described above then $\pi(\tilde{\alpha}) = \alpha$. The fundamental group $\pi_1(\mathcal{C}, v_0)$ acts on $\widetilde{\mathcal{C}}$ by preconcatenation.

The theory of coverings of categories is entirely analogous to ordinary covering theory. In fact it is a special case, as the coverings of a connected category $\mathcal{C}$ correspond bijectively to the covering spaces of its nerve.

We record the following observation.

Lemma 2.10 *Let $\phi\colon \widetilde{\mathcal{C}} \to \mathcal{C}$ be a covering of categories, and suppose $\phi(\tilde{v}) = v$ for objects v of $\mathcal{C}$ and $\tilde{v}$ of $\widetilde{\mathcal{C}}$. Any $\mathcal{C}$–path p with $i(p) = v$ has a unique lift to a $\widetilde{\mathcal{C}}$–path $\tilde{p}$ with $i(\tilde{p}) = \tilde{v}$. Moreover any elementary homotopy from p to a path p' gives a unique elementary homotopy of $\tilde{p}$ to a lift $\tilde{p}'$ of p' with the same endpoints as $\tilde{p}$.*

2.5 The category associated to a complex of groups

Any complex of groups $G(\mathcal{Y})$ has an associated category $CG(\mathcal{Y})$.

Definition 2.11 [8, III.$\mathcal{C}$.2.8] The objects of $CG(\mathcal{Y})$ are the objects of the scwol $\mathcal{Y}$. Arrows of $CG(\mathcal{Y})$ are pairs (g, a) such that a is an arrow of $\mathcal{Y}$ and $g \in G_{t(a)}$. Composition is defined by $(g, a) \circ (h, b) = (g\psi_a(h)z(a, b), ab)$.

Recall that if a is a trivial arrow then ψ_a is the identity homomorphism and $z(a, x)$ and $z(x, a)$ are always trivial.

Remark 2.12 The map $CG(\mathcal{Y}) \to \mathcal{Y}$ given by $(g, a) \to a$ is the scwolification functor (see Definition 2.2), and is a bijection on objects. This functor has an obvious section $a \mapsto (1, a)$. If there are nontrivial twisting elements, this is not a functor, but it does allow $\mathcal{Y}$–paths to be "unscwolified" to $CG(\mathcal{Y})$–paths. In Definition 2.22, we explain how to go back and forth between paths in covers of $CG(\mathcal{Y})$ and their associated scwols.

Theorem 2.13 [8, III.$\mathcal{C}$.3.15 and text before III.$\mathcal{C}$.A.13] *Suppose that the group G acts on the simply connected complex X, giving rise to an action of G on the scwol $\mathcal{X}$ as in Definition 2.7, and that v_0 is an object in $\mathcal{Y} = G\backslash\mathcal{X}$. Let $CG(\mathcal{Y})$ be the category associated to $G(\mathcal{Y})$. Then there is an isomorphism from $\pi_1(CG(\mathcal{Y}), v_0)$ to G taking any loop of the form $(g_1, a_1)^{\epsilon_1} \cdots (g_n, a_1)^{\epsilon_1}$ to the product $(g_1 h_{a_1})^{\epsilon_1} \cdots (g_n h_{a_1})^{\epsilon_n}$.*

The exponents ϵ_i in the statement are taken from the set $\{+, -\}$. We use the mild abuse of notation that if g is a group element then $g^+ = g$ and $g^- = g^{-1}$.

Definition 2.14 Let a be a nontrivial arrow of $\mathcal{Y}$. The arrow $(1, a)$ of $CG(\mathcal{Y})$ is called a *scwol arrow*. Let $g \in G_v$ where v is a vertex of the scwol $\mathcal{Y}$. The arrow $(g, \mathbb{1}_v)$ is called a *group arrow at v*, or just a *group arrow* if v is unimportant.

In later sections we abuse notation and refer to the edge $(g, \mathbb{1}_v)^+$ (for a group arrow $(g, \mathbb{1}_v)$) as "(g, v)" or even just "g". We also blur the difference between the scwol arrow $(1, a)$ and the $\mathcal{Y}$–arrow a, and often refer to the scwol arrow by "a". We also blur the distinction between the $CG(\mathcal{Y})$–edge $(1, a)^{\pm}$ and the $\mathcal{Y}$–edge $a^{\pm}$.

Definition 2.15 Let $\mathcal{C} \to CG(\mathcal{Y})$ be a covering of categories. We say that an arrow is *labeled by* (g, a) if its image in $CG(\mathcal{Y})$ is (g, a). An arrow of $\mathcal{C}$ is said to be a *scwol* (*resp. group*) *arrow* if its label is a scwol (resp. group) arrow of $CG(\mathcal{Y})$.

Lemma 2.16 *If $\mathcal{C} \to CG(\mathcal{Y})$ is any cover, then every $\mathcal{C}$–path is homotopic to a concatenation of group and scwol arrows.*

Proof Observe that any $CG(\mathcal{Y})$–arrow (g, a) is a composition of a group arrow and a scwol arrow; $(g, a) = (g, t(a)) \circ (1, a)$. This gives a homotopy in $CG(\mathcal{Y})$ to a path of the desired form. Lemma 2.10 says that the homotopy lifts. □

As described at the end of the last subsection, a choice of base vertex v_0 determines a universal covering map $\widetilde{CG(\mathcal{Y})} \to CG(\mathcal{Y})$ sending a homotopy class of path $[p]$ to its terminal vertex $t(p)$, and the arrow from $[c]$ to $[c \cdot (g, a)^-]$ to the arrow (g, a) of $CG(\mathcal{Y})$.

The group $\pi_1(CG(\mathcal{Y}), v_0) \cong G$ acts on the universal cover $\widetilde{CG(\mathcal{Y})}$ by preconcatenation of paths. The quotient by this action is $CG(\mathcal{Y})$. If $H < \pi_1(CG(\mathcal{Y}), v_0)$ is any subgroup, then $H\backslash\widetilde{CG(\mathcal{Y})}$ is an intermediate cover of categories. Every connected cover of $CG(\mathcal{Y})$ is of this form. Indeed, covers of $CG(\mathcal{Y})$ correspond to coverings of the nerve N of $CG(\mathcal{Y})$. Since $\pi_1(CG(\mathcal{Y}, v_0))$ is canonically isomorphic to $\pi_1(N, v_0)$, connected covers of $CG(\mathcal{Y})$ are all of the form $H\backslash\widetilde{CG(\mathcal{Y})}$ for $H < \pi_1(CG(\mathcal{Y}), v_0)$. The isomorphism $\pi_1(CG(\mathcal{Y}), v_0) \cong G$ from Theorem 2.13 allows us to identify such H as subgroups of G.

Proposition 2.17 *Suppose that $CG(\mathcal{Y})$ arises from an action of G on a simply connected scwol $\mathcal{X}$ via Definitions 2.7 and 2.11. Let $\phi\colon \pi_1(CG(\mathcal{Y}), v_0) \to G$ be the isomorphism from Theorem 2.13. There is a ϕ–equivariant functor $\Theta\colon \widetilde{CG(\mathcal{Y})} \to \mathcal{X}$ which factors through an isomorphism of categories $\widehat{\Theta}\colon \mathrm{scwol}(\widetilde{CG(\mathcal{Y})}) \to \mathcal{X}$.*

Proof sketch Consider an arrow x labeled by (g, a) from $[\sigma_1]$ to $[\sigma_2]$ in $\widetilde{CG(\mathcal{Y})}$. We may suppose

$$\sigma_1 = (g_1, a_1)^{\epsilon_1} \cdots (g_n, a_n)^{\epsilon_n} \quad \text{and} \quad \sigma_2 = (g_1, a_1)^{\epsilon_1} \cdots (g_n, a_n)^{\epsilon_n} \cdot (g, a)^{-}.$$

We define

$$\Theta(x) = \prod_{i=1}^{n} (g_i h_{a_i})^{\epsilon_i} \tilde{a}$$

Examining the elementary homotopies, it is not hard to see that $\Theta(x)$ is well defined.

The map ϕ sends the homotopy class of the loop $(h_1, b_1)^{\delta_1} \cdots (h_k, b_k)^{\delta_k}$ to the group element

$$(h_1 h_{b_1})^{\delta_1} \cdots (h_k h_{b_k})^{\delta_k}.$$

Since $\pi_1(CG(\mathcal{Y}), v_0)$ acts by preconcatenation of paths, Θ is clearly ϕ–equivariant.

To see that Θ is a functor, suppose that $x = yz$ in $\widetilde{CG(\mathcal{Y})}$, where x, y and z are labeled by (g, a), (h, b) and (k, c), respectively, and $i(x) = i(z) = [(g_1, a_1)^{\epsilon_1} \cdots (g_n, a_n)^{\epsilon_n}]$. Letting $p = \prod_{i=1}^{n} (g_i h_{a_i})^{\epsilon_i}$, we have $\Theta(x) = p\tilde{a}$, $\Theta(y) = p h_c^{-1} k^{-1} \tilde{b}$ and $\Theta(z) = p\tilde{c}$, and it is easily checked that $\Theta(y)\Theta(z) = \Theta(x)$.

Any two objects of $\widetilde{CG(\mathcal{Y})}$ identified under the scwolification map are separated by a group arrow. If x is a group arrow then $\Theta(x)$ is a trivial arrow, so Θ factors through a functor

$$\widehat{\Theta}\colon \mathrm{scwol}(\widetilde{CG(\mathcal{Y})}) \to \mathcal{X}.$$

It remains to show that $\widehat{\Theta}$ is an isomorphism of scwols. The fact that the homomorphism from Theorem 2.13 is surjective for any $v_0 \in \mathcal{Y}$ implies that $\widehat{\Theta}$ is also surjective, so the only difficult point is to see injectivity.

Assuming that $\Theta([\sigma]) = \Theta([\tau])$, we consider the images $\hat{\sigma}$ and $\hat{\tau}$ of the *paths* σ and τ, respectively. Since $\mathcal{X}$ is assumed to be simply connected, there is a sequence of elementary homotopies in $\mathcal{X}$ taking $\hat{\sigma}$ to $\hat{\tau}$. It can be shown that these homotopies all lift (nonuniquely) to homotopies in $\widetilde{CG(\mathcal{Y})}$, so we may assume that $\hat{\sigma} = \hat{\tau}$. If $\hat{\sigma} = \hat{\tau}$ is a degenerate path at $\tilde{v}_0$, then it is clear that σ and τ are separated by a group arrow. Otherwise, after a further homotopy in $\widetilde{CG(\mathcal{Y})}$, we can assume that

$$\sigma = (g_1, a_1)^{\epsilon_1} \cdots (g_n, a_n)^{\epsilon_n}, \quad \tau = (h_1, a_1)^{\epsilon_1} \cdots (h_n, a_n)^{\epsilon_n},$$

where the signs of the ϵ_i alternate and each arrow a_i is nontrivial. If k is the smallest index for which $g_k \neq h_k$ and $k < n$, one can find a short sequence of elementary homotopies taking σ to another path σ' which agrees with τ for the first k edges. If $k = n$, then there is a slightly shorter sequence of elementary homotopies taking σ to $\tau \cdot x$ for some group arrow x.

To sum up, if $\Theta([\sigma]) = \Theta([\tau])$, then there is a group arrow from $[\sigma]$ to $[\tau]$. It follows that the map $\widehat{\Theta}\colon \operatorname{scwol}(\widetilde{CG(\mathcal{Y})}) \to \mathcal{X}$ is injective, and hence an isomorphism. □

The map Θ also passes to quotients by subgroups of G:

Corollary 2.18 *Suppose that $CG(\mathcal{Y})$ arises from an action of G on a simply connected scwol $\mathcal{X}$ via Definitions 2.7 and 2.11, and that $\mathcal{Y} = G\backslash\mathcal{X}$ is simple. If $H < G$, the scwolification of $H\backslash\widetilde{CG(\mathcal{Y})}$ is canonically isomorphic to $H\backslash\mathcal{X}$.*

Proof For any small category $\mathcal{C}$, acted on by a group H, there is a canonical surjective functor $H\backslash \operatorname{scwol} \mathcal{C} \to \operatorname{scwol}(H\backslash\mathcal{C})$. This is an isomorphism if and only if $H\backslash \operatorname{scwol} \mathcal{C}$ is already a simple scwol. Since $\mathcal{Y} = G\backslash\mathcal{X}$ is a simple scwol, the intermediate quotient $H\backslash\mathcal{X}$ is also a simple scwol.

The naturality of scwolification (Remark 2.3) and the equivariance of Θ together mean that the map $\widehat{\Theta}$ from Proposition 2.17 is also equivariant, and so we get an isomorphism

$$\widehat{\Theta}_H\colon H\backslash \operatorname{scwol}(\widetilde{CG(\mathcal{Y})}) \to H\backslash\mathcal{X}.$$

Thus $\operatorname{scwol}(H\backslash\widetilde{CG(\mathcal{Y})}) = H\backslash \operatorname{scwol}(\widetilde{CG(\mathcal{Y})})$ is isomorphic to $H\backslash\mathcal{X}$. □

Remark 2.19 Although Corollary 2.18 does not explicitly appear in Bridson and Haefliger [8], it can be derived from results there, as we outline briefly in this remark.

Suppose $\mathcal{C}' = H\backslash\widetilde{CG(\mathcal{Y})} \to CG(\mathcal{Y})$ is a cover. According to [8, Proposition III.$\mathcal{C}$.A.24] there is an associated (category of a) complex of groups $CG(\mathcal{Y}')$, which is a subcategory of $\mathcal{C}'$ such that the inclusion is an equivalence $CG(\mathcal{Y}') \to \mathcal{C}'$. The construction of $CG(\mathcal{Y}')$ from $\mathcal{C}'$ is as in [8, Proposition III.$\mathcal{C}$.A.4], which uses the same equivalence relation as the definition of the scwolification functor as in Definition 2.2. From the construction of $CG(\mathcal{Y}')$, and the assumption that $\mathcal{Y}$ is a simple scwol, it follows that $\mathcal{Y}'$ is isomorphic to $\mathrm{scwol}(\mathcal{C}')$. Further, from the correspondence between coverings and subgroups (both for categories and also for complexes of groups), and the identification of G with $\pi_1(CG(\mathcal{Y}))$ as in Theorem 2.13, it follows that $\mathcal{Y}'$ is isomorphic to $H\backslash\mathcal{X}$, as required.

Rather than fully develop this approach, we chose to sketch a more direct approach using Proposition 2.17 and the naturality of the scwolification functor.

Definition 2.20 We denote the scwolification functor from $H\backslash\widetilde{CG(\mathcal{Y})}$ to $H\backslash\mathcal{X}$ by Θ_H. If $H = \{1\}$, we just write Θ, as in Proposition 2.17.

We observe the following.

Lemma 2.21 *Suppose that $CG(\mathcal{Y})$ arises from an action of G on a simply connected scwol $\mathcal{X}$ via Definitions 2.7 and 2.11, and that $\mathcal{Y} = G\backslash\mathcal{X}$ is simple.*

Let $\mathfrak{o}$ be an object of $\mathcal{X}$. Then $\mathrm{Stab}_G(\mathfrak{o})$ acts freely and transitively on $\Theta^{-1}(\mathfrak{o})$.

Definition 2.22 Let $K < G \cong \pi_1(CG(\mathcal{Y}), v_0)$, let $\mathcal{C}_K = K\backslash\widetilde{CG(\mathcal{Y})}$ and let $\mathcal{Z} = K\backslash\mathcal{X}$. Since $\Theta_K : \mathcal{C}_K \to \mathrm{scwol}(\mathcal{C}_K) = \mathcal{Z}$ is a functor, it gives a way to turn a $\mathcal{C}_K$–path p into a $\mathcal{Z}$–path p'. Deleting all the trivial arrows from p' produces a $\mathcal{Z}$–path, which we call the *scwolification* of p. Abusing the notation slightly, we denote the scwolification of p by $\Theta_K(p)$.

Conversely, if σ is a $\mathcal{Z}$–path, then any $\mathcal{C}_K$–path $\hat{\sigma}$ such that $\Theta_K(\hat{\sigma}) = \sigma$ is called an *unscwolification* of σ. The unscwolification is highly nonunique, but always exists.

The following can be deduced by examining the elementary homotopies.

Lemma 2.23 *Scwolifications of homotopic paths are homotopic.*

Given a $CG(\mathcal{Y})$–path p we can lift it to a $\mathcal{C}_K$–path $\hat{p}$, and then scwolify $\hat{p}$ to the $\mathcal{Z}$–path $\Theta_K(\hat{p})$.

Lemma 2.24 *Let p be a $CG(\mathcal{Y})$–loop at v and let $\hat{p}$ be a lift to $\mathcal{C}_K$. If $\Theta_K(\hat{p})$ is a loop, then there is a group arrow labeled by an element of G_v joining the endpoints of $\hat{p}$.*

2.6 The complex of groups coming from an action on a cube complex

Let X be a CAT(0) cube complex, and suppose that G acts on X by cubical automorphisms. The quotient $G\backslash X$ may or may not be a cube complex, depending on whether the groups $G_\sigma = \{g \mid g\sigma = \sigma\}$ and $\{g \mid gx = x \text{ for all } x \in \sigma\}$ agree for all cells σ.

Another way to phrase this issue is to note that, if $\mathcal{X}_0$ is the scwol of cells of X, then G acts by morphisms on $\mathcal{X}_0$, but the quotient map $\mathcal{X}_0 \to G\backslash\mathcal{X}_0$ may not be a nondegenerate morphism of scwols, since some isometry of X may fix the center of some cube, but permute faces of that cube. In order to obtain a complex of groups structure on G from the action $G \curvearrowright X$, we need a scwol quotient, so we replace $\mathcal{X}_0$ with $\mathcal{X}$, the scwol of cells of the first barycentric subdivision of X:

Definition 2.25 If W is a cube complex, the *idealization of* W is the scwol of cells of the first barycentric subdivision of W.

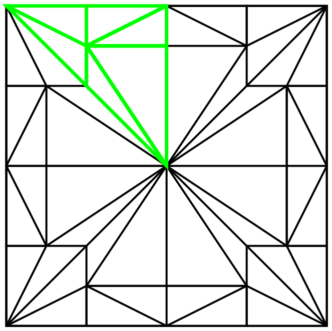
If every cube of the cube complex W embeds in W, there is another way of thinking of the idealization $\mathcal{W}$. Namely, the objects of $\mathcal{W}$ are in one-to-one correspondence with nonempty nested chains of cubes of W and there is at most one arrow in $\mathcal{W}$ between two objects: if c_1 contains c_2 as a subchain, there is an arrow from c_1 to c_2.

For example, if X is a single one-dimensional cube e with endpoints a and b, the nontrivial arrows of the idealization $\mathcal{X}$ are

$$(a) \leftarrow (a \subset e) \to (e) \leftarrow (b \subset e) \to (b). \tag{1}$$

Already a square τ with e as a face is much more complicated. The idealization is shown on the left of Figure 1 as a graph, with detail shown on the right for the highlighted portion.

Let X be a CAT(0) cube complex, and let $\mathcal{X}$ be its idealization. Any cubical automorphism of X gives a nondegenerate automorphism of $\mathcal{X}$. Moreover if such an automorphism maps a chain of cubes to itself, then it also preserves all subchains. In terms of the scwol structure this means that the stabilizer of an object also stabilizes every arrow from that object. It follows that the quotient $\mathcal{Y} = G\backslash\mathcal{X}$ is a simple scwol, and that the quotient map $\mathcal{X} \to \mathcal{Y}$ is nondegenerate morphism of scwols. Similarly, if



$$\begin{tikzcd}
(a) & & (a \subset e) \arrow[ll] \arrow[rr] & & (e) \\
& & (a \subset e \subset \tau) \arrow[u] \arrow[ull] \arrow[urr] \arrow[rr] \arrow[d] \arrow[ddrr] & & (e \subset \tau) \arrow[u] \arrow[dd] \\
& & (a \subset \tau) \arrow[uull] \arrow[drr] & & \\
& & & & (\tau)
\end{tikzcd}$$

Figure 1: Idealization of a square.

$K < G$ is any subgroup, then $K \backslash \mathcal{X}$ is a simple scwol and the maps $\mathcal{X} \to K \backslash \mathcal{X} \to \mathcal{Y}$ are nondegenerate morphisms of scwols. In particular Corollary 2.18 applies to the covers of $CG(\mathcal{Y})$.

Example 2.26 Let X be the regular 3–valent tree dual to the Farey tessellation, and let

$$G = \mathrm{PSL}(2,\mathbb{Z}) \cong \langle x \mid x^2 \rangle * \langle y \mid y^3 \rangle$$

act on X in the standard way, with y rotating around a vertex v, and x rotating around the center of an adjacent edge e. Then $CG(\mathcal{Y})$ has the following form, omitting identity arrows and most arrow labels:

(2)
$$\begin{tikzcd}
(e) \arrow[loop left, "x"] & (v \subset e) \arrow[l] \arrow[l, bend right] \arrow[r] \arrow[r, bend left] \arrow[r, bend right] & (v) \arrow[loop above, "y"] \arrow[loop right, "y^2"]
\end{tikzcd}$$

The two-fold cover corresponding to the subgroup generated by $\{y, xyx\}$ is of the following form, omitting identity arrows and all labels:

(3)
$$\begin{tikzcd}
\circ \arrow[loop left] \arrow[loop below] & \circ \arrow[l] \arrow[r] \arrow[dr] & \circ \arrow[d, bend right] & & \\
& & \circ \arrow[u, bend right] & \circ \arrow[l] \arrow[r] \arrow[ul] & \circ \arrow[loop above] \arrow[loop right]
\end{tikzcd}$$

The scwolification of this cover is isomorphic to the scwol shown in (1). In both (2) and (3), the scwol arrows are exactly the horizontal ones.

This example is one-dimensional, so there are no nontrivial compositions of scwol arrows. The reader is invited to explore a simple two-dimensional example, for example the category of the complex of groups associated to a dihedral group acting on a square with quotient scwol equal to the right-hand part of Figure 1.

Remark 2.27 Suppose that $\mathcal{C}$ is the idealization of a cube complex C, so that the realization of $\mathcal{C}$ is the second barycentric subdivision of C. In later sections, we make use of the fact that the following types of paths have canonical idealizations in $\mathcal{C}$:

(1) combinatorial paths in the 1–skeleton of the first *cubical* subdivision C^b of C (Section 3);

(2) combinatorial paths in links of cells of C (Section 4).

In both cases, this follows from the fact that subdivisions of these graphs embed naturally in the 1–skeleton of the second barycentric subdivision.

3 Quasiconvexity in the Sageev construction

In this section, we prove Theorem A. Recall that we have a hyperbolic group G acting cocompactly on a CAT(0) cube complex X, and we are required to prove that the vertex stabilizers are quasiconvex if and only if the hyperplane stabilizers are quasiconvex.

To prepare for this proof it may be useful to think about the case that X is a tree. In that case, hyperplanes are midpoints of edges, and so the statement is that edge stabilizers are quasiconvex if and only if vertex stabilizers are. Edge stabilizers are intersections of vertex stabilizers, and intersections of quasiconvex subgroups are quasiconvex, so one direction is clear. The other direction is not much harder: Consider a geodesic joining two vertices of a vertex stabilizer. The vertex stabilizer is coarsely separated from the rest of the Cayley graph by appropriate cosets of edge stabilizers. The quasiconvexity of these cosets "traps" the geodesic close to the vertex stabilizer.

Now remove the assumption that X is a tree, and suppose that vertex stabilizers are quasiconvex. It still follows that edge stabilizers are quasiconvex, but a hyperplane stabilizer is much bigger than an edge stabilizer. We will express a hyperplane stabilizer as a union of cosets of edge stabilizers, intersecting in a controlled way, and use a quasiconvexity criterion proved in the appendix to conclude that the hyperplane stabilizer is quasiconvex.

If on the other hand we assume that hyperplane stabilizers are quasiconvex, we will use them as in the tree case to control geodesics joining points in a vertex stabilizer. We inductively use more and more hyperplanes to corral points on a geodesic in an argument which terminates because of the finite-dimensionality of the cube complex.

Throughout this section we suppose that X is a CAT(0) cube complex and that $\mathcal{X}$ is its idealization (see Definition 2.25). We suppose further that G is a group acting cocompactly on this cube complex. The quotient $G\backslash\mathcal{X}$ is a scwol $\mathcal{Y}$. Making choices as in Definition 2.7, we obtain a complex of groups $G(\mathcal{Y})$, with associated category $CG(\mathcal{Y})$. Choosing a vertex $v_0 \in \mathcal{Y}$, Theorem 2.13 gives an identification of G with $\pi_1(CG(\mathcal{Y}), v_0)$. It is helpful to assume (as we may do without loss of generality) that v_0 is the orbit of some 0–cube of X.

In Section 2.4, we defined $\widetilde{CG(\mathcal{Y})}$ to be the universal covering of the category $CG(\mathcal{Y})$. Recall that the objects of $\widetilde{CG(\mathcal{Y})}$ are homotopy classes of $CG(\mathcal{Y})$–paths, starting at the basepoint $v_0 \in \mathcal{Y}$, and arrows are labeled by arrows of $CG(\mathcal{Y})$ (Definition 2.15). The basepoint of $\widetilde{CG(\mathcal{Y})}$ is $\tilde{v}_0$, the homotopy class of the constant path at v_0. As in Section 2.4 we denote the universal covering map by $\phi\colon \widetilde{CG(\mathcal{Y})} \to CG(\mathcal{Y})$. Recall from Proposition 2.17 and Definition 2.20 that $\Theta\colon \widetilde{CG(\mathcal{Y})} \to \mathcal{X}$ is the scwolification functor.

We briefly describe the contents of the remainder of this section. In Section 3.1 we explain how we consider subsets of small categories as graphs. In Section 3.2 we identify certain subsets of $\widetilde{CG(\mathcal{Y})}$ which are tuned to the cubical geometry of X^b and associate graphs with these subsets. In Section 3.3 we prove the direction (1) $\Longrightarrow$ (2) of Theorem A. In fact, we prove the more general Theorem 3.19. In Section 3.4 we prove the direction (2) $\Longrightarrow$ (1) of Theorem A. In Section 3.5 we consider various possible generalizations of Theorem A. Finally, in Section 3.6 we prove Corollary B.

3.1 Graphs from subsets of small categories

Let $\mathcal{C}$ be a (small) category, and let S be a subset of the set of arrows of $\mathcal{C}$. There is an associated graph (really a 1–complex), which we denote by $\mathrm{Gr}(S)$, with vertex set the set of objects which are either the source or target of some arrow in S, and with edges in correspondence with the arrows S.

Example 3.1 Let S be the set of *all* arrows in $\mathcal{C}$. Then $\mathrm{Gr}(\mathcal{C}) := \mathrm{Gr}(S)$ is the 1–skeleton of the nerve of $\mathcal{C}$.

Example 3.2 Suppose $\mathcal{C}$ is a group (ie $\mathcal{C}$ has a single object and each arrow of $\mathcal{C}$ is invertible), and $S_0 \subset \mathcal{C}$ is a generating set. Let S be the set of arrows in the universal cover $\widetilde{\mathcal{C}}$ with label in S_0. Then $\mathrm{Gr}(S)$ is the Cayley graph of G with respect to S_0.

3.2 Cubical paths

The group $\pi_1(CG(\mathcal{Y}), v_0)$ acts on $\widetilde{CG(\mathcal{Y})}$, with quotient the category $CG(\mathcal{Y})$. As in Theorem 2.13, we can identify G with $\pi_1(CG(\mathcal{Y}), v_0)$. The proof of each direction of Theorem A begins with choosing a certain connected G–cocompact subgraph Γ of $\mathrm{Gr}(\widetilde{CG(\mathcal{Y})})$ (see Lemma 3.9). The graph Γ admits a G–equivariant map to the 1–skeleton of the cubical subdivision of X, which we now recall.

Definition 3.3 Suppose that X is a cube complex. The (*first*) *cubical subdivision* of X, denoted by X^b, is the cube complex obtained by replacing each n–cube in X by 2^n n–cubes, found by subdividing each coordinate interval into two equal halves, and then gluing in the obvious way induced from the structure of X.

Of course, X^b is canonically homothetic to X, and X^b is NPC (respectively, CAT(0)) if and only if X is. We suppose that X is CAT(0), and therefore X^b is also.

Observation 3.4 The vertices of X^b are in bijection with the cubes of X.

The cells of X^b are in bijection with pairs $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ of cubes in X such that $\tilde{\sigma}_1 \subseteq \tilde{\sigma}_2$. The dimension of the cube corresponding to $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ is $\dim(\tilde{\sigma}_2) - \dim(\tilde{\sigma}_1)$.

Thus, a 1–cell in X^b corresponds to a pair of cubes $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ where $\tilde{\sigma}_1$ is a codimension-one face of $\tilde{\sigma}_2$. Moreover, each cell of X^b can be naturally identified with an object of $\mathcal{X}$.

As noted in Remark 2.27 any path in the 1–skeleton of X^b has a canonical idealization in $\mathcal{X}$. Each 1–cell e of X^b corresponds to some pair of cells $(\tilde{\sigma}_1 \subseteq \tilde{\sigma}_2)$ with $\tilde{\sigma}_1$ of codimension one in $\tilde{\sigma}_2$. If the path p traverses the 1–cell e, its idealization $\hat{p}$ contains consecutive arrows labeled $(\tilde{\sigma}_1 \subseteq \tilde{\sigma}_2) \to \tilde{\sigma}_1$ and $(\tilde{\sigma}_1 \subseteq \tilde{\sigma}_2) \to \tilde{\sigma}_2$, and every arrow of $\hat{p}$ has such a label. By Lemma 2.16, every $\widetilde{CG(\mathcal{Y})}$–path is homotopic to a concatenation of group arrows and scwol arrows. The graph that we use to prove Theorem A uses only scwol arrows that occur in pairs corresponding to the above description. Thus we make the following definition.

Definition 3.5 A *pair of opposable scwol arrows in* $CG(\mathcal{Y})$ is a pair of scwol arrows $(\gamma^{\downarrow}, \gamma^{\uparrow})$ such that

(1) $c = i(\gamma^{\downarrow}) = i(\gamma^{\uparrow})$ is an orbit of chain $(\tilde{\sigma}_1 \subset \tilde{\sigma}_2)$, where $\tilde{\sigma}_1$ has codimension one in $\tilde{\sigma}_2$;

(2) $\gamma^{\downarrow} = (1, a)$, where a is the arrow in $\mathcal{Y}$ corresponding to the G–orbit of the arrow $(\tilde{\sigma}_1 \subset \tilde{\sigma}_2) \to \tilde{\sigma}_1$ in $\mathcal{X}$; and

(3) $\gamma^{\uparrow} = (1, b)$, where b is the arrow in $\mathcal{Y}$ corresponding to the G–orbit of the arrow $(\tilde{\sigma}_1 \subset \tilde{\sigma}_2) \to \tilde{\sigma}_2$ in $\mathcal{X}$.

Suppose $c = i(\gamma^{\downarrow}) = i(\gamma^{\uparrow})$ for a pair of opposable scwol arrows $(\gamma^{\downarrow}, \gamma^{\uparrow})$. If $\tilde{c}$ is a lift of c to $\widetilde{CG(\mathcal{Y})}$, there are unique lifts $\tilde{\gamma}^{\downarrow}$ and $\tilde{\gamma}^{\uparrow}$ with source $\tilde{c}$. The pair $(\tilde{\gamma}^{\downarrow}, \tilde{\gamma}^{\uparrow})$ is a *pair of opposable scwol arrows in* $\widetilde{CG(\mathcal{Y})}$.

We remark that the image under Θ of a pair of opposable scwol arrows is quite restricted. In the right-hand part of Figure 1, for example, the possible images are the horizontal pair of arrows at the top of the diagram or the vertical pair at the right.

Definition 3.6 An object in $CG(\mathcal{Y})$ (equivalently, in $\mathcal{Y}$, since the objects of these two categories are the same) is *cubical* if it is an orbit of cubes in X (rather than an orbit of chains of cubes of length greater than 1). An object in $\widetilde{CG(\mathcal{Y})}$ is *cubical* if its projection to $CG(\mathcal{Y})$ is cubical.

A path p in $CG(\mathcal{Y})$ is *cubical* if

(1) the initial and terminal objects of p are cubical;

(2) p is a concatenation of group arrows and scwol arrows; and

(3) the scwol arrows occur in consecutive pairs, as pairs of opposable scwol arrows.

A path in $\widetilde{CG(\mathcal{Y})}$ is *cubical* if its projection to $CG(\mathcal{Y})$ is cubical.

It follows from the definition that all group arrows for a cubical path occur at cubical objects.

Proposition 3.7 *Suppose that v and w are cubical vertices of $CG(\mathcal{Y})$ and σ is a $CG(\mathcal{Y})$–path between v and w. Then σ is homotopic to a cubical path.*

In particular, every $g \in G = \pi_1(CG(\mathcal{Y}), v_0)$ is represented by a cubical $CG(\mathcal{Y})$–loop starting and ending at v_0.

Proof By path lifting, it suffices to prove the analogous statement for $\widetilde{CG(\mathcal{Y})}$–paths. We already observed in the proof of Proposition 2.17 that homotopies in $\mathcal{X}$ can be lifted to homotopies of $\widetilde{CG(\mathcal{Y})}$–paths, thus a given path can be homotoped to a path whose image under Θ stays in the idealization of $(X^b)^{(1)}$. Lemma 2.16 turns this into a concatenation of group arrows and scwol arrows (without changing its image under Θ). Any subpath $(1,a)^{+}\cdot(g,\mathbb{1}_{i(a)})^{\pm}$ is homotopic to $(\psi(a)(g),\mathbb{1}_{t(a)})^{\pm}\cdot(1,a)^{+}$. In particular, after a homotopy, we may assume our path contains no such subpath, and whenever a scwol arrow occurs it occurs in a pair of opposable scwol arrows. □

Definition 3.8 Suppose that for each cubical object $\mathfrak{o}$ of $\mathcal{Y}$ we choose a set $\mathbb{A}_{\mathfrak{o}} \subset G_{\mathfrak{o}}$. These determine a subset $S(\mathbb{A})$ of the arrows of $\widetilde{CG(\mathcal{Y})}$ which is the union of

(1) the set of (group) arrows with label $(g,\mathbb{1}_{\mathfrak{o}})$ for some $\mathfrak{o}$ and some $g \in \mathbb{A}_{\mathfrak{o}}$, and

(2) the set of scwol arrows occurring in some pair of opposable scwol arrows.

As discussed in Section 3.1, there is an associated graph $\mathrm{Gr}(S(\mathbb{A}))$ which we denote by $\Gamma(\mathbb{A})$. A vertex of this graph is called *cubical* if it comes from a cubical object, and otherwise it is called *central.*

Note that any central vertex of $\Gamma(\mathbb{A})$ only meets opposable scwol arrows and thus has valence exactly two, and each of its neighbors is a cubical vertex of $\Gamma(\mathbb{A})$. The valence of a cubical vertex coming from the object $\mathfrak{o}$ is equal to the number of opposable scwol arrows with terminus $\mathfrak{o}$ plus twice the cardinality of $\mathbb{A}_{\mathfrak{o}}$. Since $\mathcal{Y}$ is finite, the graph $\Gamma(\mathbb{A})$ is locally finite if and only if every $A_{\mathfrak{o}}$ is finite.

Lemma 3.9 *Suppose that $G = \pi_1(CG(\mathcal{Y}), v_0)$ is finitely generated. Then we can choose $\mathbb{A}$ so that $\Gamma(\mathbb{A})$ is locally finite, connected, and G–cocompact.*

Proof Let S be a finite generating set for G. For each $s \in S$ choose a cubical loop $p(s)$ in $CG(\mathcal{Y})$ based at v_0 representing s. For a cubical object $\mathfrak{o}$ of $\mathcal{Y}$, let $\mathbb{A}_{\mathfrak{o}}$ consist of those group arrows at $\mathfrak{o}$ which occur in some $p(s)$. There are only finitely many such, so the graph $\Gamma(\mathbb{A})$ defined in Definition 3.8 is locally finite.

To see that $\Gamma(\mathbb{A})$ is connected, let w be any vertex of $\Gamma(\mathbb{A})$. There is a path composed of opposable scwol arrows joining w to a vertex v in the G–orbit of $\tilde{v}_0$. Since S generates G, there is a concatenation of the cubical loops $p(s)$ which lifts to a path in $\Gamma(\mathbb{A})$ joining $\tilde{v}_0$ to v.

The set of G–orbits of cubical vertices of $\Gamma(\mathbb{A})$ injects into the set of objects of $\mathcal{Y}$, so it is finite and $\Gamma(\mathbb{A})$ is G–cocompact. □

Convention 3.10 For the rest of this section, we fix $\Gamma = \Gamma(\mathbb{A})$ as in the conclusion of Lemma 3.9.

The functor $\Theta\colon \widetilde{CG(\mathcal{Y})} \to \mathcal{X}$ induces a map

$$\Psi\colon \Gamma \to (X^b)^{(1)}.$$

This map is simplicial after barycentrically subdividing the target. Each pair of edges of Γ coming from a pair of opposable scwol arrows maps to a single 1–cell of $(X^b)^{(1)}$. Any central vertex of Γ maps under Ψ to an intersection of an edge of X^b with a hyperplane of X^b.

Note that the map Ψ is continuous, G–equivariant and Lipschitz.

Definition 3.11 (cubical neighborhood) Let v be a vertex of X. The *cubical neighborhood of* v is the union of those cubes of X^b which contain v. It will be denoted below by $N(v)$.

Proposition 3.12 *Let v be a vertex of X. Then $\Psi^{-1}(N(v))$ is finite Hausdorff distance from $\Psi^{-1}(v)$ in Γ.*

Proof Since G acts cocompactly on X, there are finitely many $\mathrm{Stab}(v)$–orbits of pairs of cubes $(\sigma \subset \tau)$, so σ contains v and is a codimension-one face of τ.

Let $(\sigma \subset \tau)$ be one such pair of cubes. By Lemma 2.21, $\mathrm{Stab}(\tau)$ acts freely and transitively on $\Theta^{-1}((\tau))$. The subgroup $\mathrm{Stab}(\sigma) \cap \mathrm{Stab}(\tau) = \mathrm{Stab}((\sigma \subset \tau))$ preserves the collection of pairs of opposable scwol arrows joining $\Theta^{-1}((\tau))$ to $\Theta^{-1}((\sigma))$. Moreover, $\mathrm{Stab}((\sigma \subset \tau))$ is finite index in $\mathrm{Stab}(\sigma)$. It follows that there is some $c(\tau, \sigma) > 0$ such that every vertex of $\Psi^{-1}((\tau))$ is distance at most $c(\tau, \sigma)$ from a vertex of $\Psi^{-1}((\sigma))$. As there are only finitely many $\mathrm{Stab}(v)$–orbits of pairs $(\sigma \subset \tau)$ of such faces, there is some $c > 0$ which works for every such pair.

If $x \in \Psi^{-1}(N(v))$ is a cubical vertex, it is therefore distance at most $c \cdot \dim(X)$ from a vertex of $\Psi^{-1}(v)$. If $x \in \Psi^{-1}(N(v))$ is central, then it is distance 1 from a cubical vertex of $\Psi^{-1}(N(v))$. We have shown that $\Psi^{-1}(N(v))$ is contained in the $(c \cdot \dim(X)+1)$–neighborhood of $\Psi^{-1}(v)$. Since $\Psi^{-1}(v) \subset \Psi^{-1}(N(v))$, we are finished. □

Part of our reason for working in X^b is that Proposition 3.12 would fail if we defined $N(v)$ to be the union of those cubes of X meeting v.

In the following statements we use the convention that the empty intersection of hyperplanes of X is X^b and the empty intersection of subgroups is G.

Lemma 3.13 *Suppose that $W_1, \dots, W_k$ are hyperplanes in X and that $I = \bigcap_{i=1}^k W_i$ is nonempty. For any cell τ intersecting I, the subgroup $\mathrm{Stab}(I) \cap \mathrm{Stab}(\tau)$ is finite index in $\mathrm{Stab}(\tau)$.*

Proof Finitely many hyperplanes intersect τ and these are permuted by any element of $\operatorname{Stab}(\tau)$. Thus, a finite-index subgroup of $\operatorname{Stab}(\tau)$ fixes all the hyperplanes in I. □

We observe the following consequence of the cocompactness of $G \curvearrowright X$.

Lemma 3.14 *There are finitely many G–orbits of finite sets*

$$\{W_1, \dots, W_k\}$$

of distinct hyperplanes of X with nonempty intersection $\bigcap_{i=1}^k W_i$. For each such set, the intersection $\bigcap_{i=1}^k \operatorname{Stab}(W_i)$ is finite index in $\operatorname{Stab}\bigl(\bigcap_{i=1}^k W_i\bigr)$.

Definition 3.15 Let $D > 0$. An action of a group on a metric space is *D–cobounded* if there is a set of diameter D which meets every orbit.

Proposition 3.16 *There is a constant $D > 0$ such that for any nonempty intersection I of hyperplanes of X, the subgroup $\operatorname{Stab}(I)$ acts D–coboundedly on $\Psi^{-1}(I)$.*

Proof Since there are finitely many orbits of nonempty intersections of hyperplanes, it suffices to consider a single such intersection.

Since the action of G on X is cocompact, so is the action of $\operatorname{Stab}(I)$ on I. In particular, there are finitely many $\operatorname{Stab}(I)$ orbits of cubical or central vertices in the idealization of I. Let $\mathfrak{o}$ be the object of $\mathcal{X}$ corresponding to one of these vertices. By Lemma 2.21, $\operatorname{Stab}_G(\mathfrak{o})$ acts transitively on $\Theta^{-1}(\mathfrak{o})$. By Lemma 3.13, there is a finite-index subgroup of $\operatorname{Stab}_G(\mathfrak{o})$ in $\operatorname{Stab}(I)$. Thus we see that $\Psi^{-1}(I)$ contains finitely many $\operatorname{Stab}(I)$–orbits of vertices. □

3.3 If hyperplane stabilizers are QC then cell stabilizers are QC

In this section we prove the direction (1) $\Longrightarrow$ (2) of Theorem A. As mentioned in the introduction, we prove this in greater generality than that of a hyperbolic group acting cocompactly on a CAT(0) cube complex with quasiconvex hyperplane stabilizers. The right general setting for this proof is that of *strongly quasiconvex subgroups* of finitely generated groups, as defined by Tran in [35]. (Such subgroups were also studied by Genevois [14] under the name *Morse subgroups*.)

Definition 3.17 [35, Definition 1.1] Let X be a geodesic metric space. A subset $Q \subseteq X$ is *strongly quasiconvex* if for every $K \geq 1$ and $C \geq 0$ there is some $M = M(K, C)$ such that every (K, C)–quasigeodesic in X with endpoints in Q is contained in the M–neighborhood of Q. The function $M(K, C)$ is called a *Morse gauge*.

Strong quasiconvexity persists under quasi-isometries of pairs. This is presumably known to the experts, and is closely related to [35, Proposition 4.2], but we do not see it in the literature so we provide a proof sketch.

Theorem 3.18 *Suppose X and Y are geodesic metric spaces, that $A \subset X$ is strongly quasiconvex, that $\phi\colon X \to Y$ is a quasi-isometry and that $B \subset Y$ is finite Hausdorff distance from $\phi(A)$. Then B is a strongly quasiconvex subset of Y.*

Proof sketch This is proved essentially in the same way as the corresponding fact about quasiconvex subsets of hyperbolic spaces. The difference is that instead of a single constant of quasiconvexity, we must produce a Morse gauge.

Suppose that $\phi\colon X \to Y$ and $\psi\colon Y \to X$ are (λ, ϵ)–quasi-isometries which are ϵ–quasi-inverses, and that $d_{\mathrm{Haus}}(B, \phi(A)) \le \epsilon$.

Any quasigeodesic γ joining points in B can be extended by a pair of geodesic segments of length $\le \epsilon$ to make a quasigeodesic γ' joining points in $\phi(A)$. The image of γ' under ψ can likewise be extended to a quasigeodesic γ'' between points of A. If γ was a (K, C)–quasigeodesic, then γ'' is a (K', C')–quasigeodesic where K' and C' depend only on K, C, λ and ϵ. If M is the Morse gauge for A in X, then let $M_1 = M(K', C')$. For any point p on γ, the point $\psi(p)$ is on γ'' so it is within M_1 of some point in A. Using ϕ to move back to X, we see that p is within $\lambda M_1 + 3\epsilon$ of some point of B. We can thus define a Morse gauge M' for B in Y by $M'(K, C) = \lambda M(K', C') + 3\epsilon$. □

In particular, the notion of strong quasiconvexity makes sense for subgroups of finitely generated groups.

In this subsection, we prove the following theorem.

Theorem 3.19 *Suppose that a finitely generated group G acts cocompactly on a $\mathrm{CAT}(0)$ cube complex X and that the hyperplane stabilizers are strongly quasiconvex. Then the cell stabilizers are strongly quasiconvex.*

Since quasiconvexity is equivalent to strong quasiconvexity for subgroups of hyperbolic groups, Theorem 3.19 immediately implies the direction (1) $\Longrightarrow$ (2) of Theorem A.

Note that each cell stabilizer is a finite intersection of vertex stabilizers. Tran shows that a finite intersection of strongly quasiconvex subgroups is strongly quasiconvex [35, Theorem 1.2(2)], so we only need to show that vertex stabilizers are strongly quasiconvex whenever hyperplane stabilizers are.

We will use the following general statement about intersections of strongly quasiconvex sets, analogous to [8, III.Γ.4.13].

Proposition 3.20 *For any Morse gauge M and any D and N, there is a function $R\colon [0,\infty) \to [0,\infty)$ such that the following holds: Let X be a graph of valence at most N with a free G–action, let $A, B < G$ be subgroups acting D–coboundedly on M–strongly quasiconvex subgraphs Y_A and Y_B, respectively. If $Y_C = Y_A \cap Y_B$ is nonempty, then, for any $p \in X$,*

$$d(p, Y_A \cap Y_B) \le R\bigl(\max\{d(p, Y_A), d(p, Y_B)\}\bigr).$$

Proof Let $r > 0$. We must describe $R(r)$.

Note that a concatenation of a geodesic of length r with a geodesic of any length is a $(1, 2r)$–quasigeodesic. Let $M_0 = M(1, 2r)$. Let $R(r)$ be a bound for the number of pointed oriented simplicial paths in X of length $\le 2(M_0 + D)$, up to the G–action. (A bound can be chosen depending only on N, M, D and r.)

Let $p \in X$ be chosen such that $\max\{d(p, Y_A), d(p, Y_B)\} \le r$. Let q be a closest point in Y_C to p. Suppose $d(p, q) > R(r)$, and let γ be a geodesic from p to q. Every vertex on γ lies within M_0 of both Y_A and Y_B. It follows from D–coboundedness that every vertex v on γ lies within $M_0 + D$ of some aq and some bq for $a \in A$ and $b \in B$. By our choice of $R(r)$, there must be a pair of distinct vertices v_1 and v_2 on γ and paths σ_i joining v_i to $a_i q$, and τ_i joining v_i to $b_i q$ of length at most $M_0 + D$, and an element $h \in G$ such that $hv_1 = v_2$, $h\sigma_1 = \sigma_2$ and $h\tau_1 = \tau_2$. We may assume that v_1 is closer to q than v_2 is.

Note that we have $hAq \cap Aq \neq \varnothing$. Since the action of G on X is free, this implies that $h \in A$. By the same argument $h \in B$, so $h \in C$, and thus $hq \in Y_C$. But hq is closer to p than q is, contradicting our choice of q. □

Remark 3.21 It is straightforward to see, using the above proof, that the set Y_C is $(R \circ M)$–strongly quasiconvex, and also that C acts coboundedly on Y_C (with constants depending only on D, M and R).

Towards proving Theorem 3.19, suppose that G is a finitely generated group acting cocompactly on a CAT(0) cube complex X, and suppose that hyperplane stabilizers are strongly quasiconvex in G. Recall the graph Γ from Lemma 3.9, and the continuous, G–equivariant, Lipschitz map $\Psi\colon \Gamma \to (X^b)^{(1)}$.

Lemma 3.22 *Suppose that $W_1, \dots, W_k$ are hyperplanes in X and that $I = \bigcap_{i=1}^k W_i$ is nonempty. Then $\mathrm{Stab}(I)$ is strongly quasiconvex in G.*

Proof By Lemma 3.14, the stabilizer of I is a finite-index supergroup of the intersection $\bigcap_{i=1}^k \mathrm{Stab}(W_i)$. The intersection of strongly quasiconvex subgroups is strongly quasiconvex by [35, Theorem 1.2(2)]. □

Proposition 3.23 *There exists a Morse gauge M such that for any collection of hyperplanes in X with nonempty intersection I, the set $\Psi^{-1}(I)$ is strongly quasiconvex in Γ with Morse gauge M.*

Proof The G–action on X is cocompact, so there are finitely many G–orbits of sets $\{W_1, \dots, W_k\}$ of hyperplanes of X such that $\bigcap_{i=1}^k W_i \neq \varnothing$. If I is such a set and $g \in G$ then $\Psi^{-1}(g \cdot I) = g \cdot \Psi^{-1}(I)$. Therefore, it suffices to consider a (finite) collection of representatives of G–orbits of intersections and prove that each is individually strongly quasiconvex, and then take a maximum over the finitely many Morse gauges for these subsets.

By Proposition 3.16, $\mathrm{Stab}(I)$ acts cocompactly on $\Psi^{-1}(I)$ and, by Lemma 3.22, $\mathrm{Stab}(I)$ is strongly quasiconvex in G. Therefore, by considering an orbit map $G \to \Gamma$ and applying Theorem 3.18, we see that each $\Psi^{-1}(I)$ is a strongly quasiconvex subset of Γ. □

We now give the main part of the argument of the proof of Theorem 3.19, namely that if hyperplane stabilizers are strongly quasiconvex, vertex stabilizers are also strongly quasiconvex. We therefore fix a vertex v of X.

Note that $\Psi^{-1}(v)$ is a nonempty and $\mathrm{Stab}(v)$–invariant set of vertices of Γ consisting of finitely many $\mathrm{Stab}(v)$–orbits. Thus in order to show $\mathrm{Stab}(v)$ is strongly quasiconvex in G, it suffices (by Theorem 3.18) to show that the preimage $\Psi^{-1}(v)$ is a strongly quasiconvex subset of Γ.

We fix constants $K \geq 1$ and $C \geq 0$, suppose that a and b are vertices in $\Psi^{-1}(v)$ and let γ be a (K, C)–quasigeodesic in Γ between a and b. Let y be an arbitrary vertex on γ. We have to show $d(y, \Psi^{-1}(v))$ is bounded independent of a and b. By Proposition 3.12, $\Psi^{-1}(v)$ is finite Hausdorff distance from $\Psi^{-1}(N(v))$ (recall $N(v)$ is the cubical neighborhood of v), so it is enough to show $d\big(y, \Psi^{-1}(N(v))\big)$ is bounded independent of a and b.

Here is a description of our bound: Let N be a bound for the valence of Γ and D the bound from Proposition 3.16. Let $R_0 = M(K,C)$. Assuming R_i has been defined, we let R_{i+1} be the maximum of R_i and the number $R(R_i)$, where R is the function from the conclusion of Proposition 3.20, with the Morse gauge M and the above D and N. We will prove that $d(y,\Psi^{-1}(v)) \le R_{\dim X}$.

We build a sequence of points $y = z_0, z_1, \dots, z_t$ in Γ and hyperplanes $W_1, \dots, W_t$ in X for some $t \le \dim X$ such that for each i,

$$(*_i) \qquad d(y,z_i) \le R_i, \quad \Psi(z_i) \in \bigcap_{s=1}^{i} W_s, \quad \text{and} \quad \bigcap_{s=1}^{i} W_s \cap N(v) \neq \varnothing.$$

Notice that $(*_0)$ holds, since $d(y,z_0) = 0 \le R_0$ and the empty intersection of hyperplanes is X^b.

Proposition 3.24 *Suppose $i \ge 0$ and that $z_0, \dots, z_i$ and $W_1, \dots, W_i$ have been defined and satisfy $(*_i)$. Then either $\Psi(z_i) \in N(v)$ or there is some z_{i+1} and W_{i+1} such that z_{i+1} and $W_1, \dots, W_{i+1}$ satisfy $(*_{i+1})$.*

Proof Suppose that $\Psi(z_i) \notin N(v)$. We are given hyperplanes $W_1, \dots, W_i$ such that $I = \bigcap_{s=1}^{i} W_s$ is nonempty and $I \cap N(v) \neq \varnothing$. Notice that I is a combinatorially convex subcomplex of X^b. Choose a geodesic p in the 1–skeleton of I from $I \cap N(v)$ to $\Psi(z_i)$. The first edge of p joins a vertex u_1 in $I \cap N(v)$ to a vertex u_2 which is not in $I \cap N(v)$. Thus, the vertex u_1 corresponds to a cube τ of X which contains v, and u_2 corresponds to a cube σ which is a codimension one face of τ such that σ does not contain v. Let W_{i+1} be the hyperplane of X meeting τ in a mid-cube parallel to σ. Since $u_1 \in I \cap N(v)$ we see that $I \cap W_{i+1} \cap N(v) \neq \varnothing$, since it contains the vertex u_1 of X^b. Also, because p is a geodesic in $(X^b)^{(1)}$, W_{i+1} separates $N(v)$ from $\Psi_\Gamma(z_i)$. It is clear that

$$\bigcap_{s=1}^{i+1} W_s \cap N(v) = W_{i+1} \cap I \cap N(v) \neq \varnothing,$$

so it remains to find z_{i+1} satisfying the first two conditions of $(*_{i+1})$.

Claim $\qquad d(y,\Psi^{-1}(W_{i+1})) \le R_i.$

Proof If $\Psi(y) \in W_{i+1}$ then $d(y,\Psi^{-1}(W_{i+1})) = 0$, so we suppose $\Psi(y) \notin W_{i+1}$. There are two cases.

Suppose first W_{i+1} separates v from $\Psi(y)$. In this case, we know that γ must cross $\Psi^{-1}(W_{i+1})$ between a and y. However, $\Psi_\Gamma(\gamma)$ is a loop, so γ must also cross $\Psi^{-1}(W_{i+1})$ in the segment of γ between y and b. Thus, there is a (quasigeodesic) subsegment γ_1 of γ which contains y and which starts and finishes on $\Psi^{-1}(W_{i+1})$. Since M is a Morse gauge for $\Psi^{-1}(W_{i+1})$, and $M(K,C) \le R_i$, the claim follows in this case.

Now suppose W_{i+1} does not separate v from $\Psi(y)$. In this case, since W_{i+1} *does* separate v from $\Psi(z_i)$ we know that W_{i+1} must separate $\Psi(y)$ from $\Psi(z_i)$. (In particular $i > 0$ in this case.) Since $d(y, z_i) \le R_i$, and any path from y to z_i must intersect $\Psi^{-1}(W_{i+1})$, we must have $d(y, \Psi^{-1}(W_{i+1})) \le R_i$, as required. □

Let $I = \bigcap_{s=1}^{i} W_s$. It follows immediately from the first two conditions of $(*_i)$ that $d(y, \Psi^{-1}(I)) \le R_i$. Moreover, by the claim, $d(y, \Psi_\Gamma^{-1}(W_{i+1})) \le R_i$. Furthermore,

$$\Psi^{-1}(I \cap W_{i+1}) = \Psi^{-1}(I) \cap \Psi^{-1}(W_{i+1}).$$

By Proposition 3.16 each of the sets $\Psi^{-1}(I)$, $\Psi^{-1}(W_{i+1})$ and $\Psi^{-1}(I \cap W_{i+1})$ is D–cobounded under the action of its stabilizer. Proposition 3.20 thus gives

$$d(y, \Psi^{-1}(I \cap W_{i+1})) \le R(R_i) \le R_{i+1}.$$

We choose z_{i+1} to be any point of $\Psi^{-1}(I \cap W_{i+1})$ which is closest to y. The point z_{i+1} satisfies the first two conditions of $(*_{i+1})$ so the proof is complete. □

For $j > \dim X$, there cannot exist a point z_j satisfying $(*_j)$, since there are no j–tuples of hyperplanes with nonempty intersection. Therefore, Proposition 3.24 asserts that for some $i \le \dim X$, $\Psi(z_i) = v$. We conclude that $d(y, \Psi^{-1}(v)) \le R_i \le R_{\dim X}$, as desired.

This completes the proof of Theorem 3.19.

3.4 If cell stabilizers are QC then hyperplane stabilizers are QC

In this section we prove the direction (2) $\Longrightarrow$ (1) of Theorem A. Therefore, suppose that G is a hyperbolic group acting cocompactly on a CAT(0) cube complex X, and suppose that the vertex stabilizers are quasiconvex in G.

Let W be a hyperplane in X. Then as we have noted W is a subcomplex of X^b. Given w in $W \cap (X^b)^{(0)}$, let $Y(w)$ be the (closed) 1–neighborhood in $\Psi^{-1}(W)$ of $\Psi^{-1}(w)$.

Lemma 3.25 *The sets $Y(w)$ are quasiconvex subsets of Γ with constants which do not depend on w.*

Proof Since $\mathrm{Stab}(W)$ acts cocompactly on W, there are finitely many $\mathrm{Stab}(W)$–orbits of sets $Y(w)$, so the uniformity of constants will follow immediately if we can prove each $Y(w)$ is a quasiconvex subset of Γ. We therefore fix such a w.

The stabilizer $\mathrm{Stab}(w)$ is equal to the stabilizer of some cube σ of X. Thus $\mathrm{Stab}(w)$ is virtually the intersection of the stabilizers of the vertices of σ. The vertex stabilizers are assumed to be quasiconvex, so $\mathrm{Stab}(w)$ is also quasiconvex.

Since G acts freely and cocompactly on Γ and Ψ is equivariant, $\mathrm{Stab}(w)$ acts freely and cocompactly on $\Psi^{-1}(w)$. By Lemma 3.13, $\mathrm{Stab}(W) \cap \mathrm{Stab}(w)$ is a finite-index subgroup of $\mathrm{Stab}(w)$, so it is also quasiconvex and acts freely and cocompactly on $\Psi^{-1}(w)$. It moreover acts cocompactly on $Y(w)$.

The result follows (for example, by Theorem 3.18). □

We are ready to prove the direction (2) $\Longrightarrow$ (1) of Theorem A, which is the content of the following theorem. For this result, we assume Theorem A.3, which is proved in the appendix.

Theorem 3.26 *Suppose that the hyperbolic group G acts cocompactly on the cube complex X, and that for every vertex v of X, the stabilizer $\mathrm{Stab}(v)$ is quasiconvex. Then, for every hyperplane $W \subset X$, the stabilizer $\mathrm{Stab}(W)$ is a quasiconvex subgroup of G.*

Proof As we have already remarked, quasiconvexity of vertex stabilizers implies quasiconvexity of all cell stabilizers.

Let Γ, $\Psi^{-1}(W)$ and the $Y(w)$ be as discussed above. Since G acts freely and cocompactly on Γ, we know that Γ is δ–hyperbolic for some δ. Let ϵ be a constant such that $Y(w)$ is ϵ–quasiconvex for every w (Lemma 3.25).

Since $\mathrm{Stab}(W)$ acts freely and cocompactly on $\Psi^{-1}(W)$, in order to prove the theorem it suffices to prove that $\Psi^{-1}(W)$ is quasiconvex in Γ, so let $p, q \in \Psi^{-1}(W)$.

Consider a geodesic γ in $(X^b)^{(1)}$ between $\Psi(p)$ and $\Psi(q)$. Both $\Psi(p)$ and $\Psi(q)$ lie in W. Since W is combinatorially convex in X^b, the geodesic γ is entirely contained in the 1–skeleton of W (considered as a subcomplex of X^b). The vertices $w_1, \dots, w_n$ on γ

correspond to cells of X contained in W. The sets $Y(w_i)$ corresponding to these cells satisfy the hypotheses of Theorem A.3 with $m = 2$, $c = 1$, and ϵ the quasiconvexity constant chosen above. Theorem A.3 then implies that $Y(w_1) \cup \cdots \cup Y(w_n)$ is ϵ'–quasiconvex, for a constant ϵ' depending only on ϵ and δ.

In particular, a Γ–geodesic between p and q lies within ϵ' of $Y(w_1) \cup \cdots \cup Y(w_n)$. Since each of these $Y(w_i)$ is contained in $\Psi^{-1}(W)$, the Γ–geodesic between p and q stays uniformly close to $\Psi^{-1}(W)$, as required. □

Together with Theorem 3.19, this completes the proof of Theorem A.

3.5 On generalizations of Theorem A

For a subgroup H of a hyperbolic group G, the following three conditions are equivalent:

(a) H is strongly quasiconvex in G.

(b) H is quasiconvex in G.

(c) H is undistorted in G.

Dropping the condition that G is hyperbolic, condition (b) ceases to be a useful notion, but conditions (a) and (c) still make sense.

One can ask for versions of Theorem A where the hypothesis of hyperbolicity is removed and condition (b) is replaced by either condition (a) or (c).

3.5.1 Strong quasiconvexity Replacing quasiconvexity with strong quasiconvexity we can ask about the following conditions for a finitely generated group G acting cocompactly on a CAT(0) cube complex:

(1S) Hyperplane stabilizers are strongly quasiconvex.

(2S) Vertex stabilizers are strongly quasiconvex.

(3S) All cell stabilizers are strongly quasiconvex.

As remarked earlier, (2S) $\iff$ (3S) follows from [35, Theorem 1.2(2)]. Theorem 3.19 states that (1S) $\implies$ (2S).

The remaining implication (3S) $\implies$ (1S) is *false*, as shown for example by $\mathbb{Z}^2$ acting freely on a cubulated $\mathbb{R}^2$.

3.5.2 Undistortedness The situation when replacing quasiconvexity with quasi-isometric embeddedness is murkier. We consider the following conditions, for a finitely generated group G acting cocompactly on a CAT(0) cube complex X:

(1U) Hyperplane stabilizers are undistorted.

(2U) Vertex stabilizers are undistorted.

(3U) All cell stabilizers are undistorted.

If X is a tree, (1U) and (3U) each imply (2U), but not conversely. For example, the double of a finitely generated group over a distorted group acts on a tree with undistorted vertex stabilizers but distorted edge/hyperplane stabilizers.

We do not know the relationship between (1U) and (3U) in general, so we ask the question.

Question 3.27 *For finitely generated groups acting cocompactly on CAT(0) cube complexes does (1U) $\implies$ (3U)? Does (3U) $\implies$ (1U)?*

3.6 Height of families and the proof of Corollary B

The *height* of a subgroup was introduced in [16]. We need a generalization of this notion to families of subgroups.

Definition 3.28 (height of a family) Suppose that G is a group and $\mathcal{H}$ is a collection of subgroups. The *height* of $\mathcal{H}$ is the minimum number n such that for every tuple of distinct cosets $(g_0H_0, g_1H_1, \dots, g_nH_n)$ with $H_i \in \mathcal{H}$ (and $g_i \in G$), the intersection $\bigcap_{i=0}^n H_i^{g_i}$ is finite. If there is no such n then we say the height of $\mathcal{H}$ is infinite.

When $\mathcal{H} = \{H\}$ is a single subgroup, we recover the familiar notion of the height of a subgroup from [16].

The following result for a single subgroup is part of [16, Main Theorem]. The proof of that result from [1, Corollary A.40] can be adapted in the obvious way to prove the result for finite families. This result was proved in the more general setting of strongly quasiconvex subgroups by Tran [35, Theorem 1.2(3)].

Proposition 3.29 *Let G be a hyperbolic group and $\mathcal{H}$ a finite collection of quasiconvex subgroups of G. Then the height of $\mathcal{H}$ is finite.*

We also use a special case of a theorem of Charney and Crisp [9, Theorem 5.1]:

Theorem 3.30 *Suppose that G acts cocompactly on a cube complex X. Then X is quasi-isometric to the space obtained from the Cayley graph of G by coning cosets of stabilizers of vertices to points.*

We now prove Corollary B. For convenience, we recall the statement.

Corollary B *Suppose that G is a hyperbolic group acting cocompactly on a CAT(0) cube complex X with quasiconvex hyperplane stabilizers. Then*

(1) *X is δ–hyperbolic for some δ;*

(2) *there exists a $k \geq 0$ such that the fixed-point set of any infinite subgroup of G intersects at most k distinct cells; and*

(3) *the action of G on X is acylindrical* (in the sense of Bowditch [6, page 284]).

Proof If G is a hyperbolic group acting cocompactly on a CAT(0) cube complex, and if the stabilizers in G of vertices in X are quasiconvex, then [7, Theorem 7.11], due to Bowditch, implies that this coned graph is δ–hyperbolic for some δ. Theorem 3.30 then implies that the cube complex X is δ–hyperbolic for some (possibly different) δ. Thus, we have the first statement from Corollary B.

Now we prove the statement about fixed-point sets of infinite subgroups. Let I be a collection of orbit representatives of cells in X. For $i \in I$, let $Q_i = \{g \in G \mid gi = i\}$, and let $\mathcal{Q} = \{Q_i\}_{i \in I}$. Then $\mathcal{Q}$ is a finite collection of quasiconvex subgroups of G, so it has some finite height k by Proposition 3.29. If $H < G$ is infinite with nonempty fixed-point set and σ is a cell meeting the fixed-point set of H, then $H < Q_i^g$, where $\sigma = gi$. Since the height of $\mathcal{Q}$ is k, at most k such cells appear.

In [15], Genevois studies actions of groups on hyperbolic CAT(0) cube complexes and shows in Theorem 8.33 that, in this setting, acylindricity is equivalent to the condition:

$$\text{(G)} \qquad \exists L, R \;\; \forall x, y \in X^{(0)} \;\; d(x,y) \geq L \implies \#(\mathrm{Stab}(x) \cap \mathrm{Stab}(y)) \leq R.$$

We take R to be the maximum size of a finite subgroup and $L = k$. Suppose $d(x,y) \geq L$. Then the union of the combinatorial geodesics joining x to y contains finitely many (but at least $k+1$) vertices. There is a finite-index subgroup of $\mathrm{Stab}(x) \cap \mathrm{Stab}(y)$ which fixes all of these vertices. This finite-index subgroup fixes more than k cells, so it is finite. This implies $\mathrm{Stab}(x) \cap \mathrm{Stab}(y)$ is finite, as desired. □

Remark 3.31 In the context where G is a finitely generated group acting cocompactly on a cube complex X with strongly quasiconvex hyperplane stabilizers, the same proof of conclusion (2) works as written, replacing the reference to Proposition 3.29 with a reference to [35, Theorem 1.2(3)].

4 Conditions for quotients to be CAT(0)

As noted in the introduction, Theorem D follows quickly from Theorems A and F, Agol's Theorem [1, Theorem 1.1] and Wise's Quasiconvex Hierarchy Theorem [36, Theorem 13.3]. Thus, other than Theorem A.3 in the appendix (which is independent of everything else in this paper), it remains to prove Theorem F. Therefore, we are interested in conditions on a group G acting on a CAT(0) cube complex X and a normal subgroup $K \trianglelefteq G$ which ensure that the quotient $K\backslash X$ is a CAT(0) cube complex. In this section we develop criteria in terms of complexes of groups to ensure this. In the next section, we translate these conditions into algebraic conditions on $K \trianglelefteq G$.

Three conditions need to be ensured in order for the complex $Z = K\backslash X$ to be a CAT(0) cube complex:

(1) Z must be simply connected;

(2) Z must be a cube complex (rather than a complex made out of cells which are quotients of cubes); and

(3) Z must be nonpositively curved.

We investigate these three properties in turn.

4.1 Ensuring the quotient is simply connected

First, we give a sufficient condition for $K\backslash X$ to be simply connected.

Since X is a finite-dimensional cube complex, it has finitely many shapes, and we can use the following application of a theorem of Armstrong:

Theorem 4.1 *Let X be a simply connected metric polyhedral complex with finitely many shapes, and let K be a group of isometries of X respecting the polyhedral structure, generated by elements with fixed points. Then $K\backslash X$ is simply connected.*

Proof sketch A theorem of Armstrong [4, Theorem 3] shows that $K\backslash X$ is simply connected with the CW topology. We have to show it is still simply connected with the metric topology.

Because X has finitely many shapes, there is an equivariant triangulation $\mathcal{T}$ and an $\epsilon > 0$ such that for every finite subcomplex Y, the ϵ–neighborhood of Y deformation retracts to Y. If $f : S^1 \to X$ is any loop, then a compactness argument shows that it lies in an ϵ–neighborhood of some such finite complex. We can then homotope f to have image in Y and apply the simple connectedness of $K\backslash X$ with the CW topology. □

We remark that the hypothesis of finitely many shapes is necessary even when X is CAT(0) as the following example shows:

Example 4.2 For $n \in \{2, 3, \dots\}$, let D_n be the Euclidean cone of radius 1 on a loop σ_n of length $2\pi/n^2$. For each n mark a point on σ_n. Let Y be obtained from $\bigcup_n D_n$ by identifying the marked points. Unwrapping all the cones to Euclidean discs gives a tree of Euclidean discs of radius 1. We call this CAT(0) space $\widetilde{Y}$. There is a discrete group of isometries $\Gamma = \langle \gamma_2, \gamma_3, \dots \rangle$ acting on $\widetilde{Y}$ with quotient Y such that each γ_n fixes the center of some disc and rotates it by an angle of $2\pi/n^2$. Nonetheless Y is not simply connected, as the infinite concatenation of the loops σ_n has finite length, but cannot be contracted to a point.

4.2 Ensuring the quotient is a cube complex

We now turn to the question of when $K\backslash X$ is a cube complex.

In order that the quotient $Z = K\backslash X$ be a cube complex, there needs to be no element of K which fixes a cell of X setwise but not pointwise.

Suppose that σ is a cube of X. The stabilizer G_σ has a finite-index subgroup Q_σ consisting of those elements which fix σ pointwise. Let $\{\sigma_1, \dots, \sigma_k\}$ be a set of representatives of G–orbits of cubes in X. The following result is straightforward:

Proposition 4.3 *Suppose that G acts cocompactly on the cube complex X and that K is a normal subgroup of G such that for each i we have $G_{\sigma_i} \cap K \le Q_{\sigma_i}$. Then the quotient $K\backslash X$ is a cube complex and the links of vertices in $K\backslash X$ inherit a cellular structure from the simplicial structure of cells in X.*

4.3 Ensuring the quotient is nonpositively curved

The most complicated condition to ensure is that $K\backslash X$ is nonpositively curved.

Throughout this subsection we suppose that X is a CAT(0) cube complex and that $\mathcal{X}$ is its idealization (see Definition 2.25). We suppose further that G is a group acting cocompactly on this cube complex. The induced action of G on $\mathcal{X}$ has quotient a scwol $\mathcal{Y}$. Making choices as in Definition 2.7, we obtain a complex of groups $G(\mathcal{Y})$, with associated category $CG(\mathcal{Y})$. Choosing a vertex $v_0 \in \mathcal{Y}$, there is then an identification of G with $\pi_1(CG(\mathcal{Y}), v_0)$. Moreover, we choose a normal subgroup $K \trianglelefteq G$ such that $K \backslash X$ is a cube complex. Throughout this section we let $\mathcal{Z} = K\backslash\mathcal{X}$.[5]

In Section 4.2 we discussed how to find subgroups K such that $K\backslash X$ is a cube complex, but for this section we just assume that this is the case.

Let $\mathcal{C}_K$ be the cover of the category $CG(\mathcal{Y})$ corresponding to the subgroup K. Observe that $CG(\mathcal{Y})$–loops lift to $\mathcal{C}_K$ if and only if they represent elements of K. (Basepoints are mostly omitted in this section, since we deal with a normal subgroup K.)

Standing Assumption 4.4 Lemma 2.16 tells us that any $CG(\mathcal{Y})$–path is homotopic to a concatenation of group and scwol arrows. Throughout this section we assume all $CG(\mathcal{Y})$–paths have been homotoped to this form, though we do not assume that the arrows *alternate* between group and scwol arrows. We blur the distinction between the scwol arrow $(1, a_i)$ in $CG(\mathcal{Y})$ and the $\mathcal{Y}$–arrow a_i, and also between the group arrow $(g, \mathbb{1}_v)$ and the element $g \in G_v$.

Thus we may write a $CG(\mathcal{Y})$ path for example as a list $g_1 \cdot e_1 \cdot e_2 \cdot g_2 \cdots$, where each g_i is an element of a local group G_v and represents the edge $(g, \mathbb{1}_v)^+$, corresponding to the group arrow $(g, \mathbb{1}_v)$, and each e_i is equal to some $a_i^{\pm}$ for a scwol arrow $(1, a_i)$. We implicitly assume that each concatenation we write defines a path, which forces the group arrows labeled g_i to be elements of particular local groups. Whenever we consider a $CG(\mathcal{Y})$–path of length 1 consisting of a single group arrow we are either explicit about the local group or else it is clear from the context.

If we have an edge of the form $(1, \mathbb{1}_v)^{\pm}$, we often implicitly (or explicitly) omit this arrow from our path. Again this only changes the path by an elementary homotopy.

Definition 4.5 (link of a cube) Given an n–cube σ in a cube complex Z, let b_σ be the barycenter. For sufficiently small $\epsilon > 0$, the sphere $\{x \in Z \mid d_Z(b_\sigma, x) = \epsilon\}$ is the join of an $(n-1)$–sphere with some simplicial complex (here we take the join of a (-1)–sphere with K to be equal to K). This simplicial complex is what we refer to as the *link* of σ, or link(σ). It is naturally triangulated by simplices corresponding to inclusions of σ as a face of some higher-dimensional cube. In particular link(σ) is a Δ*–complex* [21, Chapter 2.1] (though it may not be simplicial).

[5] We do not make any further assumptions than these about G and X in this subsection. This may be an important observation for future applications.

Though the link of a cube naturally has the structure of a piecewise spherical complex, we will think of it as just a combinatorial object. When we refer to paths or loops in a link, these will always be concatenations of 1–cells. The *length* of such a path is the number of 1–cells traversed.

Theorem 4.6 (Gromov's cubical link condition [8, Theorem II.5.20]) *The cube complex Z is a nonpositively curved cube complex if and only if the link of each vertex in Z is flag.*

In this section, we provide a set of conditions on the subgroup K which imply the link condition for $Z = K\backslash X$.

We record two elementary observations:

Lemma 4.7 *Let v be a vertex of a cube complex, and let $L = \mathrm{link}(v)$. Then L is simplicial if and only if for every cell σ containing v, the 1–skeleton of $\mathrm{link}(\sigma)$ contains no immersed loop of length 1 or 2.*

Proof If L fails to be simplicial, there is either a nonembedded simplex, or a pair of simplices which intersect in a set which is not a face of both. If a simplex is nonembedded, we obtain a loop of length 1 in L. If two embedded simplices τ_1 and τ_2 of L intersect in a set which is not a single face, let F_1 and F_2 be different maximal faces in the intersection, and let $f = F_1 \cap F_2$. For $i \in \{1, 2\}$, let v_i be a vertex in $F_i - f$. Then the simplices spanned by $v_1 \cup f$ and $v_2 \cup f$ correspond to points in $\mathrm{link}(f)$ which lie on an immersed loop of length 2. But f corresponds to some cube containing σ, and $\mathrm{link}(f) \subset L$ is isomorphic to the link of that cube. □

Lemma 4.8 *Let v be a vertex of a cube complex, and suppose $L = \mathrm{link}(v)$ is simplicial. Then L is a flag complex if and only if for every cell σ containing v, every loop of length 3 in $\mathrm{link}(\sigma)$ is filled by a 2–cell.*

Proof If σ is a cube and ϕ is a cube with σ as a face, of dimension one higher, then ϕ corresponds to a vertex f of the link L of σ. The link of ϕ is isomorphic to the link in L of f. The result now follows from [8, Remark II.5.16(4)]. □

Therefore, in order to ensure Z is nonpositively curved, for each cell σ in Z we must rule out loops of length 1 and 2 in $\mathrm{link}(\sigma)$ and also ensure that any loop of length 3 in $\mathrm{link}(\sigma)$ is filled by a 2–cell. We first explain how we translate between 1–cells in links in Z and $CG(\mathcal{Y})$–paths. Then we develop the required conditions to rule out loops, finally dealing with loops of length 3 which must be filled by 2–cells.

4.3.1 $CG(\mathcal{Y})$–paths associated to 1–cells in link(σ) Below we choose, for each cube σ in X and each oriented 1–cell α in the link of σ, a $CG(\mathcal{Y})$–path $p_{[\![\alpha]\!]}$ which is the label of an unscwolification of the idealization of α. As indicated by the notation, this label is the same for two such 1–cells in the same G–orbit. (In fact we will choose these paths for slightly more general objects than 1–cells in links of cubes.) If α is an oriented 1–cell, we write $\overleftarrow{\alpha}$ for the same 1–cell with the opposite orientation.

We fix a cube σ of X. The second barycentric subdivision of the link of σ embeds naturally in the geometric realization of $\mathcal{X}$. The vertices of the image of link(σ) are precisely the length ≥ 2 chains of cubes whose minimal element is σ.

We first consider an oriented 1–cell β of link(σ). The 1–cell β corresponds to a triple of cubes $\epsilon_1, \epsilon_2, \phi$ in X such that

$$\dim(\sigma) = \dim(\epsilon_i) - 1 = \dim(\phi) - 2,$$

with $\epsilon_1, \epsilon_2 \subset \phi$ and $\epsilon_1 \cap \epsilon_2 = \sigma$.

In particular, β has idealization an $\mathcal{X}$–path of length 4, made up of arrows

$$(\sigma \subset \epsilon_1) \leftarrow (\sigma \subset \epsilon_1 \subset \phi) \rightarrow (\sigma \subset \phi) \leftarrow (\sigma \subset \epsilon_2 \subset \phi) \rightarrow (\sigma \subset \epsilon_2). \tag{4}$$

In order to formulate Definition 4.9 and Lemma 4.10 below we must consider a slightly more general situation: we suppose ϵ_1, ϵ_2 and ϕ are cubes in X with ϵ_1 and ϵ_2 codimension-one subcubes of ϕ, and γ is a chain of cubes in X such that each element of γ is contained in each of ϵ_1, ϵ_2 and ϕ. We can naturally extend γ to chains which we denote by $(\gamma \subset \epsilon_1)$, $(\gamma \subset \epsilon_2)$, $(\gamma \subset \phi)$, and $(\gamma \subset \epsilon_i \subset \phi)$. This sequence of chains corresponds to a 1–cell α in an "iterated link" (a link of a cell in a link, etc), which we fix from now through Definition 4.9. The 1–cell α has idealization an $\mathcal{X}$–path of length 4,

$$(\gamma \subset \epsilon_1) \leftarrow (\gamma \subset \epsilon_1 \subset \phi) \rightarrow (\gamma \subset \phi) \leftarrow (\gamma \subset \epsilon_2 \subset \phi) \rightarrow (\gamma \subset \epsilon_2). \tag{5}$$

The $\mathcal{X}$–path (5) may not embed in $\mathcal{Y}$. There are two ways this could happen. The first is that there is an element of Stab(γ) which sends ϵ_1 to ϵ_2, but no such element fixes ϕ. In this case, the image in $\mathcal{Y}$ is a nonbacktracking loop. The second possibility is that there is an element $g \in G$ sending each of γ and ϕ to itself, but exchanging ϵ_1 and ϵ_2. If there is such a g, the idealization of the 1–cell α backtracks in $\mathcal{Y}$, forming a "half 1–cell". Since we are assuming Z is a cube complex, the second possibility does not arise for the image of (5) in $\mathcal{Z}$, though the first possibility may.

Let $y_{[\![\alpha]\!]} = a_1^+ \cdot a_2^- \cdot a_3^+ \cdot a_4^-$ be the $\mathcal{Y}$–path which is the image of the $\mathcal{X}$–path (5). For more compact notation, we define

$$\nu = [\![(\gamma \subset \phi)]\!], \quad \mu_i = [\![(\gamma \subset \epsilon_i \subset \phi)]\!], \quad \xi_i = [\![(\gamma \subset \epsilon_i)]\!] \quad \text{for } i \in \{1,2\}.$$

Then we have the injective homomorphisms

$$\psi_{a_2}\colon G_{\mu_1} \to G_\nu \quad \text{and} \quad \psi_{a_3}\colon G_{\mu_2} \to G_\nu.$$

The images of ψ_{a_2} and ψ_{a_3} are equal. The projection to $\mathcal{Y}$ of the path (5) associated to α depends only on $[\![\alpha]\!]$. We denote the common image of ψ_{a_2} and ψ_{a_3} in G_ν by G_ν^+. Note that G_ν^+ either has index 2 in G_ν (if there is a g fixing γ and ϕ and exchanging ϵ_1 with ϵ_2) or else $G_\nu^+ = G_\nu$ (if there is no such g). If G_ν^+ has index 2 in G_ν, we fix a choice of $g_\nu \in G_\nu - G_\nu^+$. We make this choice once and for all for each orbit of $(\gamma, \epsilon_1, \epsilon_2, \phi)$, so that the choice depends only on the orbit and not on the representative.

In the sequel, we refer to the vertex groups by $G_{i([\![\alpha]\!])}$ (for G_{ξ_1}) and $G_{t([\![\alpha]\!])}$ (for G_{ξ_2}). We further define "edge-inclusions" $\psi_{[\![\alpha]\!]}\colon G_\nu^+ \to G_{t([\![\alpha]\!])}$ and $\psi_{[\![\bar\alpha]\!]}\colon G_\nu^+ \to G_{i([\![\alpha]\!])}$ by

$$\psi_{[\![\alpha]\!]} = \psi_{a_4} \circ \psi_{a_3}^{-1} \quad \text{and} \quad \psi_{[\![\bar\alpha]\!]} = \psi_{a_1} \circ \psi_{a_2}^{-1}.$$

Let $E_{[\![\alpha]\!]}$ denote the image of $\psi_{[\![\alpha]\!]}$ in $G_{t([\![\alpha]\!])}$, and $E_{[\![\bar\alpha]\!]}$ denote the image of $\psi_{[\![\bar\alpha]\!]}$ in $G_{i([\![\alpha]\!])}$.

Definition 4.9 If $G_\nu^+ \neq G_\nu$, associate to α the $CG(\mathcal{Y})$–path

$$p_\alpha = a_1^+ \cdot a_2^- \cdot g_\nu \cdot a_3^+ \cdot a_4^-,$$

where $g_\nu \in G_\nu - G_\nu^+$ is the element fixed above. Note that in this case $a_2 = a_3$ and $a_1 = a_4$. If $G_\nu^+ = G_\nu$, let

$$p_\alpha = a_1^+ \cdot a_2^- \cdot a_3^+ \cdot a_4^-.$$

In either case some lift of p_α to $\widetilde{CG(\mathcal{Y})}$ is an unscwolification of the idealization of α. The $CG(\mathcal{Y})$–path p_α depends only on $[\![\alpha]\!]$; if there is a g with $g\sigma = \sigma$ and $g\alpha' = \alpha$, then $p_{\alpha'} = p_\alpha$. Therefore, we have a well-defined $CG(\mathcal{Y})$–path $p_{[\![\alpha]\!]}$.

Next we consider $CG(\mathcal{Y})$–paths whose scwolifications traverse $y_{[\![\alpha]\!]}$ and whose lifts to $\widetilde{CG(\mathcal{Y})}$ have nonbacktracking scwolifications. (Since $Z = K\backslash X$ is assumed to be a cube complex, these paths also have lifts in $\mathcal{C}_K$ whose scwolifications are nonbacktracking in $\mathcal{Z}$.) We observe in the following lemma that such paths can be converted by a sequence of elementary homotopies to paths which consist of a copy of $p_{[\![\alpha]\!]}$ bookended by group arrows.

Lemma 4.10 *Suppose that $G_\nu^+ = G_\nu$. Then any $CG(\mathcal{Y})$–path*

$$g_0 \cdot a_1^+ \cdot g_1 \cdot a_2^- \cdot g_2 \cdot a_3^+ \cdot g_3 \cdot a_4^- \cdot g_4$$

is homotopic to a $CG(\mathcal{Y})$–path of the form

$$g_0' \cdot p_{[\![\alpha]\!]} \cdot g_1'.$$

Suppose that $G_\nu^+ \neq G_\nu$. Then any $CG(\mathcal{Y})$–path

$$g_0 \cdot a_1^+ \cdot g_1 \cdot a_2^- \cdot g_2 \cdot a_3^+ \cdot g_3 \cdot a_4^- \cdot g_4$$

such that $g_2 \notin E_{[\![\alpha]\!]}$ is homotopic to a $CG(\mathcal{Y})$–path of the form

$$g_0' \cdot p_{[\![\alpha]\!]} \cdot g_1'.$$

In both cases, the scwolification of the path is fixed during the homotopy. Moreover any lift of the homotopy to a cover of $CG(\mathcal{Y})$ gives a sequence of paths with constant scwolification.

Notation 4.11 We fix some notation in order to study paths in Z and also $\mathcal{Y}$–paths and $CG(\mathcal{Y})$–paths. As above, we use $[\![\cdot]\!]$ to denote a G–orbit in $\mathcal{Z}$, which corresponds to its image in $\mathcal{Y}$ under the projection $\pi\colon \mathcal{Z} \to \mathcal{Y}$.

Let $p_{[\![\alpha]\!]}$ be one of the $CG(\mathcal{Y})$–paths fixed in Definition 4.9, corresponding to a 1–cell α in some link (or iterated link) of a cube of Z. The $CG(\mathcal{Y})$–path $p_{[\![\alpha]\!]}$ has an underlying $\mathcal{Y}$–path, which we denote by $y_{[\![\alpha]\!]}$. Define $t([\![\alpha]\!]) = t(y_{[\![\alpha]\!]})$ and $i([\![\alpha]\!]) = i(y_{[\![\alpha]\!]})$. This is so we can denote the corresponding local groups by $G_{i([\![\alpha]\!])}$ and $G_{t([\![\alpha]\!])}$.

We will also need to refer to the subgroups $E_{[\![\alpha]\!]} < G_{t([\![\alpha]\!])}$ and $E_{[\![\bar{\alpha}]\!]} < G_{i([\![\alpha]\!])}$ defined just before Definition 4.9. Each of these subgroups can be thought of as the pointwise stabilizer of some translate of a lift of α to X.

4.3.2 Loops in link($\breve{\sigma}$)

We are now ready to formulate the conditions on K which characterize whether or not $K\backslash X$ is nonpositively curved. We use Lemmas 4.7 and 4.8 repeatedly.

Recall that we have fixed a $K \lhd G$ such that $Z = K\backslash X$ is a cube complex. We also fix a cube $\breve{\sigma}$ of Z, and a lift σ of $\breve{\sigma}$ to X. If α is a 1–cell in link(σ), then as described above there is a corresponding $\mathcal{X}$–path of length 4 (its idealization). Similarly a 1–cell in the link of $\breve{\sigma}$ has a corresponding $\mathcal{Z}$–path of length 4. We sometimes conflate a concatenation of 1–cells with a concatenation of these idealizations.

We next give an algebraic characterization of 1–cells in $\mathrm{link}(\sigma)$ which project to loops of length 1 in $\mathrm{link}(\breve{\sigma})$. Recall $\mathcal{C}_K = K\backslash\widetilde{CG(\mathcal{Y})}$.

Lemma 4.12 *Let α be a 1–cell in $\mathrm{link}(\sigma)$, and $\breve{\alpha}$ the projection of α to $\mathrm{link}(\breve{\sigma})$. The endpoints of $\breve{\alpha}$ are equal if and only if there is a $CG(\mathcal{Y})$–loop of the form $p_{[\![\alpha]\!]}\cdot g$ that represents a conjugacy class in K.*

Proof Thinking of X as the geometric realization of $\mathcal{X}$, the 1–cell α is the realization of an $\mathcal{X}$–path q_α of length 4, which projects to a $\mathcal{Y}$–path $a_1^+\cdot a_2^-\cdot a_3^+\cdot a_4^-$. Let $\hat{q}_\alpha$ be an unscwolification of q_α in $\widetilde{CG(\mathcal{Y})}$, which we may choose to have label

$$a_1^+\cdot a_2^-\cdot g_1\cdot a_3^+\cdot a_4^- \tag{6}$$

for some group arrow g_1.

Suppose first that the endpoints of $\breve{\alpha}$ coincide. Then the path (6) projects in $\mathcal{C}_K$ to a path with endpoints separated by a group arrow, and so there is a $\mathcal{C}_K$–loop with label

$$a_1^+\cdot a_2^-\cdot g_1\cdot a_3^+\cdot a_4^-\cdot g_2.$$

Lemma 4.10 implies that this loop is homotopic to a loop with label $g_0\cdot p_{[\![\alpha]\!]}\cdot g_1$ for some g_0 and g_1, and starting this loop at a different place gives the required $CG(\mathcal{Y})$–loop $p_{[\![\alpha]\!]}\cdot g$ representing a conjugacy class of K.

Conversely, suppose a conjugacy class in K is represented by a $CG(\mathcal{Y})$–loop of the form $p_{[\![\alpha]\!]}\cdot g$. Then $p_{[\![\alpha]\!]}\cdot g$ lifts to a loop in $\mathcal{C}_K$ whose scwolification is a projection of a translate of q_α by some element of G. In particular, q_α must project to a loop, and so the endpoints of $\breve{\alpha}$ coincide. □

Definition 4.13 If $\mathfrak{o}$ is an object of $\mathcal{Y}$, let $K_{\mathfrak{o}}\lhd G_{\mathfrak{o}}$ be $K\cap G_{\mathfrak{o}}$.

Definition 4.14 A $CG(\mathcal{Y})$–path p is *K–nonbacktracking* if for some (equivalently any) lift $\hat{p}$ to $\mathcal{C}_K$, the scwolification $\Theta_K(\hat{p})$ is nonbacktracking. A $CG(\mathcal{Y})$–loop can be thought of as a path starting at any of its vertices. The loop is K–nonbacktracking if all these paths are K–nonbacktracking.

Lemma 4.15 *A $CG(\mathcal{Y})$–path $g_0\cdot p_{[\![\alpha_1]\!]}\cdot g_1\cdot p_{[\![\alpha_2]\!]}\cdots g_{k-1}\cdot p_{[\![\alpha_k]\!]}\cdot g_k$ is K–nonbacktracking if and only if*

(1) *for $i\in\{1,\dots,k-1\}$, if $[\![\alpha_{i+1}]\!]=[\![\bar{\alpha}_i]\!]$ then $g_i\notin E_{[\![\alpha_i]\!]}K_{t([\![\alpha_i]\!])}\subset G_{t([\![\alpha_i]\!])}$.*

Furthermore, a $CG(\mathcal{Y})$–loop with such a label is K–nonbacktracking if and only if (1) *and*

(2) *if $[\![\alpha_1]\!]=[\![\bar{\alpha}_k]\!]$, then $g_kg_0\notin E_{[\![\alpha_k]\!]}K_{t([\![\alpha_k]\!])}\subset G_{t([\![\alpha_k]\!])}$.*

The following result algebraically characterizes immersed loops of length 2 in $\operatorname{link}(\sigma)$.

Lemma 4.16 *Let p be a path in $\operatorname{link}(\breve{\sigma})$ which is a concatenation of two 1–cells, $\breve{\alpha}$ and $\breve{\beta}$, which lift respectively to 1–cells α and β in X. The following are equivalent:*

1. *There is a path $p' = \breve{\alpha}' \cdot \breve{\beta}'$ in $\operatorname{link}(\breve{\sigma})$ with $[\![\breve{\alpha}']\!] = [\![\breve{\alpha}]\!]$ and $[\![\breve{\beta}']\!] = [\![\breve{\beta}]\!]$ such that p' is an immersed loop.*
2. *There is a K–nonbacktracking $CG(\mathcal{Y})$–loop $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2$ that represents a conjugacy class in K.*

Moreover, when these conditions hold, the path p' can be chosen to be the scwolification of a lift of $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2$ (and conversely $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2$ is the $CG(\mathcal{Y})$–path which labels the unscwolification of p').

Proof Suppose that there is an immersed loop $p' = \breve{\alpha}' \cdot \breve{\beta}'$ as in (1). The idealization of p' is a $\mathcal{Z}$–path $q_{p'}$ of length 8, labeled by a $\mathcal{Y}$–path $a_1^+ \cdot a_2^- \cdot a_3^+ \cdot a_4^- \cdot b_1^+ \cdot b_2^- \cdot b_3^+ \cdot b_4^-$ as discussed above. Using Lemma 4.10, we can choose an unscwolification $\hat{q}_{p'}$ of $q_{p'}$ in $\mathcal{C}_K$ with label

$$p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]},$$

where g_1 is a group arrow. But the unscwolification $\hat{q}_{p'}$ has endpoints separated by a group arrow g_2, so there is a loop labeled $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2$ as desired. It is K–nonbacktracking since its scwolification is the path $q_{p'}$.

Conversely, suppose that there is a K–nonbacktracking $CG(\mathcal{Y})$–loop

$$p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2,$$

which represents an element of K. Then $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2$ lifts to a loop in $\mathcal{C}_K$. The scwolification of this loop gives a path p' as in condition (1). □

The following is elementary.

Lemma 4.17 *Let Q be a complex such that there are no loops of length 1 or 2 in its 1–skeleton. Any loop in Q of length 3 is nonbacktracking.*

The utility of Lemma 4.17 is that once we have found conditions to ensure that links in $K\backslash X$ have no loops of length 1 or 2 then loops of length 3 are automatically nonbacktracking.

Given Lemma 4.17, the following is proved in the same way as Lemma 4.16.

Lemma 4.18 *Suppose that* $\operatorname{link}(\breve{\sigma})$ *is simplicial, and suppose that* p *is a path in* $\operatorname{link}(\breve{\sigma})$ *which is a concatenation of three* 1*–cells,* $\breve{\alpha}$, $\breve{\beta}$ *and* $\breve{\gamma}$, *with lifts* α, β *and* γ *to* X. *The following are equivalent*:

1. *There is a path* $p' = \breve{\alpha}' \cdot \breve{\beta}' \cdot \breve{\gamma}'$ *in* $\operatorname{link}(\breve{\sigma})$ *such that* $[\![\breve{\alpha}']\!] = [\![\breve{\alpha}]\!]$, $[\![\breve{\beta}']\!] = [\![\breve{\beta}]\!]$ *and* $[\![\breve{\gamma}']\!] = [\![\breve{\gamma}]\!]$, *and* p' *is an immersed loop.*
2. *There is a* $CG(\mathcal{Y})$*–loop of the form* $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2 \cdot p_{[\![\gamma]\!]} \cdot g_3$ *that represents a conjugacy class in* K.

Moreover, when these conditions hold, the path p' *can be chosen to be the scwolification of a lift of* $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2 \cdot p_{[\![\gamma]\!]} \cdot g_3$ (*and conversely* $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2 \cdot p_{[\![\gamma]\!]} \cdot g_3$ *is the* $CG(\mathcal{Y})$*–path which labels the unscwolification of* p').

If X has dimension greater than 2, there are certainly some σ such that there are loops of length 3 in $\operatorname{link}(\sigma)$. This introduces some subtleties, which we discuss in the next subsection.

4.3.3 Loops of length 3 filled by 2–cells

We assume for the rest of the section that cubes of Z are embedded, so that objects of $\mathcal{Z}$ can be unambiguously described by chains of cubes of Z. The phenomenon we are concerned with in this section is illustrated by the following example.

Example 4.19 Let $\mathcal{Y}$ be a single 2–simplex, and consider the complex of groups $G(\mathcal{Y})$ such that $G_v \cong \mathbb{Z}$ for each vertex v, and all the other local groups are trivial. Let $x, y, z \in \pi_1(G(\mathcal{Y}))$ generate the three vertex groups. The universal cover X of $G(\mathcal{Y})$ is an infinite-valence "tree of triangles". Let $K = \langle\!\langle x^3, y^3, z^3, xyz \rangle\!\rangle$. Then $K \backslash X$ can be realized as a subset of the Euclidean plane, consisting of every other triangle of a tessellation by equilateral triangles. Moreover, if $\alpha\beta\gamma$ is the path in the 1–skeleton of $\mathcal{Y}$ labeling the boundary of $\mathcal{Y}$, there are paths in $K \backslash X$ projecting to $\alpha\beta\gamma$, but which are not filled by a 2–cell in $K \backslash X$. The issue here, as we will see, is that $xyz \in K$ is not an element of $K_x K_y K_z$, where $K_x = K \cap \langle x \rangle$, and so on.

Of course X is not a cube complex, but it can be realized as the link of a vertex of a cube complex, covering a complex of groups in which $G(\mathcal{Y})$ is embedded.

Definition 4.20 Let $\breve{\sigma}$ be a cube of Z, and let $\breve{\tau}$ be a 2–cell in $\operatorname{link}(\breve{\sigma})$. Then $\partial\breve{\tau}$ is a loop, composed of three oriented 1–cells, $\breve{\alpha} \cdot \breve{\beta} \cdot \breve{\gamma}$, which we may lift to 1–cells α, β and γ, forming the boundary of a lift τ of $\breve{\tau}$ in X. These 1–cells are associated to $CG(\mathcal{Y})$–paths $p_{[\![\alpha]\!]}$, $p_{[\![\beta]\!]}$ and $p_{[\![\gamma]\!]}$ as in Definition 4.9. Consider a $CG(\mathcal{Y})$–path of

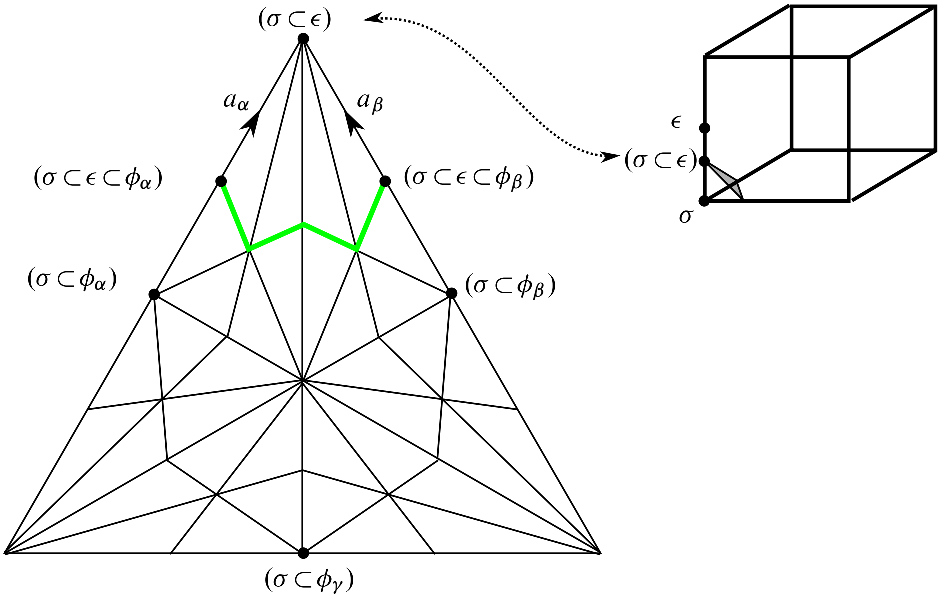

$(\sigma \subset \epsilon)$
a_α
a_β
$(\sigma \subset \epsilon \subset \phi_\alpha)$
$(\sigma \subset \epsilon \subset \phi_\beta)$
$(\sigma \subset \phi_\alpha)$
$(\sigma \subset \phi_\beta)$
$(\sigma \subset \phi_\gamma)$
ϵ
$(\sigma \subset \epsilon)$
σ

Figure 2: A part of $\mathcal{X}$ representing part of the link of σ, containing the idealization of the 1–cell ζ in green. Directions of most arrows have been omitted.

the form $q = p_{[\![\alpha]\!]} \cdot g \cdot p_{[\![\beta]\!]}$. Let $\hat{q}$ be a lift to $\mathcal{C}_K$. The realization of $\Theta_K(\hat{q})$ is a concatenation of two 1–cells $\breve{\alpha}' \cdot \breve{\beta}'$. We say that q *K–bounds a (τ, α)–corner* if there is a cube $\breve{\sigma}'$, a 2–cell $\breve{\tau}'$ in link$(\breve{\sigma}')$, and an $h \in G$ such that $\breve{\sigma}' = h\breve{\sigma}$, $\breve{\tau}' = h\breve{\tau}$, $\breve{\alpha}' = h\breve{\alpha}$ and $\breve{\beta}' = h\breve{\beta}$. If there is some (τ, α) for which the path q K–bounds a (τ, α)–corner, we may just say q *K–bounds a corner*.

If there exists a path q as above which K–bounds a (τ, α)–corner, there are X–cubes ϵ, ϕ_α, ϕ_β and ψ, all containing σ, such that $\epsilon \subset \phi_\alpha$, $\phi_\beta \subset \psi$, and

$$\dim(\psi) = \dim(\phi_\alpha) + 1 = \dim(\phi_\beta) + 1 = \dim(\epsilon) + 2 = \dim(\sigma) + 3.$$

There is a copy of the link of ϵ contained in the link of σ. The cubes ϵ, ϕ_α, ϕ_β and ψ determine an oriented 1–cell ζ in this copy of the link of ϵ. Its idealization is shown in Figure 2. The idealization of ζ begins at the object $(\sigma \subset \epsilon \subset \phi_\alpha)$ and ends at $(\sigma \subset \epsilon \subset \phi_\alpha)$. Let a_α be the arrow pointing from $(\sigma \subset \epsilon \subset \phi_\alpha)$ to $(\sigma \subset \epsilon)$, and let a_β be the arrow pointing from $(\sigma \subset \epsilon \subset \phi_\beta)$ to $(\sigma \subset \epsilon)$. These arrows project to arrows $[\![a_\alpha]\!]$ and $[\![a_\beta]\!]$ in $\mathcal{Y}$, and the path $p_{[\![\zeta]\!]}$ (defined as in Definition 4.9) travels from $i([\![a_\alpha]\!])$ to $i([\![a_\beta]\!])$.

Lemma 4.21 *The $CG(\mathcal{Y})$–loop*

$$[\![a_\alpha]\!]^+ \cdot p_{[\![\zeta]\!]} \cdot [\![a_\beta]\!]^-,$$

which is based at $[\![\sigma \subset \epsilon]\!]$, represents an element of $G_{[\![\sigma\subset\epsilon]\!]}$.

Proof All the chains which occur in this proof have the same minimal element σ, so we omit the prefix "$\sigma \subset$" from all chains until the end of the proof of the lemma. We therefore have a diagram in $\operatorname{link}(\sigma)$ in the scwol $\mathcal{X}$ as follows:

$$\begin{tikzcd}
 & & (\epsilon) & & \\
(\epsilon \subset \phi_\alpha) \arrow[urr, "a_\alpha"] & & & & (\epsilon \subset \phi_\beta) \arrow[ull, "a_\beta"'] \\
 & (\epsilon \subset \phi_\alpha \subset \psi) \arrow[ul, "a_1"] \arrow[uur, "b_1"] \arrow[dr, "a_2"] & & (\epsilon \subset \phi_\beta \subset \psi) \arrow[uul, "b_3"'] \arrow[ur, "a_4"'] \arrow[dl, "a_3"] & \\
 & & (\epsilon \subset \psi) \arrow[uuu, "b_2"] & &
\end{tikzcd}$$

We have the identities $a_\alpha a_1 = b_1 = b_2 a_2$ and $a_\beta a_4 = b_3 = b_2 a_3$ in the category $\mathcal{X}$. The path in the statement of the lemma is equal to

$$[\![a_\alpha]\!]^+ \cdot [\![a_1]\!]^+ \cdot [\![a_2]\!]^- \cdot g_{(\epsilon\subset\psi)} \cdot [\![a_3]\!]^+ \cdot [\![a_4]\!]^- \cdot [\![a_\beta]\!]^-,$$

where $g_{(\epsilon\subset\psi)}$ is the element of $G_{[\![(\epsilon\subset\psi)]\!]}$ chosen for the path p_ζ as in Definition 4.9.

Define the elements of $G_{[\![\epsilon]\!]}$

$$\begin{aligned}
h_1 &= z([\![a_\alpha]\!], [\![a_1]\!]) z([\![b_2]\!], [\![a_2]\!])^{-1}, \\
h_2 &= h_1 \psi_{[\![b_2]\!]}(g_{(\epsilon\subset\psi)}), \\
h_3 &= h_2 z([\![b_2]\!], [\![a_3]\!]) z([\![a_\beta]\!], [\![a_4]\!])^{-1},
\end{aligned}$$

where the $z([\![a]\!], [\![b]\!])$ are the twisting elements determined by the complex of groups structure on $G(\mathcal{Y})$.

We now have the sequence of elementary homotopies of $CG(\mathcal{Y})$–paths (all of which consist of applying the moves in Definition 2.1, and the rule of arrow composition in $CG(\mathcal{Y})$ from Definition 2.11),

$$[\![a_\alpha]\!]^+ \cdot p_{[\![\zeta]\!]} \cdot [\![a_\beta]\!]^- \simeq [\![a_\alpha]\!]^+ \cdot [\![a_1]\!]^+ \cdot [\![a_2]\!]^- \cdot g_{(\epsilon\subset\psi)} \cdot [\![a_3]\!]^+ \cdot [\![a_4]\!]^- \cdot [\![a_\beta]\!]^-$$

$$
\begin{aligned}
&\simeq z(\llbracket a_\alpha \rrbracket, \llbracket a_1 \rrbracket) \cdot \llbracket b_1 \rrbracket^+ \cdot \llbracket a_2 \rrbracket^- \cdot g_{(\epsilon \subset \psi)} \cdot \llbracket a_3 \rrbracket^+ \cdot \llbracket a_4 \rrbracket^- \cdot \llbracket a_\beta \rrbracket^- \\
&\simeq z(\llbracket a_\alpha \rrbracket, \llbracket a_1 \rrbracket) z(\llbracket b_2 \rrbracket, \llbracket a_2 \rrbracket)^{-1} \cdot \llbracket b_2 \rrbracket^+ \cdot \llbracket a_2 \rrbracket^+ \cdot \llbracket a_2 \rrbracket^- \\
&\qquad\qquad\qquad\qquad \cdot g_{(\epsilon \subset \psi)} \cdot \llbracket a_3 \rrbracket^+ \cdot \llbracket a_4 \rrbracket^- \cdot \llbracket a_\beta \rrbracket^- \\
&\simeq h_1 \cdot \llbracket b_2 \rrbracket^+ \cdot g_{(\epsilon \subset \psi)} \cdot \llbracket a_3 \rrbracket^+ \cdot \llbracket a_4 \rrbracket^- \cdot \llbracket a_\beta \rrbracket^- \\
&\simeq h_1 \cdot \psi_{\llbracket b_2 \rrbracket}(g_{(\epsilon \subset \psi)}) \cdot \llbracket b_2 \rrbracket^+ \cdot \llbracket a_3 \rrbracket^+ \cdot \llbracket a_4 \rrbracket^- \cdot \llbracket a_\beta \rrbracket^- \\
&\simeq h_2 z(\llbracket b_2 \rrbracket, \llbracket a_3 \rrbracket) \cdot \llbracket b_3 \rrbracket^+ \cdot \llbracket a_4 \rrbracket^- \cdot \llbracket a_\beta \rrbracket^- \\
&\simeq h_2 z(\llbracket b_2 \rrbracket, \llbracket a_3 \rrbracket) z(\llbracket a_\beta \rrbracket, \llbracket a_4 \rrbracket)^{-1} \cdot \llbracket a_\beta \rrbracket^+ \cdot \llbracket a_4 \rrbracket^+ \cdot \llbracket a_4 \rrbracket^- \cdot \llbracket a_\beta \rrbracket^- \\
&\simeq h_3 \cdot \llbracket a_\beta \rrbracket^+ \cdot \llbracket a_\beta \rrbracket^- \\
&\simeq h_3. \qquad \square
\end{aligned}
$$

Notation 4.22 The element of $G_{\llbracket \epsilon \rrbracket}$ represented by $\llbracket a_\alpha \rrbracket^+ \cdot p_{\llbracket \zeta \rrbracket} \cdot \llbracket a_\beta \rrbracket^-$ is denoted by $g_{\tau,\alpha}$.

Lemma 4.23 *A path $p_{\llbracket \alpha \rrbracket} \cdot g \cdot p_{\llbracket \beta \rrbracket}$ K–bounds a (τ,α)–corner if and only if there exists a $CG(\mathcal{Y})$–loop*

$$
\llbracket a_\alpha \rrbracket^+ \cdot g_1 \cdot p_{\llbracket \zeta \rrbracket} \cdot g_2 \cdot \llbracket a_\beta \rrbracket^- \cdot g^{-1} \tag{7}
$$

which represents an element of K.

Proof First suppose that there is a $CG(\mathcal{Y})$–loop of the form (7) representing an element of K.

Then the $CG(\mathcal{Y})$–paths

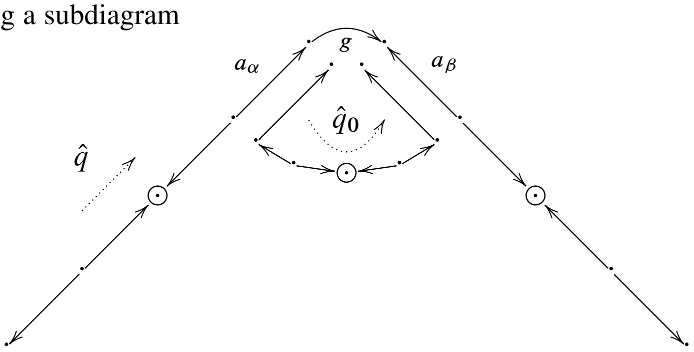
$$
p_{\llbracket \alpha \rrbracket} \cdot g \cdot p_{\llbracket \beta \rrbracket}, \quad p_{\llbracket \alpha \rrbracket} \cdot \llbracket a_\alpha \rrbracket^+ \cdot g_1 \cdot p_{\llbracket \zeta \rrbracket} \cdot g_2 \cdot \llbracket a_\beta \rrbracket^- \cdot g^{-1} \cdot g \cdot p_{\llbracket \beta \rrbracket}
$$

differ by an element of K. Thus they together form a loop which lifts to $\mathcal{C}_K$.

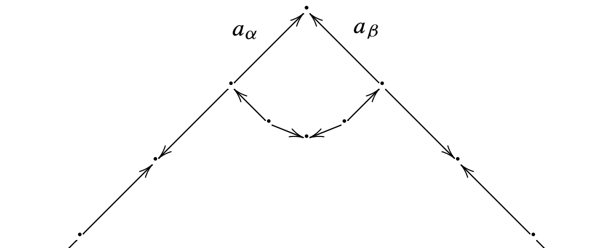
This second path is homotopic to a $CG(\mathcal{Y})$–path whose scwolification avoids the vertex $t(\llbracket \alpha \rrbracket)$ after $p_{\llbracket \alpha \rrbracket}$ but instead travels across the first three edges of $p_{\llbracket \alpha \rrbracket}$, traverses $p_{\llbracket \zeta \rrbracket}$, and then travels across the final three edges of $p_{\llbracket \beta \rrbracket}$. The homotopy lifts to $\mathcal{C}_K$, and the image in $\mathcal{Z}$ of this homotopy under the scwolification Θ_K shows that there is a 2–cell $\breve{\tau}'$ between the 1–cells $\breve{\alpha}'$ and $\breve{\beta}'$ whose idealizations are the scwolifications of the lifts of $p_{\llbracket \alpha \rrbracket}$ and $p_{\llbracket \beta \rrbracket}$, respectively. This shows that the path $p_{\llbracket \alpha \rrbracket} \cdot g \cdot p_{\llbracket \beta \rrbracket}$ K–bounds a (τ,α)–corner.

Conversely, suppose that the $CG(\mathcal{Y})$–path $q = p_{[\![\alpha]\!]} \cdot g \cdot p_{[\![\beta]\!]}$ K–bounds a (τ,α)–corner. Lift to a $\mathcal{C}_K$–path $\hat{q}$ and consider the scwolification $\Theta_K(\hat{q})$ in $\mathcal{Z}$. As in Definition 4.20, the realization of $\hat{q}$ is the concatenation of two 1–cells $\breve{\alpha}'$ and $\breve{\beta}'$ in $\operatorname{link}(\breve{\sigma}')$ for some cube $\breve{\sigma}'$ in the orbit of $\breve{\sigma}$. Moreover, there is a 2–cell $\breve{\tau}'$ with $\breve{\alpha}'$ and $\breve{\beta}'$ in the boundary of $\breve{\tau}'$ and an element h of G such that $\breve{\sigma}' = h\breve{\sigma}$, $\breve{\tau}' = h\breve{\tau}$, $\breve{\alpha}' = h\breve{\alpha}$ and $\breve{\beta}' = h\breve{\beta}$. Let v' be the vertex of $\operatorname{link}(\breve{\sigma}')$ where $\breve{\alpha}'$ and $\breve{\beta}'$ meet.

Consider the loop $q_0 = [\![a_\alpha]\!]^+ \cdot p_{[\![\zeta]\!]} \cdot [\![a_\beta]\!]^-$ as in Lemma 4.21. This represents an element of $G_{t([\![\alpha]\!])}$, and there is a lift $\hat{q}_0$ of q_0 to $\mathcal{C}_K$ such that $\Theta_K(\hat{q}_0)$ is a loop based at v' and traveling across the corner of $\breve{\tau}'$ from $\breve{\alpha}'$ to $\breve{\beta}'$. The paths q_0 and q have lifts to $\mathcal{C}_K$ forming a subdiagram



The circled dots represent either single objects or pairs of objects separated by a group arrow, depending on whether the paths $p_{[\![x]\!]}$ have length four or five for $x \in \{\alpha,\beta,\zeta\}$. The scwolification of this diagram in $\mathcal{Z}$ looks like



The edges which scwolify to a_α in $\hat{q}$ and $\hat{q}_0$ have sources connected by a group arrow labeled by some g_1. Similarly the edges which scwolify to a_β have sources connected by group arrow with some label g_2. We thus obtain a loop in $\mathcal{C}_K$ of the form (7). □

Given the criterion from Lemma 4.23, the following result is straightforward. Recall the definition of the element $g_{\tau,\alpha}$ from Notation 4.22.

Proposition 4.24 *Suppose that τ is a 2–cell in* $\operatorname{link}(\sigma)$ *and that the boundary of τ is $\alpha.\beta.\gamma$. For any $g \in G_{t[\![\alpha]\!]}$ the $CG(\mathcal{Y})$–path $p_{[\![\alpha]\!]} \cdot g \cdot p_{[\![\beta']\!]}$ K–bounds a (τ,α)–corner if and only if*

(1) $[\![\beta']\!] = [\![\beta]\!]$, *and*

(2) $g \in E_{[\![\alpha]\!]} g_{\tau,\alpha} E_{[\![\overleftharpoon{\beta}]\!]} . K_{t([\![\alpha]\!])}$

Proof Recall from Notation 4.22 that $g_{\tau,\alpha}$ is the element of $G_{t[\![\alpha]\!]}$ represented by the $CG(\mathcal{Y})$–loop

$$[\![a_\alpha]\!]^+ \cdot p_{[\![\gamma]\!]} \cdot [\![a_\beta]\!]^-.$$

Suppose that $p_{[\![\alpha]\!]} \cdot g \cdot p_{[\![\beta]\!]}$ K–bounds a (τ,α)–corner. Then consider the path

$$[\![a_\alpha]\!]^+ \cdot g_1 \cdot p_{[\![\gamma]\!]} \cdot g_2 \cdot [\![a_\beta]\!]^- \cdot g^{-1}$$

from Lemma 4.23 which represents an element of K.

We have homotopies

$$\begin{aligned}
[\![a_\alpha]\!]^+ \cdot g_1 \cdot p_{[\![\gamma]\!]} \cdot g_2 \cdot [\![a_\beta]\!]^- \cdot g^{-1} &\simeq \psi_{[\![a_\alpha]\!]}(g_1) \cdot [\![a_\alpha]\!]^+ \cdot p_{[\![\gamma]\!]} \cdot [\![a_\beta]\!]^- \cdot (\psi_{[\![a_\beta]\!]}(g_2) g^{-1}) \\
&\simeq \psi_{[\![a_\alpha]\!]}(g_1) \cdot g_{\tau,\alpha} \cdot (\psi_{[\![a_\beta]\!]}(g_2) g^{-1}) \\
&\simeq \psi_{[\![a_\alpha]\!]}(g_1) g_{\tau,\alpha} \psi_{[\![a_\beta]\!]}(g_2) g^{-1}.
\end{aligned}$$

Since $\psi_{[\![a_\alpha]\!]}(g_1) \in E_{[\![\alpha]\!]}$, $\psi_{[\![a_\beta]\!]}(g_2) \in E_{[\![\beta]\!]}$ and the whole expression above is an element of $K \cap G_{[\![v]\!]} = K_{t([\![\alpha]\!])}$,

$$g \in E_{[\![\alpha]\!]} g_{\tau,\alpha} E_{[\![\beta]\!]} K_{t([\![\alpha]\!])},$$

as required.

In order to prove the other direction, this computation may be performed in reverse. □

Lemma 4.25 *Suppose that* $\operatorname{link}(\check{\sigma})$ *is simplicial and contains 1–cells $\check{\alpha}$, $\check{\beta}$ and $\check{\gamma}$ which lift respectively to 1–cells α, β and γ in* $\operatorname{link}(\sigma)$ *in X. Let*

$$q = p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2 \cdot p_{[\![\gamma]\!]} \cdot g_3$$

be a $CG(\mathcal{Y})$–loop which represents an element of K. Suppose $\check{\eta} \subset$ $\operatorname{link}(\check{\sigma})$ *is the realization of the scwolification of some lift of q to $\mathcal{C}_K$.*

If any one of $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]}$, $p_{[\![\beta]\!]} \cdot g_2 \cdot p_{[\![\gamma]\!]}$ or $p_{[\![\gamma]\!]} \cdot g_3 \cdot p_{[\![\alpha]\!]}$ K–bounds a corner, then $\check{\eta}$ bounds a 2–cell in $\operatorname{link}(\check{\sigma})$.

Proof Note that since q represents an element of K, any lift to $\mathcal{C}_K$ is a loop, and so the realization $\check{\eta}$ is also a loop. Since $\operatorname{link}(\check{\sigma})$ is simplicial, this loop is embedded of length 3 in $\operatorname{link}(\check{\sigma})$ by Lemma 4.17.

Think of q as given by a cyclic word in the arrows of $CG(\mathcal{Y})$, and suppose that one of the three given subpaths of q K–bounds a corner. By relabeling and cyclically rotating we can assume it is the subpath $p = p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]}$, so there is some 2–cell $\check{\tau}$ in $\operatorname{link}(\check{\sigma})$ and lift τ to $\operatorname{link}(\sigma)$ and p K–bounds a (τ, α)–corner. It follows that some translate $\check{\tau}'$ of $\check{\tau}$ in $\operatorname{link}(\check{\sigma})$ has boundary given by a path $\check{\alpha}' \cdot \check{\beta}' \cdot \check{\gamma}'$, where $\alpha' \cdot \beta'$ are the first two 1–cells of the path $\check{\eta}$. If the third 1–cell of $\partial\check{\tau}'$ is not the third 1–cell of $\check{\eta}$, we obtain 1–cells in $\operatorname{link}(\check{\sigma})$ with the same endpoints, contradicting the assumption that $\operatorname{link}(\check{\sigma})$ is simplicial. So η bounds the 2–cell $\check{\tau}'$. □

Since there are finitely many $\operatorname{Stab}(\check{\sigma})$–orbits of 2–cells in $\operatorname{link}(\check{\sigma})$, we obtain the following.

Proposition 4.26 *Suppose that* $\operatorname{link}(\check{\sigma})$ *is simplicial. There are finitely many* 2*–cells* $\check{\tau}_i$ *in* $\operatorname{link}(\check{\sigma})$ (*with boundary* $\check{\alpha}_i \cdot \check{\beta}_i \cdot \check{\gamma}_i$, *and lifts* α_i, β_i *and* γ_i *to* $\operatorname{link}(\sigma)$) *such that the following holds*:

Every loop of length 3 *in* $\operatorname{link}(\check{\sigma})$ *is filled by a* 2*–cell if and only if, for every* $CG(\mathcal{Y})$*–path*

$$(*) \qquad p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\beta]\!]} \cdot g_2 \cdot p_{[\![\gamma]\!]} \cdot g_3$$

which represents an element of K, *there exists an* i *such that*

(1) $[\![\alpha]\!] = [\![\alpha_i]\!]$, $[\![\beta]\!] = [\![\beta_i]\!]$ *and* $[\![\gamma]\!] = [\![\gamma_i]\!]$,

(2) $g_1 \in E_{[\![\alpha_i]\!]} g_{\tau_i,\alpha_i} E_{[\![\overleftharpoon{\beta}_i]\!]} K_{t([\![\alpha_i]\!])}$,

(3) $g_2 \in E_{[\![\beta_i]\!]} g_{\tau_i,\beta_i} E_{[\![\overleftharpoon{\gamma}_i]\!]} K_{t([\![\beta_i]\!])}$, *and*

(4) $g_3 \in E_{[\![\gamma_i]\!]} g_{\tau_i,\gamma_i} E_{[\![\overleftharpoon{\alpha}_i]\!]} K_{t([\![\gamma_i]\!])}$.

Proof Choose the 2–cells $\check{\tau}_i$ to be representatives of the $\operatorname{Stab}(\check{\sigma})$–orbits of 2–cells (together with a fixed vertex to label the boundary — so that a single orbit may appear up to three times in the list).

Suppose first that the condition about paths of the form $(*)$ representing elements of K is satisfied, and suppose that p is a loop of length 3 in $\operatorname{link}(\check{\sigma})$ which is labeled by 1–cells $\check{\alpha}'$, $\check{\beta}'$ and $\check{\gamma}'$, in order. By Lemma 4.18 there exists a $CG(\mathcal{Y})$–path λ of the form $(*)$ which is the label of an unscwolification of p. Because of our hypothesis, there exists an i such that conditions (1)–(4) are satisfied. By Proposition 4.24, the $CG(\mathcal{Y})$–path λ K–bounds a corner at each of its three corners, and so by Lemma 4.25 the path p bounds a 2–cell, as required.

Conversely, suppose that every loop of length 3 in $\operatorname{link}(\breve{\sigma})$ bounds a 2–cell, and consider a $CG(\mathcal{Y})$–path λ of the form $(*)$ which represents an element of K. By Lemma 4.18 the scwolification of a lift of λ is (the idealization of) an immersed path of length 3. This immersed path must then bound a 2–cell $\check{\tau}$. Suppose that $\check{\tau}_i$ is the representative in the $\operatorname{Stab}(\breve{\sigma})$–orbit of the 2–cell $\check{\tau}$, so condition (1) is satisfied. According to Lemma 4.23, applied to all three corners of this 2–cell, the path λ satisfies conditions (2)–(4). □

To summarize, given Lemma 4.8, Lemmas 4.12, 4.16 and 4.18 and Proposition 4.26 give descriptions of various types of $CG(\mathcal{Y})$–paths such that the cube complex $Z = K\backslash X$ is nonpositively curved if and only if no such path lifts to $\mathcal{C}_K$.

5 Algebraic translation

In this section, we continue to work in the context of a group G acting cocompactly on a CAT(0) cube complex X. The induced action on the associated scwol $\mathcal{X}$ has quotient scwol $\mathcal{Y}$, the underlying scwol for a complex of groups structure $G(\mathcal{Y})$ on G. We let $\mathcal{Q}(G)$ be the set of cube stabilizers for $G \curvearrowright X$; equivalently $\mathcal{Q}(G)$ is the set of conjugates of the local groups for the complex of groups $G(\mathcal{Y})$.

We translate the conditions from the previous section into algebraic statements about elements of G and of $\mathcal{Q}(G)$, with an eye toward finding conditions on $K \lhd G$ implying that $K\backslash X$ is nonpositively curved. In Section 6 we use hyperbolic Dehn filling to find K which satisfy the conditions, under certain hyperbolicity assumptions on G and $\mathcal{Q}(G)$.

We fix a basepoint v_0 for $\mathcal{Y}$ and an isomorphism $\pi_1(CG(\mathcal{Y}), v_0) \cong G$ as in Section 2. The scwolification functor

$$\Theta\colon \widetilde{CG(\mathcal{Y})} \to \mathcal{X}$$

is G–equivariant. Recall also that the objects of $\widetilde{CG(\mathcal{Y})}$ are homotopy classes of paths starting at v_0.

Fix also a maximal (undirected) tree T in $\mathcal{Y}$. For each object v of $\mathcal{Y}$ which represents an orbit of cubes in X, let c_v be the unique $\mathcal{Y}$–path in T from v_0 to v. By using scwol arrows, we also consider c_v to be a $CG(\mathcal{Y})$–path in the natural way. For an object v of $\mathcal{Y}$ which represents a chain of cubes of length longer than 1, we define a $\mathcal{Y}$–path c_v

from v_0 to v as follows: if v is represented by $(\sigma_1 \subset \sigma_2 \subset \cdots \subset \sigma_k)$ (a nested chain of cubes in X) then define c_v to be the concatenation of $c_{[\![\sigma_1]\!]}$ with the path consisting of the arrows $(\sigma_1 \subset \cdots \subset \sigma_i) \to (\sigma_1 \subset \cdots \subset \sigma_{i+1})$, for $i = 1, 2, \dots, k-1$.

We use the paths c_v to define a map from (homotopy classes rel endpoints of) $CG(\mathcal{Y})$–paths to (homotopy classes of) $CG(\mathcal{Y})$–loops based at v_0 by

$$p \mapsto c_{i(p)} \cdot p \cdot \overleftarrow{c}_{t(p)}.$$

(Here and below, $\overleftarrow{c}$ denotes the reverse of the $CG(\mathcal{Y})$–path c.)

Given a path p, let $\ell_p = [c_{i(p)} \cdot p \cdot \overleftarrow{c}_{t(p)}] \in \pi_1(CG(\mathcal{Y}), v_0)$.

The following results are straightforward:

Lemma 5.1 *For any $CG(\mathcal{Y})$–paths p and p' such that $t(p) = i(p')$,*

$$\ell_{\overleftarrow{p}} = \ell_p^{-1}, \quad \ell_{p\cdot p'} = \ell_p \ell_{p'}.$$

Lemma 5.2 *Suppose that p is a $CG(\mathcal{Y})$–path starting at v_0. Let $[p]$ be the equivalence class of p in $\widetilde{CG(\mathcal{Y})}$, and let $x = \Theta([p])$. Then*

$$\mathrm{Stab}_G(x) = \{[p \cdot g \cdot \overleftarrow{p}] \mid g \in G_{[\![x]\!]}\}.$$

Definition 5.3 Given an object v of $\mathcal{Y}$, define

$$Q_v = \{[c_v \cdot g \cdot \overleftarrow{c}_v] \mid g \in G_v\}.$$

Definition 5.3 gives an explicit identification of the local groups of the complex of groups $G(\mathcal{Y})$ with finitely many elements of $\mathcal{Q}(G)$.

5.1 Algebraic formulation of the link conditions

Suppose that $K \trianglelefteq G$. In order for $Z = K \backslash X$ to be nonpositively curved, there are five conditions that need to be ensured on links in Z. Roughly speaking, they are

(1) no loop of length 1,

(2) no loop of length 2 consisting of 1–cells in different G–orbits,

(3) no loop of length 2 consisting of 1–cells in the same G–orbit,

(4) no loop of length 3 whose image in $\mathcal{Y}$ does not bound a 2–cell, and

(5) no loop of length 3 which does not bound a 2–cell but whose image in $\mathcal{Y}$ does bound a 2–cell.

More precisely, the "image in $\mathcal{Y}$" means the image in $\mathcal{Y}$ of the idealization. And we say this image p "bounds a 2–cell" if there is an unscwolification $\hat{p}$ and a lift $\tilde{p}$ of $\hat{p}$ to $\widetilde{CG(\mathcal{Y})}$ such that the realization of the scwolification of $\tilde{p}$ bounds a 2–cell in some link of a cube in X.

If $K\backslash X$ is a simply connected cube complex and we ensure each of these conditions, then Lemmas 4.7, 4.8 and 4.17 imply that $K\backslash X$ is CAT(0).

In this subsection, we formulate five results which give algebraic conditions to enforce each of these five conditions in turn. These results follow quickly from the results in Section 4 using the translation from the beginning of this section. In each case, since G acts cocompactly on a CAT(0) cube complex, there are finitely many G–orbits of links and in each link finitely many G–orbits of each of the five kinds of paths in the above list, and we can rule out each orbit behaving badly in $K\backslash X$ in turn.

Assumption 5.4 The group G acts cocompactly on the CAT(0) cube complex X, and $\mathcal{Q}(G)$ is the collection of cell stabilizers of the action.

Terminology 5.5 Under Assumption 5.4, a normal subgroup $K \trianglelefteq G$ is *cocubical* if $K\backslash X$ is a cube complex.

The following is a straightforward translation of Lemma 4.12. We spell out the proof since we use similar techniques for other more complicated results later in the section.

Theorem 5.6 *Under Assumption 5.4, there exists a finite set $F_1 \subset \mathcal{Q}(G) \times G$ such that for each $(Q, p) \in F_1$ we have $p \notin Q$ and such that, if*

(i) *$K \trianglelefteq G$ is cocubical, and*

(ii) *for each $(Q, p) \in F_1$ we have $p \notin Q.K$,*

then no link in $K\backslash X$ contains a loop of length 1.

Proof Up to the action of G, there are finitely many pairs $(\tilde{\sigma}, \tilde{\alpha})$, where $\tilde{\sigma}$ is a cube of X and $\tilde{\alpha}$ is a 1–cell in $\operatorname{link}(\tilde{\sigma})$ whose endpoints are identified by some element of G. For each such pair we will give a pair (Q, p) as in the statement of the theorem.

For such a pair, let (σ, α) be the image in $K\backslash X$. Since K is assumed to act cocubically, α is embedded in $\operatorname{link}(\sigma)$, except that its endpoints may have been identified, making it a loop. According to Lemma 4.12, α is a loop if and only if there is a $CG(\mathcal{Y})$–loop of the form $p_{\llbracket\alpha\rrbracket}.g$ that represents a conjugacy class in K. In particular, this condition only depends on the orbit $\llbracket\alpha\rrbracket$ and not on α itself. We associate to α the element $p = \ell_{p_{\llbracket\alpha\rrbracket}}$ and the subgroup $Q = Q_{t(\llbracket\alpha\rrbracket)}$, as described in the preamble to this section.

Since X itself is a CAT(0) cube complex, the 1–cell $\tilde{\alpha}$ is not a loop. Applying Lemma 4.12 in the case $K = \{1\}$, we see that $p \notin Q$. On the other hand, to say that $p \notin Q.K$ is the same as saying there is no $CG(\mathcal{Y})$–loop of the form $p_{\llbracket\alpha\rrbracket}.g$ which represents an element of K (since in such a $CG(\mathcal{Y})$–loop the element g must be in the local group $G_{t(\llbracket\alpha\rrbracket)}$). □

The next result is an application of Lemma 4.16 to paths of length 2 consisting of 1–cells in different G–orbits (since then the K–nonbacktracking condition is vacuous).

Theorem 5.7 *Under Assumption 5.4 there exists a finite set $F_2 \subset \mathcal{Q}(G)^2 \times G^2$ such that for each $(Q_1, Q_2, p_1, p_2) \in F_2$,*

$$1 \notin p_1 Q_1 p_2 Q_2,$$

and such that, if

(i) *$K \trianglelefteq G$ is cocubical, and*

(ii) *for each $(Q_1, Q_2, p_1, p_2) \in F_2$,*

$$K \cap p_1 Q_2 p_2 Q_2 = \varnothing,$$

then every loop of length 2 in a link in $K\backslash X$ consists of 1–cells in the same G–orbit.

Proof The proof is similar to the proof of Theorem 5.6 above. Lemma 4.16 implies that it is enough to verify that no link in a cube of $K\backslash X$ contains a pair of 1–cells α and β in distinct G–orbits $\llbracket\alpha\rrbracket$ and $\llbracket\beta\rrbracket$ such that there is a $CG(\mathcal{Y})$–loop $p_{\llbracket\alpha\rrbracket} \cdot g_1 \cdot p_{\llbracket\beta\rrbracket} \cdot g_2$ representing an element of K.

There are finitely many pairs of such orbits, and to each such pair we can associate the elements $p_1 = \ell_{p_{\llbracket\alpha\rrbracket}}$, $p_2 = \ell_{p_{\llbracket\beta\rrbracket}}$, $Q_1 = Q_{t(\llbracket\alpha\rrbracket)}$ and $Q_2 = Q_{t(\llbracket\beta\rrbracket)}$.

Since X is a CAT(0) cube complex, there are no nonbacktracking loops of length 2 in any links in X, so applying Lemma 4.16 with $K = \{1\}$ we see that $1 \notin p_1 Q_1 p_2 Q_2$. The result now follows from Lemma 4.16 with our choice of K. □

For paths of length 2 consisting of 1–cells in the same G–orbit, the condition is slightly more complicated, as K–backtracking paths are possible.

Theorem 5.8 *Under Assumption 5.4 there exists a finite set $F_3 \subset \mathcal{Q}(G)^2 \times G^2$ such that, for each $(Q_1, Q_2, p_1, p_2) \in F_3$,*

$$1 \notin p_1(Q_1 - Q_2^{p_2})p_2(Q_2 - Q_1^{p_1}), \tag{8}$$

and such that, if

(i) *$K \trianglelefteq G$ is cocubical,*

(ii) *no link in $K\backslash X$ contains a loop of length 1, and*

(iii) *for every $(Q_1, Q_2, p_1, p_2) \in F_3$,*

$$K \cap p_1\big(Q_1 - (Q_2^{p_2}(K \cap Q_1))\big)p_2\big(Q_2 - (Q_1^{p_1}(K \cap Q_2))\big) = \varnothing, \tag{9}$$

then no link in $K\backslash X$ contains an immersed loop of length 2 consisting of 1–cells in the same orbit.

Proof Because of assumptions (i) and (ii) we only need to be concerned with the following situation: there is some cube $\tilde{\sigma}$ of X and some 1–cell $\tilde{\alpha}$ in its link such that

(1) there is some $g \in G$ such that g fixes $t(\tilde{\alpha})$ but not $\tilde{\alpha}$;

(2) there is some $h \in G$ such that $h(i(\tilde{\alpha})) = i(g\tilde{\alpha})$ but $h^{-1}g\tilde{\alpha} \neq \tilde{\alpha}$.

There are finitely many orbits of pairs $(\tilde{\sigma}, \tilde{\alpha})$ of this type. For each orbit we pick a representative, and describe an element of $\mathcal{Q}(G)^2 \times G^2$ as in the theorem. If (9) is satisfied for this element, then $K\backslash X$ will contain no immersed loop of length 2 consisting of 1–cells in the orbit of $\tilde{\alpha}$.

We apply Lemma 4.16 to a path of length 2 of the form $\alpha.\alpha'$ where α is the image of $\tilde{\alpha}$ in $K\backslash X$ and α' is the (oppositely oriented) image of a translate of $\tilde{\alpha}$ by an element of the stabilizer of $\tilde{\sigma}$. Any immersed loop of the type we are trying to rule out gives rise to a K–nonbacktracking $CG(\mathcal{Y})$–loop $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\bar{\alpha}]\!]} \cdot g_2$ representing a conjugacy class in K. We let $p_1 = l_{p_{[\![\alpha]\!]}}$, $p_2 = l_{p_{[\![\bar{\alpha}]\!]}}$, $Q_1 = Q_{t([\![\alpha]\!])}$ and $Q_2 = Q_{i([\![\alpha]\!])}$. Using Lemma 4.15, the loop $p_{[\![\alpha]\!]} \cdot g_1 \cdot p_{[\![\bar{\alpha}]\!]} \cdot g_2$ is K–nonbacktracking if and only if $g_1 \notin E_{[\![\alpha]\!]}K_{t([\![\alpha]\!])}$ and $g_2 \notin E_{[\![\bar{\alpha}]\!]}K_{i([\![\alpha]\!])}$. The subgroup of Q_1 corresponding to $E_{[\![\bar{\alpha}]\!]}$ is equal to $Q_1 \cap Q_2^{p_2}$, and the subgroup of Q_2 corresponding to $E_{[\![\bar{\alpha}]\!]}$ is $Q_2 \cap Q_1^{p_1}$. Thus an element $p_1q_1p_2q_2$ of $p_1Q_1p_2Q_2$ comes from a K–nonbacktracking $CG(\mathcal{Y})$–loop if and only if $q_1 \notin Q_2^{p_2}(K \cap Q_1)$ and $q_2 \notin Q_1^{p_1}(K \cap Q_2)$. Applying Lemmas 4.15 and 4.16 in the case $K = \{1\}$ and $K\backslash X = X$ is CAT(0), we see that our tuple satisfies (8). For an arbitrary K we see that when (9) is satisfied, there is no immersed loop of length 2 in a link in $K\backslash X$ consisting of images of translates of $\tilde{\alpha}$. □

In order to apply Lemma 4.17, in each of the following two results we make the extra assumption that K is such that no link in $K\backslash X$ contains a loop of length 1 or 2. The following result is a translation of Lemma 4.18.

Theorem 5.9 *Under Assumption 5.4 there exists a finite set $F_4 \subset \mathcal{Q}(G)^3 \times G^3$ such that, for each $(Q_1, Q_2, Q_3, p_1, p_2, p_3) \in F_4$,*

$$1 \notin p_1 Q_1 p_2 Q_2 p_3 Q_3$$

and such that if

(i) *$K \trianglelefteq G$ is cocubical,*

(ii) *no link in $K\backslash X$ contains a loop of length 1 or 2, and*

(iii) *for all $(Q_1, Q_2, Q_3, p_1, p_2, p_3) \in F_4$,*

$$K \cap p_1 Q_1 p_2 Q_2 p_3 Q_3 = \varnothing,$$

then every loop of length 3 in a link of $K\backslash X$ has image in $\mathcal{Y}$ which bounds a 2–cell.

Proof Condition (ii) and Lemma 4.17 imply that it suffices to consider immersed loops of length 3 in links in $K\backslash X$. For each choice of triple of G–orbits $[\![\alpha]\!], [\![\beta]\!], [\![\gamma]\!]$ of 1–cells in links in X whose image in $\mathcal{Y}$ forms a loop, but whose image does not bound a 2–cell in $\mathcal{Y}$ (in the sense described at the beginning of this subsection), we proceed as follows. We associate the elements $p_1 = l_{p_{[\![\alpha]\!]}}$, $p_2 = l_{p_{[\![\beta]\!]}}$, $p_3 = l_{p_{[\![\gamma]\!]}}$, $Q_1 = Q_{t([\![\alpha]\!])}$, $Q_2 = Q_{t([\![\beta]\!])}$ and $Q_3 = Q_{t([\![\gamma]\!])}$.

Since X is a CAT(0) cube complex, we can apply Lemma 4.18 to see that

$$1 \notin p_1 Q_1 p_2 Q_2 p_3 Q_3.$$

Now let $K \trianglelefteq G$ be cocubical, and satisfy conditions (i)–(iii) from the statement. Condition (iii) implies that condition (2) from Lemma 4.18 does not hold, and by that lemma there is no immersed loop of length 3 in a link in $K\backslash X$ whose image in $\mathcal{Y}$ is $[\![\alpha]\!], [\![\beta]\!], [\![\gamma]\!]$.

Since there are finitely many such triples $[\![\alpha]\!], [\![\beta]\!], [\![\gamma]\!]$, the theorem follows. □

Finally, we deal with loops of length 3 in links in $K\backslash X$ whose image in $\mathcal{Y}$ does bound a 2–cell.

Terminology 5.10 Suppose that

$$A = (Q_1, Q_2, Q_3, p_1, p_2, p_3, h_1, h_2, h_3) \in \mathcal{Q}(G)^3 \times G^6.$$

With indices read mod 3, let

$$A_i^- = Q_{i-1}^{p_{i-1}} \cap Q_i, \quad A_i^+ = Q_i \cap Q_{i+1}^{p_{i+1}}.$$

Furthermore, let

$$B_i = A_i^- h_i A_i^+.$$

Using this terminology, we have the following translation of Proposition 4.26.

Theorem 5.11 *Under Assumption 5.4 there exists a finite set $F_5 \subseteq \mathcal{Q}(G)^3 \times G^6$ such that, for each $A = (Q_1, Q_2, Q_3, p_1, p_2, p_3, h_1, h_2, h_3)$,*

$$1 \notin p_1(Q_1 - B_1) p_2(Q_2 - B_2) p_3(Q_3 - B_3)$$

and such that, if

(i) *$K \trianglelefteq G$ is cocubical,*

(ii) *no link in $K \backslash X$ contains a loop of length 1 or 2, and*

(iii) *for all $(Q_1, Q_2, Q_3, p_1, p_2, p_3, h_1, h_2, h_3) \in F_5$,*

$$K \cap p_1(Q_1 - B_1(K \cap Q_1)) p_2(Q_2 - B_2(K \cap Q_2)) p_3(Q_3 - B_3(K \cap Q_3)) = \varnothing,$$

then no link in $K \backslash X$ contains a loop of length 3 which does not bound a 2–cell but whose image in $\mathcal{Y}$ bounds a 2–cell.

Proof For each choice of triple of orbits $[\![\alpha]\!], [\![\beta]\!], [\![\gamma]\!]$ whose image in $\mathcal{Y}$ bounds a 2–cell (in the sense described at the beginning of this subsection), we proceed as follows. Without loss of generality we choose representatives α, β and γ of these orbits such that there is a 2–cell τ with boundary $\alpha \cdot \beta \cdot \gamma$. We associate the elements $p_1 = \ell_{p_{[\![\alpha]\!]}}$, $p_2 = \ell_{p_{[\![\beta]\!]}}$, $p_3 = \ell_{p_{[\![\gamma]\!]}}$, $Q_1 = Q_{t([\![\alpha]\!])}$, $Q_2 = Q_{t([\![\beta]\!])}$, $Q_3 = Q_{t([\![\gamma]\!])}$, $h_1 = [c_{t([\![\alpha]\!])} \cdot g_{\tau,\alpha} \cdot \overleftharpoon{c}_{t([\![\alpha]\!])}]$, $h_2 = [c_{t([\![\beta]\!])} \cdot g_{\tau,\beta} \cdot \overleftharpoon{c}_{t([\![\beta]\!])}]$ and $h_3 = [c_{t([\![\gamma]\!])} \cdot g_{\tau,\gamma} \cdot \overleftharpoon{c}_{t([\![\gamma]\!])}]$.

Once again, since X is a CAT(0) cube complex, we can apply Proposition 4.26 to see that

$$1 \notin p_1(Q_1 - B_1) p_2(Q_2 - B_2) p_3(Q_3 - B_3).$$

When conditions (2)–(4) from the statement of Proposition 4.26 are translated into statements about the group G, we get exactly $g_1 \in B_1(K \cap Q_1)$, etc, which gives the statement in the conclusion of the result.

Since there are finitely many such triples $[\![\alpha]\!], [\![\beta]\!], [\![\gamma]\!]$, the theorem follows. □

6 Dehn filling

In this section we prove some results about group-theoretic Dehn filling. Theorem 6.5 gives a "weak separability" of certain multicosets, and generalizations of multicosets, and is used to find subgroups K which satisfy the conditions from Theorems 5.6–5.11. Theorem 6.5 may be of independent interest, and we expect it to have applications beyond the scope of this paper. The second main result of this section is Theorem 6.9, from which Theorem F from the introduction follows quickly by induction.

6.1 Dehn fillings

Let $(G, \mathcal{P})$ be a group pair, and let $\mathcal{N} = \{N_P \lhd P \mid P \in \mathcal{P}\}$ be a choice of normal subgroups of the peripheral groups. The collection $\mathcal{N}$ determines a *(Dehn) filling* $(\overline{G}, \overline{\mathcal{P}})$ *of* $(G, \mathcal{P})$, where $\overline{G} = G/K$ for K the normal closure of $\bigcup \mathcal{N}$, and $\overline{\mathcal{P}}$ is equal to the collection of images of elements of $\mathcal{P}$ in $\overline{G}$. The elements of $\mathcal{N}$ are called *filling kernels*. We sometimes write such a filling as

$$\pi\colon (G, \mathcal{P}) \to (\overline{G}, \overline{\mathcal{P}}),$$

omitting mention of the particular filling kernels.

If $N_P \dot{<} P$ (ie N_P is finite index in P) for all $P \in \mathcal{P}$, we say that the filling is *peripherally finite*. If $H < G$ and, for all $g \in G$, $|H \cap P^g| = \infty$ implies $N_P^g \subseteq H$, then the filling is an *H–filling*. If $\mathcal{H}$ is a family of subgroups, the filling is an *$\mathcal{H}$–filling* whenever it is an H–filling for every $H \in \mathcal{H}$.

A property P holds *for all sufficiently long fillings* of $(G, \mathcal{P})$ if there is a finite set $S \subseteq \bigcup \mathcal{P} - \{1\}$ such that P holds whenever $\left(\bigcup \mathcal{N}\right) \cap S = \varnothing$. It is frequently useful to restrict attention to specific types of fillings (peripherally finite, H–fillings, etc). If A is a property of fillings we say that P holds for *all sufficiently long* A*–fillings* if, for all sufficiently long fillings, either P holds or A does not hold.

6.2 Relatively hyperbolic group pairs

We refer the reader to [17] for background on relatively hyperbolic groups. In that paper, given a group pair $(G, \mathcal{P})$ (consisting of finitely generated groups) a space called the *cusped space* is built, which is δ–hyperbolic (for some δ) if and only if $(G, \mathcal{P})$ is relatively hyperbolic. The cusped space is built by attaching *combinatorial horoballs* to a Cayley graph for G. Each combinatorial horoball H has vertex set $tP \times \mathbb{Z}_{\geq 0}$ for some coset tP of some $P \in \mathcal{P}$, and is hyperbolic. A vertex (g, n) of such a horoball is said to have *depth* n. The depth 0 vertices of the cusped space are exactly the vertices of the Cayley graph; if two vertices are connected by an edge, then their depths differ by at most one. See [17, Section 3] for more information about the construction and geometry of the cusped space. The following result is essentially contained in [7, Theorem 7.11].

Theorem 6.1 *Suppose that G is a hyperbolic group and that $\mathcal{P}$ is a finite collection of subgroups of G. Then $(G, \mathcal{P})$ is relatively hyperbolic if and only if $\mathcal{P}$ is an almost malnormal family of quasiconvex subgroups.*

Recall that $\mathcal{P} = \{P_1, \dots, P_n\}$ is *almost malnormal* if whenever $P_i \cap P_j^g$ is infinite, we have $i = j$ and $g \in P_i$.

We can use the notion of height (see Definition 3.28) to measure how far away a family of subgroups is from being almost malnormal.

We now define the *induced peripheral structure* on G associated to a finite collection of quasiconvex subgroups of a hyperbolic group, in analogy with the construction from [2, Section 3.1].

Definition 6.2 Suppose that G is a hyperbolic group and $\mathcal{H}$ is a finite collection of quasiconvex subgroups of G. The *peripheral structure on G induced by $\mathcal{H}$* is obtained as follows:

Start by taking the collection of minimal infinite subgroups of the form

$$H_1 \cap H_2^{g_2} \cap \cdots \cap H_k^{g_k}$$

where the H_i are in $\mathcal{H}$ and the cosets $\{H_1, g_2H_2, \dots, g_kH_k\}$ are all distinct. Replace each element in this collection by its commensurator in G, and then choose one from each G–conjugacy class. The resulting collection $\mathcal{P}$ is the induced peripheral structure.

If $H \in \mathcal{H}$ then the *induced peripheral structure on H with respect to $\mathcal{H}$* is a choice of H–conjugacy representatives of intersections with H of G–conjugates of elements of $\mathcal{P}$.

We remark that the fact that there is a bound on the number k of g_iH_i as above follows from Proposition 3.29.

To state the next lemma we need a definition from [2]:

Definition 6.3 Let $H < G$ and suppose $(H, \mathcal{D})$ and $(G, \mathcal{P})$ are relatively hyperbolic. Suppose furthermore that every $D \in \mathcal{D}$ is conjugate into some $P \in \mathcal{P}$. Then — see [2, Lemma 3.1] — there is an induced H–equivariant map from the cusped space of $(H, \mathcal{D})$ to the cusped space of $(G, \mathcal{P})$. The subgroup $(H, \mathcal{D})$ is *relatively quasiconvex* if this induced map has quasiconvex image.

In [28, Appendix A] it is proved that this is the same notion as the various notions of relative quasiconvexity discussed by Hruska in [23].

The following can be proved in the same way as [2, Proposition 3.12].

Lemma 6.4 *Suppose that G is hyperbolic and $\mathcal{H}$ is a finite collection of quasiconvex subgroups of G.*

1. *The induced peripheral structure $\mathcal{P}$ is a finite collection of groups. The pair $(G, \mathcal{P})$ is relatively hyperbolic.*
2. *If $H \in \mathcal{H}$ then the induced peripheral structure $\mathcal{D}$ of H with respect to $\mathcal{H}$ is finite. The pair $(H, \mathcal{D})$ is relatively hyperbolic.*
3. *For any $H \in \mathcal{H}$, the pair $(H, \mathcal{D})$ is full relatively quasiconvex in $(G, \mathcal{P})$.*

A subgroup H is *full* if whenever P is a parabolic subgroup such that $H \cap P$ is infinite we have $H \cap P \dot{<} P$.

6.3 The appropriate metacondition

The goal of this subsection is to prove Theorem 6.5 below. The special case that $n = 1$ and $S_1 = \varnothing$ is [2, Proposition 4.5], which is about keeping elements out of full quasiconvex subgroups when performing long Dehn fillings. Here we generalize to multicosets of full quasiconvex subgroups, possibly with some elements deleted. Although the present result is more general, our proof is simpler, using the more appealing "Greendlinger Lemma"-type Theorem 6.7 below in place of the somewhat technical [2, Lemmas 4.1 and 4.2].[6]

Theorem 6.5 *Let $(G, \mathcal{P})$ be relatively hyperbolic, and let $\mathcal{Q}$ be a collection of full relatively quasiconvex subgroups. For $1 \le i \le n$, let $p_i \in G$, $Q_i \in \mathcal{Q}$ and $S_i \subseteq Q_i$ be chosen to satisfy*

$$1 \notin p_1(Q_1 - S_1) \cdots p_n(Q_n - S_n). \tag{10}$$

[6] Using such a Greendlinger Lemma in place of the results of [2] was suggested to us by Alessandro Sisto while we were collaborating on [19].

Then, for sufficiently long $\mathcal{Q}$–fillings $G \to G/K$, the kernel K contains no element of the form

$$p_1 t_1 \cdots p_n t_n \tag{11}$$

where $t_i \in Q_i - ((K \cap Q_i)S_i)$.

The five conditions in the conclusions of Theorems 5.6–5.11 each fall into the scheme of the conditions in Theorem 6.5. Therefore, we may apply Theorem 6.5 to obtain the following result. We remark that the following result is stated in the generality of relatively hyperbolic groups acting cocompactly on cube complexes with full relatively quasiconvex subgroups. This is greater generality than is strictly required for the proof of Theorem F. However, we believe that this extra generality will be of use in future work, and should be of independent interest.

Corollary 6.6 *Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and that G acts cocompactly on the CAT(0) cube complex X. Suppose every parabolic element of G fixes some point of X, and that cell stabilizers are full relatively quasiconvex. Let $\sigma_1, \dots, \sigma_k$ be representatives of the G–orbits of cubes of X. For each i let Q_i be the finite-index subgroup of $\mathrm{Stab}(\sigma_i)$ consisting of elements which fix σ_i pointwise. Let $\mathcal{Q} = \{Q_1, \dots, Q_k\}$.*

For sufficiently long $\mathcal{Q}$–fillings

$$G \to \overline{G} = G(N_1, \dots, N_m)$$

of $(G, \mathcal{P})$, with kernel K, the quotient $K\backslash X$ is a CAT(0) cube complex.

Proof The kernels of Dehn fillings are always generated by parabolic elements, and the parabolic elements act elliptically by assumption. Thus the kernel of *any* Dehn filling is generated by elliptic elements, so $K\backslash X$ is simply connected by Theorem 4.1. For sufficiently long $\mathcal{Q}$–fillings the fact that $G_{\sigma_i} \cap K \leq Q_i$ follows from [2, Proposition 4.4], so by Proposition 4.3 for such fillings $K\backslash X$ is a cube complex. Therefore, we may assume that the subgroup K is cocubical (in the sense of Terminology 5.5).

It remains to show that for sufficiently long $\mathcal{Q}$–fillings $K\backslash X$ is nonpositively curved. It follows from Theorems 5.6–5.8 and 6.5 that for sufficiently long $\mathcal{Q}$–fillings each link of each cell in $K\backslash X$ is simplicial. Thus it follows from Theorems 5.9, 5.11 and 6.5 that for sufficiently long $\mathcal{Q}$–fillings, each link of each cell in $K\backslash X$ is also flag, which means that $K\backslash X$ is nonpositively curved by Theorem 4.6. □

To prove Theorem 6.5, we use the following "Greendlinger Lemma" — cf [19, Lemma 2.26].

Theorem 6.7 *Let $C_1, C_2 > 0$. Suppose that $(G, \mathcal{P})$ is relatively hyperbolic, with cusped space X. For all sufficiently long fillings $G \to G/K$, and any geodesic γ in X joining 1 to $g \in K - \{1\}$, there is a horoball A such that*

1. *γ contains a depth C_1 vertex of A, and*
2. *there is an element k of K stabilizing A such that, for two points $a, b \in A$ and lying on γ at depth at least C_1, $d(a, kb) < d(a, b) - C_2$ (in particular, $d(1, kg) < d(1, g) - C_2$).*

Proof Let $\delta > 0$ be such that X is δ–hyperbolic, and so are the cusped spaces for sufficiently long fillings (that there exists such a δ is [2, Proposition 2.3]). We only consider such fillings, without further mention of this assumption.

Now choose L and ϵ such that every L–local $(1, C_2)$–quasigeodesic lies within an ϵ–neighborhood of any geodesic with the same endpoints. (Such L and ϵ only depend on δ and C_2; see [11, Chapter 3].)

Now choose a filling long enough that every $(2L+C_1+2\epsilon)$–ball centered on the Cayley graph embeds in the quotient cusped space. Let K be the kernel of the filling, and choose $g \in K - \{1\}$. Let γ be a geodesic from 1 to g, and let $\bar{\gamma}$ be the projection to the cusped space $K\backslash X$ for G/K. Within an $(L+C_1+2\epsilon)$–neighborhood of the Cayley graph, $\bar{\gamma}$ is an L–local geodesic. But $\bar{\gamma}$ cannot be an L–local $(1, C_2)$–quasigeodesic everywhere, since it is a loop with diameter larger than ϵ.

In particular, there is a subsegment σ of $\bar{\gamma}$ of length $l \le L$ such that the endpoints $\bar{a}$ and $\bar{b}$ of σ are less than $l - C_2$ apart. This subsegment σ must moreover lie in the image of a single horoball.

The corresponding points a and b on γ lie at depth at least C_1 in a horoball A of X. Since $d(\bar{a}, \bar{b}) < l - C_2$, there is some element $k \in K$ stabilizing A such that $d(a, kb) < l - C_2$, as desired. □

The following result follows immediately from [28, A.6].

Lemma 6.8 *Suppose that $(G, \mathcal{P})$ is relatively hyperbolic with cusped space X and that $(H, \mathcal{D}) \le (G, \mathcal{P})$ is a full relatively quasiconvex subgroup. There exists a constant κ satisfying the following:*

Suppose that $g \in G$ and that $x_1, x_2 \in gH$. Suppose that γ is a geodesic in X between x_1 and x_2. Further, suppose that aP (for $a \in G$ and $P \in \mathcal{P}$) is a coset such that γ intersects the horoball corresponding to aP to depth at least κ. Then P is infinite and $P^a \cap H^g$ has finite index in P^a.

Proof of Theorem 6.5 Let X be the cusped space associated to $(G, \mathcal{P})$ and suppose that X is δ–hyperbolic. Let C_2 be any positive number, and let

$$C_1 = \max\{|p_i|, \kappa\} + 2(n+100)\delta,$$

where κ is the constant from Lemma 6.8 above. Suppose that K is the kernel of a filling which is long enough to satisfy the conclusion of Theorem 6.7 with these constants.

In order to obtain a contradiction, suppose that there is an element $g \in K$ which is of the form

$$g = p_1 t_1 \cdots p_n t_n,$$

where $t_i \in Q_i - ((K \cap Q_i) S_i)$, and suppose that g is chosen such that $d_X(1, g)$ is minimal amongst all such choices.

Since for each i we have $Q_i - ((K \cap Q_i) S_i) \subseteq Q_i - S_i$, the assumption of the theorem implies that $g \neq 1$. We can represent the equation $g = p_1 t_1 \cdots t_n p_n$ by a geodesic $(2n+1)$–gon in X, joining the appropriate elements of the Cayley graph in turn by X–geodesics. Let γ be the geodesic for g, ρ_i the geodesic for p_i and τ_i the geodesic for t_i.

Since $g \in K - \{1\}$, by Theorem 6.7 there exist a horoball A in X, an element $k \in K$ stabilizing A, and points a and b on γ at depth at least C_1 such that k stabilizes A and $d(a, kb) < d(a, b) - C_2$. In particular, we have $d(x, kgx) < d(x, gx) - C_2$. The geodesic $(2n+1)$–gon is $(2n-1)\delta$–thin, so b lies within distance $(2n-1)\delta$ of some side other than γ. The paths ρ_i do not go deep enough into any horoballs to be this close to b, so b lies within $(2n-1)\delta$ of some point b' on some τ_i. By the choice of C_1, b' lies at depth at least κ in A.

Write $A = aP$ for some $P \in \mathcal{P}$. Note that τ_i is a geodesic between two points in the coset $p_1 t_1 \cdots p_i Q_i$. By Lemma 6.8, $P^a \cap Q_i^{p_1 t_1 \cdots p_i}$ has finite index in P^a. Since the filling is a $\mathcal{Q}$–filling, we have that $k \in Q_i^{p_1 t_1 \cdots p_i}$.

Let $k' = k^{(p_1 t_1 \cdots p_i)^{-1}}$, and let $t_i' = k' t_i$. Then $k' \in K \cap Q_i$.

Note that $kg = p_1 t_1 \cdots p_i (k' t_i) p_{i+1} \cdots p_n t_n$. Since $t_i \notin (K \cap Q_i) S_i$, we have that $t_i' \notin (K \cap Q_i) S_i$. Therefore, the element kg is another element of the required form, contradicting the choice of g as the shortest such. □

6.4 Dehn fillings which induce CAT(0) quotient cube complexes

Theorem 6.9 *Suppose that the hyperbolic group G acts cocompactly on the $CAT(0)$ cube complex X, and that cell stabilizers are virtually special and quasiconvex. Let $\sigma_1, \dots, \sigma_k$ be representatives of the G–orbits of cubes of X, and for each i let Q_i be the finite-index subgroup of $\mathrm{Stab}(\sigma_i)$ consisting of elements which fix σ_i pointwise. Let $\mathcal{Q} = \{Q_1, \dots, Q_k\}$, and let $\mathcal{P}$ be the peripheral structure on G induced by $\mathcal{Q}$, as in Definition 6.2.*

If some element of $\mathcal{Q}$ is infinite, then there exists a Dehn filling

$$G \twoheadrightarrow \overline{G} = G(N_1, \dots, N_m)$$

of $(G, \mathcal{P})$, with kernel $K_{\mathcal{P}}$ such that

1. *$\overline{G}$ is hyperbolic*;
2. *$\overline{\mathcal{Q}}$ consists of virtually special quasiconvex subgroups of $\overline{G}$*;
3. *$K_{\mathcal{P}}$ is generated by elements in cell stabilizers*;
4. *for each i, we have $K_{\mathcal{P}} \cap \mathrm{Stab}(\sigma_i) \le Q_i$*;
5. $\mathrm{height}(\overline{\mathcal{Q}}) < \mathrm{height}(\mathcal{Q})$;
6. *$K_{\mathcal{P}} \backslash X$ is a $CAT(0)$ cube complex.*

Proof Let G, X, $\mathcal{Q}$ and $\mathcal{P}$ be as in the statement of the theorem. By Lemma 6.4, $(G, \mathcal{P})$ is relatively hyperbolic. Moreover, for each $Q \in \mathcal{Q}$, the induced structure $\mathcal{D}_Q$ on Q makes $(Q, \mathcal{D}_Q)$ relatively hyperbolic, and Q is full relatively quasiconvex in $(G, \mathcal{P})$. Note that the assumption that some element of $\mathcal{Q}$ is infinite implies (by the definition of $\mathcal{P}$) that some element of $\mathcal{P}$ is infinite.

Property (1) holds for sufficiently long peripherally finite fillings of $(G, \mathcal{P})$ by the basic result of relatively hyperbolic Dehn fillings [31, Theorem 1.1]. We always assume that we have taken a filling such that $\overline{G}$ is hyperbolic.

We remark that, because each element of $\mathcal{Q}$ is finite-index in a cell stabilizer, each element of $\mathcal{Q}$ is hyperbolic and virtually special. Moreover, since each element of $\mathcal{P}$ has a finite-index subgroup which is a quasiconvex subgroup of some element of $\mathcal{Q}$ by construction, each element of $\mathcal{P}$ is also hyperbolic and virtually special. In particular, each element of $\mathcal{P}$ is residually finite. We choose particular fillings with $N_i \dot{<} P_i$, and residual finiteness guarantees the existence of the fillings that we seek.

We now explain how to ensure the properties of the conclusion of the result.

Suppose that $Q \in \mathcal{Q}$. Since $\mathcal{P}$ is the peripheral structure induced by $\mathcal{Q}$, we can choose finite-index subgroups of elements of $\mathcal{P}$ which induce $\mathcal{Q}$–fillings, and any such filling $\overline{G}$ of G naturally induces a filling $\overline{Q}$ of Q. By the Malnormal Special Quotient Theorem [36, Theorem 12.3] — see also [3, Corollary 2.8] — for each $P_i \in \mathcal{P}$ there is a subgroup $\dot{P}_i(Q) \mathrel{\dot{\lhd}} P_i$ such that if each filling kernel N_i satisfies $N_i \leq \dot{P}_i(Q)$ then the induced filling $\overline{Q}$ is virtually special (and hyperbolic). Let $\dot{P}_i$ be the intersection of the $\dot{P}_i(Q)$ for all $Q \in \mathcal{Q}$. Thus, if we choose filling kernels $N_i \leq \dot{P}_i$ then each of the induced fillings of each element of $\mathcal{Q}$ is virtually special. By [18, Proposition 4.6], the natural map from $\overline{Q}$ to $\overline{G}$ is injective for all sufficiently long fillings.[7] If we choose a sufficiently long peripherally finite filling of $(G, \mathcal{P})$ with $N_i \leq \dot{P}_i$ then [18, Proposition 4.5] implies that each $\overline{Q}$ is quasiconvex in $\overline{G}$. This ensures property (2).

For the remaining properties, we show that they hold for sufficiently long peripherally finite $\mathcal{Q}$–fillings of $(G, \mathcal{P})$. Therefore, to ensure that all of the properties hold, it suffices to take a sufficiently long $\mathcal{Q}$–filling with each $N_i \mathrel{\dot{<}} \dot{P}_i$.

Property (3) holds automatically for any $\mathcal{Q}$–filling, since $K_{\mathcal{P}}$ is generated by conjugates of elements in $\mathcal{Q}$, and each such conjugate lies in a cell stabilizer.

We now explain how to ensure each of the remaining properties in turn for sufficiently long $\mathcal{Q}$–fillings.

For property (4), suppose that $\mathcal{F}_i \sqcup \{1\}$ is a set of coset representatives for Q_i in $\mathrm{Stab}(\sigma_i)$. To ensure that (4) holds, it suffices to keep (the image of) each element of $\mathcal{F}_i$ out of the image of $\mathrm{Stab}(\sigma_i)$ in $\overline{G}$. This is true for sufficiently long $\mathcal{Q}$–fillings by [1, Theorem A.43.4], because Q_i has finite index in $\mathrm{Stab}(\sigma_i)$.

Property (5) holds for sufficiently long peripherally finite $\mathcal{Q}$–fillings of $(G, \mathcal{P})$ by an entirely analogous argument to that of [1, Theorem A.47].

Finally, property (6) holds for sufficiently long $\mathcal{Q}$–fillings by Corollary 6.6. □

The group $\overline{G}$ as above acts isometrically on $\overline{X} = K_{\mathcal{P}} \backslash X$ with quotient naturally isomorphic (as a topological space, but not as a complex of groups) to $G \backslash X$. Therefore, if the action of $\overline{G}$ on $\overline{X}$ is not proper, we can apply Theorem 6.9 to this action, to obtain a further quotient. By induction on height, we obtain the following result from the introduction.

[7] Lemma 3.7 of [18] ensures that sufficiently long $\mathcal{Q}$–fillings are sufficiently wide, in the terminology of that paper.

Theorem F *Suppose that the hyperbolic group G acts cocompactly on a CAT(0) cube complex X and that cell stabilizers are virtually special and quasiconvex. There exists a quotient $\overline{G} = G/K$ such that*

(1) *the quotient $K\backslash X$ is a CAT(0) cube complex*;

(2) *the group $\overline{G}$ is hyperbolic*; *and*

(3) *the action of $\overline{G}$ on $K\backslash X$ is proper* (*and cocompact*).

Appendix A quasiconvexity criterion

In this appendix, we give a criterion (Theorem A.3) for a possibly infinite union of quasiconvex sets in a hyperbolic space to be quasiconvex. This criterion is used in the forward direction of Theorem A: quasiconvex cell stabilizers imply quasiconvex hyperplane stabilizers. This criterion may be of independent interest.

Since any subset is a union of points, clearly some assumptions are needed.

We begin with a basic lemma about finite unions of quasiconvex subsets.

Lemma A.1 *Suppose that Y is δ–hyperbolic*, *and $P \subset Y$ is a union of k ϵ–quasiconvex subsets $P_1, \dots, P_k$ such that $P_i \cap P_{i+1} \neq \varnothing$ for each i. Then P is ρ–quasiconvex*, *where*

$$\rho = \delta(\log_2(k) + 1) + \epsilon.$$

Proof Consider a pair of points $x \in P_r$ and $y \in P_s$. Without loss of generality, assume that $r < s$ (the case $r = s$ being straightforward).

Now choose a sequence of points $p_i \in P_i \cap P_{i+1}$ for $r \le i < s$, let σ be a geodesic between x and y and let u be a point on σ. Our task is to bound the distance from u to P.

Consider the broken geodesic $\gamma = [x, p_r, p_{r+1}, \dots, p_{s-1}, y]$. Since the P_i are ϵ–quasiconvex, γ is contained in an ϵ–neighborhood of $P_r \cup \dots \cup P_s \subset P$.

Consider the geodesic polygon with one side the geodesic $\sigma = [x, y]$ and the other sides the geodesics forming γ. Let $r_0 = \left\lfloor \frac{1}{2}(r + s) \right\rfloor$, and consider the geodesic triangle $\sigma, [x, p_{r_0}], [p_{r_0}, y]$. By δ–hyperbolicity, u lies within δ of one of $[x, p_{r_0}]$ and $[p_{r_0}, y]$. Suppose it is $[x, p_{r_0}]$ (the other case being entirely similar), and suppose that $u_1 \in [x, p_{r_0}]$ is within δ of u.

Now let $r_1 = \lfloor \frac{1}{2}(r+r_0) \rfloor$ and consider the geodesic triangle $[x, p_{r_1}]$, $[p_{r_1}, p_{r_0}]$, $[p_{r_0}, x]$. By δ–hyperbolicity, u_1 is within δ of one of $[x, p_{r_1}]$ or $[p_{r_1}, p_{r_0}]$, so there is u_2 on one of these sides within δ of u_1 and within 2δ of u.

We proceed in this manner, in each case making the interval of indices half as long. After t steps of this argument we find a point u_t which is within $t\delta$ of u.

After at most $d = \log_2(k) + 1$ steps, we have a geodesic triangle where two sides are $[p_l, p_{l+1}], [p_{l+1}, p_{l+2}]$ (or maybe one endpoint x or y), and we have u_d within $d\delta$ of u, but also within ϵ of P. □

The following straightforward instance of "linear-beats-log" is tailored for use in the proof of Theorem A.3.

Lemma A.2 *Fix $\delta, \epsilon > 0$, and let $g(x) = \delta(\log_2(x+1)+1)+\epsilon$. For any $m > 0$ and $c \geq 0$ there exists a natural number $R_{m,\epsilon,\delta}$ such that for all $R_0 > R_{m,\epsilon,\delta}$,*

$$g(R_0) < \frac{1}{200} m \left(\frac{1}{4} R_0 - \frac{2g(R_0)+1}{m} - 3c \right).$$

The next result states that under appropriate hypotheses, the union of an arbitrary number of quasiconvex subsets is itself quasiconvex, with constant not depending on the number of such subsets.

Theorem A.3 *Suppose that Υ is a δ–hyperbolic space and that $m, \epsilon > 0$ and $c \geq 0$ are real numbers. There exists a constant ϵ' such that for any (finite or countably infinite) collection of subsets $\{X_i\}_{i=1}^{\Lambda}$ of Υ for which*

1. *each X_i is ϵ–quasiconvex,*
2. *for each i we have $X_i \cap X_{i+1} \neq \varnothing$, and*
3. *for any i, j, if $x \in X_i$ and $y \in X_j$, we have $d(x,y) \geq m(|i-j|-c)$,*

the set $X = \bigcup_i X_i$ is ϵ'–quasiconvex.

Proof Let $g(x) = \delta(\log_2(x+1)+1)+\epsilon$, and let $R = R_{m,\epsilon,\delta}$ be the number from Lemma A.2. Without loss of generality we may assume that $R \geq 1$.

If $\Lambda \leq 100R$ then Lemma A.1 implies X is ρ–quasiconvex with

$$\rho = \delta(\log_2(100R)+1)+\epsilon = g(100R-1).$$

On the other hand, suppose that $\Lambda > 100R$ and fix $u, v \in X$. Let j and k be such that $u \in X_j$ and $v \in X_k$, and without loss of generality suppose that $j \leq k$. It suffices to show that any geodesic $[u,v]$ stays uniformly close to $X_j \cup \cdots \cup X_k$. If $|k-j| \le 100R$ then this follows from Lemma A.1, so suppose that $|k-j| > 100R$. Let $Y = X_j \cup \cdots \cup X_k$.

Our strategy is to build a path between u and v which is uniformly quasigeodesic and stays uniformly close to Y. The theorem then follows by quasigeodesic stability. Choose a sequence of indices $t_0 = j, t_1, \ldots, t_{s-1}, t_s = k$ such that for each $0 \le r \le s-2$,

$$t_{r+1} - t_r = 100R,$$

and

$$t_s - t_{s-1} \in \mathbb{Z} \cap [100R, \ldots, 200R].$$

Moreover, for each $0 \le r \le s$ choose some $u_r \in X_{t_r}$. We require $u_0 = u$ and $u_s = v$.

For $r \in \{0, \ldots, s-1\}$, let γ_r be a geodesic between u_r and u_{r+1}. Let

$$K = g(200R) = \delta(\log_2(200R+1)+1) + \epsilon.$$

Since we assume $R \ge 1$, we know that $K > \delta$.

Since we know that for each $r \in \{0, \ldots, s-1\}$ we have $t_{r+1} - t_r \le 200R$, we know that the set

$$Y_r = \bigcup_{k=t_r}^{t_{r+1}} X_k$$

is K–quasiconvex, by Lemma A.1. In particular, the geodesic γ_r lies in a K–neighborhood of Y_r.

For each $r \in \{0, \ldots, s-1\}$ and each $x \in \gamma_r$, let $\pi_r(x)$ denote the set of closest points on Y_r to x. Furthermore, let $I_r(x)$ be the set of indices l such that $\pi_r(x) \cap X_l \ne \varnothing$.

Claim A.3.1 *For any $v \in \{t_r, \ldots, t_{r+1}\}$ there exists $x_v \in \gamma_r$ such that*

$$d_{\mathbb{N}}(v, I_r(x_v)) \le \frac{1}{2}\Big(\frac{2K+1}{m} + c\Big).$$

Proof For any $y \in \pi_r(x)$ we have $d(x,y) \le K$. Now, if x and x' are adjacent vertices and $y \in \pi_r(x)$ with $y \in X_k$ and $z \in \pi_r(x')$ with $z \in X_l$ then

$$m(|k-l| - c) \le d(y,z) \le d(y,x) + d(x,x') + d(x',z) \le 2K+1,$$

so $|k-l| \le (2K+1)/m + c$.

The claim now follows immediately from the fact that $t_r \in I_r(u_r)$ and $t_{r+1} \in I_r(u_{r+1})$, letting x and x' run over adjacent pairs of vertices in γ_r. □

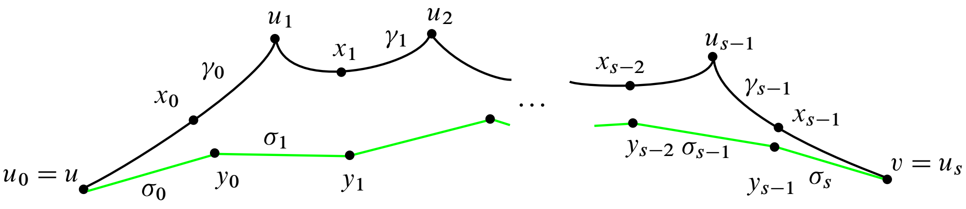


Figure 3: The σ_i forming a broken geodesic.

Suppose $0 \leq r \leq s-1$. Using Claim A.3.1, we can choose a point $x_r \in \gamma_r$ and a point $y_r \in \pi_r(x_r)$ such that $y_r \in X_{k_r}$ and

$$(\dagger) \qquad \left| k_r - \frac{t_r + t_{r+1}}{2} \right| \leq \frac{2K+1}{2m} + c.$$

Now, for each $r \in \{1, \dots, s-1\}$, let σ_r be a geodesic between y_{r-1} and y_r. Further, let σ_0 be a geodesic from u to y_0 and let σ_s be a geodesic from y_{s-1} to v (note that there is no point y_s); see Figure 3. We bound the Gromov product between σ_t and σ_{t+1} for each t. (There is no reason to expect such a bound on the Gromov product between γ_r and γ_{r+1}.)

Though we have no control on the lengths of the segments σ_0 and σ_s, the lengths of the other segments can be bounded below:

Claim A.3.2 *Suppose $0 < r < s$. The length of σ_r is at least $200K$.*

Proof By the choice of the index k_r in (†),

$$k_r - k_{r-1} \geq \frac{t_{r+1} - t_{r-1}}{2} - \frac{2K+1}{m} - 2c = 100R - \frac{2K+1}{m} - 2c > 50R - \frac{2K+1}{m} - 2c$$

(the equality follows from the choice of t_r).

Below, we apply Lemma A.2 with $R_0 = 200R$, noting that $K = g(200R)$, where g is the function from that lemma. We have

$$|\sigma_r| = d_G(y_{r-1}, y_r) \geq m(k_r - k_{r-1} - c) > m\left(50R - \frac{2K+1}{m} - 3c\right) \geq 200K.$$

The second inequality above follows from the fact that $y_i \in X_{k_i}$ so such points are at least distance $m(k_r - k_{r-1} - c)$ apart. The final inequality follows from the promised use of Lemma A.2. □

Claim A.3.3 *Let $0 \leq r \leq s-1$. The Gromov product of σ_r and σ_{r+1} is at most $8K$.*

Proof We first handle the case that $0 < r < s-1$.

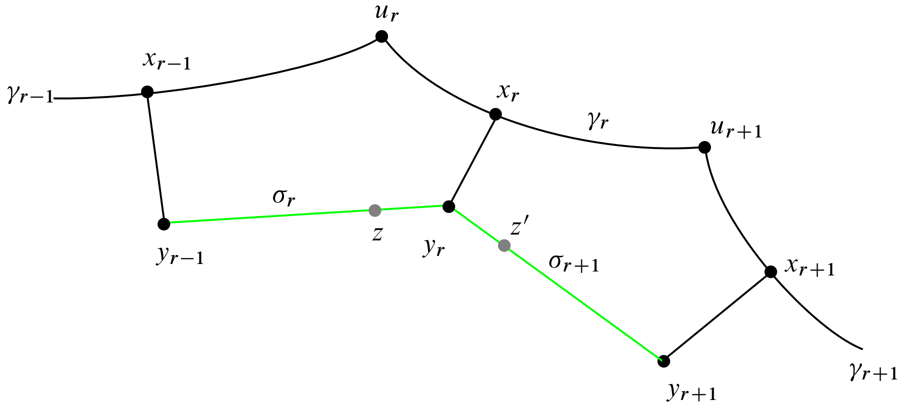


Figure 4: Computing the Gromov product of σ_r and σ_{r+1}.

For $i \in \{r, r+1\}$, the path σ_i is one side of a pentagon. The other sides are

(a) two sides of length at most K at either end of σ_i, and

(b) two "halves" of adjacent geodesics: the second "half" of γ_{i-1} and the first "half" of γ_i, joined at u_i.

See Figure 4.

By Claim A.3.2 the geodesics σ_r and σ_{r+1} have length at least $200K$. Let z be the point on σ_r at distance exactly $8K$ from y_r.

Since geodesic pentagons are 3δ–slim, we know that z must be distance at most 3δ from some point on one of the other four sides. However, it cannot be within distance 3δ of the geodesic between x_r and y_r since that geodesic has length at most K. Similarly, since $|\sigma_r| \geq 200K$, z cannot be within 3δ of the geodesic between x_{r-1} and y_{r-1}. We claim that z also cannot be within 3δ of the part of γ_{r-1} contained in the pentagon.

Indeed, suppose $w \in \gamma_{r-1}$, and choose $i_w \in I_{r-1}(w) \subset [t_{r-1}, t_r]$. There is a point w' of $\pi_{r-1}(w)$ in X_{i_w}; thus $d(w, w') \leq K$. The point x_r is likewise within K of some X_{k_r} where k_r satisfies the inequality (†). This implies that

$$|k_r - i_w| \geq \frac{t_{r+1} - t_r}{2} - \frac{2K+1}{2m} - c,$$

and so, using Lemma A.2 again,

$$d(x_r, w) \geq m\left(\frac{t_{r+1} - t_r}{2} - \frac{2K+1}{2m} - 2c\right) - 2K \geq m\left(50R - \frac{2K+1}{m} - 3c\right) - 2K \geq 198K.$$

But this contradicts $d(x_r, w) \leq d(x_r, z) + d(z, w) \leq 9K + 3\delta \leq 12K$.

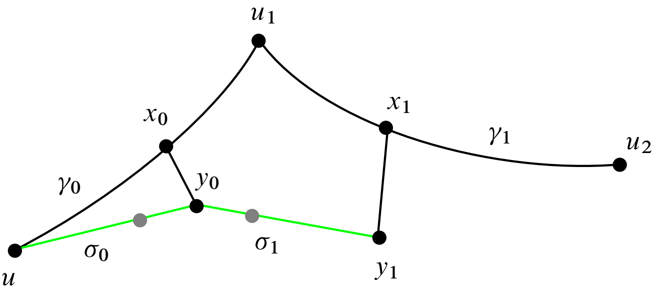


Figure 5: Computing the Gromov product of σ_0 and σ_1.

We have shown that there is some point w on γ_r between u_r and x_r within 3δ of z. Note that $d(x_r, w) \geq d(y_r, z) - K - 3\delta \geq 4K$, since $K \geq \delta$.

Now consider the pentagon formed with σ_{r+1} on one side, and the point z' on σ_{r+1} which is distance exactly $8K$ from y_r. An entirely analogous argument to the above shows that there is some w' between x_r and u_{r+1} on γ_r such that $d(z', w') \leq 3\delta$, and $d(x_r, w') \geq 4K$. Since γ_r is geodesic,

$$d(w, w') = d(w, x_r) + d(x_r, w) \geq 8K.$$

It follows that $d(z, z') \geq 8K - 6\delta \geq 2K > \delta$. Therefore, the Gromov product $(y_{r-1}, y_{r+1})_{y_r}$ is strictly less than $d(z, y_r) = d(z', y_r) = 8K$ whenever $0 < r < s$.

The cases $r = 0$ and $r = s - 1$ are symmetric, so it suffices to handle the case $r = 0$; see Figure 5. We are trying to show that $(u, y_1)_{y_0} \leq 8K$, so we may suppose without loss of generality that $d(y_0, u) > 8K$. Thus there is a point z on σ_0 at distance exactly $8K$ from y_0. Since $d(x_0, y_0) \leq K$, this point is within δ of a point w on γ_0 between u and x_0.

For the point z' on σ_1 at distance $8K$ from y_0, we argue as before. We are again able to deduce that $d(z, z') > \delta$, and so $(u, y_1)_{y_0} \leq 8K$. □

Thus, we have a collection of arcs σ_i which form a broken geodesic between u and v with segments of length at least $200K$ (except possibly the first and last) and all Gromov product at most $8K$ at the corners. Thus the union of the σ_i forms a global quasigeodesic with uniformly bounded parameters. However, each σ_i lies within a $(3\delta + K)$–neighborhood of the union of the γ_i, which in turn lie in a K–neighborhood of the union of the X_i. As explained above, this suffices to prove that the union of the X_i is ϵ'–quasiconvex with the constant ϵ' depending on the quantities δ, m and ϵ, but not on the number of the X_i, as required. □

References

[1] **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 MR Zbl With an appendix by I Agol, D Groves and J Manning

[2] **I Agol**, **D Groves**, **J F Manning**, *Residual finiteness, QCERF and fillings of hyperbolic groups*, Geom. Topol. 13 (2009) 1043–1073 MR Zbl

[3] **I Agol**, **D Groves**, **J F Manning**, *An alternate proof of Wise's malnormal special quotient theorem*, Forum Math. Pi 4 (2016) art. id. E1 MR Zbl

[4] **M A Armstrong**, *On the fundamental group of an orbit space*, Proc. Cambridge Philos. Soc. 61 (1965) 639–646 MR Zbl

[5] **N Bergeron**, **D T Wise**, *A boundary criterion for cubulation*, Amer. J. Math. 134 (2012) 843–859 MR Zbl

[6] **B H Bowditch**, *Tight geodesics in the curve complex*, Invent. Math. 171 (2008) 281–300 MR Zbl

[7] **B H Bowditch**, *Relatively hyperbolic groups*, Internat. J. Algebra Comput. 22 (2012) art. id. 1250016 MR Zbl

[8] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[9] **R Charney**, **J Crisp**, *Relative hyperbolicity and Artin groups*, Geom. Dedicata 129 (2007) 1–13 MR Zbl

[10] **D Cooper**, **D Futer**, *Ubiquitous quasi-Fuchsian surfaces in cusped hyperbolic* 3*–manifolds*, Geom. Topol. 23 (2019) 241–298 MR Zbl

[11] **M Coornaert**, **T Delzant**, **A Papadopoulos**, *Géométrie et théorie des groupes: les groupes hyperboliques de Gromov*, Lecture Notes in Math. 1441, Springer (1990) MR Zbl

[12] **Y Duong**, *On random groups*: *the square model at density $d < 1/3$ and as quotients of free nilpotent groups*, PhD thesis, University of Illinois at Chicago (2017) MR Available at `https://www.proquest.com/docview/2001345286`

[13] **E Einstein**, **D Groves**, *Relative cubulations and groups with a* 2*–sphere boundary*, Compos. Math. 156 (2020) 862–867 MR Zbl

[14] **A Genevois**, *Hyperbolicities in* CAT(0) *cube complexes*, Enseign. Math. 65 (2019) 33–100 MR Zbl

[15] **A Genevois**, *Coning-off* CAT(0) *cube complexes*, Ann. Inst. Fourier (Grenoble) 71 (2021) 1535–1599 MR Zbl

[16] **R Gitik**, **M Mitra**, **E Rips**, **M Sageev**, *Widths of subgroups*, Trans. Amer. Math. Soc. 350 (1998) 321–329 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**, *Quasiconvexity and Dehn filling*, Amer. J. Math. 143 (2021) 95–124 MR Zbl

[19] **D Groves**, **J F Manning**, **A Sisto**, *Boundaries of Dehn fillings*, Geom. Topol. 23 (2019) 2929–3002 MR Zbl

[20] **F Haglund**, **D T Wise**, *Special cube complexes*, Geom. Funct. Anal. 17 (2008) 1551–1620 MR Zbl

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

[22] **J Hempel**, 3*–manifolds*, Annals of Mathematics Studies 86, Princeton Univ. Press (1976) MR Zbl

[23] **G C Hruska**, *Relative hyperbolicity and relative quasiconvexity for countable groups*, Algebr. Geom. Topol. 10 (2010) 1807–1856 MR Zbl

[24] **G C Hruska**, **D T Wise**, *Finiteness properties of cubulated groups*, Compos. Math. 150 (2014) 453–506 MR Zbl

[25] **W Jaco**, *Lectures on three-manifold topology*, CBMS Regional Conference Series in Mathematics 43, Amer. Math. Soc., Providence, RI (1980) MR Zbl

[26] **J Kahn**, **V Markovic**, *Immersing almost geodesic surfaces in a closed hyperbolic three manifold*, Ann. of Math. 175 (2012) 1127–1190 MR Zbl

[27] **I Kapovich**, **H Short**, *Greenberg's theorem for quasiconvex subgroups of word hyperbolic groups*, Canad. J. Math. 48 (1996) 1224–1244 MR Zbl

[28] **J F Manning**, **E Martínez-Pedroza**, *Separation of relatively quasiconvex subgroups*, Pacific J. Math. 244 (2010) 309–334 MR Zbl

[29] **K Matsuzaki**, **M Taniguchi**, *Hyperbolic manifolds and Kleinian groups*, Oxford Univ. Press (1998) MR Zbl

[30] **J W Morgan**, *On Thurston's uniformization theorem for three-dimensional manifolds*, from "The Smith conjecture" (J W Morgan, H Bass, editors), Pure Appl. Math. 112, Academic, Orlando, FL (1984) 37–125 MR Zbl

[31] **D V Osin**, *Peripheral fillings of relatively hyperbolic groups*, Invent. Math. 167 (2007) 295–326 MR Zbl

[32] **M Sageev**, *Ends of group pairs and non-positively curved cube complexes*, Proc. London Math. Soc. 71 (1995) 585–617 MR Zbl

[33] **M Sageev**, *Codimension-*1 *subgroups and splittings of groups*, J. Algebra 189 (1997) 377–389 MR Zbl

[34] **G A Swarup**, *Geometric finiteness and rationality*, J. Pure Appl. Algebra 86 (1993) 327–333 MR Zbl

[35] **H C Tran**, *On strongly quasiconvex subgroups*, Geom. Topol. 23 (2019) 1173–1235 MR Zbl

[36] **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, Statistics and Computer Science, University of Illinois at Chicago
Chicago, IL, United States

Department of Mathematics, Cornell University
Ithaca, NY, United States

dgroves@uic.edu, jfmanning@cornell.edu

Proposed: Martin R Bridson
Seconded: Benson Farb, Bruce Kleiner

Received: 29 April 2019
Revised: 14 March 2021

Geometry & Topology **27**:9 (2023) 3461–3584
DOI: 10.2140/gt.2023.27.3461
Published: 5 December 2023

Cyclic homology, S^1–equivariant Floer cohomology and Calabi–Yau structures

SHEEL GANATRA

We construct geometric maps from the cyclic homology groups of the (compact or wrapped) Fukaya category to the corresponding S^1–equivariant (Floer/quantum or symplectic) cohomology groups, which are natural with respect to all Gysin and periodicity exact sequences and are isomorphisms whenever the (nonequivariant) open–closed map is. These *cyclic open–closed maps* give constructions of geometric smooth and/or proper Calabi–Yau structures on Fukaya categories, which in the proper case implies the Fukaya category has a cyclic A_∞ model in characteristic 0, and also give a purely symplectic proof of the noncommutative Hodge–de Rham degeneration conjecture for smooth and proper subcategories of Fukaya categories of compact symplectic manifolds. Further applications of cyclic open–closed maps, to counting curves in mirror symmetry and to comparing topological field theories, are the subject of joint projects with Perutz and Sheridan, and with Cohen.

53D12, 53D37; 19D55

1 Introduction

This paper concerns the compatibility between chain level S^1–actions arising in two different types of Floer theory on a symplectic manifold M. The first of these $C_{-*}(S^1)$–actions[1] is induced geometrically on the *Hamiltonian Floer homology chain complex*

[1] We will use a cohomological grading convention, so singular chain complexes are *negatively graded*.

$CF^*(M)$, formally a type of Morse complex for an action functional on the free loop space, through rotating free loops. The homological action of $[S^1]$ is known as the *BV operator* $[\Delta]$, and the $C_{-*}(S^1)$–action can be used to define *S^1–equivariant Floer*[2] *homology theories*; see eg Bourgeois and Oancea [6] and Seidel [51]. The second $C_{-*}(S^1)$–action lies on the *Fukaya category* of M, and has discrete or combinatorial origins, coming from the hierarchy of compatible cyclic $\mathbb{Z}/k\mathbb{Z}$–actions on cyclically composable chains of morphisms between Lagrangians. A (categorical analogue of a) fundamental observation of Connes [13], Tsygan [63] and Loday and Quillen [42] is that such a structure, which exists on any category $\mathcal{C}$, can be packaged into a $C_{-*}(S^1)$–action on the *Hochschild homology chain complex* $\mathrm{CH}_*(\mathcal{C})$ of the category; see also Keller [32] and McCarthy [43]. The associated operation of multiplication by (a cycle representing) $[S^1]$ is frequently called the *Connes B operator*, and the corresponding S^1–equivariant homology theories are called *cyclic homology groups*.

A relationship between the Hochschild homology of the Fukaya category $\mathcal{F}$ and Floer homology on M is provided by the so-called *open–closed string map*

$$\mathcal{OC}\colon \mathrm{CH}_*(\mathcal{F}) \to CF^{*+n}(M); \tag{1-1}$$

see Abouzaid [1]. Our main result is about the compatibility of $\mathcal{OC}$ with $C_{-*}(S^1)$–actions. Namely, we prove — under technical hypotheses detailed below the main result — that $\mathcal{OC}$ can be made (coherently homotopically) $C_{-*}(S^1)$–equivariant:

Theorem 1.1 *Suppose that M, its Fukaya category and $CF^*(M)$ satisfy the technical assumptions ($\star$). Then the map $\mathcal{OC}$ admits a geometrically defined* ***S^1–equivariant enhancement****, to an A_∞ homomorphism of $C_{-*}(S^1)$–modules,*

$$\widetilde{\mathcal{OC}} \in \mathrm{RHom}^n_{C_{-*}(S^1)}(\mathrm{CH}_*(\mathcal{F}), CF^*(M)).$$

Remark 1.2 Theorem 1.1 implies (but is not implied by) the statement (Theorem 5.14) that $[\mathcal{OC}]$ intertwines homological actions of $[S^1]$.

Remark 1.3 In the geometric settings considered here $\mathcal{OC}$ does not a priori strictly intertwine the $C_{-*}(S^1)$–actions (due to a priori nonequivariant perturbations made to moduli spaces to define operations, and further due to the potential nontriviality of Δ, which — as Δ is defined using moduli spaces but B is defined using algebra — imply that

[2] Sometimes S^1–equivariant Floer theory is instead defined as Morse theory of an action functional on the S^1–Borel construction of the loop space. For a comparison between these two definitions, see [6].

$\mathcal{OC} \circ B$ and $\Delta \circ \mathcal{OC}$ involve moduli spaces of maps from differing domains). In particular, the homomorphism $\widetilde{\mathcal{OC}}$ involves extra data recording coherently higher homotopies between the two $C_{-*}(S^1)$–actions. This explains our use of the term "enhancement".

Remark 1.4 It can be shown using usual invariance arguments that the enhancement $\widetilde{\mathcal{OC}}$ we define in this paper is uniquely determined up to homotopy: while the geometric chain-level construction requires a number of auxiliary choices (of perturbation data on moduli spaces), any two sets of such choices produce homotopic enhancements.

To explain the consequences of Theorem 1.1 to cyclic homology and equivariant Floer homology, recall that there are a variety of S^1*–equivariant homology* chain complexes (and homology groups) that one can associate functorially to an A_∞ $C_{-*}(S^1)$–module P. For instance, denote by

$$P_{\mathrm{h}S^1}, \quad P^{\mathrm{h}S^1}, \quad P^{\mathrm{Tate}} \tag{1-2}$$

the *homotopy orbit complex*, *homotopy fixed-point complex* and *Tate complex* constructions of P, described in Section 2.2. When applied to the Hochschild complex $\mathrm{CH}_*(\mathcal{C})$, the constructions (1-2) by definition recover complexes computing (*positive*) *cyclic homology*, *negative cyclic homology* and *periodic cyclic homology* groups of $\mathcal{C}$, respectively; see Section 3.2. Similarly the group $H^*(CF^*(M)_{\mathrm{h}S^1})$ is the S^1*–equivariant Floer cohomology* studied (for the symplectic homology Floer chain complex); see eg Bourgeois and Oancea [6], Seidel [51] and Viterbo [64]. The groups $H^*(CF^*(M)^{\mathrm{h}S^1})$ and $H^*(CF^*(M)^{\mathrm{Tate}})$ have also been studied in recent work in Floer theory; see Albers, Cieliebak and Frauenfelder [4], Seidel [56] and Zhao [66]. Functoriality of the constructions (1-2) and homotopy-invariance properties of $C_{-*}(S^1)$–modules (see Corollary 2.18 and Proposition 2.19) immediately imply:

Corollary 1.5 *Let $HF^{*,+/-/\infty}_{S^1}(M)$ denote the* (*cohomology of the*) *homotopy orbit complex*, *fixed-point complex*, *and Tate complex construction applied to $CF^*(M)$*, *and let $\mathrm{HC}^{+/-/\infty}(\mathcal{C})$ denote the corresponding positive/negative/periodic cyclic homology groups. Under the hypotheses* ($\star$) *of Theorem 1.1*, $\widetilde{\mathcal{OC}}$ *induces* ***cyclic open–closed maps***

$$[\widetilde{\mathcal{OC}}^{+/-/\infty}]\colon \mathrm{HC}^{+/-/\infty}_*(\mathcal{F}) \to HF^{*+n,+/-/\infty}_{S^1}(M), \tag{1-3}$$

which are naturally compatible with respect to the various periodicity/Gysin exact sequences, and which are isomorphisms whenever $\mathcal{OC}$ is. □

The map (1-1) is frequently an isomorphism, allowing one to recover in these cases closed string Floer/quantum homology groups from open string, categorical ones; see Abouzaid, Fukaya, Oh, Ohta and Ono [2], Bourgeois, Ekholm and Eliashberg [5], Ganatra [24] and Ganatra, Perutz and Sheridan [27]. In such cases, Theorem 1.1 and Corollary 1.5 allow one to further categorically recover the $C_{-*}(S^1)$ as well as the associated equivariant homology groups (in terms of the cyclic homology groups of the Fukaya category).

Remark 1.6 There are other S^1–equivariant homology functors to which our results apply tautologically as well. For instance, consider the contravariant functor $P \mapsto (P_{\mathrm{h}S^1})^\vee$; when applied to $\mathrm{CH}_*(\mathcal{C})$ this produces the *cyclic cohomology* chain complex of $\mathcal{C}$.

We have been deliberately vague about which Fukaya category and which Hamiltonian Floer homology groups Theorem 1.1 applies to, as it applies in several different geometric (compact and noncompact) settings. To keep this paper a manageable length, we implement the map $\widetilde{\mathcal{OC}}$ and prove Theorem 1.1 in the technically simplest of such settings — our technical hypotheses are detailed in ($\star$) below — for which the moduli spaces appearing in the constructions can be shown to be well behaved by classical methods. That being said, we should remark that our methods and arguments are orthogonal to the usual analytic difficulties faced in constructing Fukaya categories and open–closed maps in more general contexts, and we expect they should extend relatively directly to other settings. For instance, in the setting of relative Fukaya categories of compact projective Calabi–Yau manifolds (not considered here), an adapted version of our construction will appear in joint work with Perutz and Sheridan [26].

($\star$) Assumptions on M, $\mathcal{F}$ and $CF^*(M)$

In our main results we make technical assumptions, explained in detail in Section 3.3 for M and its Fukaya category and in Sections 4.1.1–4.1.2 for the corresponding Hamiltonian Floer homology chain complexes, which broadly encapsulate the following situations:

(1) If M is compact and satisfies suitable technical hypotheses such as being monotone or symplectically aspherical (see Section 3.3.1), one could take $\mathcal{F}$ to be the usual Fukaya category (or a summand thereof) of those compact Lagrangians also satisfying suitable technical hypotheses such as being monotone or not bounding disks with symplectic area. In this case $CF^*(M)$, the Hamiltonian

Floer complex of any (sufficiently generic) Hamiltonian, is quasi-isomorphic to the *quantum cohomology* ring with its trivial $C_{-*}(S^1)$–action.

(2) If M is noncompact and Liouville, one could take $\mathcal{F} = \mathcal{W}$ to be the *wrapped Fukaya category* and $CF^*(M) = SC^*(M)$ to be the *symplectic cohomology cochain complex* with its (typically highly nontrivial) $C_{-*}(S^1)$–action.

(3) If M is noncompact and Liouville, one could take $\mathcal{F} \subset \mathcal{W}$ to be the *Fukaya category of compact exact Lagrangians*. When restricted to $\mathrm{CH}_*(\mathcal{F})$, the map $\mathcal{OC}$ to $SC^*(M)$ of (2) factors through $H^*(M, \partial^\infty M)$, the *relative* (or *compactly supported*) *cohomology group* with its trivial $C_{-*}(S^1)$–action. In fact, as reviewed in Section 5.6.2, $\mathcal{OC}$ further factors through the symplectic *homology* chain complex $SC_*(M) \cong (SC^*(M))^\vee[-2n]$. One could take any of these groups ($SC^*(M)$, $H^*(M, \partial^\infty M)$ or $SC_*(M)$) to be $CF^*(M)$ here. For the main portion of the paper we use $CF^*(M) := H^*(M, \partial^\infty M)$.

For example, in case (2) above, when the relevant $[\mathcal{OC}]$ map is an isomorphism, Corollary 1.5 computes various S^1–equivariant symplectic cohomology groups[3] in terms of cyclic homology groups of the wrapped Fukaya category.

Remark 1.7 For the Fukaya subcategory of a single Lagrangian in a compact symplectic manifold M over a characteristic-zero (Novikov) field containing $\mathbb{R}$, a variant of the (positive) cyclic open–closed map has also been constructed by Fukaya, Oh, Ohta and Ono [23] (and will be generalized to multiple Lagrangians in Abouzaid, Fukaya, Oh, Ohta and Ono [2]). Their construction, which requires the target group ($H^*(M)$) to have trivial $C_{-*}(S^1)$–action, uses Connes' small ("coinvariants of cyclic group action bar") complex for (in characteristic zero only) positive cyclic homology, along with cyclically symmetric (necessarily virtual) perturbations of all moduli spaces (building on work of Fukaya [20] described in Remark 1.14), to directly construct a geometric map bypassing the higher A_∞ $C_{-*}(S^1)$–action homotopies constructed here. It does not seem possible to generalize the methods of [23] to the (possibly noncompact M with arbitrary coefficients, eg integral/rational/finite characteristic) settings considered here; see for instance the discussion in Remark 1.11. Also, the perspective of $C_{-*}(S^1)$–modules taken here makes it simpler to talk about (and describe) all cyclic homology theories at once, as well to study the compatibility of additional structures, eg exact sequences, semi-infinite/noncommutative Hodge structures.

[3] In particular, it computes the usual equivariant symplectic cohomology $SH^*_{S^1}(M) = H^*(SC^*(M)_{\mathrm{h}S^1})$; see Bourgeois and Oancea [6] but note differing conventions, eg regarding homology vs cohomology.

Remark 1.8 There are other settings in which Fukaya categories are now well studied, for instance Fukaya categories of Lefschetz fibrations (and more general LG models), or more generally partially wrapped Fukaya categories (such as wrapped Fukaya categories of *Liouville sectors*). We do not discuss these situations in our paper, but expect suitable versions of Theorem 1.1 to hold in such settings too. We do note however that the target of the open–closed map from Hochschild homology in such settings is usually more subtle than in the cases discussed here, eg it does not typically have the structure of a unital ring.

Remark 1.9 One can consider variations on Theorem 1.1. As a notable example, let M denote a (noncompact) Liouville manifold, and $\mathcal{F}$ the Fukaya category of compact exact Lagrangians in M. Then there is a nontrivial refinement of the map $\mathrm{HH}_*(\mathcal{F}) \to H^*(M, \partial^\infty M)$, which can be viewed as a pairing $\mathrm{HH}_*(\mathcal{F}) \times H^*(M) \to \boldsymbol{k}$, to a pairing

$$\mathrm{CH}_*(\mathcal{F}) \otimes SC^*(M) \to \boldsymbol{k}.$$

(Symplectic cohomology does not satisfy Poincaré duality, so this is *not* equivalent to a map to symplectic cohomology.) Our methods also imply that this pairing admits an S^1–equivariant enhancement, with respect to the diagonal $C_{-*}(S^1)$–action on the left and the trivial action on the right. Passing to adjoints, we obtain cyclic open–closed maps from S^1–equivariant symplectic cohomology to cyclic *cohomology* groups of $\mathcal{F}$, and from cyclic homology of $\mathcal{F}$ to equivariant symplectic *homology*. See Section 5.6.2 for more details.

Beyond computing equivariant Floer cohomology groups in terms of cyclic homology theories, we describe in the following subsection two applications of Theorem 1.1 to the structure of Fukaya categories.

Remark 1.10 We anticipate additional concrete applications of Theorem 1.1 and its homological shadow, Theorem 5.14. For instance, one can study the compatibility of open–closed maps with *dilations* in the sense of Seidel and Solomon [58], which are elements B in $SH^*(M)$ satisfying $[\Delta]B = 1$ — the existence of dilations strongly constrains intersection properties of embedded Lagrangians; see Seidel [55]. Theorem 5.14, or rather the variant discussed in Remark 1.9, implies that *if there exists a dilation, eg an element $x \in SH^1(M)$ with $[\Delta]x = 1$, then on the Fukaya category of compact Lagrangians $\mathcal{F}$, there exists $x' \in (\mathrm{HH}_{n+1}(\mathcal{F}))^\vee$ with $x' \circ [B] = [\mathrm{tr}]$*, where tr is the geometric weak proper Calabi–Yau structure on the Fukaya category; see Section 1.1.

1.1 Calabi–Yau structures on the Fukaya category

Calabi–Yau structures are a type of cyclically symmetric duality structure on a dg or A_∞ category $\mathcal{C}$ generalizing the notion of a nowhere-vanishing holomorphic volume form on a complex algebraic variety X in the case $\mathcal{C} = \mathrm{perf}(X)$. As is well understood, there are two (in some sense dual) types of Calabi–Yau structures on A_∞ categories:

(1) **Proper Calabi–Yau structures** (Kontsevich and Soibelman [37]) These can be associated to *proper* categories $\mathcal{C}$ (those which have cohomologically finite-dimensional morphism spaces), abstract and refine the notion of integration against a nowhere-vanishing holomorphic volume form. For $\mathcal{C} = \mathrm{perf}(X)$ with X a proper n–dimensional variety, the resulting structure in particular induces the Serre duality pairing with trivial canonical sheaf $\mathrm{Ext}^*(\mathcal{E},\mathcal{F}) \times \mathrm{Ext}^*(\mathcal{F},\mathcal{E}) \to \boldsymbol{k}[-n]$. Roughly, a proper Calabi–Yau structure on $\mathcal{C}$ (of dimension n) is a map $[\widetilde{\mathrm{tr}}]\colon \mathrm{HC}^+_*(\mathcal{C}) \to \boldsymbol{k}[-n]$ satisfying a nondegeneracy condition.

(2) **Smooth Calabi–Yau structures** (Kontsevich, Takeda and Vlassopoulos [39]) These can be associated to *smooth* categories $\mathcal{C}$ (those with perfect diagonal bimodule), and abstract the notion of the nowhere-vanishing holomorphic volume form itself, along with the induced identification (by contraction against the volume form) of polyvectorfields with differential forms. Roughly, a smooth Calabi–Yau structure on $\mathcal{C}$ (of dimension n) is a map $[\widetilde{\mathrm{cotr}}]\colon \boldsymbol{k}[n] \to \mathrm{HC}^-_*(\mathcal{C})$, or equivalently an element $[\tilde{\sigma}]$ or "$[\mathrm{vol}_{\mathcal{C}}]$" in $\mathrm{HC}^-_{-n}(\mathcal{C})$, satisfying a nondegeneracy condition.

In both cases, the nondegeneracy condition can be phrased purely in terms of the underlying nonequivariant shadow of the map, eg in the first case on the induced map $[\mathrm{tr}]\colon \mathrm{HH}_*(\mathcal{C}) \to \mathrm{HC}^+(\mathcal{C}) \xrightarrow{[\widetilde{\mathrm{tr}}]} \boldsymbol{k}[-n]$. Precise definitions are reviewed in Section 6. When $\mathcal{C}$ is simultaneously smooth and proper, it is a folk result that the notions are equivalent; see [27, Proposition 6.10].

In general, Calabi–Yau structures may not exist and when they do, there may be a nontrivial space of choices; see Menichi [45] for an example. A Calabi–Yau structure in either form induces nontrivial identifications between Hochschild invariants of the underlying category $\mathcal{C}$.[4] Moreover, categories with Calabi–Yau structures (should) carry induced 2–dimensional chain level TQFT operations on their Hochschild homology

[4] In the proper case, there is an induced isomorphism between Hochschild cohomology and the linear dual of Hochschild homology. In the smooth case, there is an isomorphism between Hochschild cohomology and homology without taking duals.

chain complexes, associated to moduli spaces of Riemann surfaces with marked points; see Costello [14] and Kontsevich and Soibelman [37] in the proper case, and Kontsevich, Takeda, and Vlassopoulos in the smooth case [39; 38]. If the category is proper and nonsmooth (resp. smooth nonproper) the resulting TQFT is incomplete in that every operation must have at least one input (resp. output). In the smooth and proper case in particular, Calabi–Yau structures play a central role in the mirror symmetry motivated question of recovering Gromov–Witten invariants from the Fukaya category and to the related question of categorically recovering Hamiltonian Floer homology with all of its (possibly higher homotopical) operations. See Costello [14; 15] and Kontsevich [35] for work around these questions in the setting of abstract topological field theories, and Ganatra, Perutz and Sheridan [27] for applications of Calabi–Yau structures to recovering genus-0 Gromov–Witten invariants from the Fukaya category.

Remark 1.11 A closely related to (1), and well studied, notion is that of a *cyclic A_∞ category*: this is an A_∞ category $\mathcal{C}$ equipped with a chain level perfect pairing

$$\langle -,-\rangle\colon \hom(X,Y)\times\hom(Y,X)\to \boldsymbol{k}[-n]$$

such that the induced correlation functions

$$\langle \mu^d(-,-,\dots,-),-\rangle$$

are strictly (graded) cyclically symmetric for each d; see for instance Cho and Lee [9], Costello [14] and Fukaya [20]. Although the property of being a cyclic A_∞ structure is not a homotopy-invariant notion (ie not preserved under A_∞ quasi-equivalences), cyclic A_∞ categories and proper Calabi–Yau structures turn out to be weakly equivalent *in characteristic* 0, in the following sense. Any cyclic A_∞ category carries a canonical proper Calabi–Yau structure, and Kontsevich and Soibelman [37, Theorem 10.7] proved that a proper Calabi–Yau structure on any A_∞ category $\mathcal{C}$ determines a (canonical up to quasi-equivalence) quasi-isomorphism between $\mathcal{C}$ and a cyclic A_∞ category $\widetilde{\mathcal{C}}$. When $\operatorname{char}(\boldsymbol{k})\neq 0$, the two notions of proper Calabi–Yau and cyclic A_∞ differ in general, due to group cohomology obstructions to imposing cyclic symmetry. In such instances, it seems that the notion of a proper Calabi–Yau structure is the "correct" one (as it is a homotopy-invariant notion and, by Theorem 1.12, the compact Fukaya category always has one).

As a first application of Theorem 1.1, we verify the longstanding expectation that various compact Fukaya categories possess geometrically defined canonical Calabi–Yau structures.

Theorem 1.12 *The Fukaya category of compact Lagrangians has, under technical hypotheses* ($\star$), *a canonical geometrically defined proper Calabi–Yau structure over any ground field $\boldsymbol{k}$ (over which the Fukaya category and $\widetilde{\mathcal{OC}}$ are defined).*

In fact, this proper Calabi–Yau structure is easy to describe in terms of the cyclic open–closed map (cf Corollary 1.5): it is the composition of the map[5]

$$\widetilde{\mathcal{OC}}^+ \colon \mathrm{HC}_*^+(\mathcal{F}) \to H^{*+n}(M, \partial M)((u))/uH^{*+n}(M, \partial M)[[u]]$$

with the linear map to $\boldsymbol{k}$ which sends the top class $\mathrm{PD}(\mathrm{pt}) \cdot u^0 \in H^{2n}(M, \partial M)$ to 1, and all other generators $\alpha \cdot u^{-i}$ to 0. See Section 6 for more details.

As a consequence of the discussion in Remark 1.11, specifically [37, Theorem 10.7], we deduce that

Corollary 1.13 *If* $\mathrm{char}(\boldsymbol{k}) = 0$, *then any Fukaya category of compact Lagrangians satisfying* ($\star$) *admits a (canonical up to equivalence) cyclic A_∞ (minimal) model.*

Remark 1.14 In the case of compact symplectic manifolds and over $\boldsymbol{k} =$ a Novikov field containing $\mathbb{R}$, Fukaya [20] constructed a cyclic A_∞ model of the Floer cohomology algebra of a single compact Lagrangian, which will be extended to multiple objects by Abouzaid, Fukaya, Oh, Ohta and Ono [2].

Remark 1.15 In order to construct (chain level) 2d–TFTs on the Hochschild chain complexes of categories, Kontsevich and Soibelman [37] partly show (on the closed sector) that a proper Calabi–Yau structure can be used instead of the (weakly equivalent in characteristic 0) cyclic A_∞ structures considered in Costello [14]. One might similarly hope that, for applications of cyclic A_∞ structures to disc-counting/open Gromov–Witten invariants developed in Fukaya [21], a proper Calabi–Yau structure is in fact sufficient. See Cho and Lee [9] for related work.

Turning to smooth Calabi–Yau structures, in Section 6.2, we will establish the following existence result for smooth Calabi–Yau structures, which applies to wrapped Fukaya categories of noncompact (Liouville) manifolds as well as to Fukaya categories of compact manifolds.

[5] Recall that $C^*(M, \partial M)$ has the trivial $C_{-*}(S^1)$–module structure; the homology of the associated homotopy orbit complex is $H^{*+n}(M, \partial M)((u))/uH^{*+n}(M, \partial M)[[u]]$, where $|u| = 2$, as described in Section 2.

Theorem 1.16 *Under the technical hypotheses* ($\star$), *suppose further that our symplectic manifold M is* ***nondegenerate*** *in the sense of* [24], *meaning that the map* $[\mathcal{OC}]\colon \mathrm{HH}_{*-n}(\mathcal{F}) \to HF^*(M)$ *hits the unit* $1 \in HF^*(M)$. *Then, its* (*compact or wrapped*) *Fukaya category $\mathcal{F}$ possesses a canonical, geometrically defined* ***strong smooth Calabi–Yau structure****.*

Once more, the cyclic open–closed map gives an efficient description of this structure: it is the unique element $\mathrm{HC}^{-}_{-n}(\mathcal{F})$ mapping via $\widetilde{\mathcal{OC}}^{-}$ to the geometrically canonical lift $\tilde{1} \in H^*(CF^*(M)^{\mathrm{h}S^1})$ of the unit $1 \in CF^*(M)$ described in Section 4.4.[6]

Remark 1.17 In contrast to compact Fukaya categories or wrapped Fukaya categories of Liouville manifolds, the Fukaya categories of noncompact Lagrangians discussed in Remark 1.8 are typically not Calabi–Yau in either sense,[7] even if they are smooth or proper categories; indeed they typically arise as homological mirrors to perfect/coherent complexes on non-Calabi–Yau varieties. Instead, one might expect such categories to admit *pre-Calabi–Yau structures* in the sense of Kontsevich, Takeda and Vlassopoulos [38] (see also Yeung [65] and Seidel [57] for a construction of related structures), or *relative Calabi–Yau structures* in the sense of Brav and Dyckerhoff [7].

The notion of a smooth Calabi–Yau structure, or sCY structure, will be studied further in forthcoming joint work with R Cohen [12], and used to compare the wrapped Fukaya category of a cotangent bundle and string topology category of its zero section as *categories with sCY structures* (in order to deduce a comparison of topological field theories on both sides).

1.2 Noncommutative Hodge–de Rham degeneration for smooth and proper Fukaya categories

For a $C_{-*}(S^1)$–module P, there is a canonical Tor spectral sequence converging to $H^*(P_{\mathrm{h}S^1})$ with first page $H^*(P) \otimes_{\boldsymbol{k}} H^*(\boldsymbol{k}_{\mathrm{h}S^1}) \cong H^*(P) \otimes_{\boldsymbol{k}} H_*(\mathbb{CP}^\infty)$. When applied to the Hochschild complex $P = \mathrm{CH}_*(\mathcal{C})$ of a (dg/A_∞) category $\mathcal{C}$, the resulting spectral sequence, from (many copies of) $\mathrm{HH}_*(\mathcal{C})$ to $\mathrm{HC}^+(\mathcal{C})$ is called the *Hochschild-to-cyclic* or *noncommutative Hodge–de Rham* (*ncHDR*) *spectral sequence.* The latter name comes from the fact that when $\mathcal{C} = \mathrm{perf}(X)$ is perfect complexes on a complex

[6] As shown in [24; 27], if $[\mathcal{OC}]$ hits 1, then $[\mathcal{OC}]$ is an isomorphism, and hence by Corollary 1.5, $[\widetilde{\mathcal{OC}}^{-}]$ is too. Hence one can speak about the unique element.

[7] One manifestation of this is the failure of the target of the open–closed map to have a distinguished unit element, as also discussed in Remark 1.8.

variety X, this spectral sequence is equivalent (via Hochschild–Kostant–Rosenberg (HKR) isomorphisms) to the usual Hodge-to-de Rham spectral sequence from Hodge cohomology to de Rham cohomology

$$H^*(X, \Omega_X^*) \Rightarrow H^*_{\mathrm{dR}}(X),$$

which degenerates (as we are in characteristic 0) whenever X is smooth and proper. Motivated by this, Kontsevich [37; 35] formulated the *noncommutative Hodge–de Rham* (*ncHDR*) *degeneration conjecture*: for any smooth and proper category $\mathcal{C}$ in characteristic 0, its ncHDR spectral sequence degenerates. A general proof of this fact for $\mathbb{Z}$–graded categories was recently given by Kaledin [30; 29], following earlier work establishing it in the coconnective case.

Using the cyclic open–closed map, we can give a purely symplectic proof of the ncHDR degeneration property for those smooth and proper $\mathcal{C}$ arising as Fukaya categories, including in non-$\mathbb{Z}$–graded cases:

Theorem 1.18 *Let $\mathcal{A} \subset \mathcal{F}(M)$ be a smooth and proper subcategory of any Fukaya category of any compact symplectic manifold satisfying the technical assumptions* ($\star$), *over any field $\boldsymbol{k}$* (*over which the Fukaya category and the cyclic open–closed map are defined*). *Then*, *the noncommutative Hodge–de Rham spectral sequence for $\mathcal{A}$ degenerates.*

Proof The noncommutative Hodge–de Rham spectral sequence for $\mathcal{A}$ degenerates at page 1 if and only if $P = CH_*(\mathcal{A})$ is isomorphic (in the category of $C_{-*}(S^1)$–modules) to a trivial $C_{-*}(S^1)$–module, for instance, if the $C_{-*}(S^1)$–action is trivializable; see Dotsenko, Shadrin and Vallette [16, Theorem 2.1]. For compact symplectic manifolds M, recall that $CF^*(M) \cong H^*(M)$ has a canonically trivial(izable) $C_{-*}(S^1)$–action. (See Corollary 4.16; this comes from, for instance, the fact that we can choose a C^2–small Hamiltonian to compute the complex, all of the orbits of which are constant loops on which geometric rotation acts trivially. Or more directly, we can modify the definition of $\widetilde{\mathcal{OC}}$ to give a map directly to $H^*(M)$ with its trivial $C_{-*}(S^1)$–action, as described in Section 5.6.1.)

By earlier work [27; 25], whenever $\mathcal{A}$ is smoooth, $\mathcal{OC}|_{\mathcal{A}}$ is an isomorphism from $\mathrm{HH}_{*-n}(\mathcal{A})$ onto a nontrivial summand S of $HF^*(M) \cong \mathrm{QH}^*(M)$; the $C_{-*}(S^1)$–action on this summand is trivial too. Theorem 1.1 shows that $\widetilde{\mathcal{OC}}|_{\mathcal{A}}$ induces an isomorphism in the category of $C_{-*}(S^1)$–modules between $\mathrm{CH}_*(\mathcal{A})$ and $S[n]$ with its trivial action, so we are done. □

Remark 1.19 Theorem 1.18 holds for a field $\boldsymbol{k}$ of any characteristic over which the Fukaya category and relevant structures (satisfy ($\star$) and) are defined, for any grading structure that can be defined on the given Fukaya category; eg it holds for the $\mathbb{Z}/2$–graded Fukaya category of a monotone symplectic manifold over a field of any characteristic. In contrast, for an arbitrary smooth and proper $\mathbb{Z}/2$–graded dg category in characteristic zero, the noncommutative Hodge–de Rham degeneration is not yet established (though it is expected). And it is not always true in finite characteristic.

An incomplete explanation for the degeneration property holding for finite characteristic smooth and proper Fukaya categories may be that the Fukaya category over a characteristic p field $\boldsymbol{k}$ (whenever Lagrangians are monotone or tautologically unobstructed at least) may always admit a lift to second Witt vectors $W_2(\boldsymbol{k})$.[8]

As is described in joint work (partly ongoing) with Perutz and Sheridan [27; 26], the cyclic open–closed map $\widetilde{\mathcal{OC}}^-$ can further be shown to be a *morphism of semi-infinite Hodge structures*, a key step (along with the above degeneration property and construction of Calabi–Yau structure) in recovering Gromov–Witten invariants from the Fukaya category and enumerative mirror predictions from homological mirror theorems.

1.3 Outline of paper

In Section 2, we recall a convenient model for the category of A_∞–modules over $C_{-*}(S^1)$ and various equivariant homology functors from this category. In Section 3, we review the (compact and wrapped) Fukaya category along with $C_{-*}(S^1)$–action on its (and more generally, any cohomologically unital A_∞ category's) *nonunital Hochschild chain complex* (a variant on usual cyclic bar complex that has usually appeared in the symplectic literature, eg in Abouzaid [1]). In Section 4, we recall the construction of the A_∞ $C_{-*}(S^1)$–module structure on the (Hamiltonian) Floer chain complex, following Bourgeois and Oancea [6] and Seidel [51]; note that our technical setup is slightly different, though equivalent. Then we prove our main results in Section 5. Some technical and conceptual variations on the construction of $\widetilde{\mathcal{OC}}$ (including Remark 1.9) are discussed at the end of this section; see Section 5.6. Finally, in Section 6 we apply our results to construct proper and smooth Calabi–Yau structures, proving Theorems 1.12 and 1.16.

[8] The author wishes to thank Mohammed Abouzaid for discussions regarding this point.

1.4 Conventions

We work over a ground field $\boldsymbol{k}$ of arbitrary characteristic, though we note that all of our geometric constructions are valid over an arbitrary ring, eg $\mathbb{Z}$. All chain complexes will be graded *cohomologically*, including singular chains of any space, which hence have negative the homological grading and are denoted by $C_{-*}(X)$. All gradings are either in $\mathbb{Z}$ or $\mathbb{Z}/2$ (in the latter case, degrees of maps are implicitly mod 2).

Acknowledgements I'd like to thank Paul Seidel for a very helpful conversation and Nick Sheridan for several helpful discussions about technical aspects of this paper such as signs. I would also like to thank Zihong Chen and Yasin Uskuplu for corrections. Part of this work was revised during a visit at the Institut Mittag-Leffler in 2015, which I'd like to thank for its hospitality. Finally, I'd like to thank a referee for a number of helpful suggestions, comments and corrections which improved the exposition of this article.

The author was partially supported by the National Science Foundation through a postdoctoral fellowship (grant DMS-1204393) and agreement DMS-1128155. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

2 Complexes with circle action

In this section, we review a convenient model for the category of A_∞ $C_{-*}(S^1)$–modules, for which the A_∞ $C_{-*}(S^1)$–action can be described by a single hierarchy of maps satisfying equations. We also describe various equivariant homology complexes in this language in terms of simple formulae. This model appears elsewhere in the literature as ∞*–mixed complexes* or S^1*–complexes* or *multicomplexes* (we will sometimes adopt the second term); see eg [6; 66; 16], but note that the first and third references use homological grading conventions.

2.1 Definitions

Let $C_{-*}(S^1)$ denote the dg algebra of chains on the circle with coefficients in $\boldsymbol{k}$, graded cohomologically, with multiplication induced by the Pontryagin product $S^1 \times S^1 \to S^1$. This algebra is *formal*, or quasi-isomorphic to its homology, an exterior algebra on one generator Λ of degree -1 with no differential. Henceforth, by abuse of notation we take this exterior algebra as our working model for $C_{-*}(S^1)$,

$$C_{-*}(S^1) := \boldsymbol{k}[\Lambda]/\Lambda^2, \quad \text{where } |\Lambda| = -1, \tag{2-1}$$

and use the terminology $C_{-*}^{\mathrm{sing}}(S^1)$ to refer to usual singular chains on S^1.

Definition 2.1 A *strict S^1–complex*, or a *chain complex with strict/dg S^1–action*, is a unital differential graded module over $\boldsymbol{k}[\Lambda]/\Lambda^2$.

Let (M, d) be a strict S^1–complex; by definition (M, d) is a cochain complex (recall our conventions for complexes from Section 1.4) and the unital dg $\boldsymbol{k}[\Lambda]/\Lambda^2$–module structure is equivalent to the data of the single additional operation of multiplying by Λ,

$$\Delta = \Lambda \cdot -\colon M_* \to M_{*-1}, \tag{2-2}$$

which must square to zero and anticommute with d. In other words, (M, d, Δ) is what is known as a *mixed complex*; see eg [8; 31; 41].

We will need to work with the weaker notion of an A_∞–action, or rather an A_∞–module structure over $C_{-*}(S^1) = \boldsymbol{k}[\Lambda]/\Lambda^2$. Recall that a *(left) A_∞–module M* [33; 52; 50; 24] over the associative graded algebra $A = \boldsymbol{k}[\Lambda]/\Lambda^2$ is a graded $\boldsymbol{k}$–module M equipped with maps

$$\mu^{k|1}\colon A^{\otimes k} \otimes M \to M, \quad \text{for } k \geq 0, \tag{2-3}$$

of degree $1-k$, satisfying the A_∞–module equations described in [50] or [24, (2.35)]. Since $A = \boldsymbol{k}[\Lambda]/\Lambda^2$ is unital, we can work with modules that are also *strictly unital* (see [50, (2.6)]); this implies that all multiplications by a sequence with at least one unit element is completely specified,[9] and hence the only nontrivial structure maps to define are the operators

$$\delta_k := \mu_M^{k|1}(\underbrace{\Lambda, \dots, \Lambda}_{k \text{ copies}}, -)\colon M \to M[1-2k] \quad \text{for } k \geq 0. \tag{2-4}$$

The A_∞–module equations are equivalent to the relations

$$\sum_{i=0}^{s} \delta_i \delta_{s-i} = 0 \tag{2-5}$$

for (2-4), for each $s \geq 0$. We summarize the discussion so far with the following definition.

[9] More precisely, $\mu^{1|1}(1, \boldsymbol{m}) = \boldsymbol{m}$ and $\mu^{k|1}(\dots, 1, \dots, \boldsymbol{m}) = 0$ for $k > 1$.

Definition 2.2 An *S^1–complex*, or a *chain complex with an A_∞ S^1–action*, is a strictly unital (left) A_∞–module M over $\boldsymbol{k}[\Lambda]/\Lambda^2$. Equivalently, it is a graded $\boldsymbol{k}$–module M equipped with operations $\{\delta_k\colon M \to M[1-2k]\}_{k\geq 0}$ satisfying, for each $s \geq 0$, the hierarchy of equations (2-5).

Remark 2.3 If X is a topological space with S^1–action, then $C_{-*}(X)$ carries a dg $C_{-*}^{\mathrm{sing}}(S^1)$–module structure, with module action induced by the action $S^1 \times X \to X$. Under the A_∞ equivalence $C_{-*}^{\mathrm{sing}}(S^1) \cong \boldsymbol{k}[\Lambda]/\Lambda^2$, it follows that $C_{-*}(X)$ carries an A_∞ (not necessarily dg) $\boldsymbol{k}[\Lambda]/\Lambda^2$–module structure, which can further be made strictly unital, by [40, Theorem 3.3.1.2] or by passing to normalized chains. If one wishes, one can then appeal to abstract strictification results to produce a dg $\boldsymbol{k}[\Lambda]/\Lambda^2$–module which is quasi-isomorphic as A_∞ $\boldsymbol{k}[\Lambda]/\Lambda^2$–modules to $C_{-*}(X)$. More directly, it turns out [12] that one can find an equivalent dg $\boldsymbol{k}[\Lambda]/\Lambda^2$–module by taking a suitable quotient of the normalized singular chain complex $C_{-*}(X)$ to form *unordered normalized singular chains* of X (identifying simplices differing by permuting vertices and quotienting by those that are degenerate).

Remark 2.4 There are multiple sign conventions for A_∞–modules over an A_∞–algebra; the most common two conventions appear in [50, (2.6)] and [52, (1j)], as well as many other places. These conventions are completely irrelevant for strictly unital $A = \boldsymbol{k}[\Lambda]/\Lambda^2$–modules, as the reduced degree of any element in $\bar{A} = \mathrm{span}_{\boldsymbol{k}}(\Lambda)$ is zero; hence the (Koszul) signs in various formulae are $+1$ in either convention.

For $s=0$, equation (2-5) says simply that the differential $d = \delta_0$ squares to 0; for $s=1$, equation (2-5) implies $\delta := \delta_1$ anticommutes with d, and for $s=2$, $(\delta)^2 = -(d\delta_2 + \delta_2 d)$, or that δ^2 is chain-homotopic to zero, but not strictly zero, as measured by the chain homotopy δ_2.

S^1–complexes, as strictly unital A_∞–modules over the augmented algebra $A = \boldsymbol{k}[\Lambda]/\Lambda^2$, are the objects of a dg category, which we will call

$$S^1\text{–mod} := uA\text{–mod}, \tag{2-6}$$

whose morphisms and compositions we now recall.[10] Denote by $\epsilon\colon A \to \boldsymbol{k}$ the augmentation map, and $\bar{A} = \ker\epsilon = \mathrm{span}_{\boldsymbol{k}}(\Lambda)$ the augmentation ideal. Let M and N be two strictly unital A_∞ A–modules. A *unital premorphism of degree k* from M to N

[10] For the definition of this category compare [50, pages 90, 94], where it is called $\mathrm{mod}(A) = \mathrm{mod}(A, \boldsymbol{k})$.

is a collection of maps $F^{d|1}\colon \bar{A}^{\otimes d} \otimes M \to N$ for $d \geq 0$, of degree $k-d$, or equivalently, since $\dim_{\boldsymbol{k}}(\bar{A}) = 1$ in degree -1, a collection of operators

$$F = \{F^d\}_{d \geq 0}, \quad F^d := F^{d|1}(\underbrace{\Lambda, \dots, \Lambda}_{d \text{ copies}}, -)\colon M \to N[k-2d]. \tag{2-7}$$

If $T(\bar{A}[1]) = \bigoplus_{d \geq 0} \bar{A}[1]^{\otimes d}$ denotes the tensor algebra of $\bar{A}[1]$, then F can be alternatively packaged into the data of a single map $F := \bigoplus_{d \geq 0} F^d\colon T\bar{A}[1] \otimes M \to N$ of degree k. The space of premorphisms of each degree form the graded space of morphisms in S^1–mod, which we will denote by $\mathrm{Rhom}_{S^1}(-,-)$:

$$\begin{aligned}\mathrm{Rhom}_{S^1}(M,N) &:= \bigoplus_{k \in \mathbb{Z}} \mathrm{Rhom}^k_{S^1}(M,N) \\ &:= \bigoplus_{k \in \mathbb{Z}} \hom_{\mathrm{grVect}}(T(\bar{A}[1]) \otimes M, N[k]) \\ &= \biggl(\bigoplus_{k \in \mathbb{Z}} \hom_{\mathrm{grVect}}\biggl(\bigoplus_{d \geq 0} M[2d], N[k]\biggr)\biggr).\end{aligned} \tag{2-8}$$

There is a differential ∂ on (2-8) described in [50, page 90]; in terms of the simplified form of premorphisms (2-7), one has

$$(\partial F)^s = \sum_{i=0}^{s} F^i \circ \delta^M_{s-i} - (-1)^{\deg(F)} \sum_{j=0}^{s} \delta^N_{s-j} \circ F^j. \tag{2-9}$$

An A_∞ $\boldsymbol{k}[\Lambda]/\Lambda^2$*–module homomorphism*, or S^1*–complex homomorphism*, is a premorphism $F = \{F^d\}$ which is closed, ie $\partial F = 0$. In particular, F is an A_∞–module homomorphism if the following equations are satisfied for each s:

$$\sum_{i=0}^{s} F^i \circ \delta^M_{s-i} = (-1)^{\deg(F)} \sum_{j=0}^{s} \delta^N_{s-j} \circ F^j. \tag{2-10}$$

Note that the $s = 0$ equation reads $F^0 \circ \delta^M_0 = (-1)^{\deg(F)} \delta^N_0 \circ F^0$, so (if $\partial F = 0$) F^0 induces a cohomology level map $[F^0]\colon H^*(M) \to H^{*+\deg(F)}(N)$. A module homomorphism (or closed morphism) F is said to be a *quasi-isomorphism* if $[F^0]$ is an isomorphism on cohomology. A *strict* module homomorphism F is one for which $F^k = 0$ for $k > 0$.

Remark 2.5 There is an enlarged notion of a *nonunital* premorphism (used for modules which are not necessarily strictly unital), which is a collection of maps $\{\hat{F}^d\colon A^{\otimes d} \otimes M \to N\}_d$ instead of $\{F^d\colon \bar{A}^{\otimes d} \otimes M \to N\}_d$. Any premorphism

$F = \{F^d\}_d$ as we have defined it extends to a nonunital premorphism $\widehat{F} = \{\widehat{F}^d\}$ by declaring $\widehat{F}^d(\dots, 1, \dots, \boldsymbol{m}) = 0$. For strictly unital modules, the resulting inclusion from the complex of premorphisms to the complex of nonunital premorphisms is a quasi-isomorphism.

Remark 2.6 When M and N are dg modules, or strict S^1–complexes, the complex $\operatorname{Rhom}_{S^1}(M, N)$ is a *reduced bar model* of the chain complex of derived $\boldsymbol{k}[\Lambda]/\Lambda^2$–module homomorphisms, which is one of the reasons we have adopted the terminology “Rhom”. In the A_∞ setting, we recall that there is no sensible “nonderived” notion of a $\boldsymbol{k}[\Lambda]/\Lambda^2$–module map; compare [50].

The composition in the category S^1–mod,

$$\operatorname{Rhom}_{S^1}(N, P) \otimes \operatorname{Rhom}_{S^1}(M, N) \to \operatorname{Rhom}_{S^1}(M, P), \tag{2-11}$$

is defined by

$$(G \circ F)^s = \sum_{j=0}^{s} G^{s-j} \circ F^j. \tag{2-12}$$

Remark 2.7 If M is any S^1–complex, then its endomorphisms $\operatorname{Rhom}_{S^1}(M, M)$, equipped with composition, form a dg algebra. As an example, consider $M = \boldsymbol{k}$, with trivial module structure determined by the augmentation $\epsilon\colon \boldsymbol{k}[\Lambda]/\Lambda^2 \to \boldsymbol{k}$. It is straightforward to compute that, as a dga,

$$\operatorname{Rhom}_{S^1}(\boldsymbol{k}, \boldsymbol{k}) \cong \boldsymbol{k}[u], \quad \text{with } |u| = 2. \tag{2-13}$$

In terms of the definition of morphism spaces (2-8), u corresponds to the unique morphism $G = \{G^d\}_{d \geq 0}$ of degree $+2$ with $G^1 = \mathrm{id}$ and $G^s = 0$ for $s \neq 1$.

In addition to taking the morphism spaces, one can define the (derived) *tensor product* of S^1–complexes N and M: using the isomorphism $A \cong A^{\mathrm{op}}$ coming from commutativity of $A = \boldsymbol{k}[\Lambda]/\Lambda^2$, first view N as a *right A_∞ A–module* (see [50, pages 90, 94], where the category of right A–modules is called $\operatorname{mod}(\boldsymbol{k}, A)$, see also [52, (1j)] and [24, Section 2]), and then take the usual (necessarily derived) tensor product of N and M over A (see [50, page 91] or [24, Section 2.5]). The resulting chain complex — which we will, by abuse of notation, indicate as the derived tensor product over S^1 — has underlying graded vector space

$$N \otimes_{S^1}^{\mathbb{L}} M := N \otimes_A^{\mathbb{L}} M := \bigoplus_{d \geq 0} N \otimes \bar{A}[1]^{\otimes d} \otimes M = \bigoplus_{d \geq 0} (N \otimes_{\boldsymbol{k}} M)[2d], \tag{2-14}$$

where the degree s part is $\bigoplus_{d\geq 0}\bigoplus_t N_t\otimes M_{s+2d-t}$. Let us refer to an element $n\otimes m$ of the d^{th} summand of this complex by suggestive notation

$$n\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d\text{ times}}\otimes m$$

as in the first line of (2-14). With this notation, the differential on (2-14) acts as

$$\begin{aligned}\text{(2-15)}\quad &\partial(n\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d}\otimes m)\\ &=\sum_{i=0}^{d}\bigl((-1)^{|m|}\delta_i^N n\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d-i}\otimes m+n\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d-i}\otimes\delta_i^M m\bigr).\end{aligned}$$

Here our sign convention follows [24, Section 2.5] rather than [50], though the sign difference is minimal.

Remark 2.8 Analogously to Remark 2.6, if M and N are unital dg modules over $A=\boldsymbol{k}[\Lambda]/\Lambda^2$, the chain complex described above computes their derived tensor product, whose homology is $\operatorname{Tor}_A(M,N)$. While we have therefore opted for the notation $N\otimes_A^{\mathbb{L}}M$, or rather the abbreviation $N\otimes_{S^1}^{\mathbb{L}}M$, we note that the (derived) tensor product of A_∞–modules is often written in the A_∞ literature without the superscript $\mathbb{L}$ as simply $N\otimes_A M$; compare [50, equation (2.6)].

The pairing (2-14) is suitably functorial with respect to morphisms of the S^1–complexes involved, meaning that $-\otimes_{S^1}N$ and $M\otimes_{S^1}-$ both induce dg functors from S^1–mod to chain complexes; compare [50, page 92]. For instance, if $F=\{F^j\}\colon M_0\to M_1$ is a premorphism of S^1–complexes, then there are induced maps

$$\begin{aligned}&F_\sharp\colon N\otimes_{S^1}^{\mathbb{L}}M_0\to N\otimes_{S^1}^{\mathbb{L}}M_1,\\ \text{(2-16)}\quad &n\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d}\otimes m\mapsto\sum_{j=0}^{d}n\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d-j}\otimes F^j(m),\end{aligned}$$

$$\begin{aligned}&F_\sharp\colon M_0\otimes_{S^1}^{\mathbb{L}}N\to M_1\otimes_{S^1}^{\mathbb{L}}N,\\ \text{(2-17)}\quad &m\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d}\otimes n\mapsto\sum_{j=0}^{d}(-1)^{\deg(F)\cdot|n|}F^j(m)\otimes\underbrace{\Lambda\otimes\cdots\otimes\Lambda}_{d-j}\otimes n,\end{aligned}$$

which are chain maps if $\partial(F)=0$.

Hom and tensor complexes of S^1–complexes, as in any category of A_∞–modules, satisfy the following strong homotopy-invariance properties.

Proposition 2.9 (homotopy invariance) *If $F\colon M \to M'$ is any quasi-isomorphism of S^1–complexes (meaning $\partial(F)=0$ and $[F^0]\colon H^*(M) \xrightarrow{\cong} H^*(M')$ is an isomorphism), then composition with F induces quasi-isomorphisms of hom and tensor complexes*:

$$\begin{aligned} F\circ\cdot : \mathrm{Rhom}_{S^1}(M', P) &\xrightarrow{\sim} \mathrm{Rhom}_{S^1}(M, P), \\ \cdot\circ F\colon \mathrm{Rhom}_{S^1}(P, M) &\xrightarrow{\sim} \mathrm{Rhom}_{S^1}(P, M'), \\ F_\sharp\colon N \otimes^{\mathbb{L}}_{S^1} M &\xrightarrow{\sim} N \otimes^{\mathbb{L}}_{S^1} M', \\ F_\sharp\colon M \otimes^{\mathbb{L}}_{S^1} N &\xrightarrow{\sim} M' \otimes^{\mathbb{L}}_{S^1} N. \end{aligned} \tag{2-18}$$

The proof is a standard argument (though we do not know a specific reference): one exhibits acyclicity of the cone of each of the above maps by studying the spectral sequence with respect to the length filtration (with respect to the number of $\bar{A}^{\otimes d}$ factors in the bar model of the complexes); the first page of the associated spectral sequence is the cone of the map associated to the derived homs/tensor products of the associated homology-level modules by the homology level map $[F^0]$, which is acyclic by hypothesis; hence the second page vanishes and the cone is acyclic; compare analogous arguments in [52, Lemma 2.12] or [24, Proposition 2.2].

Let $(P, \{\delta_i^P\})$ and $(Q, \{\delta_j^Q\}_j)$ be S^1–complexes, and $f\colon P \to Q$ a chain map of some degree $\deg(f)$ (with respect to the δ_0^P and δ_0^Q differentials). An *S^1–equivariant enhancement of f* is a degree $\deg(f)$ homomorphism $\boldsymbol{F} = \{\boldsymbol{F}^i\}_{i\geq 0}$ of S^1–complexes — eg a closed morphism, so $\boldsymbol{F}$ satisfies (2-10) — with $[\boldsymbol{F}^0] = [f]$.

Remark 2.10 There are a series of obstructions to the existence of an S^1–equivariant enhancement of a given chain map f; for instance, a first necessary condition is the vanishing of the cohomology class $[f]\circ[\delta_1^P] - [\delta_1^Q]\circ[f]$.

Finally, we note that, just as the product of S^1–spaces $X\times Y$ possesses a diagonal action, the (linear) tensor product of S^1–complexes is again an S^1–complex.

Lemma 2.11 *If*

$$\left(M, \delta_{\mathrm{eq}}^M = \sum_{i=0}^{\infty} \delta_j^M u^j\right) \quad \textit{and} \quad \left(N, \delta_{\mathrm{eq}}^N = \sum_{i=0}^{\infty} \delta_i^N u^i\right)$$

are S^1–complexes, then the graded vector space $M\otimes N$ is naturally an S^1–complex with $\delta_{\mathrm{eq}}^{M\otimes N} = \sum_{i=0}^{\infty} \delta_k^{M\otimes N} u^k$, where

$$\delta_k^{M\otimes N}(\boldsymbol{m}\otimes\boldsymbol{n}) := (-1)^{|\boldsymbol{n}|}\delta_k^M \boldsymbol{m}\otimes\boldsymbol{n} + \boldsymbol{m}\otimes\delta_k^N\boldsymbol{n}. \tag{2-19}$$

We call the resulting S^1–action on $M\otimes N$ the *diagonal S^1–action.*

Proof We compute

$$\begin{aligned}\text{(2-20)}\quad &\delta_j^{M\otimes N}\delta_k^{M\otimes N}(\boldsymbol{m}\otimes\boldsymbol{n})\\ &=\delta_j^M\delta_k^M\boldsymbol{m}\otimes\boldsymbol{n}+(-1)^{|\boldsymbol{n}|+1}\delta_j^M\boldsymbol{m}\otimes\delta_k^N\boldsymbol{n}+(-1)^{|\boldsymbol{n}|}\delta_k^M\boldsymbol{m}\otimes\delta_j^N\boldsymbol{n}+\boldsymbol{m}\otimes\delta_j^N\delta_k^N\boldsymbol{n}.\end{aligned}$$

Summing over all $j+k=s$, the middle two terms cancel in pairs and the sums of the leftmost terms (resp. rightmost) terms respectively vanish because M (resp. N) is an S^1–complex. □

Definition 2.12 Let $M := (M, d)$ be a chain complex over $\boldsymbol{k}$. The pullback of M along the (augmentation) map $\boldsymbol{k}[\Lambda]/\Lambda^2 \to \boldsymbol{k}$ is called the *trivial* S^1*–complex*, or *chain complex with trivial* S^1*–action* associated to M, and denoted by $\underline{M}^{\text{triv}}$. Concretely, $\underline{M}^{\text{triv}} := (M, \delta_0 = d, \delta_k = 0 \text{ for } k > 0)$.

2.2 Equivariant homology groups

Let M be an S^1–complex. Let $\boldsymbol{k} = \underline{\boldsymbol{k}}^{\text{triv}}$ denote the strict trivial rank-1 S^1–complex concentrated in degree 0.

Definition 2.13 The *homotopy orbit complex* of M is the (derived) tensor product of M with $\boldsymbol{k}$ over $C_{-*}(S^1)$:

$$M_{\mathrm{h}S^1} := \boldsymbol{k} \otimes^{\mathbb{L}}_{S^1} M. \tag{2-21}$$

The (strict) morphism of S^1–complexes $\epsilon\colon \boldsymbol{k}[\Lambda]/\Lambda^2 \to \boldsymbol{k}$ (here $\boldsymbol{k}[\Lambda]/\Lambda^2$ comes equipped with structure maps $\delta_k = 0$ for $k \neq 1$, and $\delta_1 = \Lambda\cdot -)$ induces by functoriality a chain map from M to $M_{\mathrm{h}S^1}$ called the *projection to homotopy orbits*,

$$\mathrm{pr}\colon M \cong \boldsymbol{k}[\Lambda]/\Lambda^2 \otimes^{\mathbb{L}}_{S^1} M \to \boldsymbol{k} \otimes^{\mathbb{L}}_{S^1} M = M_{\mathrm{h}S^1}. \tag{2-22}$$

Remark 2.14 When $M = C_{-*}(X)$, with S^1–complex induced by a topological S^1–action on X as in Remark 2.3, the complex (2-21) computes the Borel equivariant homology of X, by the following reasoning: first, the A_∞ equivalence between $\boldsymbol{k}[\Lambda]/\Lambda^2$ and $C_{-*}^{\text{sing}}(S^1)$ induces an equivalence

$$M_{\mathrm{h}S^1} \simeq C_{-*}(\mathrm{pt}) \otimes^{\mathbb{L}}_{C_{-*}^{\text{sing}}(S^1)} C_{-*}(X).$$

Next, one observes that $C_{-*}(ES^1) \to C_{-*}(\mathrm{pt})$ is a quasi-isomorphism of dg $C_{-*}^{\text{sing}}(S^1)$–modules, where the $C_{-*}^{\text{sing}}(S^1)$–actions are induced by the S^1–actions on ES^1 and pt, respectively. Hence, there is a quasi-isomorphism of derived tensor products

$$C_{-*}(\mathrm{pt}) \otimes^{\mathbb{L}}_{C_{-*}^{\text{sing}}(S^1)} C_{-*}(X) \simeq C_{-*}(ES^1) \otimes^{\mathbb{L}}_{C_{-*}^{\text{sing}}(S^1)} C_{-*}(X).$$

Finally, it is a standard fact in algebraic topology (used in the construction of Eilenberg–Moore-type spectral sequences, eg [44, Theorem 7.27] and [17, Proposition 6.13]) that, as ES^1 is a principal S^1–bundle,

$$C_{-*}(ES^1) \otimes^{\mathbb{L}}_{C^{\mathrm{sing}}_{-*}(S^1)} C_{-*}(X) \simeq C_{-*}(ES^1 \times_{S^1} X) = C_{-*}(X_{\mathrm{h}S^1}),$$

which is the usual chain complex computing (Borel) equivariant homology. This gives some justification for the usage of the subscript $\mathrm{h}S^1$ notation in Definition 2.13.

Definition 2.15 The *homotopy fixed-point complex* of M is the chain complex of morphisms from $\boldsymbol{k}$ to M in the category of S^1–complexes,

$$M^{\mathrm{h}S^1} := \mathrm{Rhom}_{S^1}(\boldsymbol{k}, M). \tag{2-23}$$

The morphism of modules $\epsilon\colon \boldsymbol{k}[\Lambda]/\Lambda^2 \to \boldsymbol{k}$ induces a chain map $M^{\mathrm{h}S^1} \to M$, called the *inclusion of homotopy fixed points*,

$$\iota\colon M^{\mathrm{h}S^1} = \mathrm{Rhom}_{S^1}(\boldsymbol{k}, M) \to \mathrm{Rhom}_{S^1}(\boldsymbol{k}[\Lambda]/\Lambda^2, M) \cong M. \tag{2-24}$$

Remark 2.16 To motivate the usage "homotopy fixed points", in the topological category, the usual fixed points of a G–action can be described as $\mathrm{Maps}_G(\mathrm{pt}, X)$. When $M = C_{-*}(X)$ for X an S^1–space, there is a canonical map $C_{-*}(X^{\mathrm{h}S^1}) \to (C_{-*}(X))^{\mathrm{h}S^1}$. However, in contrast to the case of homotopy orbits discussed in Remark 2.14, this map need not be an equivalence.

Composition in the category S^1–mod induces a natural action of

$$\begin{aligned} \mathrm{Rhom}_{S^1}(\boldsymbol{k}, \boldsymbol{k}) &= \boldsymbol{k}[u] \quad \text{with } |u| = 2 \\ &= H^*(BS^1) \end{aligned} \tag{2-25}$$

on the homotopy fixed-point complex. There is a third important equivariant homology complex, called the *periodic cyclic*, or *Tate* complex of M, defined as the localization of $M^{\mathrm{h}S^1}$ away from $u = 0$,

$$M^{\mathrm{Tate}} := M^{\mathrm{h}S^1} \otimes_{\boldsymbol{k}[u]} \boldsymbol{k}[u, u^{-1}]. \tag{2-26}$$

The Tate construction sits in an exact sequence between the homotopy orbits and fixed points.

Remark 2.17 (Gysin sequences) It is straightforward from the viewpoint of A_∞ $C_{-*}(S^1)$–modules to explain the appearance of various Gysin and periodicity sequences. Take for instance the *Gysin exact triangle*

$$M \xrightarrow{\mathrm{pr}} M_{\mathrm{h}S^1} \to M_{\mathrm{h}S^1}[2] \xrightarrow{[1]} M.$$

This is a manifestation of a canonical exact triangle of objects in S^1–mod,

$$\boldsymbol{k}[\Lambda]/\Lambda^2 \xrightarrow{\epsilon} \boldsymbol{k} \xrightarrow{u} \boldsymbol{k}[2] \xrightarrow{[1]} \boldsymbol{k}[\Lambda]/\Lambda^2$$

(recall in Remark 2.7 it was shown that $\operatorname{Rhom}_{S^1}(\boldsymbol{k},\boldsymbol{k}) \cong \boldsymbol{k}[u]$), pushed forward by the functor $(-) \otimes^{\mathbb{L}}_{S^1} M$. The other exact sequences arise similarly.

As a special case of the general homotopy-invariance properties of A_∞–modules stated in Proposition 2.9, we have:

Corollary 2.18 *If $F\colon M \to N$ is a homomorphism of S^1–complexes (meaning a closed morphism), it induces chain maps between equivariant theories:*

$$F^{hS^1}\colon M^{hS^1} \to N^{hS^1}, \tag{2-27}$$

$$F_{hS^1}\colon M_{hS^1} \to N_{hS^1}, \tag{2-28}$$

$$F^{\mathrm{Tate}}\colon M^{\mathrm{Tate}} \to N^{\mathrm{Tate}}. \tag{2-29}$$

If F is a quasi-isomorphism of S^1–complexes (meaning simply that $[F^0]$ is a homology isomorphism), then (2-27)–(2-29) are quasi-isomorphisms of chain complexes. □

Functoriality further tautologically implies:

Proposition 2.19 *If $F\colon M \to N$ is a homomorphism of S^1–complexes, then the various induced maps (2-27)–(2-29) intertwine all of the long exact sequences for (equivariant homology groups of) M with those for N.* □

2.3 u–linear models for S^1–complexes

It is convenient to package the data described in the previous two sections into "u–linear generating functions", in the following way: Let u be a formal variable of degree $+2$. Let us use the abuse of notation

$$M[\![u]\!] := M \mathbin{\widehat{\otimes}_{\boldsymbol{k}}} \boldsymbol{k}[u]$$

for the u–adically completed tensor product in the category of graded vector spaces; in other words $M[\![u]\!] := \bigoplus_k M[\![u]\!]_k$, where $M[\![u]\!]_k = \bigl\{\sum_{i=0}^{\infty} m_i u^i \mid m_i \in M_{k-2i}\bigr\}$. Then, we frequently write an S^1–complex $(M, \{\delta_k\}_{k\geq 0})$ as a $\boldsymbol{k}$–module M equipped with a map

$$\delta_{\mathrm{eq}}^{(M)} = \sum_{i=0}^{\infty} \delta_i^M u^i \colon M \to M[\![u]\!] \tag{2-30}$$

of total degree 1, satisfying $\delta_{\mathrm{eq}}^2 = 0$. Here we are implicitly conflating δ_{eq} with its u–linear extension to a map $M[\![u]\!] \to M[\![u]\!]$ in order to u–linearly compose and obtain a map $M \to M[\![u]\!]$.

Premorphisms from M to N of degree k can similarly be recast as maps $F_{\mathrm{eq}} = \sum_{i=0}^{\infty} F_i u^i \colon M \to N[\![u]\!]$ of pure degree k (so each F_i has degree $k - 2i$). The differential on premorphisms can be described u–linearly as

$$\partial(F_{\mathrm{eq}}) = F_{\mathrm{eq}} \circ \delta_{\mathrm{eq}}^M - (-1)^{\deg(F)} \delta_{\mathrm{eq}}^N \circ F_{\mathrm{eq}}, \tag{2-31}$$

and composition is simply the u–linear composition $G_{\mathrm{eq}} \circ F_{\mathrm{eq}}$ (again, one implicitly u–linearly extends G_{eq} and then u–linearly composes); explicitly,

$$\Bigl(\sum_{i\geq 0} G_i u^i\Bigr) \circ \Bigl(\sum_{j\geq 0} F_j u^j\Bigr) = \sum_{k\geq 0} \Bigl(\sum_{i+j=k} G^i \circ F^j\Bigr) u^k.$$

With respect to this packaging, the formulae for various equivariant homology chain complexes can be given the more readable forms

$$M_{\mathrm{h}S^1} = (M(\!(u)\!)/uM[\![u]\!], \delta_{\mathrm{eq}}), \tag{2-32}$$

$$M^{\mathrm{h}S^1} = (M[\![u]\!], \delta_{\mathrm{eq}}), \tag{2-33}$$

$$M^{\mathrm{Tate}} = (M(\!(u)\!), \delta_{\mathrm{eq}}), \tag{2-34}$$

where, again, we use the abuse of notation $M(\!(u)\!) = M[\![u]\!] \otimes_{\boldsymbol{k}[u]} \boldsymbol{k}[u, u^{-1}]$. (On the other hand, note that (2-32) is *not* completed.) As before, any homomorphism (that is, closed morphism) of S^1–complexes $F_{\mathrm{eq}} = \sum_{i=0}^{\infty} F^i u^i$ induces a $\boldsymbol{k}[u]$–linear chain map between homotopy fixed-point complexes by u–linearly extended composition, and hence, by tensoring over $\boldsymbol{k}[u]$ with $\boldsymbol{k}(\!(u)\!)/u\boldsymbol{k}[\![u]\!]$ or $\boldsymbol{k}(\!(u)\!)$, chain maps between homotopy orbit and Tate complex constructions. With respect to these explicit complexes, the projection to homotopy orbits (2-22) and inclusion of fixed points (2-24) chain maps have simple explicit descriptions

$$\mathrm{pr} \colon M \to M_{\mathrm{h}S^1}, \qquad \alpha \mapsto \alpha \cdot u^0, \tag{2-35}$$

$$\iota \colon M^{\mathrm{h}S^1} \to M, \qquad \sum_{i=0}^{\infty} \alpha_i u^i \mapsto \alpha_0. \tag{2-36}$$

Remark 2.20 This u–linear lossless packaging of the data describing an S^1–complex is a manifestation of *Koszul duality*; in the case of $A = \boldsymbol{k}[\Lambda]/\Lambda^2$, it posits that there is a fully faithful embedding, $\mathrm{Rhom}(\boldsymbol{k}, -) = (-)^{\mathrm{h}S^1}$ from A–modules into $B := \mathrm{Rhom}_A(\boldsymbol{k}, \boldsymbol{k}) = \boldsymbol{k}[u]$–modules.

From the u–linear point of view, it is easier to observe that the exact triangle of $\boldsymbol{k}[u]$–modules $\boldsymbol{k}[\![u]\!] \to \boldsymbol{k}(\!(u)\!) \to \boldsymbol{k}(\!(u)\!)/\boldsymbol{k}[\![u]\!] = u^{-1}(\boldsymbol{k}(\!(u)\!)/u\boldsymbol{k}[\![u]\!])$ induces an exact triangle (functorial in M) between equivariant homology chain complexes

$$M^{\mathrm{h}S^1} \to M^{\mathrm{Tate}} \to M_{\mathrm{h}S^1}[2] \xrightarrow{[1]} M^{\mathrm{h}S^1}.$$

3 Circle action on the open sector

3.1 The usual and nonunital Hochschild chain complex

Recall that an A_∞ category over $\boldsymbol{k}$, $\mathcal{C}$ is specified by the following data:

- A collection of objects $\operatorname{ob}\mathcal{C}$.
- For each pair of objects X, X', a graded vector space $\hom_{\mathcal{C}}(X, X')$ over $\boldsymbol{k}$.
- For any set of $d+1$ objects $X_0, \dots, X_d$, higher multilinear (over $\boldsymbol{k}$) composition maps

(3-1) $$\mu^d \colon \hom_{\mathcal{C}}(X_{d-1}, X_d) \times \cdots \times \hom_{\mathcal{C}}(X_0, X_1) \to \hom_{\mathcal{C}}(X_0, X_d)$$

(sometimes equivalently viewed as a map from the tensor product) of degree $2-d$, satisfying for each $k > 0$ the (quadratic) A_∞ relations

(3-2) $$\sum_{i,l} (-1)^{\maltese_i} \mu_{\mathcal{C}}^{k-l+1}(x_k, \dots, x_{i+l+1}, \mu_{\mathcal{C}}^l(x_{i+l}, \dots, x_{i+1}), x_i, \dots, x_1) = 0,$$

with sign

(3-3) $$\maltese_i := \|x_1\| + \cdots + \|x_i\|,$$

where $|x|$ denotes degree and $\|x\| := |x| - 1$ denotes reduced degree.

The first two equations of (3-2) imply that μ^1 is a differential, and the cohomology level maps $[\mu^2]$ are a genuine composition for the (nonunital) category $H^*(\mathcal{C})$ with the same objects and morphisms,

(3-4) $$\operatorname{Hom}_{H^*(\mathcal{C})}(X, Y) := H^*(\hom_{\mathcal{C}}(X, Y), \mu^1).$$

We say that $\mathcal{C}$ is *cohomologically unital* if there exist cohomology-level identity morphisms $[e_X] \in \operatorname{Hom}_{H^*(\mathcal{C})}(X, X)$ for each object X, making $H^*(\mathcal{C})$ into a genuine category. We say that $\mathcal{C}$ is *strictly unital* if there exist elements $e_X^+ \in \hom_{\mathcal{C}}(X, X)$, for every object X, satisfying

(3-5) $$\begin{cases} \mu^1(e_X^+) = 0, & \\ (-1)^{|y|}\mu^2(e_{X_1}^+, y) = y = \mu^2(y, e_{X_0}^+) & \text{for any } y \in \hom_{\mathcal{C}}(X_0, X_1), \\ \mu^d(\dots, e_X^+, \dots) = 0 & \text{for } d > 2. \end{cases}$$

We call such elements the chain-level, or strict, identity elements.

The *Hochschild chain complex*, or *cyclic bar complex*, of an A_∞ category $\mathcal{C}$ is the direct sum of all cyclically composable sequences of morphism spaces in $\mathcal{C}$,

$$\text{(3-6)}\quad \mathrm{CH}_*(\mathcal{C}) := \bigoplus_{\substack{k\geq 0\\ X_{i_0},\dots,X_{i_k}\in \mathrm{ob}\,\mathcal{C}}} \hom_{\mathcal{C}}(X_{i_k},X_{i_0})\otimes \hom_{\mathcal{C}}(X_{i_k-1},X_{i_k})\otimes\cdots\otimes \hom_{\mathcal{C}}(X_{i_0},X_{i_1}).$$

The (cyclic bar) differential b acts on Hochschild chains by summing over ways to cyclically collapse elements by any of the A_∞ structure maps:

$$\begin{aligned}\text{(3-7)}\quad & b(\boldsymbol{x}_d\otimes x_{d-1}\otimes\cdots\otimes x_1)\\ &= \sum (-1)^{\#_k^d}\mu^{d-i}(x_k,\dots,x_1,\boldsymbol{x}_d,x_{d-1},\dots,x_{k+i+1})\otimes x_{k+i}\otimes\cdots\otimes x_{k+1}\\ &\qquad + \sum (-1)^{\maltese_1^s}\boldsymbol{x}_d\otimes\cdots\otimes \mu^j(x_{s+j+1},\dots,x_{s+1})\otimes x_s\otimes\cdots\otimes x_1,\end{aligned}$$

with signs

$$\text{(3-8)}\qquad \maltese_i^k := \sum_{j=i}^{k} \|x_i\|,$$

$$\text{(3-9)}\qquad \#_k^d := \maltese_1^k\cdot(1+\maltese_{k+1}^d)+\maltese_{k+1}^{d-1}+1.$$

In this complex, Hochschild chains are (cohomologically) graded as

$$\text{(3-10)}\quad \deg(\boldsymbol{x}_d\otimes x_{d-1}\otimes\cdots\otimes x_1) := \deg(\boldsymbol{x}_d)+\sum_{i=1}^{d-1}\deg(x_i)-d+1 = |\boldsymbol{x}_d|+\sum_{i=1}^{d-1}\|x_i\|.$$

Remark 3.1 Frequently the notation $\mathrm{CH}_*(\mathcal{C},\mathcal{C})$ is used for (3-6) to emphasize that Hochschild homology is taken here with *diagonal coefficients*, rather than coefficients in another bimodule.

If $\mathcal{C}$ is a strictly unital A_∞ category, then the chain complex (3-6) carries a strict S^1–action $B\colon \mathrm{CH}_*(\mathcal{C})\to \mathrm{CH}_{*-1}(\mathcal{C})$, involving summing over ways to cyclically permute chains and insert identity morphisms; see Remark 3.7 below. However, there is a quasi-isomorphic nonunital Hochschild complex of $\mathcal{C}$ which always carries a strict S^1–action (even if $\mathcal{C}$ is not strictly unital), which we will now describe.

As a graded vector space, the *nonunital Hochschild complex* consists of two copies of the cyclic bar complex, the second copy shifted down in grading by 1:

$$\text{(3-11)}\qquad \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C}) := \mathrm{CH}_*(\mathcal{C})\oplus \mathrm{CH}_*(\mathcal{C})[1].$$

With respect to the decomposition (3-11), we sometimes refer to elements as $\sigma := (\check{\alpha}, \widehat{\beta})$, with the notation $\check{\alpha}$ or $\widehat{\beta}$ indicating that a given element α and β belong to the left or right factor respectively. Similarly, we refer to the left and right factors as the *check factor* and the *hat factor*, respectively.

Let b' denote a version of the differential (3-7) omitting the "wrap-around terms" (often simply called the *bar differential*):

$$\begin{aligned}(3\text{-}12)\quad & b'(x_d \otimes x_{d-1} \otimes \cdots \otimes x_1) \\ &= \sum (-1)^{\maltese_1^s} x_d \otimes \cdots \otimes x_{s+j+1} \otimes \mu^j(x_{s+j}, \dots, x_{s+1}) \otimes x_s \otimes \cdots \otimes x_1 \\ &\qquad + \sum (-1)^{\maltese_1^{d-j}} \mu^j(x_d, x_{d-1}, \dots, x_{d-j+1}) \otimes x_{d-j} \otimes \cdots \otimes x_1.\end{aligned}$$

For an element $\widehat{\beta} = x_d \otimes \cdots \otimes x_1$ in the hat (right) factor of (3-11), define an element $d_{\wedge\vee}(\widehat{\beta})$ in the check (left) factor of (3-11) by

$$(3\text{-}13)\quad d_{\wedge\vee}(\widehat{\beta}) := (-1)^{\maltese_2^d + \|x_1\| \cdot \maltese_2^d + 1} x_1 \otimes x_d \otimes \cdots \otimes x_2 + (-1)^{\maltese_1^{d-1}} x_d \otimes \cdots \otimes x_1.$$

In this language, the differential on the nonunital Hochschild complex can be written

$$(3\text{-}14)\qquad b^{\mathrm{nu}} \colon (\check{\alpha}, \widehat{\beta}) \mapsto (b(\check{\alpha}) + d_{\wedge\vee}(\widehat{\beta}), b'(\widehat{\beta})),$$

or equivalently can be expressed via the matrix

$$(3\text{-}15)\qquad b^{\mathrm{nu}} = \begin{pmatrix} b & d_{\wedge\vee} \\ 0 & b' \end{pmatrix}.$$

The left factor $\mathrm{CH}_*(\mathcal{C}) \hookrightarrow \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C})$ is by definition a subcomplex. Moreover, since the quotient complex is the standard A_∞ bar complex with differential b', which is acyclic for cohomologically unital $\mathcal{C}$ (by a standard length-filtration spectral sequence argument, compare [52, Lemma 2.12] or [24, Proposition 2.2]), it follows that:

Lemma 3.2 *The inclusion map* $\iota \colon \mathrm{CH}_*(\mathcal{C}) \hookrightarrow \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C})$ *is a quasi-isomorphism* (*when* $\mathcal{C}$ *is cohomologically unital*). □

Remark 3.3 The nonunital Hochschild complex of $\mathcal{C}$ can be conceptually explained in terms of cyclic bar complexes as follows; cf [41, Section 1.4.1; 61, Section 3.5]. First, *augment* the category $\mathcal{C}$ by adjoining strict units; meaning, consider the A_∞ category $\mathcal{C}^+$ with $\operatorname{ob}\mathcal{C}^+ = \operatorname{ob}\mathcal{C}$ and

$$(3\text{-}16)\qquad \hom_{\mathcal{C}^+}(X, Y) = \begin{cases} \hom_{\mathcal{C}}(X, Y) & \text{when } X \neq Y, \\ \hom_{\mathcal{C}}(X, X) \oplus \boldsymbol{k}\langle e_X^+ \rangle & \text{when } X = Y, \end{cases}$$

whose A_∞ structure maps are completely determined by the fact that $\mathcal{C}$ is an A_∞ subcategory, and the elements e_X^+ act as strict units in the sense of (3-5). Next, consider the *normalized* (*or reduced*) *Hochschild complex* of the strictly unital category $\mathcal{C}^+$, $\mathrm{CH}_*^{\mathrm{red}}(\mathcal{C}^+)$, by definition the quotient of $\mathrm{CH}_*(\mathcal{C}^+)$ by the acyclic subcomplex consisting of e^+ terms in any position but the first. Now, take the further quotient of $\mathrm{CH}_*^{\mathrm{red}}(\mathcal{C}^+)$ by the subcomplex of length one Hochschild chains of the form e_X^+ for some X. The resulting complex, denoted by $\widetilde{\mathrm{CH}}_*(\mathcal{C}^+)$, can be identified as a chain complex with $\mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C})$ via the map

$$\begin{aligned} f\colon \widetilde{\mathrm{CH}}_*(\mathcal{C}^+) &\xrightarrow{\cong} \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C}), \\ y_k \otimes \cdots \otimes y_1 &\longmapsto \begin{cases} (0, y_{k-1} \otimes \cdots \otimes y_1) & \text{if } y_k = e_X^+ \text{ for some } X, \\ (y_k \otimes \cdots \otimes y_1, 0) & \text{otherwise.} \end{cases} \end{aligned} \tag{3-17}$$

In particular, the differential in $\mathrm{CH}^{\mathrm{nu}}(\mathcal{C})$ on a Hochschild chain $\hat{\beta}$ in the right factor (of the decomposition (3-11)) agrees with the (usual cyclic bar) Hochschild differential applied to $e_X^+ \otimes \beta$ under the correspondence f.

3.2 Circle action on the Hochschild complex

The S^1–action on the nonunital (or usual) Hochschild complex is built out of several intermediate operations. First, let $t\colon \mathrm{CH}_*(\mathcal{C}) \to \mathrm{CH}_*(\mathcal{C})$ denote the (signed) cyclic permutation operator on the cyclic bar complex generating the $\mathbb{Z}/k\mathbb{Z}$ cyclic action on the length-k expressions

$$t\colon x_k \otimes \cdots \otimes x_1 \mapsto (-1)^{\|x_1\| \cdot \maltese_2^k + \|x_1\| + \|x_k\|} x_1 \otimes x_k \otimes \cdots \otimes x_2. \tag{3-18}$$

(This is not a chain map.)

Let N denote the *norm* of this operation; that is, the sum of all powers of t (this depends on k, the length of a given Hochschild chain),

$$N\colon x_k \otimes \cdots \otimes x_1 = \sigma \mapsto (1 + t + t^2 + \cdots + t^{k-1})\sigma. \tag{3-19}$$

Let $s^{\mathrm{nu}}\colon \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C}) \to \mathrm{CH}_{*-1}^{\mathrm{nu}}(\mathcal{C})$ be the linear map which sends check chains to the corresponding hat chains, and hat chains to zero:

$$s^{\mathrm{nu}}(x_d \otimes \cdots \otimes x_1, y_t \otimes \cdots \otimes y_1) := (-1)^{\maltese_1^d + \|x_d\| + 1}(0, x_d \otimes \cdots \otimes x_1). \tag{3-20}$$

(Again, this is not a chain map.)

Finally define $B^{\mathrm{nu}}\colon \mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}) \to \mathrm{CH}^{\mathrm{nu}}_{*-1}(\mathcal{C})$ by

$$
\begin{aligned}
\text{(3-21)}\quad B^{\mathrm{nu}}&(x_k \otimes \cdots \otimes x_1, y_l \otimes \cdots \otimes y_1)\\
&:= \sum_{i=1}^{k} (-1)^{\maltese_1^i \maltese_{i+1}^k + \|x_k\| + \maltese_1^k + 1}(0, x_i \otimes \cdots \otimes x_1 \otimes x_k \otimes \cdots \otimes x_{i+1}))\\
&= s^{\mathrm{nu}}(N(x_k \otimes \cdots \otimes x_1), y_l \otimes \cdots \otimes y_1)\\
&= \sum_{i=0}^{k-1} s^{\mathrm{nu}}(t^i(x_k \otimes \cdots \otimes x_1), y_l \otimes \cdots \otimes y_1).
\end{aligned}
$$

Lemma 3.4 *We have* $(B^{\mathrm{nu}})^2 = 0$ *and* $b^{\mathrm{nu}}B^{\mathrm{nu}} + B^{\mathrm{nu}}b^{\mathrm{nu}} = 0$. *That is*, $\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C})$ *is a strict* S^1*–complex*, *with the action of* $\Lambda = [S^1]$ *given by* B^{nu}. □

Let $b_{\mathrm{eq}} = b^{\mathrm{nu}} + uB^{\mathrm{nu}}$ be the strict S^1–complex structure on the nonunital Hochschild complex $\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C})$, u–linearly packaged as in Section 2.3. Using this, we can define cyclic homology groups, as follows.

Definition 3.5 The (*positive*) *cyclic* chain complex, the *negative cyclic* chain complex, and the *periodic cyclic* chain complexes of $\mathcal{C}$ are the *homotopy orbit complex*, *homotopy fixed-point complex*, and *Tate constructions* of the S^1–complex $(\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}), b_{\mathrm{eq}})$, respectively. That is,

$$\text{(3-22)}\qquad \mathrm{CC}^+_*(\mathcal{C}) := (\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}))_{\mathrm{h}S^1} = (\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}) \otimes_{\boldsymbol{k}} \boldsymbol{k}(\!(u)\!)/u\boldsymbol{k}[\![u]\!], b_{\mathrm{eq}}),$$

$$\text{(3-23)}\qquad \mathrm{CC}^-_*(\mathcal{C}) := (\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}))^{\mathrm{h}S^1} = (\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}) \mathbin{\widehat{\otimes}}_{\boldsymbol{k}} \boldsymbol{k}[\![u]\!], b_{\mathrm{eq}}),$$

$$\text{(3-24)}\qquad \mathrm{CC}^\infty_*(\mathcal{C}) := (\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}))^{\mathrm{Tate}} = (\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}) \mathbin{\widehat{\otimes}}_{\boldsymbol{k}} \boldsymbol{k}(\!(u)\!), b_{\mathrm{eq}}),$$

with grading induced by setting $|u| = +2$, and where, as in Section 2.3, $\widehat{\otimes}$ refers to the u–adically completed tensor product in the category of graded vector spaces. The cohomologies of these complexes, denoted by $\mathrm{HC}^{+/-/\infty}_*(\mathcal{C})$, are called the (*positive*), *negative* and *periodic cyclic homologies* of $\mathcal{C}$, respectively.

The $C_{-*}(S^1)$–module structure on $\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C})$ is suitably functorial, in the following sense. Let $\boldsymbol{F}\colon \mathcal{C} \to \mathcal{C}'$ be an A_∞ functor. There is an induced chain map on nonunital Hochschild complexes

$$\text{(3-25)}\qquad \boldsymbol{F}^{\mathrm{nu}}_\sharp\colon \mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}) \to \mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C}', \mathcal{C}'), \quad (x, y) \mapsto (\boldsymbol{F}_\sharp(x), \boldsymbol{F}'_\sharp(y)),$$

where

$$F'_\sharp(x_k\otimes\cdots\otimes x_0) := \sum_{i_1,\dots,i_s} F^{i_1}(x_k\cdots)\otimes\cdots\otimes F^{i_s}(\cdots x_0), \tag{3-26}$$

$$\begin{aligned} F_\sharp(x_k\otimes\cdots\otimes x_0) := \sum_{i_1,\dots,i_s,j} & F^{j+1+i_1}(x_j,\dots,x_0,x_k,\dots,x_{k-i_1+1}) \\ & \otimes F^{i_2}(\cdots)\otimes\cdots\otimes F^{i_s}(x_{j+i_s},\dots,x_{j+1}), \end{aligned} \tag{3-27}$$

which is an isomorphism on homology if F is a quasi-isomorphism (indeed, even a Morita equivalence). This functoriality preserves S^1 structures:

Proposition 3.6 *$F_\sharp^{\mathrm{nu}}$ gives a strict morphism of strict S^1–complexes, meaning*

$$F_\sharp^{\mathrm{nu}}\circ b^{\mathrm{nu}} = b^{\mathrm{nu}}\circ F_\sharp^{\mathrm{nu}} \quad \text{and} \quad F_\sharp^{\mathrm{nu}}\circ B^{\mathrm{nu}} = B^{\mathrm{nu}}\circ F_\sharp^{\mathrm{nu}}.$$

In other words, the premorphism of A_∞ $\mathbf{k}[\Lambda]/\Lambda^2$–modules defined as

$$F_*^d(\underbrace{\Lambda,\dots,\Lambda}_{d},\sigma) := \begin{cases} F_\sharp^{\mathrm{nu}}(\sigma) & \text{if } d=0, \\ 0 & \text{if } d\geq 1, \end{cases} \tag{3-28}$$

is closed, ie an A_∞–module homomorphism.

Sketch of proof It is well known that $F_\sharp^{\mathrm{nu}}$ is a chain map, so it suffices to verify that $F_\sharp^{\mathrm{nu}}\circ B^{\mathrm{nu}} = B^{\mathrm{nu}}\circ F_\sharp^{\mathrm{nu}}$, or in terms of (3-25),

$$F'_\sharp\circ s^{\mathrm{nu}}N = s^{\mathrm{nu}}N\circ F_\sharp. \tag{3-29}$$

We leave this an exercise, noting that applying either side to a Hochschild chain $x_k\otimes\cdots\otimes x_1$, the sums match identically. □

Remark 3.7 If $\mathcal{C}$ is strictly unital, one can also define an operator $B\colon \mathrm{CH}_*(\mathcal{C})\to \mathrm{CH}_{*-1}(\mathcal{C})$ on the usual cyclic bar complex by

$$B = (1-t)sN,$$

where, up to a sign, s denotes the operation of inserting, at the beginning of a chain, the unique strict unit e_X^+ preserving cyclic composability:

$$s\colon x_k\otimes\cdots\otimes x_1 \mapsto (-1)^{\|x_k\|+\maltese_1^k+1} e_{X_{i_k}}^+\otimes x_k\otimes\cdots\otimes x_1, \tag{3-30}$$

where $x_k\in\hom_{\mathcal{C}}(X_{i_k},X_{i_0})$. It can be shown that $B^2=0$ and $Bb+bB=0$, $\mathrm{CH}_*(\mathcal{C})$ is a strict S^1–complex; moreover that the quasi-isomorphism $\mathrm{CH}_*(\mathcal{C})\cong\mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C})$ is one of S^1–complexes. In fact, B descends to the *reduced Hochschild complex* $\mathrm{CH}_*^{\mathrm{red}}(\mathcal{C})$ described in Remark 3.3, where it takes the simpler form

$$B^{\mathrm{red}} = sN,$$

as applying tsN results in a Hochschild chain with a strict unit not in the first position, which becomes zero in $\mathrm{CH}^{\mathrm{red}}_*(\mathcal{C})$. If $\mathcal{C}$ was not necessarily strictly unital, following Remark 3.3 one can still consider the quotient of the reduced Hochschild complex of the augmented category $\mathcal{C}^+$, which we called $\widetilde{\mathrm{CH}}_*(\mathcal{C}^+)$. The discussion here equips this complex with an S^1–action $\widetilde{B}^{\mathrm{red}}$. Under the bijection f of (3-17), $\widetilde{B}^{\mathrm{red}}$ is sent to B^{nu} and s is sent to s^{nu}.

Remark 3.8 Continuing Remark 3.3, suppose we have constructed $\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{C})$ as $\widetilde{\mathrm{CH}}_*(\mathcal{C}) := \mathrm{CH}^{\mathrm{red}}_*(\mathcal{C}^+)/\bigoplus_X \boldsymbol{k}\langle e^+_X\rangle$, the quotient of the reduced Hochschild complex of the augmented category $\mathcal{C}^+$. Given any $\boldsymbol{F}$ as above, extend $\boldsymbol{F}$ to an augmented functor $\boldsymbol{F}^+$ by mandating that

$$(\boldsymbol{F}^+)^1(e^+_X) = e^+_{\boldsymbol{F}X} \quad \text{and} \quad (\boldsymbol{F}^+)^d(\dots, e^+_X, \dots) = 0. \tag{3-31}$$

It is easy to see that the map $(\boldsymbol{F}^+)_*\colon \mathrm{CH}_*(\mathcal{C}^+) \to \mathrm{CH}_*((\mathcal{C}')^+)$ descends to a map $\widetilde{\boldsymbol{F}}\colon \widetilde{\mathrm{CH}}_*(\mathcal{C}^+) \to \widetilde{\mathrm{CH}}_*((\mathcal{C}')^+)$. Under the bijection (3-17), this precisely corresponds to $\boldsymbol{F}^{\mathrm{nu}}_\sharp$ described above. In particular, the fact that strictly unital functors induce strict S^1–morphisms between (usual) Hochschild complexes immediately implies Proposition 3.6.

Remark 3.9 There are options besides the nonunital Hochschild complex for seeing the $C_{-*}(S^1)$–action on a Hochschild complex of the Fukaya category. For instance, one could:

(1) Perform a strictly unital replacement (via homological algebra as in [52, Section 2] or [40, Theorem 3.2.1.1]), and work with the Hochschild complex of the replacement. However, this doesn't retain a relationship between the A_∞ operations and geometric structure, and hence is difficult to use with open–closed maps.

(2) Geometrically construct a strictly unital structure on the Fukaya category via constructing *homotopy units* [22], which roughly involves building a series of geometric higher homotopies between the operation of A_∞ multiplying by a specified geometrically defined cohomological unit, and the operation of A_∞ multiplying by a strict unit (which is algebraically defined, but may also be geometrically characterized in terms of forgetful maps). From this one defines a strictly unital A_∞ category $\mathcal{F}^{\mathrm{hu}}$ with $\hom_{\mathcal{F}^{\mathrm{hu}}}(X,X) = \hom_{\mathcal{F}}(X,X) \oplus \boldsymbol{k}\langle e^+_X, f_X\rangle$ and $\hom_{\mathcal{F}^{\mathrm{hu}}}(X,Y) = \hom_F(X,Y)$ for $X \neq Y$, extending the A_∞ structure on $\mathcal{F}$, such that each e^+_X is a strict unit and $\mu^1(f_X) = e^+_X - e_X$ for e_X a chosen a cohomological unit. The geometric higher

homotopies alluded to above give operations used to define for instance, μ^k of a sequence of elements containing one or more f_X terms.

Remark 3.7 then equips the usual Hochschild complex $\mathrm{CH}_*(\mathcal{F}^{\mathrm{hu}}, \mathcal{F}^{\mathrm{hu}})$ with a strict S^1–action. Using this one can construct a cyclic open–closed map with source $\mathrm{CH}_*(\mathcal{F}^{\mathrm{hu}}, \mathcal{F}^{\mathrm{hu}})$, in a manner completely analogous to the construction of $\mathcal{F}^{\mathrm{hu}}$ and the cyclic open–closed map here. This option is equivalent to the one we have chosen (and has some benefits), but requires additional technicalities/moduli spaces beyond the route taken here — both in constructing and defining the category $\mathcal{F}^{\mathrm{hu}}$, and then in defining further "higher homotopies" between inserting a cohomological unit asymptotic and imposing a strict unit (ie forgettable) constraint into the cyclic open–closed map in various places, which give operations that correspond to applying the cyclic open–closed map to a Hochschild chain with one or more f_X terms.

A construction of homotopy units was introduced in the work of Fukaya, Oh, Ohta and Ono [22, Chapter 7, Section 3.1]. See [24] for an implementation in the (possibly wrapped) exact (or otherwise tautologically unobstructed), multiple Lagrangians setting.

3.3 The Fukaya category

The goal of this subsection is to review (under simplifying technical hypotheses) the definition of the Fukaya category of a symplectic manifold. The outcome, a (homologically unital but not necessarily strictly unital) A_∞ category, will in particular carry a circle action on its nonunital Hochschild complex.

In Section 3.3.1 below, we detail a set of simplifying assumptions imposed on all of the moduli spaces of Floer curves considered in this paper (mostly pertaining to transversality and compactness), and recall examples of the variety of geometric situations in which they are satisfied. Such assumptions are in particular satisfied in the technically simplest cases in which (compact or wrapped) Fukaya categories can be defined, namely exact (Liouville) and monotone or aspherical symplectic manifolds. In Section 3.3.2 we will quickly review the construction of the Fukaya category under such hypotheses. The initial thread of discussion will focus on compact Lagrangians, but immediately extends to wrapped Fukaya categories of Liouville manifolds as described in a series of remarks; here we are using the framework of quadratic Hamiltonians as defined in [1] for wrapped Fukaya categories, whose construction is nearly as simple as that of compact Fukaya categories and requires only a few minor modifications.

3.3.1 Geometric setup and assumptions about moduli spaces of Floer trajectories

To simplify technicalities, the main assumption we make about moduli spaces in this paper is as follows.

Assumption 3.10 (main assumption about moduli spaces) All (semistably compactified) moduli spaces of Floer trajectories considered in this paper of virtual dimension ≤ 1 are — for generic choices of complex structure and Hamiltonian ("perturbation data") — compact transversally cut-out manifolds with boundary of dimension equal to virtual dimension. Moreover, the union of any such moduli space with fixed "input" asymptotic conditions over all possible "output" asymptotic conditions remains compact, and in particular is empty for all but finitely many possible output conditions (vacuous when there are only finitely many possible outputs).

Let $M = (M^{2n}, \omega)$ denote our target symplectic manifold and fix a collection of (always properly embedded) Lagrangian submanifolds $\{L_i\}$ in M, which we wish to be the objects of our Fukaya category. We will call any M, $\{L_i\}$, and choices of Floer perturbation data used to define moduli spaces for which Assumption 3.10 holds *admissible*. We will say M and/or $M, \{L_i\}$ are admissible if they possess an ample supply of Floer data for which Assumption 3.10 holds for the moduli spaces considered below involving these targets. Examples of admissible M include:

- Any Liouville manifold (in particular noncompact), which is to say that ω is exact with a fixed choice of primitive λ, such that flowing out by the Liouville vector field Z (defined by $\iota_Z \omega = \lambda$) induces a diffeomorphism

$$M \backslash \mathring{\overline{M}} \cong \partial \overline{M} \times [0, \infty) \tag{3-32}$$

  for some codimension-zero manifold-with-boundary $\overline{M}$, called a Liouville domain whose completion is M.

- Any compact symplectic manifold which is either monotone, ie $[\omega] = 2\tau c_1(M)$ for some constant $\tau > 0$, or symplectically aspherical, ie $\omega(\pi_2(M)) = 0$.

If M is Liouville, we henceforth fix a cylindrical end (3-32), and use r to refer to the corresponding $[0, \infty)$ coordinate. Examples of (properly embedded) admissible Lagrangian submanifolds $L \subset M$ in admissible M include:

- In Liouville M, one can take any exact L, ie with $\lambda_L = df$, equipped with fixed choice of primitive which vanishes outside a compact set, which implies, as in (3-32), that L is modeled on the cone of a Legendrian near infinity.

- In (compact) monotone M, one can take monotone L, in the sense that $\omega(-) = \rho\mu_L(-)\colon H_2(M,L) \to \mathbb{R}$ for some constant $\rho > 0$, where ω is symplectic area and μ_L is the Maslov class.
- In (compact) symplectically aspherical M one can take L to be tautologically unobstructed, ie L bounds no J–holomorphic discs for some J, which holds for all J if $\omega(\pi_2(M,L)) = 0$.

The conditions above on M and L serve to rule out "bad" (unstable) breakings (such as J–holomorphic sphere bubbles in M or J–holomorphic disc bubbles in M with boundary on L) from arising in the limit of a sequence of curves in the moduli spaces considered, which could obstruct compactness and/or simultaneously complicate transversality arguments.

Remark 3.11 (more general examples of admissible M and L) More generally, one could impose that the possible noncompactness of M and (if M is noncompact) L must be of the geometrically tame variety and that M/L have no/bound no J–holomorphic spheres/discs, or if they do, that such spheres/discs can either be shown (using classical methods) either not to arise in the compactifications of virtual one-dimensional moduli spaces, or to arise but only contribute canceling contributions to the resulting algebraic formulae.

For noncompact M and $\{L_i\}$, on any given moduli space of trajectories considered, further (nongeneric) assumptions on the profile of Floer perturbation data near ∞ are required to ensure Assumption 3.10 holds, to preclude sequences of curves escaping to ∞ so that usual Gromov compactness techniques apply, and also to obtain the second finiteness statement of Assumption 3.10, which is trivial in the compact case due to there being a finite list of outputs. We will say a few words about this in Remarks 3.15–3.18; the verification of Assumption 3.10 for the A_∞ structure maps (by citing established works) appears in Lemma 3.19. The verification of Assumption 3.10 for other moduli spaces considered in the paper is identical and hence omitted. However, we will in various places point out that the restrictions are needed on Floer data in noncompact cases to preclude curves escaping to infinity and obtain finiteness along the lines of Lemma 3.19.

3.3.2 Admissible Fukaya categories For an admissible M, we review the definition of the Fukaya category associated to an admissible collection of Lagrangians in M, which we will term an *admissible Fukaya category*. Examples of admissible Fukaya categories, in light of the examples given above, include:

(1) In a compact aspherical M, the Fukaya category of tautologically unobstructed Lagrangians.

(2) In a monotone M, the Fukaya category of monotone Lagrangians.

(3) In Liouville M, the Fukaya category of compact exact Lagrangians.

(4) In Liouville M, the wrapped Fukaya category of exact (cylindrical at infinity) Lagrangians.

Fix first an underlying ground field $\boldsymbol{k}$ and grading structure ($\mathbb{Z}$ or $\mathbb{Z}/2$ here, but see Remark 3.12) that we wish to use when defining the category. If $2c_1(M) = 0$ and we wish to define a $\mathbb{Z}$–graded category, we begin by equipping M with a *grading structure*, which is a trivialization of the square of the canonical bundle $(\Lambda^n_{\mathbb{C}} T^* M)^{\otimes 2}$. Next, one equips the Lagrangian submanifolds under consideration with some extra structure depending on the ground field $\boldsymbol{k}$ and the grading structure. Concretely, we say an *admissible Lagrangian brane* consists of a properly embedded admissible Lagrangian submanifold $L \subset M$ which is equipped with the following extra two optional pieces of data (which are only required if one wants to work with char $\boldsymbol{k} \neq 2$ or with $\mathbb{Z}$–gradings, respectively, the latter in particular is always excluded in the monotone case):

(3-33) an orientation and Spin structure, and

(3-34) a grading in the sense of [48] (with respect to the fixed grading structure on M).

These choices of extra data respectively require L to be Spin and satisfy $2c_1(M, L) = 0$, where $c_1(M, L) \in H^2(M, L)$ is the relative first Chern class.

Remark 3.12 There are other possible grading structures on M and L that one can use to equip the Fukaya category with suitable gradings (under geometric hypotheses), for instance $\mathbb{Z}/2k$–gradings, homology class gradings or hybrids thereof; cf [48; 59]. We suppress discussion of these, but, seeing as such matters are largely orthogonal to our arguments, note that our results apply in such contexts as well.

Henceforth, by abuse of notation all Lagrangians are implicitly admissible Lagrangian branes. Denote by ob $\mathcal{F}$ a finite collection of such (admissible) Lagrangians. Choose a (potentially time-dependent) Hamiltonian $H = H_t \colon M \to \mathbb{R}$ satisfying the following genericity condition:

Assumption 3.13 All time-1 chords of X_{H_t} between any pair of Lagrangians in ob $\mathcal{F}$ are nondegenerate.

Remark 3.14 It is straightforward to adapt all of our constructions to larger collections of Lagrangians by, for instance, choosing a different Hamiltonian H_{L_0,L_1} for each pair of Lagrangians L_0, L_1 (as well as a different H for closed orbits), and by choosing Floer perturbation data depending on corresponding sequences of objects; see eg [52, Section 9j]. We have opted to use a single H_t simply to keep the notation simpler.

Remark 3.15 (admissible Hamiltonians in the Liouville case) When M is Liouville, we need to impose further restrictions on the profile of H near ∞ in order to satisfy Assumption 3.10. If $\operatorname{ob}\mathcal{F}$ consists solely of compact exact Lagrangians, it suffices to impose that H is compactly supported, or more generally of the form $f(r)$ near infinity for some function with nonnegative first and second derivatives. If $\operatorname{ob}\mathcal{F}$ contains any noncompact Lagrangians, we will impose, following [1], that H satisfies the following *quadratic at* ∞ condition: $H = r^2$ on the cylindrical end (3-32), outside a compact subset.

For any pair of Lagrangians $L_0, L_1 \in \operatorname{ob}\mathcal{F}$ the set $\chi(L_0, L_1)$ of time-1 Hamiltonian flow lines of H from L_0 to L_1 can be thought of as the critical points of an action functional $\mathcal{P}_{L_0,L_1}$ on the *path space* from L_0 to L_1; this functional is, a priori, multivalued, but it is certainly $\mathbb{R}$–valued in the presence of primitives λ for ω and f_i for $\lambda|_{L_i}$. Given a choice of grading structure on M and grading for each L_i above, elements of $\chi(L_0, L_1)$ can be graded by the *Maslov index*

$$\deg\colon \chi(L_0, L_1) \to \mathbb{Z}. \tag{3-35}$$

In the absence of grading structures this is always well defined mod 2, provided our Lagrangians are oriented, which is automatic if they are Spin. As a graded $\boldsymbol{k}$–module, the morphism space in the Fukaya category between L_0 and L_1, also known as the (*wrapped* if M is Liouville) *Floer homology cochain complex of* L_0 *and* L_1 *with respect to* H, has one (free) generator for each element of $\chi(L_0, L_1)$; concretely,

$$\hom^i_{\mathcal{F}}(L_0, L_1) = CF^*(L_0, L_1, H_t, J_t) := \bigoplus_{\substack{x\in\chi(L_0,L_1)\\ \deg(x)=i}} |o_x|_{\boldsymbol{k}}, \tag{3-36}$$

where the *orientation line* o_x is the real vector space associated to x by index theory[11] and for any one-dimensional real vector space V and any ring $\boldsymbol{k}$, the $\boldsymbol{k}$*–normalization*

$$|V|_{\boldsymbol{k}} \tag{3-37}$$

[11] See [52, Section 11h]; a priori, o_x depends on a choice of trivialization of x^*TM compatible with the grading structure. However, there is a unique such choice in the presence of a $\mathbb{Z}$–grading, and in the $\mathbb{Z}/2$–graded case any two choices made induce canonically isomorphic orientation lines.

is the $\boldsymbol{k}$–module generated by the two possible orientations on V, with the relationship that their sum vanishes. If one does not want to worry about signs, note that $|V|_{\mathbb{Z}/2} \cong \mathbb{Z}/2$ canonically.

The A_∞ structure maps arise as counts of parametrized families of (suitably coherently perturbed) solutions to Floer's equation with source a disc with d inputs and one output. We will quickly summarize the definition and relevant choices required, referring the reader to standard references for more details. The basic reference we follow is [52] for Fukaya categories of compact exact Lagrangians in Liouville manifolds; see also [60] for the mostly straightforward generalization to the monotone case. In the main body of exposition, we focus on the (slightly simpler) case of compact (admissible) Lagrangians in compact (admissible) symplectic manifolds; we detail the additional data and variations required for Fukaya categories of exact Lagrangians in Liouville manifolds (which are simpler if one is working only with compact exact Lagrangians) in Remarks 3.15–3.18.

For $d \geq 2$, we use the notation $\overline{\mathcal{R}}^d$ for the (Deligne–Mumford compactified) moduli space of discs with $d+1$ marked points modulo reparametrization, with one point z_0^- marked as negative and the remainder $z_1^+, \dots, z_d^+$ (labeled counterclockwise from z_0^-) marked as positive. Orient the open (interior) locus $\mathcal{R}^d$ as in [52, Section 12g] and [1]. $\overline{\mathcal{R}}^d$ can be given the structure of a manifold-with-corners, and its higher strata are trees of stable discs with a total of d exterior positive marked points and 1 exterior negative marked point. Denote the positive and negative semi-infinite strips by

$$Z_+ := [0,\infty) \times [0,1], \tag{3-38}$$

$$Z_- := (-\infty,0] \times [0,1]. \tag{3-39}$$

One first equips the spaces $\overline{\mathcal{R}}^d$ for each d with a *consistent collection of strip-like ends* $\mathfrak{S}$; that is, for each component S of $\overline{\mathcal{R}}^d$, a collection of maps $\epsilon_k^{\pm} \colon Z_\pm \to S$ all with disjoint image in S, chosen so that positive/negative strips map to neighborhoods of positively/negatively labeled boundary marked points respectively, smoothly varying with respect to the manifold-with-corners structures and compatible with choices made on boundary and corner strata, which are products of lower-dimensional copies of spaces $\overline{\mathcal{R}}^k$.

In order to associate transversely cut out moduli spaces of such maps, one studies a parametric family of solutions to Floer's equation depending on a choice of "Floer (or perturbation) data" over the parameter space. Concretely, a *Floer datum* for a family of domains (in this case $\overline{\mathcal{R}}^d$) is a choice, for each domain S in the parametric family, of

- an S–dependent (domain-dependent) almost-complex structure J_S and Hamiltonian H_S,
- a one-form α on S,

which depend (smoothly) on the particular domain in $\overline{\mathcal{R}}^d$ (and the position in that domain), and are *compatible with strip-like ends*, meaning α pulls back to dt and (H_S, J_S) pull back to a fixed choice (H_t, J_t) in coordinates (3-38)–(3-39). One inductively chooses a *Floer datum for the A_∞ structure*, which is a choice of Floer data for the collection of domains $\{\overline{\mathcal{R}}^d\}_{d\geq 2}$ which is *consistent*, meaning that the Floer data chosen on a given family of domains $\overline{\mathcal{R}}^d$ agree smoothly along the boundary and corners (which are products of lower-dimensional spaces $\overline{\mathcal{R}}^k$) with previous choices of Floer data made. Such consistent choices exist essentially because spaces of Floer data are contractible.

Remark 3.16 (Floer data for compact exact Lagrangians in Liouville manifolds) If M is Liouville and we are studying the Fukaya category of compact exact Lagrangians, there is an additional requirement imposed on any Floer datum one uses; namely one requires that J_S be of *contact type* in a neighborhood of infinity in the sense of [52, (7.3)], and H_S be either 0 or of the form $f(r)$ near infinity for some function with nonnegative first and second derivatives. The more restrictive types of Floer data chosen for wrapped Fukaya categories in Remark 3.17 of course suffice as well.

Remark 3.17 (Floer data for wrapped Fukaya categories) Following [1], we recall the additional information and constraints appearing in Floer data for wrapped Floer theory (with quadratic Hamiltonians). If M is a Liouville manifold let $\psi^\rho\colon M \to M$ denote the time $\log(\rho)$ (outward) Liouville flow. One fixes for each S, in addition to (H_S, J_S, α_S), a collection of constants $w_k \in \mathbb{R}_{>0}$ for each end, called *weights* (so w_k is the weight associated to the k^{th} end), and a map $\rho_S\colon \partial S \to \mathbb{R}_{>0}$, called the *time-shifting map*, where:

(1) ρ_S should be constant and equal to the weight w_k on the k^{th} strip-like end.

(2) The one-form α_S should be *subclosed* (meaning $d\alpha_S \leq 0$), equal to $w_k\,dt$ in the local coordinates on each strip-like end, and 0 when restricted to ∂S. By Stokes' theorem, this condition implies the sum of weights over all negative ends is greater than or equal to the sum of weights over all positive ends, and there should therefore be at least one negative end always (in this case there is one).

(3) The Hamiltonian should be quadratic at infinity and pull back to $H \circ \psi^{w_k}/w_k^2$ in coordinates on each end. Note that such a Hamiltonian is quadratic if H is, by an elementary computation [1, Lemma 3.1].

(4) The almost-complex structure should be of contact type at infinity and equal to $(\psi^{w_k})^* J_t$ in coordinates on each end.

There is a *rescaling action* by $(\mathbb{R}_{>0}, \cdot)$ on the space of such surface dependent data, which sends

$$(\rho_S, \{w_k\}, \alpha_S, H_S, J_S) \mapsto \left(\lambda\rho_S, \{\lambda w_k\}, \lambda\alpha_S, \frac{H_S \circ \psi^\lambda}{\lambda^2}, (\psi^\lambda)^* J_S\right) \quad \text{for } \lambda \in \mathbb{R}_{>0}.$$

Using this action, one also relaxes the consistency requirement imposed: The Floer datum on $\overline{\mathcal{R}}^d$ must agree smoothly, on a boundary or corner stratum, with *some rescaling* of the previously made choice; compare [1, Definition 4.1].

Given our choices of Floer data, we can define the moduli spaces appearing in the A_∞ operations. First for any pair of objects L_0, L_1, and any pair of chords $x_0, x_1 \in \chi(L_0, L_1)$, define $\widetilde{\mathcal{R}}^1(x_0; x_1)$ to be the moduli space of maps $u\colon \mathbb{R}_s \times [0,1]_t \to M$ with boundary condition and asymptotics $u(s,0) \in L_0$, $u(s,1) \in L_1$, $\lim_{s\to+\infty} u(s,t) = x_1$ and $\lim_{s\to-\infty} u(s,t) = x_0$ satisfying Floer's equation for (H_t, J_t),

$$(du - X \otimes dt)^{0,1} = 0, \tag{3-40}$$

where X is the Hamiltonian vector field associated to H_t and $(0,1)$ is taken with respect to J_t. The translation action on $\mathbb{R}_s$ descends to a map on this moduli space (as the equation satisfied is s–independent), and we define the moduli space of (unparametrized) Floer strips to be

$$\mathcal{R}^1(x_0; x_1) := \widetilde{\mathcal{R}}^1(x_0; x_1)/\mathbb{R}, \tag{3-41}$$

with the added convention that whenever we are in a component of $\widetilde{\mathcal{R}}^1(x_0; x_1)$ with expected dimension 0, this quotient is replaced by the empty set. Now for $d \geq 2$ let $L_0, \dots, L_d$ be objects of $\mathcal{F}$ and fix any sequence of chords $\vec{x} = \{x_k \in \chi(L_{k-1}, L_k)\}$ as well as another chord $x_0 \in \chi(L_0, L_d)$. We write $\mathcal{R}^d(x_0; \vec{x})$ for the space of maps

$$u\colon S \to M$$

with source an arbitrary element $S \in \mathcal{R}^d$, satisfying boundary conditions and asymptotics

$$\begin{cases} u(z) \in L_k & \text{if } z \in \partial S \text{ lies between } z^k \text{ and } z^{k+1}, \\ \lim\limits_{s\to\pm\infty} u \circ \epsilon^k(s, \cdot) = x_k, \end{cases} \tag{3-42}$$

where the limit above is taken as $s \to +\infty$ if the k^{th} end is positive and $-\infty$ if it is negative, and differential equation

$$(du - X_S \otimes \alpha_S)^{0,1} = 0, \tag{3-43}$$

where X_S is the Hamiltonian vector field associated to H_S and where $0, 1$ is taken with respect to the complex structure J_S (for the choice of consistent Floer datum we have fixed).

The consistency condition imposed on Floer data over the abstract moduli spaces $\overline{\mathcal{R}}^d$, along with the compatibility with strip-like ends, implies that the (Gromov-type) compactification of the space of maps $\overline{\mathcal{R}}^d(x_0; \vec{x})$ can be formed by adding the images of the natural inclusions of products of lower-dimensional such moduli spaces,

$$\overline{\mathcal{R}}^{d_2}(y; \vec{x}_2) \times \overline{\mathcal{R}}^{d_1}(x_0; \vec{x}_1) \to \overline{\mathcal{R}}^d(x_0; \vec{x}), \tag{3-44}$$

where y agrees with one of the elements of $\vec{x}_1$ and $\vec{x}$ is obtained by removing y from $\vec{x}_1$ and replacing it with the sequence $\vec{x}_2$. Here, we let d_1 range from 1 to d, with $d_2 = d - d_1 + 1$, with the stipulation that $d_1 = 1$ or $d_2 = 1$ is the semistable case (3-41).

Remark 3.18 (operations for wrapped Fukaya categories) In the setting of the wrapped Fukaya category (continuing Remark 3.17), one needs to incorporate the map ρ_S into the Lagrangian boundary conditions and asymptotics specified in Floer's equation; namely, instead of (3-42), we require the moving boundary condition $u(z) \in (\psi^{\rho_S(z)})^* L_k$ if $z \in \partial S$ lies between z^k and z^{k+1}, where $(\psi^\rho)^* L_i$ denotes the pullback by ψ^ρ (or application of $(\psi^\rho)^{-1} = \psi^{1/\rho}$). We similarly impose that on the k^{th} end, $\lim_{s \to \pm\infty} u \circ \epsilon^k(s, \cdot) = (\psi^{\rho_S(z) := w_k})^* x_k$. The point is that Liouville flow for time $\log(\rho)$ defines a canonical identification between Floer complexes,

$$CF^*(L_0, L_1; H, J_t) \simeq CF^*\Big((\psi^\rho)^* L_0, (\psi^\rho)^* L_1; \frac{H}{\rho} \circ \psi^\rho, (\psi^\rho)^* J_t\Big). \tag{3-45}$$

The right-hand object is equivalently the (wrapped) Floer complex for

$$((\psi^\rho)^* L_0, (\psi^\rho)^* L_1)$$

for a strip with one-form $\rho\, dt$ using Hamiltonian $(H/\rho^2) \circ \psi^\rho$ and $(\psi^\rho)^* J_t$. Up to Liouville flow, the Floer equation and boundary conditions satisfied on the k^{th} strip-like end therefore coincides with the usual Floer equation for (H_t, J_t) between L_{k-1} and L_k. In light of this condition and the weakened consistency requirement for Floer data described in Remark 3.17, one can again deduce (3-44), that lower-dimensional strata of the Gromov bordification of the space of maps can be identified (now possibly using a nontrivial Liouville rescaling) with products of previously defined moduli spaces.

In the graded setting, every connected component of the moduli space $\overline{\mathcal{R}}^d(x_0;\vec{x})$ has expected (or virtual) dimension $\deg(x_0)+d-2-\sum_{1\le k\le d}\deg(x_k)$; more generally, this moduli space consists of components of varying expected dimension (a number which can be computed using index theory in terms of the underlying homotopy class of u) all of whose mod 2 reductions are $\deg(x_0)+d-2-\sum_{1\le k\le d}\deg(x_k)$. The following lemma is the prototypical method of verifying Assumption 3.10 for the various moduli spaces considered throughout the paper.

Lemma 3.19 *Assumption 3.10 holds for the moduli spaces $\overline{\mathcal{R}}^d(x_0;\vec{x})$ for admissible M, $\{L_i\}$ and generic choices of a Floer datum for the A_∞ structure* (*satisfying the constraints detailed in Remarks 3.15–3.17 in the Liouville case*). *Namely*: *components of these moduli spaces of virtual dimension ≤ 1 are* (*for generic choices*) *compact manifolds-with-boundary of the given expected dimension. Moreover, given a fixed $\vec{x}$ these moduli spaces are empty for all but finitely many x_0*; *this is automatic if there are only finitely many possible x_0 to begin with*, *for instance if all of the Lagrangians being considered are compact.*

Proof If M is compact (and admissible), these assertions (the last of which is automatic) follow from standard Gromov compactness and transversality methods as in [52, (9k), (11h) and Proposition 11.13]. In the case that M and possibly also its Lagrangians are noncompact, there is an additional concern that solutions could escape to infinity in the target. To address this one can, for instance, appeal to the *integrated maximum principle* (compare [3, Lemma 7.2] or [1, Section B]), which implies that elements of $\mathcal{R}(x_0;\vec{x})$ have image contained in a compact subset of M dependent on x_0 and $\vec{x}$, from where one can again appeal to standard Gromov compactness techniques. (This is strongly dependent on the form of H, J and α chosen for our Floer data as in Remarks 3.15–3.17.) The same result can be used to show that solutions do not exist for x_0 of sufficiently negative *action* compared to $\vec{x}$ (with our conventions, action is bounded above and there are finitely many x_0 with action above any fixed level), verifying the last assertion. □

Choose a generic Floer datum for the A_∞ structure satisfying Lemma 3.19 and let $u\in\overline{\mathcal{R}}^d(x_0;\vec{x})$ be a rigid curve, meaning for us an element of the virtual dimension-0 component (which has dimension 0 in this case). By [52, (11h), (12b),(12d)], given the fixed orientation[12] of $\mathcal{R}^d$, any such element $u\in\overline{\mathcal{R}}^d(x_0;\vec{x})$ determines an isomorphism

[12] In the case $d\ge 2$, that is. For $d=1$, one instead needs to "orient the operation of quotienting by $\mathbb{R}$" as in [52, (12f)].

of orientation lines

$$\mathcal{R}^d_u\colon o_{x_d}\otimes\cdots\otimes o_{x_1}\to o_{x_0}. \tag{3-46}$$

Now for any one-dimensional vector spaces $V_1,\dots,V_k$ and W, an isomorphism

$$f\colon V_k\otimes\cdots\otimes V_1\to W$$

induces a canonical map between $\boldsymbol{k}$–normalizations,

$$|V_1|_{\boldsymbol{k}}\otimes\cdots\otimes|V_k|_{\boldsymbol{k}}\cong|V_1\otimes\cdots\otimes V_k|_{\boldsymbol{k}}\to|W|_{\boldsymbol{k}},$$

which by abuse of notation, to simplify notation, we also call f (rather than $|f|_{\boldsymbol{k}}$). Using this, for $d\ge 1$ define the d^{th} A_∞ operation

$$\mu^d\colon \hom^*_{\mathcal{F}}(L_{d-1},L_d)\otimes\cdots\otimes\hom^*_{\mathcal{F}}(L_0,L_1)\to\hom^*_{\mathcal{F}}(L_0,L_d) \tag{3-47}$$

as a sum

$$\mu^d([x_d],\dots,[x_1]):=\sum_{u\in\overline{\mathcal{R}}^d(x_0;\vec{x})\ \text{rigid}}(-1)^{\bigstar_d}\mathcal{R}^d_u([x_d],\dots,[x_1]), \tag{3-48}$$

where $[x_i]\in|o_{x_i}|_{\boldsymbol{k}}$ is an arbitrary element, $\mathcal{R}^d_u$ is the map (on $\boldsymbol{k}$–normalizations induced by) (3-46), and the sign is given by

$$\bigstar_d=\sum_{i=1}^d i\cdot\deg(x_i). \tag{3-49}$$

Note that this sum is finite by Lemma 3.19. An analysis of the codimension-1 boundary of one-dimensional moduli spaces along with their induced orientations establishes that the maps μ^d satisfy the A_∞ relations; see [52, Proposition 12.3].

We record here two abuses of notation which will systematically appear in definitions and usage of operations such as μ^d. First, as above, we will frequently use the same symbol for a multilinear map $F\colon V_1\times\cdots\times V_k\to W$ and its corresponding linear map $F\colon V_1\otimes\cdots\otimes V_k\to W$. Second, we will frequently use x_i to refer to the arbitrary element $[x_i]\in|o_{x_i}|_{\boldsymbol{k}}$ to simplify expressions; for instance, above we might write $\mu^d(x_d,\dots,x_1)$ in place of $\mu^d([x_d],\dots,[x_1])$.

4 Circle action on the closed sector

4.1 Floer cohomology and symplectic cohomology

Let M be admissible as in Section 3.3.1. Given a (potentially time-dependent) Hamiltonian $H\colon M\to\mathbb{R}$, *Hamiltonian Floer cohomology* when it is defined is formally the

Morse cohomology of the H–perturbed action functional $\mathcal{A}_H : \mathcal{L}M \to \mathbb{R}$ on the free loop space $\mathcal{L}M$ of M. If ω is exact and comes with a fixed primitive λ, this functional can be written as

$$x \mapsto -\int_x \lambda + \int_0^1 H_t(x(t))\,dt.$$

In general, $\mathcal{A}_H$ may be multivalued, but $d\mathcal{A}_H$ is always well defined, leading at least to a Morse–Novikov-type theory. Recall that the set of *critical points* of $\mathcal{A}_{H_t}$ (when H_t is implicit) is precisely the set $\mathcal{O}$ of time-1 orbits of the associated (time-dependent) Hamiltonian vector field X_H, and we assume H_t is chosen sufficiently generically that:

Assumption 4.1 The elements of $\mathcal{O}$ are nondegenerate.

Optionally, given the data of a grading structure on M in the sense of Section 3.3.2 one can define an absolute $\mathbb{Z}$–grading on orbits by $\deg(y) := n - \mathrm{CZ}(y)$, where CZ is the Conley–Zehnder index of y (and such a grading is always well defined mod 2).

Fix a (potentially S^1–dependent) almost-complex structure J_t. In the formal picture, this induces a metric on $\mathcal{L}M$. A *Floer trajectory* is formally a gradient flowline of $\mathcal{A}_{H_t}$ using the metric induced by J_t; concretely it is a map $u : (-\infty, \infty) \times S^1 \to M$ satisfying Floer's equation (3-40) (which is formally the gradient flow equation for $\mathcal{A}_{H_t}$), and converging exponentially near $\pm\infty$ to a pair of specified orbits $y^\pm \in \mathcal{O}$. In standard coordinates s, t on the cylinder (ie $s \in \mathbb{R}$, $t \in \mathbb{R}/\mathbb{Z} = S^1$) this reads as

$$\partial_s u = -J_t(\partial_t u - X). \tag{4-1}$$

The space of nonconstant Floer trajectories between a fixed y^+ and y^- modulo the free $\mathbb{R}$–action given by translation in the s direction is denoted by $\mathcal{M}(y^-; y^+)$. As in Morse theory, one should compactify this space by allowing *broken trajectories*,

$$\overline{\mathcal{M}}(y^-; y^+) = \coprod \mathcal{M}(y^k; y^+) \times \mathcal{M}(y^{k-1}; y^k) \times \cdots \times \mathcal{M}(y^1; y^2) \times \mathcal{M}(y^-; y^1). \tag{4-2}$$

In the graded situation, every component of $\overline{\mathcal{M}}(y^-; y^+)$ has expected/virtual dimension $\deg(y^-) - \deg(y^+) - 1$; in general, $\overline{\mathcal{M}}(y^-; y^+)$ has components of varying virtual dimension, of fixed parity $\deg(y^-) - \deg(y^+) - 1$, depending on the underlying homotopy class of the cylinder. By Assumption 3.10 for $\overline{\mathcal{M}}(y^-; y^+)$, for generic choices of (time-dependent) J_t, the virtual dimension ≤ 1 components of the moduli spaces $\overline{\mathcal{M}}(y^-; y^+)$ are compact manifolds (with boundary) of the given expected dimension; fix such a J_t.

Putting this all together, the *Floer cochain complex* for (H_t, J_t) over $\boldsymbol{k}$ has generators corresponding to orbits of H_t,

$$CF^i(M) := CF^i(M; H_t, J_t) := \bigoplus_{\substack{y \in \mathcal{O} \\ \deg(y)=i}} |o_y|_{\boldsymbol{k}}, \tag{4-3}$$

where the *orientation line* o_y is a real vector space associated to every orbit in $\mathcal{O}$ via index theory[13] (see eg [1, Section C.6]), and $|V|_{\boldsymbol{k}}$ is the $\boldsymbol{k}$–normalization of V as in (3-37). The differential $d\colon CF^*(M; H_t, J_t) \to CF^*(M; H_t, J_t)$ counts rigid elements of the compactified moduli spaces (4-2). To fix sign issues, we recall that for a rigid element $u \in \mathcal{M}(y_0; y_1)$ (meaning u belongs to a component of virtual, hence actual, dimension 0) there is a natural isomorphism between orientation lines induced by index theory (see eg [52, (11h), (12b),(12d)] and [1, Lemma C.4]),

$$\mu_u\colon o_{y_1} \to o_{y_0}. \tag{4-4}$$

Then, one defines the differential as

$$d([y_1]) = \sum_{u \in \overline{\mathcal{M}}(y_0; y_1) \text{ rigid}} (-1)^{\deg(y_1)} \mu_u([y_1]), \tag{4-5}$$

where $[y_1] \in |o_{y_1}|_{\boldsymbol{k}}$ is an arbitrary element and μ_u is the map (on $\boldsymbol{k}$–normalizations induced by) (4-4). One can show that $d^2 = 0$ (under the assumptions made), and we call the resulting cohomology group $HF^*(H_t, J_t)$.

If M is compact (and admissible), Assumption 3.10 holds for all (suitably generic) J_t, and all H_t whose time-1 orbits are nondegenerate as in Assumption 4.1. If M is noncompact and admissible then further hypotheses are needed on the profile of (H_t, J_t) at ∞ to obtain admissibility, in particular to prevent curves from escaping to ∞ in M and ensure compactness of $\bigcup_{y^-} \overline{\mathcal{M}}(y^-; y^+)$; we recall the two most relevant possible hypotheses for our purposes in Sections 4.1.1–4.1.2, which can lead to distinct Floer cohomology groups. For simplicity, the discussion in Section 4.1.2 subsumes the case of compact M as well.

Remark 4.2 Our (cohomological) grading convention for Floer cohomology follows [51; 47; 1; 24].

4.1.1 Symplectic cohomology

Symplectic cohomology [10; 11; 19; 64], the target of the open–closed map for wrapped Fukaya categories, is Hamiltonian Floer cohomology

[13] As before, this index-theoretic definition a priori depends on a choice of trivialization of y^*TM compatible with the grading structure, but any two choices induce isomorphic lines.

for a particular class of Hamiltonians on noncompact convex symplectic manifolds. There are several methods for defining this group. We define it here by making the following specific choices of target, Hamiltonian, and almost-complex structure:

- M is a Liouville manifold *equipped with a conical end*, meaning that it comes equipped with a choice of (3-32). (This serves primarily as a technical device; the resulting invariants are independent of the specific choice.)
- The Hamiltonian term H_t is a sum $H + F_t$ of an *autonomous Hamiltonian* $H\colon M \to \mathbb{R}$ which is *quadratic at* ∞, namely

$$H|_{M\setminus\overline{M}}(r,y) = r^2, \tag{4-6}$$

and a time-dependent perturbation F_t such that on the collar (3-32) of M, we have:

(4-7) For any $r_0 \gg 0$, there exists an $R > r_0$ such that $F(t,r,y)$ vanishes in a neighborhood of R.

For instance, F_t could be supported near nontrivial orbits of H, where it is modeled on a Morse function on the circle. We denote by $\mathcal{H}(M)$ the class of Hamiltonians satisfying (4-6).

- The almost-complex structure should belong to the class $\mathcal{J}(M)$ of complex structures which are (*rescaled*) *contact type* on the cylindrical end (3-32), meaning that for some $c > 0$,

$$\lambda \circ J = f(r)\,dr, \tag{4-8}$$

where f is any function with $f(r) > 0$ and $f'(r) \geq 0$.

A well-known result [47; 1] asserts that Assumption 3.10 holds for the resulting spaces of broken Floer trajectories (4-2). Hence if M, H_t and J_t are as above, one has a well-defined Floer chain complex $CF^*(M, H_t, J_t)$, which we refer to as the *symplectic cochain complex* $SC^*(M)$; this will be the Floer chain complex we use when working with wrapped Fukaya categories. We call the resulting cohomology group *symplectic cohomology* $SH^*(M)$.

4.1.2 Relative cohomology

We review here the Floer cohomology group that is the target of the open–closed map for an admissible symplectic manifold M when working with a Fukaya category of *compact* admissible Lagrangian submanifolds in the sense of Section 3.3.1. Fix a (nondegenerate, generic) pair (H_t, J_t) which is arbitrary for compact M and which satisfies the following additional properties if M is Liouville:

- H is linear of very small negative slope near infinity:

$$H_t|_{M\setminus\overline{M}}(r,y) = -\lambda r, \tag{4-9}$$

where r is the cylindrical coordinate and $\lambda \ll 1$ is a sufficiently small number (smaller than the length of any Reeb orbit on $\partial\overline{M}$).

- J_t is (rescaled) contact type near infinity as before.

It is well known that Assumption 3.10 holds for the moduli spaces (4-2) for generic (H_t, J_t) as above [47], and also that:

Proposition 4.3 *For generic (H_t, J_t) as above, there is an isomorphism*

$$HF^*(H_t, J_t) \cong H^*(\overline{M}, \partial\overline{M}).$$

$H^*(\overline{M}, \partial\overline{M})$ equals $H^*(M)$ in the case that M is compact, using the convention then that $\overline{M} = M$ and $\partial\overline{M} = \varnothing$). □

The isomorphism can be realized in one of two ways:

- Choose H_t as above to be a C^2–small (time-independent) Morse function, in which case a well-known argument of Floer [18] equates $HF^*(H_t, J_t)$ with the Morse complex of H by showing that all Floer trajectories must in fact be Morse trajectories of $H|_{\overline{M}}$ (which in turn, as H is inward pointing near $\overline{M}$, compute the relative cohomology).
- Construct a geometric *PSS morphism* [46]

$$\mathrm{PSS}\colon H^*(\overline{M}, \partial\overline{M}) \cong H_{2n-*}(M) \to HF^*(H_t, J_t).$$

4.2 The cohomological BV operator

The first-order BV operator is a Floer analogue of a natural operator that exists on the Morse cohomology of any manifold with a smooth S^1–action. Like the case of ordinary Morse theory, this operator exists even when the Hamiltonian and complex structure (cf Morse function and metric) are not S^1–equivariant.

For $p \in S^1$, consider the collection of cylindrical ends on $\mathbb{R} \times S^1$

$$\begin{aligned} \epsilon_p^+ &: (s,t) \mapsto (s+1, t+p) \quad &&\text{for } s \geq 0,\\ \epsilon_p^- &: (s,t) \mapsto (s-1, t) &&\text{for } s \leq 0. \end{aligned} \tag{4-10}$$

Pick $K\colon S^1 \times (\mathbb{R} \times S^1) \times M \to \mathbb{R}$ dependent on p, satisfying

$$(\epsilon_p^{\pm})^* K(p, s, \cdot, \cdot) = H(t, m), \tag{4-11}$$

meaning that

$$K_p(s,t,m) = \begin{cases} H(t+p,m) & \text{if } s \geq 1, \\ H(t,m) & \text{if } s \leq -1, \end{cases} \tag{4-12}$$

so, in the range $-1 \leq s \leq 1$, $K_p(s,t,m)$ interpolates between $H_{t+p}(m)$ and $H_t(m)$ (and outside of this interval is independent of s).

Similarly, pick a family of almost-complex structures $J\colon S^1 \times (\mathbb{R} \times S^1) \times M \to \mathbb{R}$ satisfying

$$(\epsilon_p^{\pm})^* J(p,s,t,m) = J(t,m). \tag{4-13}$$

Now, for $x^+, x^- \in \mathcal{O}$, define

$$\mathcal{M}_1(x^-;x^+) \tag{4-14}$$

to be the *parametrized moduli space* of Floer cylinders

$$\Big\{(p,u) \mid p \in S^1, u\colon S \to M \text{ is such that } \lim_{s\to\pm\infty} (\epsilon_p^{\pm})^* u(s,\cdot) = x^{\pm} \text{ and } (du - X_K \otimes dt)^{0,1} = 0\Big\}. \tag{4-15}$$

There is a natural bordification by adding broken Floer cylinders to either end,

$$\overline{\mathcal{M}}_1(x^-;x^+) = \coprod \mathcal{M}(a_0;x^+) \times \cdots \times \mathcal{M}(a_k;a_{k-1}) \times \mathcal{M}_1(b_1;a_k) \times \mathcal{M}(b_2;b_1) \times \cdots \times \mathcal{M}(x^-;b_l). \tag{4-16}$$

Remark 4.4 (choices of K and J when M is noncompact) When M is noncompact and Liouville, further constraints on the profile of K and J are required near ∞ (beyond genericity) in order to satisfy Assumption 3.10. In the case of symplectic cohomology described in Section 4.1.1, it suffices to choose K carefully as follows. Given that $H_t(M) = H + F_t$ is a sum of an autonomous term and a time-dependent term that is zero at infinitely many levels tending towards infinity, we can ensure that

(4-17) at infinitely many levels tending towards infinity, $K_p(s,t,m)$ is equal to r^2,

and in particular is autonomous. In the setting of Section 4.1.2 (when M is noncompact), we can similarly ensure a version of (4-17) with r^2 replaced by $-\lambda r$ (in this case we could also more simply ensure that $K_p(s,t,m) = -\lambda r$ outside a compact set). In either case, one can take J to be (rescaled) contact type on the cylindrical end. As usual, the verification of Assumption 3.10 for the moduli spaces (4-16) on Liouville M follows by combining the results [1, Section B] or [3, Lemma 7.2] — which prevent curves escaping to ∞ and ensure $\mathcal{M}_1(x^+, x^-)$ is empty for all but finitely many x^- given the constraints near ∞ fixed in this remark — with classical transversality and compactness arguments.

As before, $\overline{\mathcal{M}}_1(x^-; x^+)$ contains components of varying expected dimension depending on the underlying homotopy class β of a map. Due to the fact that we are studying one-parameter families of domains and not quotienting by $\mathbb{R}$, the relevant expected dimension is 2 more than the expected dimension of the components of $\overline{\mathcal{M}}(x^-; x^+)$ underlying the same homotopy class β. In particular, in the graded case, this expected dimension is $\deg(x^+) - \deg(x^-) + 1$ for every component. By Assumption 3.10 for admissible choices of the above data, ie generic choices satisfying Remark 4.4 in the noncompact case, every component of $\overline{\mathcal{M}}_1(x^-; x^+)$ of virtual dimension ≤ 1 is a compact manifold-with-boundary of dimension equal to its virtual dimension. (In particular, the boundary of the one-dimensional components consists of the once-broken trajectories in (4-16).) In the usual fashion, counting rigid elements of this compactified moduli space of maps with the right sign (explained more carefully in the next section) gives an operation $\delta_1 \colon CF^*(M) \to CF^{*-1}(M)$ satisfying

$$d\delta_1 + \delta_1 d = 0,$$

which comes from the fact that the codimension-1 boundary of $\overline{\mathcal{M}}_1(x^-; x^+)$ is

$$\coprod_y \overline{\mathcal{M}}(y; x^+) \times \overline{\mathcal{M}}_1(x^-; y) \cup \overline{\mathcal{M}}_1(y; x^+) \times \overline{\mathcal{M}}(x^-; y).$$

It would be desirable for δ_1 to square to zero on the chain level, which would give $(CF^*(M), \delta_0 = d, \delta_1)$ the structure of a *strict S^1–complex*, or *mixed complex*. However, the S^1–dependence of our Hamiltonian and almost-complex structure prevent this, in a manner we now explain.

Typically, one attempts to prove that a geometric/Floer-theoretic operation (such as δ_1^2) is zero by exhibiting that the relevant moduli problem has no zero-dimensional solutions (due to, say, extra symmetries in the equation), or otherwise arises as the boundary of a one-dimensional moduli space. To that end, we first describe a moduli space parametrized by $S^1 \times S^1$ which looks like two of the previous parametrized spaces naively superimposed, leading us to call the associated operation $\delta_2^{\mathrm{naive}}$. The extra symmetry involved in this definition will allow us to easily conclude:

Lemma 4.5 *The operation $\delta_2^{\mathrm{naive}}$ is the zero operation.*

For $(p_1, p_2) \in S^1 \times S^1$, consider the collection of cylindrical ends

$$\begin{aligned} \epsilon^+_{(p_1,p_2)}&\colon (s,t) \mapsto (s+1, t+p_1+p_2) && \text{for } s \geq 0, \\ \epsilon^-_{(p_1,p_2)}&\colon (s,t) \mapsto (s-1, t) && \text{for } s \leq 0. \end{aligned} \tag{4-18}$$

Pick $K\colon (S^1 \times S^1) \times (\mathbb{R} \times S^1) \times M \to \mathbb{R}$ dependent on (p_1, p_2), satisfying

$$\epsilon^{\pm}_{(p_1,p_2)} K(p_1, p_2, s, \cdot, \cdot) = H(t,m), \tag{4-19}$$

meaning that

$$K_{(p_1,p_2)}(s,t,m) = \begin{cases} H(t+p_1+p_2, m) & \text{if } s \geq 1, \\ H(t,m) & \text{if } s \leq -1, \end{cases} \tag{4-20}$$

so in the range $-1 \leq s \leq 1$, $K_{p_1+p_2}(s,t,m)$ interpolates between $H_{t+p_1+p_2}(m)$ and $H_t(m)$.

Similarly, pick a family of almost-complex structures $J\colon S^1 \times S^1 \times (\mathbb{R} \times S^1) \times M \to \mathbb{R}$,

$$\epsilon^{\pm}_{(p_1,p_2)} J(p_1, p_2, s, t, m) = J(t,m), \tag{4-21}$$

such that

$$J \text{ only depends on the sum } p_1 + p_2. \tag{4-22}$$

Now, for $x^+, x^- \in \mathcal{O}$, define

$$\mathcal{M}_2^{\text{naive}}(x^-; x^+) \tag{4-23}$$

to be the *parametrized moduli space* of Floer cylinders

$$\bigl\{(p_1, p_2, u) \mid (p_1, p_2) \in S^1 \times S^1, u\colon S \to M \text{ is such that } \lim_{s \to \pm\infty} (\epsilon^{\pm}_{(p_1,p_2)})^* u(s, \cdot) = x^{\pm} \text{ and } (du - X_K \otimes dt)^{0,1} = 0\bigr\}. \tag{4-24}$$

For generic choices of K and J (again bearing in mind the extra impositions of Remark 4.4 in the noncompact case), this moduli space, suitably compactified by adding broken trajectories, will be (for components of virtual dimension ≤ 1) a manifold with boundary of the correct (expected) dimension; the dimension agrees mod 2 in the $\mathbb{Z}/2$–graded case and exactly in the graded case with $\deg(x^+) - \deg(x^-) + 2$. The details are similar to the previous section, and will be omitted. Counts of rigid elements in this moduli space will thus, in the usual fashion, give a map of degree -2, which we call δ_2^{naive}.

Proof of Lemma 4.5 Let (p_1, p_2, u) be an element of $\mathcal{M}_2^{\mathrm{naive}}(x^-; x^+)$. Then, for any $r \in S^1$, $(p_1 - r, p_2 + r, u)$ is an element too, as the equation satisfied by the map u only depends on the sum $p_1 + p_2$. We conclude that elements of $\mathcal{M}_2^{\mathrm{naive}}(x^-; x^+)$ are never rigid, and thus that the resulting operation $\delta_2^{\mathrm{naive}}$ is zero. □

We would like $\delta_2^{\mathrm{naive}}$ to be genuinely equal to δ_1^2, which would imply that $\delta_1^2 = 0$. However, this is only true on the homology level; the lack of S^1 invariance of our Hamiltonian and almost-complex structure, and the corresponding family of choices of homotopy between $\theta^* H_t$ and H_t, over varying $\theta \in S^1$, breaks symmetry and ensures that $\delta_1^2 \neq \delta_2^{\mathrm{naive}}$ as geometric chain maps. However, there is a geometric chain homotopy, δ_2, between δ_1^2 and $\delta_2^{\mathrm{naive}}$, along with a hierarchy of higher homotopies δ_k forming the S^1–complex structure on $CF^*(M)$, which we define in the next section. See in particular Lemma 4.11 for the proof of the S^1–complex equations, one of which recovers the chain homotopy between δ_1^2 and $\delta_2^{\mathrm{naive}} = 0$.

4.3 The A_∞ circle action

We turn to a "coordinate-free" definition of the relevant parametrized moduli spaces, which will help us incorporate the construction into open–closed maps.

Definition 4.6 An *r–point angle-decorated cylinder* consists of a semi-infinite or infinite cylinder $C \subseteq (-\infty, \infty) \times S^1$, along with a collection of auxiliary points $p_1, \dots, p_r \in C$, satisfying

$$(p_1)_s \le \cdots \le (p_r)_s, \tag{4-25}$$

where $(a)_s$ denotes the $s \in (-\infty, \infty)$ coordinate. The *heights* associated to this data are the s coordinates

$$h_i = (p_i)_s \quad \text{for } i = 1, \dots, r, \tag{4-26}$$

and the *angles* associated to C are the S^1 coordinates

$$\theta_i := (p_i)_t \quad \text{for } i \in 1, \dots, r. \tag{4-27}$$

The *cumulative rotation* of an r–point angle-decorated cylinder is the first angle:

$$\eta := \eta(C, p_1, \dots, p_r) = \theta_1. \tag{4-28}$$

The *i^{th} incremental rotation* of an r–point angle-decorated cylinder is the difference between the i^{th} and $i - 1^{\mathrm{st}}$ angles,

$$\kappa_i^{\mathrm{inc}} := \theta_i - \theta_{i+1}, \quad \text{where } \theta_{r+1} = 0. \tag{4-29}$$

Inductively, each θ_i can be expressed as a sum of all incremental rotations from i to r,

$$\theta_i = \sum_{j=i}^{r} \kappa_j^{\text{inc}}. \tag{4-30}$$

Definition 4.7 The *moduli space of r–point angle-decorated cylinders*

$$\mathcal{M}_r \tag{4-31}$$

is the space of r–point angle-decorated infinite cylinders, modulo translation.

Remark 4.8 (orientation for $\mathcal{M}_r$) The space C_r of all r–point angle-decorated infinite cylinders (not modulo translation) has a canonical complex orientation. Thus, to orient the quotient space $\mathcal{M}_r := C_r/\mathbb{R}$ it is sufficient to give a choice of trivialization of the action of $\mathbb{R}$ on C_r. We choose ∂_s to be the vector field inducing said trivialization.

For an element of this moduli space, the angles and relative heights of the auxiliary points continue to be well defined, so there is a noncanonical isomorphism

$$\mathcal{M}_r \simeq (S^1)^r \times [0,\infty)^{r-1}. \tag{4-32}$$

The moduli space $\mathcal{M}_r$ thus possesses the structure of an open manifold-with-corners, with boundary and corner strata given by the various loci where heights of the auxiliary points p_i are coincident.[14] Given an arbitrary representative C of $\mathcal{M}_r$ with associated heights $h_1, \dots, h_r$, we can always find a translation $\widetilde{C}$ satisfying $\widetilde{h}_r = -\widetilde{h}_1$; we call this the *standard representative* associated to C.

Given a representative C of this moduli space, and a fixed constant δ, we fix a positive cylindrical end around $+\infty$,

$$\epsilon^+ \colon [0,\infty) \times S^1 \to C, \quad (s,t) \mapsto (s + h_r + \delta, t), \tag{4-33}$$

and a negative cylindrical end around $-\infty$ (note the angular rotation in t!),

$$\epsilon^- \colon (-\infty, 0] \times S^1 \to C, \quad (s,t) \mapsto (s - (h_1 - \delta), t + \theta_1). \tag{4-34}$$

These ends are disjoint from the p_i and vary smoothly with C; via thinking of C as a sphere with two points with asymptotic markers removed, these cylindrical ends correspond to the positive asymptotic marker having angle 0 and the negative asymptotic marker having angle $\theta_1 = \kappa_1^{\text{inc}} + \kappa_2^{\text{inc}} + \cdots + \kappa_r^{\text{inc}}$.

[14] We allow the points p_i themselves to coincide; one alternative is to first Deligne–Mumford compactify, and then collapse all sphere bubbles containing multiple points p_i. That the result still forms a smooth manifold-with-corners is a standard local calculation near any such stratum.

There is a compactification of $\mathcal{M}_r$ consisting of *broken r–point angle-decorated cylinders*,

$$\overline{\mathcal{M}}_r = \coprod_s \coprod_{\substack{j_1,\dots,j_s \\ j_i>0, \sum j_i = r}} \mathcal{M}_{j_1} \times \cdots \times \mathcal{M}_{j_s}. \tag{4-35}$$

The stratum consisting of s–fold broken configurations lies in the codimension-s boundary, with the manifold-with-corners structure explicitly defined by local gluing maps using the ends (4-33) and (4-34). The gluing maps, which rotate the bottom cylinder of the gluing in order to match its top end (4-33) with the bottom end (4-34) of the upper cylinder, induce cylindrical ends on the glued cylinders, which agree with the choices of ends made in (4-33)–(4-34). Concretely, for a 1–fold broken configuration of the form $\mathcal{M}_{r-k} \times \mathcal{M}_k$, the gluing map, for any choice of sufficiently small gluing parameter, has the following effect on angles:

$$\big((\theta_1,\dots,\theta_{r-k}),(\bar{\theta}_1,\dots,\bar{\theta}_k)\big) \mapsto \big(\bar{\theta}_1+\theta_1,\bar{\theta}_2+\theta_1,\dots,\bar{\theta}_k+\theta_1,\theta_1,\dots,\theta_{r-k}\big), \tag{4-36}$$

where we have denoted coordinates in the second, bottom factor by $\bar{\theta}_j$ for $1 \le j \le k$, and in the first, top factor by θ_i for $1 \le i \le r-k$; see Figure 1. More simply, in the glued surface, the list of incremental angles $(\kappa_1^{\mathrm{inc,glued}},\dots,\kappa_r^{\mathrm{inc,glued}})$ is equal to the concatenation of the lists of incremental angles of the original bottom and top surfaces, $(\bar{\kappa}_1^{\mathrm{inc}},\bar{\kappa}_2^{\mathrm{inc}},\dots,\bar{\kappa}_k^{\mathrm{inc}},\kappa_1^{\mathrm{inc}},\kappa_2^{\mathrm{inc}},\dots,\kappa_{r-k}^{\mathrm{inc}})$.

The compactification $\overline{\mathcal{M}}_r$ thus has codimension-1 boundary covered by the images of the natural inclusion maps

$$\overline{\mathcal{M}}_{r-k} \times \overline{\mathcal{M}}_k \to \partial\overline{\mathcal{M}}_r \quad \text{for } 0<k<r, \tag{4-37}$$

$$\overline{\mathcal{M}}_r^{i,i+1} \to \partial\overline{\mathcal{M}}_r \quad \text{for } 1 \le i < r, \tag{4-38}$$

where $\overline{\mathcal{M}}_r^{i,i+1}$ denotes the compactification of the locus where i^{th} and $i+1^{\text{st}}$ heights are coincident,

$$\mathcal{M}_r^{i,i+1} := \{C \in \mathcal{M}_r \mid h_i = h_{i+1}\}. \tag{4-39}$$

With regards to the above stratum, for $r>1$ there is a projection map which will be relevant, a version of the forgetful map which remembers only the first of the angles with coincident heights:

$$\begin{gathered} \pi_i : \mathcal{M}_r^{i,i+1} \to \mathcal{M}_{r-1}, \\ (h_1,\dots,h_i,h_{i+1}=h_i,h_{i+2},\dots,h_r) \mapsto (h_1,\dots,h_i,h_{i+2},\dots,h_r), \\ (\theta_1,\dots,\theta_i,\theta_{i+1},\dots,\theta_r) \mapsto (\theta_1,\dots,\theta_{i-1},\theta_i,\widehat{\theta}_{i+1},\theta_{i+2},\dots,\theta_r). \end{gathered} \tag{4-40}$$

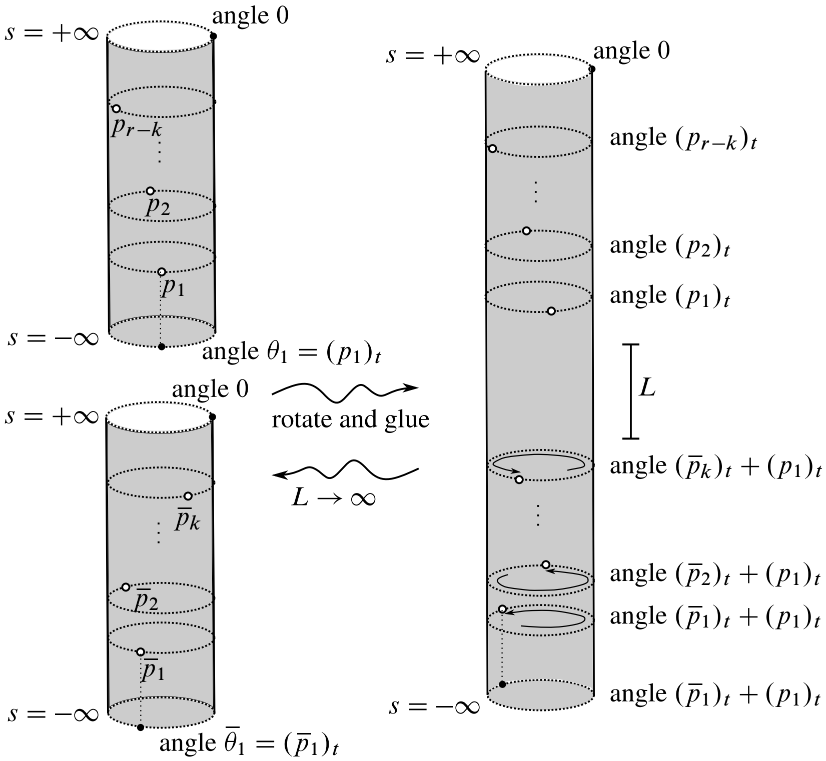


Figure 1: The gluing map for an angle-decorated cylinder rotates all of the angles of the bottom cylinder by the first angle of the top cylinder as in (4-36).

The map π_i is compatible with the choice of positive and negative ends (4-33)–(4-34) and hence π_i extends to compactifications

$$\pi_i\colon \overline{\mathcal{M}}_r^{i,i+1} \to \overline{\mathcal{M}}_{r-1}. \tag{4-41}$$

We will equip each r–point angle-rotated cylinder $\widetilde{C} := (C, p_1, \dots, p_r)$ with perturbation data for Floer's equation or a *Floer datum* in the sense of the last section, which consists of

- the positive and negative cylindrical ends on $\epsilon^{\pm}\colon C^{\pm} \to C$ chosen in (4-33)–(4-34),
- the one-form on C given by $\alpha = dt$,
- a surface-dependent Hamiltonian $H_{\widetilde{C}}\colon C \to \mathcal{H}(M)$ compatible with the positive and negative cylindrical ends, meaning that

$$(\epsilon^{\pm})^* H_{\widetilde{C}} = H_t, \tag{4-42}$$

where H_t was the previously chosen Hamiltonian, and

- a surface-dependent almost-complex structure $J_{\tilde{C}}\colon C \to \mathcal{J}_1(M)$ also compatible with $\epsilon^{\pm}$, meaning that

$$(\epsilon^{\pm})^* J_{\tilde{C}} = J_t \tag{4-43}$$

for our previously fixed choice J_t.

A choice of *Floer data for the S^1–action* is an inductive (smoothly varying) choice of Floer data, for each r and each representative $S = (C, p_1, \dots, p_r)$ of $\overline{\mathcal{M}}_r$, satisfying the following consistency conditions at boundary strata:

(4-44) At a boundary stratum (4-37), the datum chosen coincides with the product of Floer data already chosen on lower-dimensional spaces.

(4-45) At a boundary stratum (4-38), the Floer data coincides with the pullback, via the forgetful map π_i defined in (4-41) of the Floer data chosen on $\overline{\mathcal{M}}_{r-1}$.

Inductively, since the space of choices at each level is nonempty and contractible (and since the consistency conditions are compatible along overlapping strata), universal and consistent choices of Floer data exist. From the gluing map, a representative S sufficiently near the boundary strata (4-37) inherits cylindrical regions, also known as *thin parts*, which are the surviving images of the cylindrical ends of lower-dimensional strata. Together with the cylindrical ends of S, this determines a collection of cylindrical regions.

Definition 4.9 Given a fixed positive constant δ, the (δ*–spaced*) *rotated cylindrical regions* for an r–point angle-decorated cylinder $(C, p_1, \dots, p_r)$ consist of the following cylindrical ends and finite cylinders, where $h_i = (p_i)_s$ and $\theta_i = (p_i)_t$:

- The *top cylinder* $\epsilon^+\colon [0, \max(\mathrm{top}(C) - (h_r + \delta), 0)] \times S^1 \to C$, defined by

$$(s,t) \mapsto \bigl(\min(s + h_r + \delta, \mathrm{top}(C)), t\bigr). \tag{4-46}$$

- The *bottom cylinder* $\epsilon^-\colon [\min(\mathrm{bottom}(C) - (h_1 - \delta), 0), 0] \times S^1 \to C$, defined by

$$\begin{aligned}(s,t) &\mapsto \bigl(\max(s - (h_1 - \delta), \mathrm{bottom}(C)), t + \theta_1\bigr)\\ &= \Bigl(\max(s - (h_1 - \delta), \mathrm{bottom}(C)), t + \sum_{j=1}^{r} \kappa_i\Bigr).\end{aligned}$$

- For any $1 \le i \le r-1$ satisfying $h_{i+1} - h_i > 2\delta$, the i^{th} *thin part* is

$$\epsilon_i\colon [h_i + \delta, h_{i+1} - \delta] \times S^1 \to C, \quad (s,t) \mapsto (s, t + \eta_i). \tag{4-47}$$

Note that a given r–point angle-decorated cylinder may not have an i^{th} thin part for a given $i \in [1, r-1]$, and indeed may not have any thin parts. The consistency conditions at boundary strata can be ensured in particular by requiring that for any (δ–spaced) rotated cylindrical region $\epsilon\colon C' \to C$ of sufficiently large length (greater than some fixed L, say $L = 3\delta$) associated to $(C, p_1, \dots, p_r)$ and δ, we have that

$$\epsilon^*(K_C, J_C) = (K_t, J_t). \tag{4-48}$$

Given the cylindrical regions of Definition 4.9, this would imply the following condition on (K_C, J_C) (assuming $L \gg 3\delta$): for $z = (s,t) \in C$,

$$(K_z, J_z) = \begin{cases} (K_t, J_t) & \text{for } s > h_r + \delta, \\ (\epsilon^-)^*(K_t, J_t) = (K_{t+\theta_1}, J_{t+\theta_1}) & \text{for } s < h_1 - \delta, \\ \epsilon_i^*(K_z, J_z) = (K_{t+\theta_i}, J_{t+\theta_i}) & \text{if } h_{i+1} - h_i > 3\delta \\ & \quad \text{and } s \in [h_i + \delta, h_{i+1} - \delta]. \end{cases} \tag{4-49}$$

Given a choice of Floer data for the S^1–action and a pair of asymptotics $(x^+, x^-) \in \mathcal{O}$ for each $k \geq 1$, there is an associated parametrized moduli space of Floer cylinders with source an arbitrary element of $S \in \mathcal{M}_r$, where the Floer equation is with respect to the Hamiltonian H_S and J_S, with asymptotics (x^+, x^-):

$$\begin{aligned} &\mathcal{M}_r(x^-; x^+) := \\ &\quad \Big\{S = (C, p_1, \dots, p_r) \in \mathcal{M}_r,\ u\colon C \to M \mid \lim_{s \to \pm\infty} (\epsilon^\pm)^* u(s, \cdot) = x^\pm \text{ and} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad (du - X_{H_S} \otimes dt)^{(0,1)_S} = 0\Big\}. \end{aligned} \tag{4-50}$$

The consistency condition imposes that the boundary of the Gromov bordification $\overline{\mathcal{M}}_r(x^-; x^+)$ is covered by the images of the natural inclusions

$$\overline{\mathcal{M}}_{r-k}(y; x^+) \times \overline{\mathcal{M}}_k(x^-; y) \to \partial\overline{\mathcal{M}}_r(x^-; x^+), \tag{4-51}$$

$$\overline{\mathcal{M}}_r^{i,i+1}(x^-; x^+) \to \partial\overline{\mathcal{M}}_r(x^-; x^+), \tag{4-52}$$

along with the usual semistable strip-breaking boundaries

$$\begin{aligned} \overline{\mathcal{M}}(y; x^+) \times \overline{\mathcal{M}}_r(x^-; y) &\to \partial\overline{\mathcal{M}}_r(x^-; x^+), \\ \overline{\mathcal{M}}_r(y; x^+) \times \overline{\mathcal{M}}(x^-; y) &\to \partial\overline{\mathcal{M}}_r(x^-; x^+). \end{aligned} \tag{4-53}$$

Remark 4.10 (Floer data in the Liouville case) Continuing Remark 4.4, when M is Liouville we impose the following further constraint on Floer data:

(4-54) $H_{\widetilde{C}}$ is equal to r^2 or $-\lambda r$ (depending on whether we are in the setting of Section 4.1.1 or Section 4.1.2) at infinitely many levels of r tending to ∞, and $J_{\widetilde{C}}$ is (rescaled) contact type near ∞.

(In fact, in the setting of Section 4.1.2 we can take $H_{\widetilde{C}}$ to be simply equal to $-\lambda r$ for all r outside of a compact set.) By [3, Lemma 7.2] or [1, Section B], this hypothesis implies that sequences of curves with fixed asymptotics cannot escape to ∞ in M, and that $\mathcal{M}_r(x^-;x^+)$ given a fixed x^+ is nonempty for only finitely many x^-, both necessary inputs to verifying Assumption 3.10.

In the $\mathbb{Z}$–graded case, the virtual dimension of (every component of) $\overline{\mathcal{M}}_r(x^-;x^+)$ is

$$\deg(x^+)-\deg(x^-)+(2r-1), \tag{4-55}$$

while in the $\mathbb{Z}/2$–graded case every component has virtual dimension of the above parity. By Assumption 3.10, for a generic fixed choice of Floer data for the S^1–action (satisfying Remark 4.10 in the Liouville case), the components of virtual dimension ≤ 1 of the moduli spaces $\overline{\mathcal{M}}_r(x^-;x^+)$ are compact manifolds-with-boundary of the correct (expected) dimension. As usual, signed counts of rigid elements of this moduli space for varying x^+ and x^- (using induced maps on orientation lines, twisted as in the differential by $(-1)^{\deg(x_+)}$ — see (4-5)) give the matrix coefficients for the overall map

$$\delta_r\colon CF^*(M)\to CF^{*-2r+1}(M). \tag{4-56}$$

In the degenerate case $r=0$ we define δ_0 to be the (already defined) differential,

$$\delta_0:=d\colon CF^*(M)\to CF^{*+1}(M). \tag{4-57}$$

Lemma 4.11 *For each r,*

$$\sum_{i=0}^{r}\delta_i\delta_{r-i}=0. \tag{4-58}$$

Proof The counts of rigid elements associated to the boundary of one-dimensional components of $\partial\overline{\mathcal{M}}_r(x^+;x^-)$, along with a description of this codimension-1 boundary (4-51)–(4-53) immediately implies that

$$\biggl(\sum_{i=1}^{r}\delta_i\delta_{r-i}\biggr)+\biggl(\sum_{i}\delta_r^{i,i+1}\biggr)+(d\delta_r+\delta_r d)=0, \tag{4-59}$$

where $\delta_r^{i,i+1}$ for each i is the operation associated to the moduli space of maps (4-52). (Observe that $\delta_2^{1,2}$ is precisely the operation δ_2^{naive} from Section 4.2.) Note that the consistency condition (4-45) implies that the Floer datum chosen for any element $S\in\mathcal{M}_r^{i,i+1}$ only depends on $\pi_i(S)$, where the forgetful map $\pi_i\colon\mathcal{M}_r^{i,i+1}\to\mathcal{M}_{r-1}$ has one-dimensional fibers. Hence given an element $(S,u)\in\overline{\mathcal{M}}_r^{i,i+1}(x^-;x^+)$, it follows that $(S',u)\in\overline{\mathcal{M}}_r^{i,i+1}(x^-;x^+)$ for all $S'\in\pi_i^{-1}\pi_i(S)$. In other words, elements of $\overline{\mathcal{M}}_r^{i,i+1}(x^-;x^+)$ are never rigid, so the associated operation $\delta_r^{i,i+1}$ is zero. □

By definition we conclude:

Corollary 4.12 *The pair* $(CF^*(M; H_t, J_t), \{\delta_r\}_{r\geq 0})$ *as defined above forms an* S^1*–complex, in the sense of Definition 2.2.* □

By using continuation maps parametrized by various $(S^1)^r \times (0,1]^r$ (or equivalently, by spaces of angle-decorated cylinders that are not quotiented by overall $\mathbb{R}$–translation), one can prove:

Proposition 4.13 *Any continuation map* $f\colon CF^*(M, H_1) \to CF^*(M, H_2)$ *enhances to a homomorphism* $\boldsymbol{F}$ *of* S^1*–complexes (which is, in particular, a quasi-isomorphism if* f *is). Moreover, this homomorphism is canonical up to homotopy, in the sense that any two homomorphisms* $\boldsymbol{F}$ *and* $\boldsymbol{F}'$ *enhancing* f *constructed geometrically from parametrized continuation maps differ by an exact premorphism of* S^1*–complexes (also constructed geometrically).* □

We omit the proof, which is standard; see eg [66], but note some notational differences. In particular, the S^1–complex defined on the symplectic cochain complex $SC^*(M)$ or the Hamiltonian Floer complex (with small negative slope if M is noncompact) is an invariant of M, up to quasi-isomorphism.

Remark 4.14 (relation to earlier definitions in the literature) In [6], three different definitions of S^1–equivariant symplectic cohomology are considered and shown to be equivalent. One of the definitions involves taking the S^1–equivariant homology associated to a certain S^1–complex defined on $CF^*(M) = SC^*(M)$ [6, Proposition 2.19]. After normalizing for differing conventions (eg homological versus cohomological conventions for Floer theory, and the fact that their u^{-1} is our u), it is direct to see that the S^1–complex constructed therein coincides up to equivalence with the one here — and even agrees on the chain level, seeing as the choices of Floer data chosen in that paper constitute a choice of Floer data for the S^1–action in our sense; compare, for instance, [6, Figure 1] with (4-49).

4.4 The circle action on the interior

From the formal point of view of Floer homology of M as the Morse homology of an action functional on the free loop space $\mathcal{L}M$, one would expect the contributions coming from constant loops to be acted on trivially by the $C_{-*}(S^1)$–action, which comes from rotation of free loops. This is indeed the case, as we now review.

Suppose that the Hamiltonian H_t defining $CF^*(M)$ is chosen to be C^2–small, time-independent, and Morse in the compact region of $\overline{M}$ (which equals M if M is compact). Then, Floer [18] proved that all orbits of H_t inside $\overline{M}$ are (constant orbits at) Morse critical points of H, and all Floer cylinders between such orbits which remain in $\overline{M}$ are in fact Morse trajectories of H.

Let $C_{\mathrm{Morse}}(H)$ denote the Morse complex of H. In the setting where H is as in Section 4.1.2 (M can be Liouville or compact), all contributions to $CF^*(M)$ (both orbits and cylinders) come from $\overline{M}$, so Floer's argument gives an isomorphisms

$$C_{\mathrm{Morse}}(H) \cong CF^*(M). \tag{4-60}$$

In the setting where H is quadratic at infinity (and M is Liouville) as in Section 4.1.1, one can ensure the collection of orbits coming from $\overline{M}$ is an action-filtered subcomplex — and, for instance, the integrated maximum principle will ensure that all cylinders between such orbits lie in $\overline{M}$. Hence, there is an inclusion of subcomplexes

$$C_{\mathrm{Morse}}(H) \to SC^*(M), \tag{4-61}$$

which, under smallness constraints on the Floer data for the S^1–action gives an S^1–*subcomplex* [66, Lemma 5.4], meaning the operators δ_k preserve the subcomplex and in fact the action filtration; hence a morphism of S^1–complexes. We will discuss both of the above cases at once: in either case, by considering a Hamiltonian which is C^2–small on $\overline{M}$, we obtain an inclusion of S^1–subcomplexes

$$C_{\mathrm{Morse}}(H) \to CF^*(M) \tag{4-62}$$

with the understanding that in the former case this inclusion is the whole complex.

Lemma 4.15 *There exists a choice of Floer data for the S^1–action so that $C_{\mathrm{Morse}}(H)$ becomes a trivial S^1–subcomplex; meaning that the various operators δ_r, $r \geq 1$, associated to the $C_{-*}(S^1)$–action strictly vanish on the subcomplex.*

Proof By the integrated maximum principle, any Floer trajectory with asymptotics along two generators in $C_{\mathrm{Morse}}(H)$ remains in the interior of $\overline{M}$. We can choose the Hamiltonian term of our Floer data on $\mathcal{M}_r$ in this region of M to be autonomous (ie t– and s–independent on the cylinder), C^2–small and Morse — in fact equal to H; then Floer's theorem [18] again guarantees that any Floer trajectory in $\mathcal{M}_r(x^-; x^+)$ between Morse critical points $x^\pm$ is in fact a Morse trajectory of H. It follows that for critical points x^+, x^- of H, any element $u = (C, \vec{p})$ in the parametrized moduli space of maps $\widetilde{\mathcal{M}}_r(x^-; x^+)$ solves an equation that is independent of the choice of parameter

$\vec{p} \in (S^1)^r \times (0,1]^{r-1}$. Namely, u lives in a family of solutions of dimension at least $2r-1$ (given by varying $\vec{p}$), and hence u cannot be rigid. The associated operation δ_r, which counts rigid solutions, is therefore zero. □

By invariance of the S^1–complex structure on $CF^*(M)$ (up to homotopically canonical quasi-isomorphism as in Proposition 4.13), we conclude:

Corollary 4.16 *For M compact and admissible, or Liouville with (H, J) as in Section 4.1.2, $CF^*(M)$ is quasi-isomorphic to a trivial S^1–complex, canonically up to homotopy.*

Corollary 4.17 *For M Liouville with (H, J) as in Section 4.1.1, the inclusion chain map*

$$C^*_{\mathrm{Morse}}(M) \to SC^*(M) \tag{4-63}$$

lifts (cohomologically) canonically to a chain map

$$\begin{aligned}(C_{\mathrm{Morse}}(H)[\![u]\!], d_{\mathrm{Morse}}) = (C_{\mathrm{Morse}}(H))^{\mathrm{h}S^1} \\ \to (SC^*(M))^{\mathrm{h}S^1} = (SC^*(M)[\![u]\!], \delta_{\mathrm{eq}}),\end{aligned} \tag{4-64}$$

inducing a cohomological map

$$H^*(M)[\![u]\!] \to H^*(SC^*(M)^{\mathrm{h}S^1}). \tag{4-65}$$

Remark 4.18 Another possibly more direct way of producing the map $H^*(M)[\![u]\!] \to H^*(SC^*(M)^{\mathrm{h}S^1})$ is via an S^1–equivariant enhancement $\widetilde{\mathrm{PSS}}$ of the PSS morphism $\mathrm{PSS}\colon C^*(M) \to SC^*(M)$. We omit a further description here, and simply note that the resulting map can be shown to coincide cohomologically with the map defined above.

Since the S^1–complex structure on $C^*_{\mathrm{Morse}}(H)$ is trivial, one can (canonically) split the inclusion of homotopy fixed points map (2-36) $H^*(C^*_{\mathrm{Morse}}(H)^{\mathrm{h}S^1}) \to H^*(C^*_{\mathrm{Morse}}(H))$ by the map

$$H^*(M) \xrightarrow{[x \mapsto x \cdot 1]} H^*(M)[\![u]\!]. \tag{4-66}$$

The associated composition

$$H^*(M) \xrightarrow{[x \mapsto x \cdot 1]} H^*(M)[\![u]\!] \to H^*(SC^*(M)^{\mathrm{h}S^1}) \xrightarrow{[\iota]} SH^*(M) \tag{4-67}$$

coincides with the usual map $H^*(M) \to SH^*(M)$. In particular, we note that the homotopy fixed-point complex of $SC^*(X)$ possesses a canonical (geometrically defined) cohomological element,

$$\widetilde{1} \in H^*(SC^*(M)^{\mathrm{h}S^1}), \tag{4-68}$$

lifting the usual unit $1 \in SH^*(M)$ (under the map $[\iota]$), defined as the image of 1 under the map (4-65).

5 Cyclic open–closed maps

5.1 Open–closed Floer data

Here we review the sort of Floer perturbation data that needs to be specified on the domains appearing in the open–closed map and their cyclic analogues. The main body of our treatment, following Section 3.3, consists of a (slightly modified) simplification of the setup from [1] tailored to the case of Fukaya categories of compact admissible M; in Remarks 5.1 and 5.2 below we will indicate the modifications we need to make — following [1] and building on Remarks 3.15–3.18 above — in the case of compact Fukaya categories of Liouville manifolds (minor modifications) or wrapped Fukaya categories (slightly more involved modifications). There is one notable deviation from [1], in that we allow our interior marked point to have a varying asymptotic marker and choose Floer data depending on this choice, as is done in constructions of BV-type operations in Hamiltonian Floer theory involving such asymptotic markers; see eg [58; 55].

Let S be a disc with d boundary punctures $z_1, \dots, z_d$ (labeled in counterclockwise order) marked as positive, and an interior marked point p removed, marked as either positive or negative; for the main body of the construction p is negative. We also equip the interior marked point p with an *asymptotic marker*, that is, a half-line $\tau_p \in T_pS$ (or equivalently an element of the unit tangent bundle, defined with respect to some metric). Call any such $S = (S, z_1, \dots, z_d, p, \tau_p)$ an *open–closed framed disc.*

In addition to the notation for semi-infinite strips (3-38)–(3-39), we use the following notation to refer to the positive and negative semi-infinite cylinder:

$$A_+ := [0, \infty) \times S^1, \tag{5-1}$$

$$A_- := (-\infty, 0] \times S^1. \tag{5-2}$$

A *Floer datum* on a stable open–closed framed disc S consists of the following choices on each component:

(1) A collection of *strip-like or cylindrical ends* $\mathfrak{S}$ around each boundary or interior marked point, of sign matching the sign of the marked point; strip-like ends were defined in Section 3.3 and a positive (resp. negative) cylindrical end is a map

$$\delta_j^{\pm}\colon A_{\pm} \to S.$$

(So for the main body of the construction, we use a negative cylindrical end around p.) All of the strip-like ends around each of the z_i should be positive, and all (strip-like or cylindrical) ends should have disjoint image in S. The cylindrical end around p should further should be *compatible with the asymptotic marker*, meaning the points with angle zero should asymptotically approach the marker,

$$\lim_{s\to\pm\infty} \delta^{\pm}(s,0) = \tau_p. \tag{5-3}$$

(2) A one-form α_S on S, an S–dependent Hamiltonian function H_S on M, and an S–dependent almost-complex structure J_S on M, such that on each strip-like end these data pull back to a given fixed (dt, H_t, J_t), (which we used to define Lagrangian Floer homology chain complexes) and on the cylindrical end this data pulls back to a given fixed $(dt, H_t^{\mathrm{cyl}}, J_t^{\mathrm{cyl}})$ which we used to define our Hamiltonian Floer homology chain complex. (Note that in many cases we could further simplify and choose $(H_t^{\mathrm{cyl}}, J_t^{\mathrm{cyl}}) = (H_t, J_t)$, given a sufficiently generic choice of (H_t, J_t).)

Given a stable open–closed framed disc S equipped with a Floer datum F_S, a collection of Lagrangians $\{L_0,\dots,L_{d-1}\}$ and asymptotics $\{x_1,\dots,x_d; y\}$ with x_i a chord between L_{i-1} and $L_{i \bmod d}$, a map $u\colon S\to M$ satisfies *Floer's equation for F_S with boundary $\{L_0,\dots,L_{d-1}\}$ and asymptotics $\{x_1,\dots,x_d; y\}$* if

$$(du - X_S \otimes \alpha_S)^{0,1} = 0 \quad \text{using the Floer data given by } F_S \tag{5-4}$$

(meaning X_S is the Hamiltonian vector field associated to H_S, and 0, 1 parts are taken with respect to J_S), and

$$\begin{cases} u(z)\in L_i \text{ if } z\in\partial S \text{ lies counterclockwise from } z_i, \text{ clockwise from } z_{i+1 \bmod d}, \\ \lim\limits_{s\to+\infty} u\circ\epsilon^k(s,\cdot) = x_k, \\ \lim\limits_{s\to\mp\infty} u\circ\delta(s,\cdot) = y. \end{cases} \tag{5-5}$$

Here ϵ^k denotes the k^{th} strip-like end, δ denotes the cylindrical end, and the sign $\mp$ in the last line is $-$ if δ is a negative end — which is the case for the main body of the construction — and $+$ if δ is a positive end.

Remark 5.1 (Floer data for compact Lagrangians in Liouville manifolds) If M is Liouville and we are studying the Fukaya category of compact exact Lagrangians, then we take H_t^{cyl}, J_t^{cyl} (the data required to define Floer cohomology) as in Section 4.1.2 and we again impose the additional requirements on Floer data described in Remark 3.16, As before the H_t^{cyl}, J_t^{cyl} and more restrictive types of Floer data chosen for wrapped Fukaya categories in Remark 5.2 below would also work. The ability to choose H_t^{cyl} and J_t^{cyl} as in Section 4.1.2 is indicative of a more general freedom in the Floer data here, which also will allow us later to define operations in which the interior marked point (and all boundary marked points) are positive; see Section 5.6.2.

Remark 5.2 (Floer data and Floer's equation for wrapped Fukaya categories) Almost exactly as in Remark 3.17, and following [1], in order to associate operations between the wrapped Fukaya category and symplectic cohomology one needs to make the following modifications to the notion of Floer data. First, one takes H_t^{cyl}, J_t^{cyl} to be the data defining the symplectic cochain complex as in Section 4.1.1. Then one equips S with strip-like and cylindrical ends as above. Let ψ^ρ as before denote the time $\log(\rho)$ Liouville flow on M. The modifications to the Floer data are:

- **Extra choices of weights and time-shifting maps** Exactly as in Remark 3.17, one associates a *weight* $w_k \in \mathbb{R}_{>0}$ to each boundary or interior marked point and a time-shifting map $\rho_S\colon \partial S \to \mathbb{R}_{>0}$ agreeing with w_k near the k^{th} strip-like end.
- **Modified requirements on the one-form** The one-form α_S should be subclosed, meaning $d\alpha_S \leq 0$, should restrict to 0 along ∂S, and restrict to $w_k\,dt$ on each (strip-like or cylindrical) end, as in Remark 3.17(2). It follows by Stokes' theorem that the weight at the (output) cylindrical end should be greater than the sum of weights over all (input) strip-like ends. In particular, it is not possible for α_S to be subclosed and restrict to 0 along ∂S, conditions necessary to appeal to the integrated maximum principle if the interior marked point were also positive. (This is a reflection of the fact that wrapped Fukaya categories do not admit geometric operations with no outputs.)
- **Modified requirements on Hamiltonians**, as in Remark 3.17(3) The Hamiltonian term should pull back to $H \circ \psi^{w_k}/w_k^2$ along any strip-like end, and to $H^{\mathrm{cyl}} \circ \psi^{w_k}/w_k^2$ along the cylindrical end. The Hamiltonian term should also be quadratic at infinitely many levels of (3-32) tending to infinity. (This is a slight weakening of Remark 3.17 coming from the fact that the Hamiltonian used to define $SC^*(X)$ is not quadratic at every level near infinity due to (4-7).)

- **Modified requirements on almost-complex structures**, as in Remark 3.17(4) The almost-complex structure should be contact type at infinity and pull back to $(\psi^{w_k})^* J_t$ along each strip-like end and $(\psi^{w_k})^* J_t^{\mathrm{cyl}}$ along the cylindrical end.

Exactly as in Remark 3.17, there is a rescaling action on the space of such Floer data, and we will relax any consistency requirement imposed on Floer data to allow for an arbitrary rescaling when equating different choices of Floer data. Finally, we note the slight modifications to the boundary and asymptotic conditions of Floer's equation (5-5), following Remark 3.18: on the boundary component of ∂S lying counterclockwise from z_i and clockwise from $z_{i+1 \bmod d}$ we impose the moving boundary condition $u(z) \in (\psi^{\rho_S(z)})^* L_i$, on the k^{th} strip-like end we impose $\lim_{s \to +\infty} u \circ \epsilon^k(s, \cdot) = (\psi^{w_k})^* x_k$, and on the cylindrical end we impose $\lim_{s \to -\infty} u \circ \delta(s, \cdot) = (\psi^{w})^* y$, where w is the weight associated to the interior puncture p.

Exactly as in the proof of Lemma 3.19, the constraints to Floer data in the Liouville case made in the above two remarks help ensure Assumption 3.10 holds for associated moduli spaces.

5.2 Nonunital open–closed maps

We begin by constructing a variant of the open–closed map of [1] with source the nonunital Hochschild complex of (3-11), which we call the *nonunital open–closed map* and indicate by $\mathcal{OC}$ or $\mathcal{OC}^{\mathrm{nu}}$:

(5-6) $$\mathcal{OC} := \mathcal{OC}^{\mathrm{nu}} \colon \mathrm{CH}^{\mathrm{nu}}_{*-n}(\mathcal{F}) \to CF^*(M).$$

This map actually has a straightforward explanation from the perspective of Remark 3.3: we define the map $\mathcal{OC}$ from $\widetilde{\mathrm{CH}}_*(\mathcal{F}^+)$ by counting discs with an arbitrary number of boundary punctures and one interior puncture asymptotic to an orbit, as in [1], with the proviso that we treat the formal elements e_L^+ as "fundamental class $[L]$ point constraints (ie empty constraints)": we fill back in the relevant boundary puncture and impose no constraints on that marked point. With respect to the decomposition (3-11), we define a pair of maps

(5-7) $$\check{\mathcal{OC}} \oplus \hat{\mathcal{OC}} \colon \mathrm{CH}_*(\mathcal{F}) \oplus \mathrm{CH}_*(\mathcal{F})[1] \to CF^*(M)$$

giving the left and right components of the nonunital open–closed map

(5-8) $$\mathcal{OC} \colon \mathrm{CH}^{\mathrm{nu}}_{*-n}(\mathcal{F}) \to SC(M), \quad (x, y) \mapsto \check{\mathcal{OC}}(x) + \hat{\mathcal{OC}}(y).$$

Since the left (check) factor is equal to the usual cyclic bar complex for Hochschild homology, $\check{\mathcal{OC}}$ will be defined exactly as the open–closed map is defined in [1] (briefly recalled below), and the new content is the map $\hat{\mathcal{OC}}$. We will define $\hat{\mathcal{OC}}$ below (and recall the definition of $\check{\mathcal{OC}}$) and prove, extending [1]:

Lemma 5.3 *$\mathcal{OC}$ is a chain map of degree n.*

We note a notational difference from [1], which uses $\mathcal{OC}$ to refer to what we call here $\check{\mathcal{OC}}$; in contrast, in this paper we use $\mathcal{OC}$ exclusively to refer to the (nonunital) open–closed map $\mathcal{OC} = \mathcal{OC}^{\mathrm{nu}} := \check{\mathcal{OC}} \oplus \hat{\mathcal{OC}}$ with domain the nonunital Hochschild complex. Of course, the two maps $\mathcal{OC}$ and $\check{\mathcal{OC}}$ are homologically the same. That is, assuming Lemma 5.3:

Corollary 5.4 *As homology-level maps,* $[\mathcal{OC}] = [\check{\mathcal{OC}}]$.

Proof By construction, the chain level map $\check{\mathcal{OC}}$ constructed in [1] factors as

$$\mathrm{CH}_{*-n}(\mathcal{F}) \subset \mathrm{CH}^{\mathrm{nu}}_{*-n}(\mathcal{F}) \xrightarrow{\mathcal{OC}} CF^*(M). \tag{5-9}$$

The first inclusion is a quasi-isomorphism by Lemma 3.2, since $\mathcal{F}$ is known to be cohomologically unital. □

The moduli space controlling the operation $\check{\mathcal{OC}}$, denoted by

$$\overline{\check{\mathcal{R}}}{}^1_d, \tag{5-10}$$

is the (Deligne–Mumford compactification of the) abstract moduli space of discs with d boundary positive punctures $z_1, \dots, z_d$ labeled in counterclockwise order and one interior negative puncture z_{out}, with an asymptotic marker τ_{out} at z_{out} pointing towards z_d. The space (5-10) has a manifold-with-corners structure, with boundary strata described in [1, Section C.3] — there, the space is called $\overline{\mathcal{R}}{}^1_d$ — in short, codimension-one strata consist of disc bubbles containing any cyclic subsequence of k inputs attached to an element of $\check{\mathcal{R}}^1_{d-k+1}$ at the relative position of this cyclic subsequence. Orient the top stratum $\check{\mathcal{R}}^1_d$ by trivializing it, sending $[S]$ to the unit disc representative S with z_d and z_{out} fixed at 1 and 0, and taking the orientation induced by the (angular) positions of the remaining marked points:

$$-dz_1 \wedge \dots \wedge dz_{d-1}. \tag{5-11}$$

The moduli space controlling the new map $\hat{\mathcal{OC}}$ is nearly identical to $\check{\mathcal{R}}^1_d$, but there is additional freedom in the direction of the asymptotic marker at the interior puncture z_{out}. The top (open) stratum is easiest to define: let

$$\mathcal{R}^{1,\mathrm{free}}_d \tag{5-12}$$

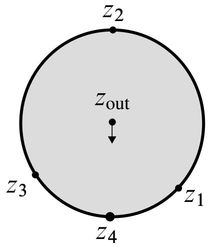


Figure 2: A representative of an element of the moduli space $\check{\mathcal{R}}^1_4$ with special points at 0 (output), $-i$.

be the moduli space of discs with d positive boundary punctures and one interior negative puncture as in $\check{\mathcal{R}}^1_d$, but with the asymptotic marker τ_{out} pointing anywhere between z_1 and z_d.

Remark 5.5 There is a delicate point in naively compactifying $\mathcal{R}_d^{1,\mathrm{free}}$: on any formerly codimension-one stratum in which z_1 and z_d bubble off, the position of τ_{out} becomes fixed too, and so the relevant stratum actually should have codimension two (and hence does not contribute to the codimension-one boundary equation for $\widehat{\mathcal{OC}}$; moreover, there is no nice corner chart near this stratum). For technical convenience, we pass to an alternative, larger (blown-up) model for the compactification in which these strata have codimension one but consist of degenerate contributions.

In light of Remark 5.5, we use (5-12) as motivation and instead define

$$\hat{\mathcal{R}}^1_d \tag{5-13}$$

to be the abstract moduli space of discs with $d+1$ boundary punctures $z_f, z_1, \dots, z_d$ and an interior puncture z_{out} with asymptotic marker τ_{out} pointing towards the boundary point z_f, modulo automorphism. We mark z_f as "auxiliary", but otherwise the space is abstractly isomorphic to $\check{\mathcal{R}}^1_{d+1}$. Identifying $\hat{\mathcal{R}}^1_d$ with the space of unit discs with z_{out} and z_f fixed at 1 and 0, the remaining (angular) positions of $z_1, \dots, z_d$ determine an orientation

$$-dz_1 \wedge \cdots \wedge dz_d. \tag{5-14}$$

The *forgetful map*

$$\pi_f \colon \hat{\mathcal{R}}^1_d \to \mathcal{R}_d^{1,\mathrm{free}} \tag{5-15}$$

puts back in the point z_f and forgets it. Since the point z_f is recoverable from the direction of the asymptotic marker at z_{out}, we get:

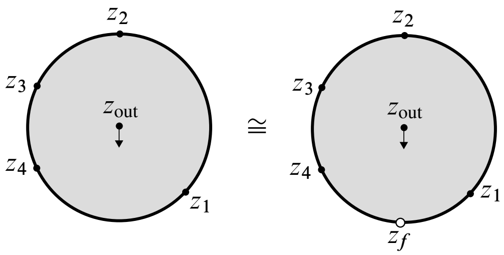


Figure 3: A representative of an element of the moduli space $\mathcal{R}^1_{4,\text{free}}$ and the corresponding element of $\widehat{\mathcal{R}}^1_4$.

Lemma 5.6 *The map π_f is a diffeomorphism.* □

The perspective of the former space (5-13) gives us a model for the compactification

$$\overline{\mathcal{R}}^{1,\text{free}}_d \tag{5-16}$$

as the ordinary Deligne–Mumford compactification

$$\overline{\widehat{\mathcal{R}}}{}^1_d. \tag{5-17}$$

We call a component T of a representative S of (5-17) the *main component* if it contains the interior marked point, and a *secondary component* if its output is attached to the main component. As a manifold with corners, (5-17) is equal to the compactification $\overline{\widecheck{\mathcal{R}}}{}^1_{d+1}$ except from the point of view of assigning Floer datum, as we will be forgetting the point z_f instead of fixing asymptotics for it. It is convenient therefore (for the purpose of indicating choices of Floer data made) to name components of strata containing z_f differently. At any stratum:

- We treat the main component (containing z_{out} and k boundary marked points) as belonging to $\overline{\widehat{\mathcal{R}}}{}^1_{k-1}$ if it contains z_f and $\overline{\widecheck{\mathcal{R}}}{}^1_k$ otherwise.
- If the i^{th} boundary marked point of any nonmain component was z_f, we view it as an element of $\mathcal{R}^{k,f_i}$, the space of discs with one output and k input marked points removed from the boundary, with the i^{th} point marked as "forgotten," constructed in Section A.2.
- We treat any other nonmain component as belonging to $\mathcal{R}^k$ as usual.

Thus, the codimension-one boundary of the Deligne–Mumford compactification is covered by the natural inclusions of the strata

$$\overline{\mathcal{R}}^m \times_i \overline{\widehat{\mathcal{R}}}{}^1_{d-m+1}, \quad \text{with } 1 \le i < d-m+1, \tag{5-18}$$

$$\overline{\mathcal{R}}^{m,f_k} \times_{d-m+1} \overline{\widecheck{\mathcal{R}}}{}^1_{d-m+1}, \quad \text{with } 1 \le j \le m,\ 1 \le k \le m, \tag{5-19}$$

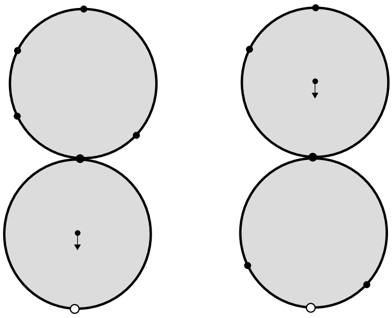


Figure 4: A schematic of the two distinct types of codimension-one boundary strata of (5-17) in codimension one. On the left side, corresponding to (5-18), a disc bubble forms involving any collection of boundary marked points not including z_f. On the right side, corresponding to (5-19), a disc bubble forms involving the point z_f.

where the notation $\times_j$ means that the output of the first component is identified with the j^{th} boundary input of the second. See Figure 4.

The forgetful map π_f extends to a map $\overline{\pi}_f$ from the compactification $\overline{\hat{\mathcal{R}}}{}^1_d$ (to the space of stable framed open–closed discs with d marked points) as follows: $\overline{\pi}_f$ puts the auxiliary point z_f back in, eliminates any component which is not main or secondary and which has only one nonauxiliary marked point p, and labels the positive marked point below this component by p. Given a representative S of $\overline{\hat{\mathcal{R}}}{}^1_d$, we call $\overline{\pi}_f(S)$ the *associated reduced surface*. We will study maps from the associated reduced surfaces $\overline{\pi}_f(S)$, parametrized by S. To this end, define a *Floer datum* on a stable disc S in $\overline{\hat{\mathcal{R}}}{}^1_d$ to consist of a Floer datum for the underlying reduced surface $\overline{\pi}_f(S)$ in the sense of Section 5.1.

First, in Section A.2 we describe an inductive construction of Floer data for (the underlying reduced surfaces of) the compactified moduli space of discs with a forgotten point $\overline{\mathcal{R}}^{d,f_i}$, for every d and i, with the following properties:

(5-20) $\begin{cases} \text{For } d > 2, \text{ the choice of Floer datum on } \mathcal{R}^{d,f_i} \text{ should be pulled back from the forgetful map } \mathcal{R}^{d,f_i} \to \mathcal{R}^{d-1}. \\ \text{For } d = 2, \text{ the Floer datum on the surface } S \text{ (with } z_i \text{ forgotten) should be translation invariant.} \end{cases}$

Next, we choose a *Floer datum for the nonunital open–closed map*, which is an inductive set of choices $(\boldsymbol{D}_{\check{\mathcal{OC}}}, \boldsymbol{D}_{\hat{\mathcal{OC}}})$, for each $d \geq 1$ and every representative $S \in \overline{\check{\mathcal{R}}}{}^1_d$, $T \in \overline{\hat{\mathcal{R}}}{}^1_d$, of a Floer datum for S and (the associated reduced surface of) T, respectively. As usual, these choices should be smoothly varying, and restrict smoothly to previously chosen Floer data on boundary strata. Note that for a given d the boundary strata have components that are either $\overline{\check{\mathcal{R}}}^1_{d'}$ or $\overline{\hat{\mathcal{R}}}^1_{d'}$ for $d' < d$, a stratum $\mathcal{R}^{d'}$ (over which we have chosen a Floer datum for the A_∞ structure), or a stratum $\mathcal{R}^{d,f_i}$ where we have chosen a Floer datum in Section A.2 as described above. (As usual for Liouville manifolds we use the notion of Floer data and consistency described in Remark 5.1 or Remark 5.2 in the wrapped case.) Contractibility of the space of choices at every stage (and consistency of the compatibility conditions imposed at corners) ensures as usual that a Floer datum for the nonunital open–closed map exists.

Fixing such a choice, we obtain, for any d–tuple of Lagrangians $L_0, \dots, L_{d-1}$, and asymptotic conditions $\vec{x} = (x_d, \dots, x_1)$ with $x_i \in \chi(L_{i-1}, L_{i \bmod d})$ and $y_{\mathrm{out}} \in \mathcal{O}$, a pair of moduli spaces

$$\check{\mathcal{R}}^1_d(y_{\mathrm{out}}; \vec{x}), \tag{5-21}$$

$$\hat{\mathcal{R}}^1_d(y_{\mathrm{out}}; \vec{x}), \tag{5-22}$$

of parametrized families of solutions to Floer's equation,

$$\bigl\{(S,u) \mid S \in \check{\mathcal{R}}^1_d, u\colon S \to M \text{ such that } (du - X \otimes \alpha)^{0,1} = 0 \\ \text{using the Floer datum given by } \boldsymbol{D}_{\check{\mathcal{OC}}}(S)\bigr\}, \tag{5-23}$$

$$\bigl\{(S,u) \mid S \in \hat{\mathcal{R}}^1_d, u\colon \pi_f(S) \to M \text{ such that } (du - X \otimes \alpha)^{0,1} = 0 \\ \text{using the Floer datum given by } \boldsymbol{D}_{\hat{\mathcal{OC}}}(S)\bigr\}, \tag{5-24}$$

satisfying asymptotic and boundary conditions (in either case) as in (5-5), with the modification for wrapped Fukaya categories involving Liouville rescalings described in Remark 5.2. The expected dimensions of every component of (5-21) and (5-22), respectively, in the $\mathbb{Z}$–graded case are

$$\deg(y_{\mathrm{out}}) - n + d - 1 - \sum_{k=1}^{d} \deg(x_k), \tag{5-25}$$

$$\deg(y_{\mathrm{out}}) - n + d - \sum_{k=1}^{d} \deg(x_k), \tag{5-26}$$

and mod 2 these in the $\mathbb{Z}/2$–graded case.

As usual there are Gromov-type bordifications

$$\overline{\check{\mathcal{R}}}^1_d(y_{\mathrm{out}}; \vec{x}), \tag{5-27}$$

$$\overline{\hat{\mathcal{R}}}^1_d(y_{\mathrm{out}}; \vec{x}), \tag{5-28}$$

which allow semistable breakings, as well as maps from strata corresponding to the boundary strata of $\overline{\check{\mathcal{R}}}{}^1_d$ and $\overline{\hat{\mathcal{R}}}{}^1_d$.

By Assumption 3.10, for generic choices of Floer datum for the nonunital open–closed map, the components of (5-27) and (5-28) of virtual dimension ≤ 1 are compact manifolds-with-boundary of dimension agreeing with virtual dimension. Fix such a Floer datum. At a rigid element u of each of the above moduli spaces, we obtain, using the fixed orientations of moduli spaces of domains (5-11)–(5-14) and [1, Lemma C.4], isomorphisms of orientation lines

$$(\check{\mathcal{R}}^1_d)_u \colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_{y_{\mathrm{out}}}, \tag{5-29}$$

$$(\hat{\mathcal{R}}^1_d)_u \colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_{y_{\mathrm{out}}}. \tag{5-30}$$

These isomorphisms in turn define the $|o_{y_{\mathrm{out}}}|_{\boldsymbol{k}}$ component of the check and hat components of the nonunital open–closed map with d inputs in the lines $|o_{x_d}|_{\boldsymbol{k}}, \dots, |o_{x_1}|_{\boldsymbol{k}}$, up to a sign twist:

$$\check{\mathcal{OC}}_d([x_d], \dots, [x_1]) := \sum_{u \in \overline{\check{\mathcal{R}}}{}^d_1(y; x_d, \dots, x_1) \text{ rigid}} (-1)^{\check{\star}_d} (\check{\mathcal{R}}^1_d)_u([x_d], \dots, [x_1]), \tag{5-31}$$

where $\check{\star}_d := \deg(x_d) + \sum_{k=1}^d k \deg(x_k)$, and

$$\hat{\mathcal{OC}}_d([x_d], \dots, [x_1]) := \sum_{u \in \overline{\hat{\mathcal{R}}}{}^1_d(y_{\mathrm{out}}; \vec{x}) \text{ rigid}} (-1)^{\hat{\star}_d} (\hat{\mathcal{R}}^1_d)_u([x_d], \dots, [x_1]), \tag{5-32}$$

where $\hat{\star}_d := \sum_{i=1}^d i \cdot \deg(x_i)$.

By analyzing the boundary of one-dimensional components of the moduli spaces $\overline{\mathcal{R}}{}^1_d(y_{\mathrm{out}}; \vec{x})$, the consistency condition imposed on Floer data, and a sign analysis, in [1] it was proved that:

Lemma 5.7 [1, Lemma 5.4] *The map $\mathcal{OC} := \check{\mathcal{OC}}$ is a chain map of degree n; that is, $(-1)^n d_{CF} \circ \check{\mathcal{OC}} = \check{\mathcal{OC}} \circ b$.* □

Similarly, we prove the following, completing the proof of Lemma 5.3:

Lemma 5.8 *The following equation holds:*

$$(-1)^n d_{CF} \circ \hat{\mathcal{OC}} = \check{\mathcal{OC}} \circ d_{\wedge\vee} + \hat{\mathcal{OC}} \circ b'. \tag{5-33}$$

Proof The consistency condition imposed on Floer data implies that the boundary of the one-dimensional components of $\overline{\check{\mathcal{R}}}{}^1_d(y;\vec{x})$ are covered by the images of the natural inclusions of the rigid (zero-dimensional) components of the moduli spaces of maps coming from the boundary strata (5-18) and (5-19) along with (the rigid components of) semistable breakings,

$$\overline{\hat{\mathcal{R}}}{}^1_d(y_1;\vec{x})\times\overline{\mathcal{M}}(y_{\mathrm{out}};y_1)\to\partial\overline{\check{\mathcal{R}}}{}^1_d(y_{\mathrm{out}};\vec{x}), \tag{5-34}$$

$$\overline{\mathcal{R}}^1(x;x_i)\times\overline{\hat{\mathcal{R}}}{}^1_d(y_{\mathrm{out}};\tilde{\vec{x}})\to\partial\overline{\check{\mathcal{R}}}{}^1_d(y_{\mathrm{out}};\vec{x}), \tag{5-35}$$

where $\tilde{\vec{x}}$ denotes the collection of inputs $\vec{x}$ with x_i replaced with x. Let $\mu^{d,i}$ be the operation associated to the space of discs with i^{th} point marked as forgotten $\mathcal{R}^{d,f_i}$, which is described in detail in Section A.2. The operation $\mu^{d,i}$ takes a composable sequence of $d-1$ inputs, separated into an $i-1$ tuple and a $d-i$ tuple; in line with Remark 3.3 we will use the suggestive notation[15]

$$\mu^d(x_d,\dots,x_{i+1},e^+,x_{i-1},\dots,x_1):=\mu^{d,i}(x_d,\dots,x_{i+1};x_{i-1},\dots,x_1). \tag{5-36}$$

(Recall the abuse of notation $x_i:=[x_i]$.) Then, up to sign, by the standard codimension-one boundary principle for Floer-theoretic operations, we have shown that

$$\begin{aligned}0={}&d_{CF}\hat{\mathcal{OC}}(x_d,\dots,x_1)\\&-\sum_{i,j}(-1)^{\maltese_1^i}\hat{\mathcal{OC}}(x_d,\dots,x_{i+j+1},\mu^j(x_{i+j},\dots,x_{i+1}),x_i,\dots,x_1)\\&\qquad-\sum_{i,j,k}(-1)^{\sharp_j^k}\check{\mathcal{OC}}\big(\mu^{j+k+1}(x_j,\dots,x_1,e^+,x_d,\dots,x_{d-k+1}),\\&\hspace{20em}x_{d-k},\dots,x_{j+1}\big),\end{aligned} \tag{5-37}$$

with desired signs

$$\maltese_m^n=\sum_{j=m}^{n}\|x_i\|, \tag{5-38}$$

$$\sharp_j^k=\maltese_1^j\maltese_{j+1}^d+\maltese_{j+1}^d+1. \tag{5-39}$$

However, as shown in Section A.2,

$$\mu^{j+k+1}(x_j,\dots,x_1,e^+,x_d,\dots,x_{d-k+1})=\begin{cases}x_1&\text{if } j=1,\ k=0,\\(-1)^{|x_d|}x_d&\text{if } j=0,\ k=1,\\0&\text{otherwise.}\end{cases} \tag{5-40}$$

[15] In fact, when the Fukaya category is equipped with homotopy units, one can ensure that there is a strict unit element e^+ in each self-hom space for which μ^k with an e^+ element admits a geometric description as above. See eg [22] or [24].

(In this manner, e^+, though a formal element, behaves as a strict unit.) So if equation (5-37) held, it would follow that

$$
\begin{aligned}
\text{(5-41)}\quad & d_{CF} \circ \hat{\mathcal{OC}}(x_d \otimes \cdots \otimes x_1) \\
&= (-1)^{\|x_1\|\maltese_2^d + \maltese_2^d + 1} \check{\mathcal{OC}}(x_1 \otimes x_d \otimes \cdots \otimes x_2) \\
&\qquad + (-1)^{|x_d| + \maltese_1^d + 1} \check{\mathcal{OC}}(x_d \otimes \cdots \otimes x_1) + \hat{\mathcal{OC}} \circ b'(x_d \otimes \cdots \otimes x_1) \\
&= \check{\mathcal{OC}}((-1)^{\maltese_1^d + \|x_d\|}(1-t)(x_d \otimes \cdots \otimes x_1)) + \hat{\mathcal{OC}} \circ b'(x_d \otimes \cdots \otimes x_1) \\
&= (\check{\mathcal{OC}} \circ d_{\wedge\vee} + \hat{\mathcal{OC}} \circ b')(x_d \otimes \cdots \otimes x_1).
\end{aligned}
$$

So we are done if we establish that the signs are exactly (5-38)–(5-39).

Using the notation

$$\text{(5-42)}\qquad \mathcal{OC}(e^+ \otimes x_d \otimes \cdots \otimes x_1) := \hat{\mathcal{OC}}(x_d \otimes \cdots \otimes x_1),$$

where again e^+ is simply a formal symbol referring to the position of the auxiliary (forgotten) input point, we observe that the equation (5-37) is exactly the equation for $\mathcal{OC}$ being a chain map on inputs of the form $(e^+ \otimes x_d \otimes \cdots \otimes x_1)$, where we treat an "$e^+$" input as an auxiliary unconstrained point on our domain. The sign verification therefore follows from that of $\check{\mathcal{OC}}$ being a chain map (in [1, Lemma 5.4]), for we have used identical orientations on the abstract moduli space $\hat{\mathcal{R}}_d^1$ as on $\check{\mathcal{R}}_{d+1}^1$ (identifying z_f with z_{d+1}), and on $\mathcal{R}^{d,f_i}$ as on $\mathcal{R}^d$, and we can even insert a formal degree zero orientation line o_{e^+} into the procedure for orienting moduli spaces of open–closed maps (see [1, Section C.6]), corresponding to the marked point (obtained by filling in) z_f. Note that o_{e^+}, being of degree zero, commutes with everything, and is just used as a placeholder as if we had an asymptotic condition at z_f. □

Proof of Lemma 5.3 As $\check{\mathcal{OC}}$ is already known to be a chain map by [1, Lemma 5.4], repeated as Lemma 5.7 above, the new part to check is that

$$(-1)^n d_{CF} \circ \hat{\mathcal{OC}} = \check{\mathcal{OC}} d_{\wedge\vee} + \hat{\mathcal{OC}} \circ b'.$$

This is the content of Lemma 5.8 above. □

5.3 An auxiliary operation

It will be technically convenient to define an auxiliary operation

$$\text{(5-43)}\qquad \mathcal{OC}^{S^1} : \mathrm{CH}_{*-n}(\mathcal{F}) \to CH^{*+1}(M)$$

from the left factor of the nonunital Hochschild complex to Floer cochains, in which the asymptotic marker τ_{out} varies freely around the circle. This operation is more easily comparable to the BV operator on Floer cohomology, and moreover, we will show that $\mathcal{OC}^{S^1}$ (and $\hat{\mathcal{OC}}$) can be chosen to satisfy the following crucial identity:

Proposition 5.9 *There is an equality of chain-level operations,*

$$\mathcal{OC}^{S^1} = \hat{\mathcal{OC}} \circ B^{\mathrm{nu}}. \tag{5-44}$$

To define (5-43), let

$$\mathcal{R}_d^{S^1} \tag{5-45}$$

be the abstract moduli space of discs with d boundary positive punctures $z_1, \dots, z_d$ labeled in counterclockwise order and one interior negative puncture z_{out}, with an asymptotic marker τ_{out} at z_{out} (or choice of real half-line in $T_{z_{\mathrm{out}}}D$) which is free to vary. Equivalently,

(5-46) $\mathcal{R}_d^{S^1}$ is the space of discs with $z_1, \dots, z_d$ and z_{out} as before, and an extra auxiliary interior marked point p_1 such that, for a representative with (z_{out}, z_1) fixed at $(0, -i)$, $|p_1| = \frac{1}{2}$, and the asymptotic marker τ_{out} points towards p_1.

By using a representative with fixed (z_{out}, z_1) as above, the argument of p_1 produces an abstract identification

$$\mathcal{R}_d^{S^1} = \check{\mathcal{R}}_d^1 \times S^1. \tag{5-47}$$

Using this identification, fix an orientation of (5-47) given by negative the product orientation of (5-11) with the standard counterclockwise orientation on S^1. The Deligne–Mumford-type compactification can thus be thought of as

$$\overline{\mathcal{R}}_d^{S^1} = \overline{\check{\mathcal{R}}}_d^1 \times S^1. \tag{5-48}$$

Given an element S of $\mathcal{R}_d^{S^1}$ and a choice of marked point z_i on the boundary of S, we say that τ_{out} *points at* z_i, if, when S is reparametrized so that z_1 fixed at $-i$ and z_{out} fixed at 0, the vector τ_{out} is tangent to the straight line from z_{out} to z_i. Equivalently, for this representative, z_{out}, p_1 and z_i are colinear. For each i, the locus where τ_{out} points at z_i forms a codimension-one submanifold, denoted by

$$\mathcal{R}_d^{S_i^1}. \tag{5-49}$$

The notion compactifies well; if z_i is not on the main component of (5-48), we say that τ_{out} *points at* z_i if it points at the root of the bubble tree z_i is on. This compactified locus $\overline{\mathcal{R}}_d^{S_i^1}$ can be identified with $\overline{\mathcal{R}}_d^1$ via the map

$$\tau_i \colon \overline{\mathcal{R}}_d^{S_i^1} \to \overline{\mathcal{R}}_d^1 \tag{5-50}$$

which cyclically permutes the labels of the boundary marked points so that z_i is now labeled z_d.

In a similar fashion, we have an invariant notion of what it means for τ_{out} to point *between z_i and z_{i+1}*; this is a codimension-zero submanifold with corners of (5-47), denoted by

$$\mathcal{R}_d^{S^1_{i,i+1}}. \tag{5-51}$$

The compactification has some components that are codimension-one submanifolds with corners of (5-48), when z_i and z_{i+1} both lie on a bubble tree.

Finally, there is a *free* $\mathbb{Z}_d$–action generated by the map

$$\kappa\colon \overline{\mathcal{R}}_d^{S^1} \to \overline{\mathcal{R}}_d^{S^1} \tag{5-52}$$

which cyclically permutes the labels of the boundary marked points; for concreteness, κ changes the label z_i to z_{i+1} for $i < d$, and z_d to z_1. Note that if, on a given S, τ_{out} points between z_i and z_{i+1}, then on $\kappa(S)$, τ_{out} points between $z_{i+1 \bmod d}$ and $z_{i+2 \bmod d}$.

Lemma 5.10 *The action generated by* (5-52) *is free and properly discontinuous.*

Sketch The basic observation arises on the level of uncompactified moduli spaces: since any element of $\mathcal{R}_d^{S^1}$ has a unit disk representative with (z_{out}, p_1) fixed at $\bigl(0, \frac{1}{2}\bigr)$, the positions of the remaining points identify $\mathcal{R}_d^{S^1}$ with the space of tuples $(z_1, \dots, z_d)$ of disjoint (cyclically ordered) points on S^1 (without any further quotienting by automorphism). The action of κ, which cyclically permutes the labels $z_1, \dots, z_d$ in this identification, evidently acts freely and properly discontinuously on this locus. Similarly, an element of a boundary stratum consists of an element of $\mathcal{R}_k^{S^1}$ for some $k \le d$ with some collection of stable disc bubble trees attached to some or all of the marked points of $\mathcal{R}_k^{S^1}$, so that there are d leaf (nonnodal) boundary marked points, along with a counterclockwise ordered labeling of these marked points by $z_1, \dots, z_d$ (note that there is a well-defined cyclic counterclockwise ordering of boundary marked points on any such stable configuration). By using a representative of the main component $\mathcal{R}_k^{S^1}$ with (z_{out}, p_1) fixed at $\bigl(0, \frac{1}{2}\bigr)$, an explicit analysis shows that the action of (5-52) remains free and properly discontinuous — for instance, to see free, note that there is a well-defined "first boundary nonnodal marked point at or counterclockwise from the argument of p_1"; the action of (5-52) freely permutes the label of this first boundary marked point hence cannot have a fixed point. □

The quotient of the action of κ consists of the space of discs with z_{out} and p_1 as before,[16] equipped with d *cyclically unordered* or *unlabeled* boundary marked points. Note that on the open-locus $\mathring{\mathcal{R}}_d^{S^1}$, where τ_{out} does not point at a boundary marked point, one can choose a labeling by setting the boundary point immediately clockwise of where τ_{out} points to be z_d. This induces a diffeomorphism

(5-53) $$\mathring{\mathcal{R}}_d^{S^1}/\kappa \cong \mathcal{R}_d^{1,\mathrm{free}}.$$

Similarly, on the complementary locus where τ_{out} points at a boundary marked point, we can similarly choose a labeling by declaring this boundary marked point to be z_d, giving a diffeomorphism (of this locus) with $\check{\mathcal{R}}_d^1$.

We now choose Floer perturbation data for the family of moduli spaces $\mathcal{R}_d^{S^1}$; in fact, it will be helpful to rechoose Floer data for the moduli spaces appearing in the nonunital open–closed map to have extra compatibility. To that end, a *BV compatible Floer datum for the nonunital open–closed map* is an inductive choice $(\boldsymbol{D}_{\check{\mathcal{OC}}}, \boldsymbol{D}_{\hat{\mathcal{OC}}}, \boldsymbol{D}_{S^1})$ of Floer data where $\boldsymbol{D}_{\check{\mathcal{OC}}}$ and $\boldsymbol{D}_{\hat{\mathcal{OC}}}$ is a universal and consistent choice of Floer data for the nonunital open–closed map as before, and $\boldsymbol{D}_{S^1}$ consists of, for each $d \geq 1$ and every representative $S \in \overline{\mathcal{R}}_d^{S^1}$, a Floer datum for S varying smoothly over the moduli space. Again, these satisfy the usual consistency condition with respect to previously made choices along lower-dimensional strata. Moreover, there are two additional inductive constraints on the Floer data chosen:

(5-54) On the codimension-one loci $\overline{\mathcal{R}}_d^{S_i^1}$ where τ_{out} points at z_i, the Floer datum should agree with the pullback by τ_i of the existing Floer datum for the (check) open–closed map.

(5-55) The Floer datum should be κ–equivariant, where κ is the map (5-52).

Also, there is a final a posteriori constraint on the Floer data for the nonunital open–closed map $\boldsymbol{D}_{\hat{\mathcal{OC}}}$: for $S \in \overline{\hat{\mathcal{R}}}_d^1$:

(5-56) The Floer datum on the main component S_0 of $\bar{\pi}_f(S)$ should coincide with the existing datum chosen on $S_0 \in \mathcal{R}_d^{1,\mathrm{free}} \subset \mathcal{R}_d^{S^1}$.

By an inductive argument as before, a BV compatible Floer datum for the nonunital open–closed map exists.

To explain the way choices are made (which ensures both existence at every stage and that the requirements above are satisfied): we choose the data for $\mathcal{R}_d^{S^1}$ prior to

[16] Meaning z_{out} is a negative interior puncture, and p_1 is an auxiliary interior marked point such that for any representative with z_{out} fixed at 0, $|p_1| = \frac{1}{2}$.

choosing that of $\bar{\hat{\mathcal{R}}}^1_d$ and note that the condition (5-56) specifies the Floer datum on $\hat{\mathcal{R}}^1_d$ entirely. In particular, the conditions (5-20) required on the latter Floer datum are compatible with consistency and the condition (5-54). With regards to choosing the data for $\mathcal{R}^{S^1}_d$, the equivariance constraint (5-55), which is compatible with both (5-54) (a κ–equivariant condition) and with the consistency condition, is also unproblematic in light of Lemma 5.10: one can pull back a Floer datum from the quotient of $\overline{\mathcal{R}}^{S^1}_d$ by κ.

Fixing a BV compatible Floer datum for the nonunital open–closed map we obtain, for any d–tuple of Lagrangians $L_0, \dots, L_{d-1}$, and asymptotics $\vec{x} = (x_d, \dots, x_1)$ with $x_i \in \chi(L_{i-1}, L_{i \bmod d})$, and $y_{\mathrm{out}} \in \mathcal{O}$, a moduli space

$$\mathcal{R}^{S^1}_d(y_{\mathrm{out}}; \vec{x}) \tag{5-57}$$

of parametrized families of solutions to Floer's equation, with respect to the Floer data chosen,

$$\{(S, u) \mid S \in \mathcal{R}^{S^1}_d, u\colon \pi_f(S) \to M \text{ such that } (du - X \otimes \alpha)^{0,1} = 0\}, \tag{5-58}$$

satisfying asymptotic and boundary conditions as in (5-5) (again with the modifications of Remarks 5.1 or 5.2 for compact or wrapped Fukaya categories of Liouville manifolds). Generically the Gromov–Floer compactifications

$$\overline{\mathcal{R}}^{S^1}_d(y_{\mathrm{out}}; \vec{x}) \tag{5-59}$$

of the components of virtual dimension ≤ 1 are compact manifolds-with-boundary of the expected dimension; this dimension coincides (mod 2 or exactly depending on whether we are in a $\mathbb{Z}/2$– or $\mathbb{Z}$–graded setting) with

$$\deg(y_{\mathrm{out}}) - n + d - \sum_{k=1}^{d} \deg(x_k). \tag{5-60}$$

Each rigid $u \in \overline{\mathcal{R}}^{S^1}_d(y_{\mathrm{out}}; \vec{x})$ gives by the orientation from (5-48) and [1, Lemma C.4] an isomorphism of orientation lines

$$(\mathcal{R}^{S^1}_d)_u\colon o_{x_d} \otimes \dots \otimes o_{x_1} \to o_{y_{\mathrm{out}}}, \tag{5-61}$$

which gives the $|o_{y_{\mathrm{out}}}|_{\boldsymbol{k}}$ component of the S^1 open–closed map with d inputs in the lines $|o_{x_d}|_{\boldsymbol{k}}, \dots, |o_{x_1}|_{\boldsymbol{k}}$, up to a sign twist given below: define

$$\mathcal{OC}^{S^1}([x_d], \dots, [x_1]) := \sum_{u \in \overline{\mathcal{R}}^{S^1}_d(y_{\mathrm{out}}; \vec{x}) \text{ rigid}} (-1)^{\clubsuit_d} (\mathcal{R}^{S^1}_d)_u([x_d], \dots, [x_1]), \tag{5-62}$$

where $\clubsuit_d = \sum_{i=1}^d (i+1) \cdot \deg(x_i) + \deg(x_d) + d - 1$.

The proof of Proposition 5.9, which equates $\mathcal{OC}^{S^1}$ with $\hat{\mathcal{OC}} \circ B^{\mathrm{nu}}$, appears below and is composed of two steps. First, we decompose the moduli space $\mathcal{R}_d^{S^1}$ into sectors in which τ_{out} points between a pair of adjacent boundary marked points. It will follow that the sum of the corresponding "sector operations" is exactly $\mathcal{OC}^{S^1}$. The sector operations in turn can be compared to $\hat{\mathcal{OC}}$ via cyclically permuting inputs and an orientation analysis.

We begin by defining the relevant sector operations: For $i \in \mathbb{Z}/(d+1)\mathbb{Z}$, define

$$\hat{\mathcal{R}}^1_{d,\tau_i} \tag{5-63}$$

to be the abstract moduli space of discs with $d+1$ boundary punctures $z_1, \dots, z_i$, z_f, $z_{i+1}, \dots, z_d$ arranged in counterclockwise order and interior puncture z_{out} with asymptotic marker pointing towards the boundary point z_f, which is also marked as "auxiliary". There is a bijection

$$\tau_i : \hat{\mathcal{R}}^1_{d,\tau_i} \simeq \hat{\mathcal{R}}^1_d \tag{5-64}$$

given by cyclically permuting labels, inducing a model for the compactification $\overline{\hat{\mathcal{R}}}^1_{d,\tau_i}$. However, we will use a different orientation than the one induced by pullback: on a slice with fixed position of z_d and z_{out}, we take the volume form

$$dz_1 \wedge \cdots \wedge dz_{d-1} \wedge dz_f. \tag{5-65}$$

By construction, the induced "forgetful map"

$$\pi_f^i : \hat{\mathcal{R}}^1_{d,\tau_i} \to \mathcal{R}^{S^1_{i,i+1}} \tag{5-66}$$

is an oriented diffeomorphism that extends to a map between compactifications. Note as before that strictly speaking this map does not forget any information, at least on the open locus.

Remark 5.11 In the case $i = 0$, this orientation agrees with the previously chosen orientation (5-14) on $\hat{\mathcal{R}}^1_d$. We previously defined the orientation on $\hat{\mathcal{R}}^1_d$ in terms of a different slice of the group action. To compare the forms $dz_1 \wedge \cdots \wedge dz_{d-1} \wedge dz_f$ (coming from the slice with fixed z_d and z_{out}) and $-dz_1 \wedge \cdots \wedge dz_d$ (coming from the slice with fixed z_f and z_{out}), note that either orientation is induced by the following procedure:

- Fixing an orientation on the space of discs as above with fixed position of z_{out} (but not z_f or z_d): we shall fix the canonical orientation $dz_1 \wedge \cdots \wedge dz_d \wedge dz_f$.

- Fixing a choice of trivializing vector field for the remaining S^1–action on this space of discs with fixed z_{out}: we shall fix $S = (-\partial_{z_f} - \partial_{z_1} - \cdots - \partial_{z_d})$.
- Fixing a convention for contracting orientation forms along slices of the action: to determine the orientation on a slice of an S^1–action, we will contract the orientation on the original space on the right by the trivializing vector field.

Moreover, this data induces an orientation on the quotient by the S^1–action, and also an oriented isomorphism between the induced orientation on any slice and that of the quotient. It follows that on the quotient, the orientation $-dz_1 \wedge \cdots \wedge dz_d$ (from the slice where z_f is fixed) and the orientation $dz_1 \wedge \cdots \wedge dz_{d-1} \wedge dz_f$ (from the slice where z_d is fixed) agree. We conclude that these two orientations agree. The author thanks Nick Sheridan for relevant discussions about orientations of moduli spaces.

Choose as a Floer datum for each $\overline{\mathcal{R}}^1_{d,\tau_i}$ the Floer datum pulled back from $\overline{\hat{\mathcal{R}}}{}^1_d$ via (5-64); this system of choices is automatically inductively consistent with choices made on lower strata, inheriting this property from the Floer data on the collection of $\overline{\hat{\mathcal{R}}}{}^1_d$. Using this choice, for any d–tuple of Lagrangians $L_0, \dots, L_{d-1}$, and asymptotic conditions $\vec{x} = (x_d, \dots, x_1)$, with $x_i \in \chi(L_{i-1}, L_{i \bmod d})$, and $y_{\mathrm{out}} \in \mathcal{O}$, we obtain a moduli space

$$\mathcal{R}^1_{d,\tau_i}(y_{\mathrm{out}}; \vec{x}) = \hat{\mathcal{R}}^1_d(y_{\mathrm{out}}; (x_{i-1}, \dots, x_1, x_d, \dots, x_i)) \tag{5-67}$$

of parametrized families of solutions to Floer's equation,

$$\bigl\{(S,u) \mid S \in \hat{\mathcal{R}}^1_d,\, u\colon \pi_f(S) \to M \text{ is such that } (du - X \otimes \alpha)^{0,1} = 0 \\ \text{using the Floer datum for } \pi_f(S)\bigr\}, \tag{5-68}$$

satisfying asymptotic and boundary conditions as in (5-5) (with the modifications as in Remarks 5.1 or 5.2 in the Liouville case), as well as its Gromov–Floer compactification

$$\overline{\mathcal{R}}^1_{d,\tau_i}(y_{\mathrm{out}}; \vec{x}) := \overline{\hat{\mathcal{R}}}{}^1_d(y_{\mathrm{out}}; (x_i, \dots, x_1, x_d, \dots, x_{i+1})), \tag{5-69}$$

whose components of virtual dimension ≤ 1 (at least) are compact manifolds-with-boundary of the correct dimension, coinciding (exactly in the graded case and mod 2 in the $\mathbb{Z}/2$–graded case) with $\deg(y_{\mathrm{out}}) - n + d - \sum_{j=0}^{d} \deg(x_j)$.

Each rigid element $u \in \overline{\mathcal{R}}^1_{d,\tau_i}(y_{\mathrm{out}}; \vec{x})$ gives, by (5-65) and [1, Lemma C.4], an isomorphism of orientation lines

$$(\mathcal{R}^1_{d,\tau_i})_u\colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_{y_{\mathrm{out}}}, \tag{5-70}$$

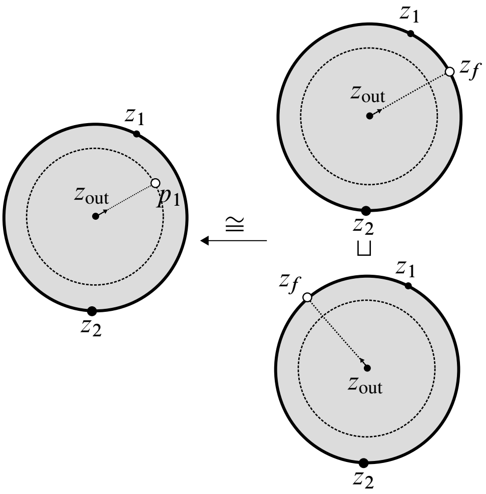


Figure 5: The diffeomorphism between $\widehat{\mathcal{R}}^1_{2,\tau_0} \cup \widehat{\mathcal{R}}^1_{2,\tau_1}$ and the open dense part of $\mathcal{R}_2^{S^1}$ given by $\mathcal{R}_2^{S^1_{0,1}} \cup \mathcal{R}_2^{S^1_{1,2}}$. The former spaces can in turn be compared to $\widehat{\mathcal{R}}^1_2$ via cyclic permutation of labels.

which defines the $|o_{y_{\mathrm{out}}}|_{\boldsymbol{k}}$ component of an operation $\widehat{\mathcal{OC}}_{d,\tau_i}$, with d inputs in the lines $|o_{x_d}|_{\boldsymbol{k}}, \dots, |o_{x_1}|_{\boldsymbol{k}}$, up to the following sign twist:

$$\widehat{\mathcal{OC}}_{d,\tau_i}([x_d], \dots, [x_1]) := \sum_{u \in \overline{\widehat{\mathcal{R}}}^1_{d,\tau_i}(y_{\mathrm{out}}; \vec{x}) \text{ rigid}} (-1)^{\clubsuit_d} (\widehat{\mathcal{R}}^1_{d,\tau_i})_u([x_d], \dots, [x_1]), \tag{5-71}$$

where $\clubsuit_d = \sum_{i=1}^d (i+1) \cdot \deg(x_i) + \deg(x_d) + d - 1$.

Lemma 5.12 *As chain-level operations,*

$$\mathcal{OC}^{S^1} = \sum_i \widehat{\mathcal{OC}}_{d,\tau_i}. \tag{5-72}$$

Proof For each d, there is an embedding of abstract moduli spaces

$$\coprod_i \widehat{\mathcal{R}}^1_{d,\tau_i} \xrightarrow{\coprod_i \pi_f^i} \coprod_i \mathcal{R}_d^{S^1_{i,i+1}} \hookrightarrow \mathcal{R}_d^{S^1}. \tag{5-73}$$

See Figure 5.

By construction, this map is compatible with Floer data (this uses the fact that the Floer data on $\mathcal{R}^{S^1_{i,i+1}}$ agrees with the data on $\widehat{\mathcal{R}}^1_d$ via the reshuffling map κ^{-i} by (5-55)), and covers all but a codimension-one locus in the target. Since, after perturbation, zero-dimensional solutions to Floer's equation can be chosen to come from the complement of any codimension-one locus in the source abstract moduli space, we conclude that the two operations in the lemma, which arise from either side of (5-73), are identical up to sign. To fix the signs, note that (5-73) is in fact an oriented embedding, and all the sign twists defining the operations $\widehat{\mathcal{OC}}_{d,\tau_i}$ are chosen to be compatible with the sign twist in the operation $\mathcal{OC}^{S^1}$. □

Next, because the Floer data used in the constructions are identical, we have that $\widehat{\mathcal{OC}}_{d,\tau_i}(x_d \otimes \cdots \otimes x_1) := \widehat{\mathcal{OC}}_{d,\tau_i}(x_d, \dots, x_1)$ (recall the abuse of notation $x_i := [x_i]$) agrees with $\widehat{\mathcal{OC}}(x_i \otimes \cdots \otimes x_1 \otimes x_d \otimes \cdots \otimes x_{i+1}) := \widehat{\mathcal{OC}}(x_i, \dots, x_1, x_d, \dots, x_{i+1})$ up to a sign difference coming from orientations of abstract moduli spaces, cyclically reordering inputs, and sign twists. The following proposition computes the sign difference, and hence completes the proof of Proposition 5.9:

Lemma 5.13 *There is an equality*

$$\widehat{\mathcal{OC}}_{d,\tau_i}(x_d \otimes \cdots \otimes x_1) = \widehat{\mathcal{OC}}^d(s^{\mathrm{nu}}(t^i(x_d \otimes \cdots \otimes x_1))), \tag{5-74}$$

where s^{nu} is the operation (3-20) *arising from changing a check term to a hat term with a sign twist.*

Proof It is evident that $\widehat{\mathcal{OC}}_{d,\tau_i}$ agrees with $\widehat{\mathcal{OC}}_d \circ s^{\mathrm{nu}} \circ t^i$ up to sign, as the Floer data used in the two constructions are identical. By an inductive argument it suffices to verify the equalities of signed operations

$$\widehat{\mathcal{OC}}_{d,\tau_0} = \widehat{\mathcal{OC}}_d \circ s^{\mathrm{nu}}, \tag{5-75}$$

$$\widehat{\mathcal{OC}}_{d,\tau_1} = \widehat{\mathcal{OC}}_{d,\tau_0} \circ t, \tag{5-76}$$

the remaining sign changes being entirely incremental. For the equality (5-75), we simply note that the signs appearing in the operations $\widehat{\mathcal{OC}}_{d,\tau_0}([x_d], \dots, [x_1])$ and $\widehat{\mathcal{OC}}_d([x_d], \dots, [x_1])$ differ in the following fashions:

- The abstract orientations on the moduli space of domains agree, as in Remark 5.11.
- The difference in sign twists is given by
$$\clubsuit_d - \hat{\star}_d = \sum_{i=1}^{d} |x_i| + |x_d| + d - 1 = \left(\sum_{i=1}^{d} \|x_i\|\right) + 1 + |x_d| = \maltese_1^d + \|x_d\|.$$

All together, the parity of difference in signs is $\maltese_1^d + \|x_d\|$, which accounts for the sign in the algebraic operation s^{nu} (see (3-20)); this verifies (5-75).

Next, the sign difference between the two operations in the equality (5-76) is a sum of three contributions:

- The two orientations of abstract moduli spaces[17] from $-dz_1 \wedge \cdots \wedge dz_d$ to $dz_2 \wedge \cdots \wedge dz_d \wedge dz_1$ differ by a sign change of parity
$$d-1.$$
- For a given collection of inputs, the change in *sign twisting data* from $\clubsuit_d = \sum_{i=1}^{d}(i+1)\cdot|x_i| + |x_d| + d-1$ to $\sum_{i=1}^{d-1}(i+1)|x_{i+1}| + (d+1)|x_1| + |x_1| + d-1 = \sum_{i=2}^{d} i|x_i| + d|x_1| + d-1$ ($\clubsuit_d$ for the sequence $(x_2,\dots,x_d,x_1)$) induces a sign change of parity
$$\begin{aligned}\sum_{i=2}^{d}|x_i| + |x_d| + d|x_1| &= \sum_{i=1}^{d}|x_i| + |x_d| + (d-1)|x_1| \\ &= \sum_{i=1}^{d}\|x_i\| + (d-1)\|x_1\| + \|x_d\| \\ &= \maltese_1^d + (d-1)\|x_1\| + \|x_d\|.\end{aligned}$$
- Finally, the reordering of determinant lines of the inputs induces a sign change of parity
$$\begin{aligned}|x_1|\cdot\Bigl(\sum_{i=2}^{d}|x_i|\Bigr) &= \|x_1\|\cdot\Bigl(\sum_{i=2}^{d}\|x_i\|\Bigr) + \sum_{i=2}^{d}\|x_i\| + (d-1)\|x_1\| + (d-1) \\ &= \|x_1\|\maltese_2^d + \maltese_1^d + d\|x_1\| + (d-1).\end{aligned}$$

The cumulative sign parity is congruent mod 2 to
$$\|x_1\|\maltese_2^d + \|x_1\| + \|x_d\|,$$
which is precisely the sign appearing in t (see (3-18)). This verifies (5-76). □

Proof of Proposition 5.9 Combine Lemmas 5.12 and 5.13; note the definition of B^{nu} given in (3-21). □

5.4 Compatibility of homology-level BV operators

Before diving into the statement of chain-level equivariance, we prove a homology-level statement. The theorem below is insufficient for studying, say, equivariant homology groups, but may be of independent interest.

[17] On the slice where z_f and z_{out} are fixed; see Remark 5.11.

Theorem 5.14 *The homology-level open–closed map* $[\mathcal{OC}]$ *intertwines the Hochchild and symplectic cohomology* BV *operators*; *that is*,

$$[\mathcal{OC}] \circ [B^{\mathrm{nu}}] = [\delta_1] \circ [\mathcal{OC}]. \tag{5-77}$$

Theorem 5.14 is an immediate consequence of the following chain-level statement:

Proposition 5.15 *The following diagram homotopy commutes*:

$$\begin{tikzcd} \mathrm{CH}_{*-n}(\mathcal{F},\mathcal{F}) \arrow[r, hook, "\iota", "\sim"'] \arrow[d, "\check{\mathcal{OC}}"'] & \mathrm{CH}^{\mathrm{nu}}_{*-n}(\mathcal{F},\mathcal{F}) \arrow[r, "B^{\mathrm{nu}}"] & \mathrm{CH}^{\mathrm{nu}}_{*-n-1}(\mathcal{F},\mathcal{F}) \arrow[d, "\mathcal{OC}"] \\ CF^*(M) \arrow[rr, "\delta_1"] & & CF^{*-1}(M) \end{tikzcd} \tag{5-78}$$

where ι *is the inclusion onto the left factor*, *which is a quasi-isomorphism by Lemma 3.2. More precisely*, *there exists an operation* $\check{\mathcal{OC}}^1 : \mathrm{CH}_{*-n}(\mathcal{F},\mathcal{F}) \to CF^{*-2}(M)$ *satisfying*

$$(-1)^{n+1} d\check{\mathcal{OC}}^1 + \check{\mathcal{OC}}^1 b = \hat{\mathcal{OC}} B^{\mathrm{nu}} \iota - (-1)^n \delta_1 \check{\mathcal{OC}}. \tag{5-79}$$

Proof of Theorem 5.14 Proposition 5.15 implies that $[\delta_1] \circ [\check{\mathcal{OC}}] = [\mathcal{OC}] \circ [B^{\mathrm{nu}}] \circ [\iota]$, where $\iota : \mathrm{CH}_{*-n}(\mathcal{F},\mathcal{F}) \to \mathrm{CH}^{\mathrm{nu}}_{*-n}(\mathcal{F},\mathcal{F})$ is the inclusion of chain complexes. But by Lemma 3.2, $[\iota]$ is an isomorphism and by Corollary 5.4, $[\check{\mathcal{OC}}] = [\mathcal{OC}]$. □

To define $\check{\mathcal{OC}}^1$, consider

$${}_1\check{\mathcal{R}}^1_d, \tag{5-80}$$

the moduli space of discs with d positive boundary marked points $z_1, \dots, z_d$ labeled in counterclockwise order, one interior negative puncture z_{out} equipped with an asymptotic marker, and one additional interior marked point p_1 (without an asymptotic marker), marked as *auxiliary*. Also, with respect to the unit disc representative of any element of this moduli space fixing z_d at 1 and z_{out} at 0 on the unit disc, p_1 should lie *inside the circle of radius* $\frac{1}{2}$, so

$$0 < |p_1| < \tfrac{1}{2}. \tag{5-81}$$

Using the above representative, one can talk about the *angle*, or *argument* of p_1

$$\theta_1 := \arg(p_1). \tag{5-82}$$

We require that with respect to the above representative:

(5-83) The asymptotic marker on z_{out} points in the direction θ_1.

For every representative $S \in {}_1\check{\mathcal{R}}^1_d$:

(5-84) Fix a negative cylindrical end around z_{out} not containing p_1, compatible with the direction of the asymptotic marker, or *equivalently compatible with the angle* θ_1.

We orient (5-80) as follows: pick, on a slice of the automorphism action which fixes the position of z_d at 1 and z_{out} at 0, the volume form

$$-r_1\, dz_1 \wedge dz_2 \wedge \cdots \wedge dz_{d-1} \wedge dr_1 \wedge d\theta_1. \tag{5-85}$$

The compactification of (5-80) is a real blow-up of the ordinary Deligne–Mumford compactification, in the sense of [34] (see [58] for a first discussion in the context of Floer theory), reviewed in Section A.1; this is the case $k = 1$ of the more general description therein. The result of this discussion is that the codimension-one boundary of the compactified check moduli space ${}_1\overline{\check{\mathcal{R}}}^1_d$ is covered by the images of the natural inclusions of the following strata:

$$\overline{\mathcal{R}}^s \times {}_1\overline{\check{\mathcal{R}}}^1_{d-s+1}, \tag{5-86}$$

$$\overline{\check{\mathcal{R}}}^1_d \times \overline{\mathcal{M}}_1, \tag{5-87}$$

$$\overline{\check{\mathcal{R}}}^{S^1}_d. \tag{5-88}$$

The stratum (5-88) describes the locus which $|p_1| = \frac{1}{2}$, which is exactly the locus we defined to be the auxiliary moduli space $\mathcal{R}^{S^1}_d$ inducing the operation $\mathcal{OC}^{S^1}$. The strata (5-86)–(5-87) have manifold-with-corners structure given by standard local gluing maps using fixed choices of strip-like ends near the boundary. For (5-86) this is standard, and for (5-87), the local gluing map uses the cylindrical ends (5-84) and (4-33) — in other words, one rotates the 1–pointed angle cylinder by an amount commensurate to the angle of the marked point z_d on the disk before gluing; see Section A.1, particularly (A-12). See Figure 6 for a schematic of (5-80) and two out of the three types of strata (5-87)–(5-88).

We will as usual fix a *Floer datum for the BV homotopy*, meaning an inductive choice, for every $d \geq 1$, of Floer data for every representative $S \in {}_1\overline{\check{\mathcal{R}}}^1_d$ varying smoothly in S, which on boundary strata is smoothly equivalent to a product of Floer data inductively chosen on lower-dimensional moduli spaces. Such a system of choices exist again by a contractibility argument, and for any such choice, one obtains, for any d–tuple of Lagrangians $L_0, \ldots, L_{d-1}$ and asymptotic conditions

$$\vec{x} = (x_d, \ldots, x_1) \text{ with } x_i \in \chi(L_{i-1}, L_{i \bmod d}) \quad \text{and} \quad y_{\text{out}} \in \mathcal{O}, \tag{5-89}$$

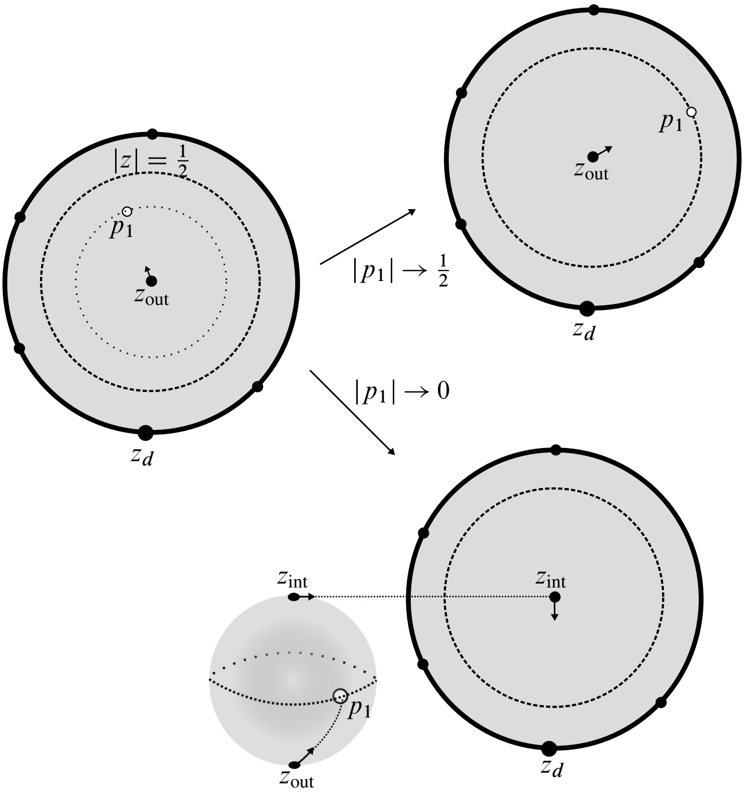


Figure 6: A schematic of an element of (5-80) on the left and a schematic of two of its three types of degenerations on the right, (5-88) (above) and (5-87) (below). The remaining type of degeneration (5-86), omitted from the figure, occurs when some boundary marked points coalesce into a disc bubble.

a compactified moduli space

$$ {}_1\overline{\check{\mathcal{R}}}{}^1_d(y_{\text{out}},\vec{x}) \tag{5-90}$$

of maps into M with source an arbitrary element S of the moduli space (5-80), satisfying Floer's equation using the Floer datum chosen for the given S as in (5-4) with asymptotics and boundary conditions as in (5-5), with the usual modifications in the Liouville case detailed in Remarks 5.1 and 5.2. The virtual dimension of every component of ${}_1\overline{\check{\mathcal{R}}}{}^1_d(y_{\text{out}},\vec{x})$ coincides (mod 2 or exactly depending on whether we are in a $\mathbb{Z}/2$– or $\mathbb{Z}$–graded setting) with

$$\deg(y_{\text{out}})-n+d+1-\sum_{i=1}^{d}\deg(x_i). \tag{5-91}$$

By Assumption 3.10, for generic choices of Floer data, the Gromov–Floer compactification of the components of virtual dimension ≤ 1 of (5-90) are compact manifolds-with-boundary of expected dimension. For rigid elements u of the moduli spaces (5-90), the orientations (5-85) and [1, Lemma C.4] induce isomorphisms of orientation lines

$$({}_1\check{\mathcal{R}}^1_d)_u \colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_y. \tag{5-92}$$

As usual "counting rigid elements u", ie summing application of these isomorphisms over all u, defines the $|o_{y_{\text{out}}}|_{\boldsymbol{k}}$ component of an operation $\check{\mathcal{OC}}^1$, up to a sign twist which we specify:

$$\check{\mathcal{OC}}^1([x_d],\ldots,[x_1]) := \sum_{u \in {}_k\check{\overline{\mathcal{R}}}^1_d(y_{\text{out}};\vec{x}) \text{ rigid}} (-1)^{\check{\star}_d} ({}_k\check{\mathcal{R}}^1_d)_u([x_d],\ldots,[x_1]), \tag{5-93}$$

where the sign is given by

$$\check{\star}_d = \deg(x_d) + \sum_i i \cdot \deg(x_i). \tag{5-94}$$

A codimension-one analysis of the moduli spaces (5-90) reveals:

Proposition 5.16 *The following equation is satisfied*:

$$(-1)^n \delta_1 \check{\mathcal{OC}} + (-1)^n d\check{\mathcal{OC}}^1 = \mathcal{OC}^{S^1} + \check{\mathcal{OC}}^1 b. \tag{5-95}$$

Proof The boundary of the one-dimensional components of (5-90) are covered by the rigid components of the following types of strata:

- Spaces of maps with domain lying on the codimension-one boundary of the moduli space, ie in (5-86)–(5-88).
- Semistable breakings, namely those of the form

$${}_1\check{\overline{\mathcal{R}}}^1_d(y_1;\vec{x}) \times \overline{\mathcal{M}}(y_{\text{out}};y_1), \tag{5-96}$$

$$\overline{\mathcal{R}}^1(x;x_i) \times {}_1\check{\overline{\mathcal{R}}}^1_d(y_{\text{out}};\tilde{\vec{x}}), \tag{5-97}$$

where $\tilde{\vec{x}}$ denotes the collection of inputs $\vec{x}$ with x_i replaced with x.

All together, this implies, up to signs, that

$$(-1)^n \delta_1 \check{\mathcal{OC}} + (-1)^n d\check{\mathcal{OC}}^1 = \mathcal{OC}^{S^1} + \check{\mathcal{OC}}^1 b. \tag{5-98}$$

Equation (5-98) is of course a shorthand for saying, for any d and any tuple of d cyclically composable morphisms $x_d, \dots, x_1$, that

$$
\begin{aligned}
\text{(5-99)}\quad (-1)^n \sum_{i \in \{0,1\}} \delta_i \check{\mathcal{OC}}_d^{k-i}(x_d, \dots, x_1) &= \mathcal{OC}_d^{S^1}(x_d, \dots, x_1) \\
&\quad + \sum_{i,s} (-1)^{\maltese_1^s} \check{\mathcal{OC}}_{d-i+1}^1(x_d, \dots, x_{s+i+1}, \mu^i(x_{s+i}, \dots, x_{s+1}), x_s, \dots, x_1) \\
&\quad\quad + \sum_{i,j} (-1)^{\sharp_j^i} \check{\mathcal{OC}}_{d-i-j}^1\big(\mu^{i+j+1}(x_i, \dots, x_1, x_d, \dots, x_{d-j}), \\
&\qquad\qquad\qquad x_{d-j-1}, \dots, x_{i+1}\big).
\end{aligned}
$$

(Recall the abuse of notation $x_i := [x_i]$.) Thus, it suffices to verify that the signs coming from the codimension-one boundary are exactly those appearing in (5-98) — in particular, that the terms in, for instance, $\check{\mathcal{OC}}^1 b$ appear with the right sign.

Let us recall broadly how the signs are computed. For any operator g defined above, such as $\mathcal{OC}$, $\mathcal{OC}^{S^1}$, μ, d, δ_1 etc, we let g_{ut} denote the *untwisted* version of the same operator, for instance, the operator whose matrix coefficients come from the induced isomorphism on orientation lines, without any sign twists by the degree of the inputs. So, for instance, $\mu^d(x_d, \dots, x_1) = (-1)^{\sum_{i=1}^d i \deg(x_i)} \mu_{\mathrm{ut}}^d(x_d, \dots, x_1)$, and so on. The methods described in [52, Proposition 12.3] and elaborated upon in [1, Section C.3, Lemma 5.3] and [24, Section B], when applied to the boundary of the one-dimensional component of the moduli space of maps, $\overline{\check{\mathcal{R}}}_d^1(y_{\mathrm{out}}, \vec{x}))$, imply the signed equality

$$
\begin{aligned}
\text{(5-100)}\quad 0 = d_{\mathrm{ut}} \check{\mathcal{OC}}_{\mathrm{ut}}^1(x_d, \dots, x_1) + (\delta_1)_{\mathrm{ut}} \check{\mathcal{OC}}_{\mathrm{ut}}(x_d, \dots, x_1) - \mathcal{OC}_{\mathrm{ut}}^{S^1}(x_d, \dots, x_1) \\
+ (-1)^{\mathfrak{f}_d} \check{\mathcal{OC}}^1 b(x_d, \dots, x_1),
\end{aligned}
$$

where

$$
\text{(5-101)}\qquad \mathfrak{f}_d := \sum_i (i+1) \deg(x_i) + \deg(x_d) = \check{\star}_d + \maltese_d - d
$$

is an auxiliary sign.

To explain equation (5-100), we note first that the signs appearing in all terms but the last are simply induced by the boundary orientation on the moduli space of domains. The sign appearing in the first term also follows from a standard boundary orientation analysis for Floer cylinders, which we omit (but see eg [52, (12.19-012.20)] for a version close in spirit). The signs for the first two terms are also exactly as in Lemma 4.11. Finally, in the last term, the sign $(-1)^{\mathfrak{f}_d} \check{\mathcal{OC}}^1 b(x_d \otimes \cdots \otimes x_1)$ (compare [52, (12.25)] and [24, (B.59)]) appears as a cumulative sum of:

- The sign twists which turn the untwisted operations $\check{\mathcal{OC}}^1_{\mathrm{ut}}$ and μ^s_{ut} into the usual operations $\check{\mathcal{OC}}^1$ and μ^s.
- The Koszul sign appearing in the Hochschild differential b.
- The boundary orientation sign appearing in the relevant (untwisted) term of $\check{\mathcal{OC}}^1 b$, for instance $\check{\mathcal{OC}}^1_{\mathrm{ut}}(x_d, \dots, x_{n+m+1}, \mu^m_{ut}(x_{n+m}, \dots, x_{n+1}), x_n, \dots, x_1)$, which itself is as a sum of two different contributions:
  - (a) The comparison between the boundary (of the chosen) orientation and the product (of the chosen orientation) on the moduli of *domains*.
  - (b) Koszul reordering signs, which measure the signed failure of the method of orienting the moduli of maps (in terms of orientations of the domain and orientation lines of inputs and outputs) to be compatible with passing to boundary strata.

See [52, (12d)] for more details in the case of the A_∞ structure, and [1, Section C] as well as [24, Section C] for the case of these computations for the open–closed map. We note in particular that the forgetful map $F_1 : {}_1\check{\mathcal{R}}^1_d \to \check{\mathcal{R}}^1_d$ which forgets the point p_1 (and changes the direction of the asymptotic marker to point at z_d) has *complex oriented fibers* (in which just the marked point p_1 varies). So the boundary analysis of these "$\check{\mathcal{OC}}^1 \circ b$" strata appearing here is identical to the analysis strata appearing in [1; 24] for the "$\mathcal{OC} \circ b$" strata, which is why we have not repeated it here.

Multiplying all terms of (5-100) by $(-1)^{\check{\star}_d + \maltese_d - d + 1}$ and noting that, for instance, $\maltese_d - d + 1 + n - 2 = \deg(\check{\mathcal{OC}}^1(x_d \otimes \cdots \otimes x_1))$, so that

$$\begin{aligned}(-1)^{\check{\star}_d + \maltese_d - d + 1} (\delta_1)_{\mathrm{ut}} \check{\mathcal{OC}}^1_{\mathrm{ut}}(x_d, \dots, x_1) &= (-1)^{\deg(\check{\mathcal{OC}}^1(x_d,\dots,x_1)) - n} (\delta_1)_{\mathrm{ut}} (-1)^{\check{\star}_d} \check{\mathcal{OC}}^1_{\mathrm{ut}}(x_d, \dots, x_1) \\ &= \delta_1 \check{\mathcal{OC}}^1(x_d, \dots, x_1),\end{aligned} \tag{5-102}$$

and similarly for the $d \circ \mathcal{OC}^1$ term, it follows that

$$\begin{aligned}0 = (-1)^n \delta_1 \check{\mathcal{OC}}(x_d, \dots, x_1) + (-1)^n d \check{\mathcal{OC}}^1(x_d, \dots, x_1) \\ - \check{\mathcal{OC}}^1 b(x_d, \dots, x_1) - (-1)^{\check{\star}_d + \maltese_d - d + 1} \mathcal{OC}^{S^1}_{\mathrm{ut}}(x_d, \dots, x_1),\end{aligned} \tag{5-103}$$

but $\check{\star}_d + \maltese_d - d + 1 = \clubsuit_d$, and hence the last term above is $-\mathcal{OC}^{S^1}(x_d, \dots, x_1)$, as desired. □

Proof of Proposition 5.15 The "sector decomposition" performed in Proposition 5.9 which compares $\mathcal{OC}^{S^1}$ to $\hat{\mathcal{OC}} \circ B^{\mathrm{nu}} \circ \iota$, along with Proposition 5.16, immediately implies the result. □

5.5 The main construction

We now turn to the definition of the (closed) morphism of S^1–complexes, and the proof of Theorem 1.1 and Corollary 1.5. The required data takes the form

$$\widetilde{\mathcal{OC}} = \bigoplus_{k\geq 0} \overline{\boldsymbol{k}[\Lambda]/\Lambda^2}^{\otimes k} \otimes \mathrm{CH}^{\mathrm{nu}}_*(\mathcal{F}) \to CF^*(M)[n], \tag{5-104}$$

which is equivalent, as recalled in Section 2.1, to defining the collection of maps $\widetilde{\mathcal{OC}} = \{\mathcal{OC}^k\}_{k\geq 0}$, or u–linearly (see Section 2.3) $\widetilde{\mathcal{OC}} = \sum_{k=0}^{\infty} \mathcal{OC}^k u^k$, where

$$\mathcal{OC}^k = (\check{\mathcal{OC}}^k + \hat{\mathcal{OC}}^k) := \widetilde{\mathcal{OC}}^{k|1}(\Lambda, \dots, \Lambda, -)\colon \mathrm{CH}^{\mathrm{nu}}_*(\mathcal{F}) \to CF^{*+n-2k}(M). \tag{5-105}$$

(Recall from Section 2.1 that $\boldsymbol{k}[\Lambda]/\Lambda^2$ is our small model for $C_{-*}(S^1)$, and S^1–complexes are by definition strictly unital A_∞–modules over $\boldsymbol{k}[\Lambda]/\Lambda^2$.) By definition, the case $k = 0$ is already covered:

$$\mathcal{OC}^0 = (\check{\mathcal{OC}}^0 \oplus \hat{\mathcal{OC}}^0) = (\check{\mathcal{OC}} \oplus \hat{\mathcal{OC}}) = \mathcal{OC}. \tag{5-106}$$

To handle the general case ($k \geq 0$), for each d we will associate operations, for each d, to compactifications of three moduli spaces of domains, in the order

$$ {}_k\check{\mathcal{R}}^1_d, \tag{5-107}$$

$$ {}_k\mathcal{R}^{S^1}_d, \tag{5-108}$$

$$ {}_k\hat{\mathcal{R}}^1_d. \tag{5-109}$$

The moduli space (5-108) will induce an auxiliary operation useful for the proof, whereas (5-107) and (5-109) will lead to the desired operations. For $k = 0$, these moduli spaces are simply $\check{\mathcal{R}}^1_d$, $\mathcal{R}^{S^1}_d$ and $\hat{\mathcal{R}}^1_d$ as defined earlier, and the $k = 1$ case of (5-107) was defined in (5-80). Inductively, we will construct and study operations from (5-107) and (5-108) simultaneously, and then finally construct (5-109). Using these moduli spaces, we will construct the maps $\check{\mathcal{OC}}^k$ and $\hat{\mathcal{OC}}^k$, as well as an auxiliary operation $\mathcal{OC}^{S^1,k}$ (which we compare to $\hat{\mathcal{OC}}^{k-1} \circ B^{\mathrm{nu}}$ in Proposition 5.20 below), and then prove:

Proposition 5.17 *The following equations hold, for each $k \geq 0$:*

$$(-1)^n \sum_{i\geq 0}^{k} \delta_i \check{\mathcal{OC}}^{k-i} = \hat{\mathcal{OC}}^{k-1} B^{\mathrm{nu}} + \check{\mathcal{OC}}^k b, \tag{5-110}$$

$$(-1)^n \sum_{i\geq 0}^{k} \delta_i \hat{\mathcal{OC}}^{k-i} = \hat{\mathcal{OC}}^k b' + \check{\mathcal{OC}}^k (1-t). \tag{5-111}$$

All at once, writing

$$\mathcal{OC}^k = (\check{\mathcal{OC}}^k + \hat{\mathcal{OC}}^k), \quad \widetilde{\mathcal{OC}} = \sum_{i=0}^{\infty} \mathcal{OC}^i u^i, \quad \delta_{\mathrm{eq}} = \sum_{j=0}^{\infty} \delta_j^{CF} u^j, \quad b_{\mathrm{eq}} = b^{\mathrm{nu}} + uB^{\mathrm{nu}},$$

as in Section 2.3, we have that

$$(-1)^n \delta_{\mathrm{eq}} \circ \widetilde{\mathcal{OC}} = \widetilde{\mathcal{OC}} \circ b_{\mathrm{eq}}. \tag{5-112}$$

This will also directly imply our main theorems, as spelled out at the end of this subsection.

The space (5-107) is the moduli space of discs with d positive boundary marked points $z_1, \dots, z_d$ labeled in counterclockwise order, one interior negative puncture z_{out} equipped with an asymptotic marker, and k additional interior marked points $p_1, \dots, p_k$ (without asymptotic markers), marked as *auxiliary*. Also, on the unit disc representative of any element of this moduli space which fixes z_d at 1 and z_{out} at 0, the p_i should be *strictly radially ordered* with norms in $\left(0, \frac{1}{2}\right)$; that is,

$$0 < |p_1| < \cdots < |p_k| < \tfrac{1}{2}. \tag{5-113}$$

Using the above representative, one can talk about the *angle*, or *argument*, of each auxiliary interior marked point,

$$\theta_i := \arg(p_i). \tag{5-114}$$

We require that with respect to the above representative:

(5-115) The asymptotic marker on z_{out} points in the direction θ_1 (or towards z_d if $k = 0$).

(Equivalently one could define $\theta_{k+1} = 0$, so that θ_1 is always defined.) See Figure 7 for a depiction. For every representative $S \in {}_k\check{\mathcal{R}}_d^1$:

(5-116) Fix a negative cylindrical end around z_{out} not containing any p_i, compatible with the direction of the asymptotic marker, or *equivalently compatible with the angle* θ_1.

The second moduli space (5-108) is the moduli space of discs with d positive boundary marked points $z_1, \dots, z_d$ labeled in counterclockwise order, 1 interior negative puncture z_{out} equipped with an asymptotic marker, and $k+1$ additional interior marked points $p_1, \dots, p_k, p_{k+1}$ (without asymptotic markers), marked as *auxiliary*. With respect to

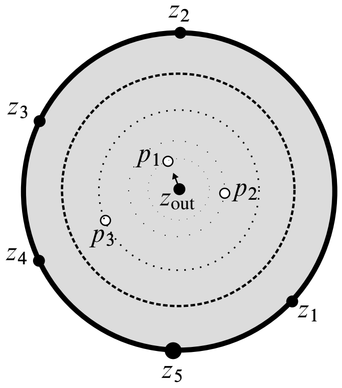


Figure 7: A representative of an element of the moduli space ${}_3\check{\mathcal{R}}^1_5$.

the unit disc representative of any element this moduli space fixing z_d at 1 and z_{out} at 0, the p_i should again be *strictly radially ordered*, this time with norms lying in $\bigl(0, \frac{1}{2}\bigr]$ and with p_{k+1} *lying on the circle of radius* $\frac{1}{2}$,

$$0 < |p_1| < \cdots < |p_k| < |p_{k+1}| = \tfrac{1}{2}. \tag{5-117}$$

The asymptotic marker on z_{out} for this representative again satisfies condition (5-115). Abstractly we have that ${}_k\mathcal{R}^{S^1}_d \cong \times_k\check{\mathcal{R}}^1_d \times S^1$, where the S^1 parameter is given by the position of p_{k+1}. See Figure 8 for a depiction of ${}_{k-1}\check{\mathcal{R}}^{S^1}_d$.

The compactification of (5-107) is a real blow-up of the ordinary Deligne–Mumford compactification, in the sense of [34] (see [58] for a first discussion in the context of Floer theory), reviewed in more detail in Section A.1. The result of the discussion there is that the codimension-one boundary of the compactified check moduli space ${}_k\overline{\check{\mathcal{R}}}^1_d$ is

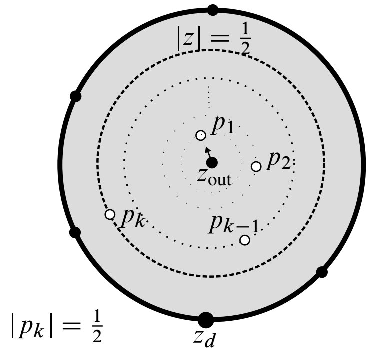


Figure 8: A representative of an element of the moduli space ${}_{k-1}\check{\mathcal{R}}^{S^1}_d$, which also arises as the boundary stratum (5-120) of ${}_k\overline{\check{\mathcal{R}}}^1_d$.

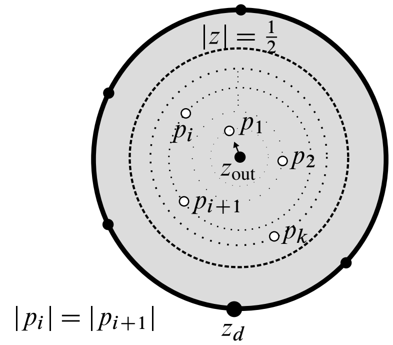


Figure 9: A representative of an element of the stratum (5-121).

covered by the images of the natural inclusions of the strata

$$\overline{\mathcal{R}}^s \times {}_k\check{\overline{\mathcal{R}}}{}^1_{d-s+1}, \tag{5-118}$$

$$ {}_s\check{\overline{\mathcal{R}}}{}^1_d \times \overline{\mathcal{M}}_{k-s}, \tag{5-119}$$

$$ {}_{k-1}\check{\overline{\mathcal{R}}}{}^{S^1}_d, \tag{5-120}$$

$$ {}^{i,i+1}_{k}\check{\overline{\mathcal{R}}}{}^1_d. \tag{5-121}$$

The strata (5-120)–(5-121), in which $|p_k| = \frac{1}{2}$ (Figure 8) and $|p_i| = |p_{i+1}|$ (Figure 9), respectively, describe the boundary loci of the ordering condition (5-113) and hence come equipped with a natural manifold-with-corners structure. The strata (5-118)–(5-119) have manifold-with-corners structure given by standard local gluing maps using fixed choices of strip-like ends near the boundary. For (5-118), depicted in Figure 10, this is standard, and for (5-119), depicted in Figure 11, the local gluing map uses the cylindrical ends (5-116) and (4-33) — in other words, one rotates the $(k-s)$–pointed angle cylinder by an amount commensurate to the angle of the first marked point p_{k-s+1} on the disk before gluing — as also described in Section A.1.

Associated to the stratum (5-121) where p_i and p_{i+1} have coincident magnitudes, there is a forgetful map

$$\check{\pi}_i \colon {}^{i,i+1}_{k}\check{\overline{\mathcal{R}}}{}^1_d \to {}_{k-1}\check{\overline{\mathcal{R}}}{}^1_d \tag{5-122}$$

which simply forgets the point p_{i+1}. Since the norm of p_{i+1} and p_i agree on this locus, this amounts to forgetting the argument of p_{i+1} (in particular, the fibers of $\check{\pi}_i$ are one-dimensional).

The compactification of the S^1 moduli space (5-108) can be modeled abstractly by ${}_k\check{\overline{\mathcal{R}}}{}^1_d \times S^1$. However, it is again preferable to give an explicit description of the boundary

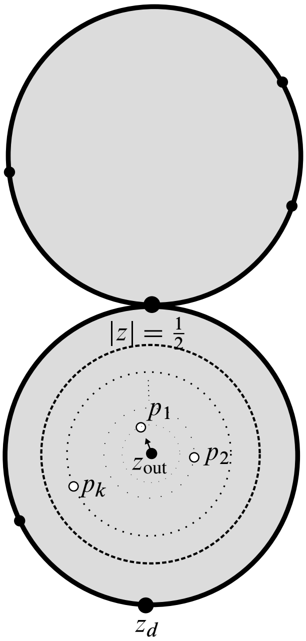


Figure 10: A representative of an element of the boundary stratum (5-118) in which a disc bubble forms (such a disc bubble is allowed to include the "first" point — z_d by our convention — but need not, and does not in the figure).

strata, which are covered in codimension one by the strata

$$\overline{\mathcal{R}}^s \times {}_k\overline{\mathcal{R}}^{S^1}_{d-s+1}, \tag{5-123}$$

$$ {}_s\overline{\mathcal{R}}^{S^1}_d \times \overline{\mathcal{M}}_{k-s}, \tag{5-124}$$

$$ {}^{i,i+1}_{k}\overline{\mathcal{R}}^{S^1}_d. \tag{5-125}$$

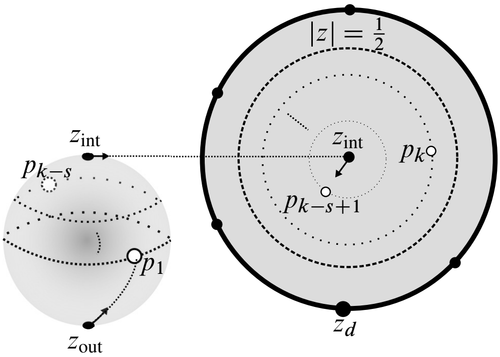


Figure 11: A representative of an element of the boundary stratum (5-119).

Here, (5-123) and (5-124) are just versions of the degenerations (5-118) and (5-119), in which a collection of boundary points bubbles off, or a collection of auxiliary points converges to z_{out} and bubbles off; the fact that the latter occurs in codimension one is part of the "real blow-up phenomenon" already discussed. The stratum (5-125) is the locus where $|p_i| = |p_{i+1}|$, for $i \le k$; so when $i = k$, $|p_k| = |p_{k+1}| = \frac{1}{2}$.

As in (5-122), on the stratum (5-125), where p_i and p_{i+1} have coincident magnitudes, define the map

$$\pi_i^{S^1} : {}_k^{i,i+1}\overline{\mathcal{R}}_d^{S^1} \to {}_{k-1}\overline{\mathcal{R}}_d^{S^1} \tag{5-126}$$

to be the one forgetting the point p_{i+1}. As before, this map has one-dimensional fibers.

For an element $S \in {}_k\overline{\mathcal{R}}_d^{S^1}$, we say that p_{k+1} *points at a boundary point* z_i if, for any unit disc representative of S with z_{out} at the origin, the ray from z_{out} to p_{k+1} intersects z_i. The locus where p_{k+1} points at z_i is denoted by

$${}_k\overline{\mathcal{R}}_d^{S^1_i}. \tag{5-127}$$

Similarly, we say that p_{k+1} *points between* z_i *and* z_{i+1} (modulo d, so this includes the case of pointing between z_d and z_1) if for such a representative, the ray from z_{out} to p_{k+1} intersects the portion of ∂S between z_i and z_{i+1}. The locus where p_{k+1} points between z_i and z_{i+1} is denoted by

$${}_k\overline{\mathcal{R}}_d^{S^1_{i,i+1}}. \tag{5-128}$$

As before in (5-52), there is a free and properly discontinuous $\mathbb{Z}_d$–action

$$\kappa : {}_k\overline{(\mathcal{R}_d^1)^{S^1}} \to {}_k\overline{(\mathcal{R}_d^1)^{S^1}} \tag{5-129}$$

which cyclically permutes the labels of the boundary marked points; as before, κ changes the label z_i to z_{i+1} for $i < d$, and z_d to z_1; compare Lemma 5.10.

Finally, we come to the third moduli space (5-109), the moduli space of discs with $d+1$ positive boundary marked points $z_1, \dots, z_d, z_f$ labeled in counterclockwise order, one interior negative puncture z_{out} equipped with an asymptotic marker, and k additional interior marked points $p_1, \dots, p_k$ (without an asymptotic marker), marked as *auxiliary*, satisfying a *strict radial ordering* condition as before: for any representative element with z_f fixed at 1 and z_{out} at 0, we require (5-113) to hold, as well as condition (5-115). The boundary marked point z_f is also marked as auxiliary, but apart from this designation we see, identifying z_f with z_{d+1}, that ${}_k\widehat{\mathcal{R}}_d^1 \cong {}_k\check{\mathcal{R}}_{d+1}^1$. See Figure 12.

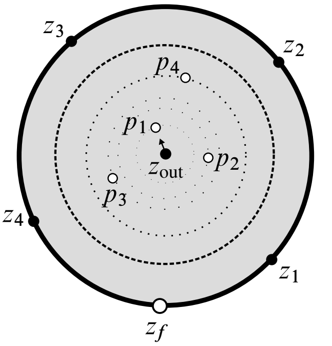


Figure 12: A representative of an element of the moduli space ${}_4\widehat{\mathcal{R}}^1_4$.

In codimension one, the compactification ${}_k\overline{\widehat{\mathcal{R}}}^1_d$ has boundary covered by inclusions of the strata

$$\overline{\mathcal{R}}^s \times {}_k\overline{\widehat{\mathcal{R}}}^1_{d-s+1}, \tag{5-130}$$

$$\overline{\mathcal{R}}^{m,f_k} \times_{d-m+1} {}_k\overline{\widehat{\mathcal{R}}}^1_{d-m+1}, \quad \text{where } 1 \le k \le m, \tag{5-131}$$

$$ {}_s\overline{\widehat{\mathcal{R}}}^1_d \times \overline{\mathcal{M}}_{k-s}, \tag{5-132}$$

$$ {}_{k-1}\overline{\widehat{\mathcal{R}}}^{S^1}_d, \tag{5-133}$$

$$ {}^{i,i+1}_{k}\overline{\widehat{\mathcal{R}}}^1_d. \tag{5-134}$$

Once more, on strata (5-134) where p_i and p_{i+1} have coincident magnitudes, depicted in Figure 13, left, define the map

$$\hat{\pi}_i \colon {}^{i,i+1}_{k}\overline{\widehat{\mathcal{R}}}^1_d \to {}_{k-1}\overline{\widehat{\mathcal{R}}}^1_d \tag{5-135}$$

to be the one forgetting the point p_{i+1}. Again, this map has one-dimensional fibers. On the stratum (5-133), which is the locus where $|p_k| = \frac{1}{2}$ (Figure 13, right), there is also a map of interest

$$\hat{\pi}_{\text{boundary}} \colon {}_{k-1}\overline{\widehat{\mathcal{R}}}^{S^1}_d \to {}_{k-1}\overline{\mathcal{R}}^{S^1}_d \tag{5-136}$$

which forgets the position of the auxiliary boundary point z_f. The stratum (5-132), depicted in Figure 14, is the locus where some subcollection of interior auxiliary points $p_1, \dots, p_{k-s}$ tend to zero and split off an angle-decorated cylinder (in the manner again described in Section A.1 for (5-119)). The strata (5-130) and (5-131), depicted in Figure 15 on the left and right, respectively, are the loci where a disc bubble forms involving some boundary marked points (not including or including z_f, respectively).

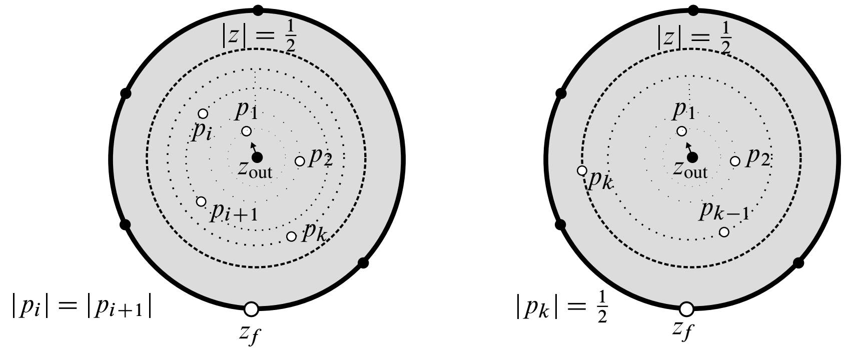


Figure 13: A representative of an element of the stratum (5-134), left, and a representative of an element of the boundary stratum (5-133), right.

Denote by ${}_k\mathcal{R}_d^{1,\text{free}} := {}_k\mathcal{R}_d^{S^1_{d,1}}$ the sector of the moduli space ${}_k\mathcal{R}_d^{S^1}$ where p_{k+1} points between z_d and z_1. The *auxiliary-rescaling map*

$$\pi_f : {}_k\widehat{\mathcal{R}}_d^1 \to {}_k\mathcal{R}_d^{1,\text{free}} \tag{5-137}$$

(our replacement of the "forgetful map") can be described as follows: given a representative S in ${}_k\widehat{\mathcal{R}}_d^1$ with z_{out} fixed at the origin, there is a unique point p with $|p| = \frac{1}{2}$ between z_{out} and z_f. The element $\pi_f(S)$ is the element of ${}_k\mathcal{R}_d^{S^1}$ obtained from S by setting p_{k+1} equal to this point p and deleting z_f. Of course, z_f is not actually forgotten, because it is determined by the position of p_{k+1}. In particular, (5-137) is a diffeomorphism. We extend this map to a map $\overline{\pi}_f$ from the compactification ${}_k\overline{\widehat{\mathcal{R}}}_d^1$ as in Section 5.2, by putting the auxiliary point z_f back in, eliminating any component which is not main or secondary which has only one (nonauxiliary) boundary marked point q, and by labeling the positive marked point below this component by q.

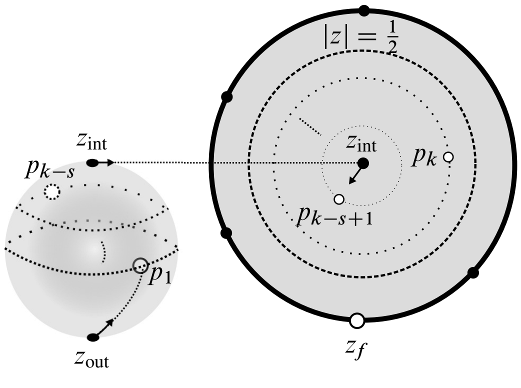


Figure 14: A representative of an element of the boundary stratum (5-132).

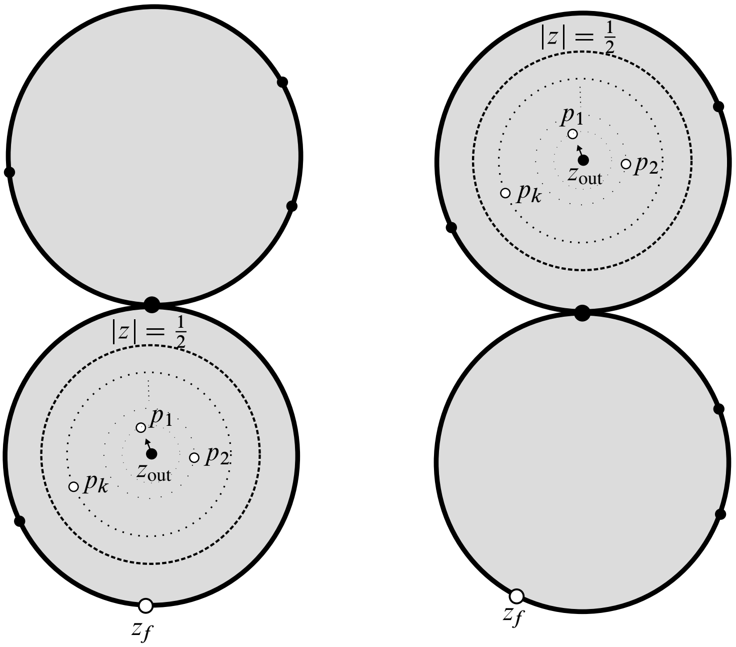


Figure 15: Left: a representative of an element of the boundary stratum (5-130) in which a disc bubble forms not including the auxiliary point z_f. Right: a representative of an element of the boundary stratum (5-131) in which a disc bubble forms including the auxiliary point z_f.

We orient the moduli spaces (5-107)–(5-109) as follows: pick, on a slice of the automorphism action which fixes the position of z_d at 1 and z_{out} at 0, the volume forms

$$-r_1 \cdots r_k \, dz_1 \wedge dz_2 \wedge \cdots \wedge dz_{d-1} \wedge dr_1 \wedge d\theta_1 \wedge \cdots \wedge dr_k \wedge d\theta_k, \tag{5-138}$$

$$r_1 \cdots r_k \, dz_1 \wedge dz_2 \wedge \cdots \wedge dz_{d-1} \wedge d\theta_{k+1} \wedge dr_1 \wedge d\theta_1 \wedge \cdots \wedge dr_k \wedge d\theta_k, \tag{5-139}$$

$$r_1 \cdots r_k \, dz_1 \wedge dz_2 \wedge \cdots \wedge dz_{d-1} \wedge dz_f \wedge dr_1 \wedge d\theta_1 \wedge \cdots \wedge dr_k \wedge d\theta_k. \tag{5-140}$$

Above, (r_i, θ_i) denote the polar coordinate positions of the point p_i. (We could equivalently use Cartesian coordinates (x_i, y_i) and substitute $dx_i \wedge dy_i$ for every instance of $r_i \, dr_i \wedge d\theta_i$, but polar coordinates are straightforwardly compatible with the boundary stratum where $|p_k| = \frac{1}{2}$.)

A *Floer datum* on a stable disc S in ${}_k\check{\overline{\mathcal{R}}}{}^1_d$ or a stable disc S in ${}_k\overline{\mathcal{R}}{}^{S^1}_d$ is simply a Floer datum for S in the sense of Section 5.1. A *Floer datum* on a stable disc $S \in {}_k\overline{\hat{\mathcal{R}}}{}^1_d$ is a Floer datum for $\overline{\pi}_f(S)$.

Again we will make a system of choices of Floer data for the above moduli spaces. A *Floer datum for the cyclic open–closed map* is an inductive sequence of choices, for every $k \geq 0$ and $d \geq 1$, of Floer data for every representative

$$S_0 \in {}_k\overline{\check{\mathcal{R}}}{}^1_d, \quad S_1 \in {}_k\overline{\mathcal{R}}{}^{S^1}_d \quad \text{and} \quad S_2 \in {}_k\overline{\hat{\mathcal{R}}}{}^1_d,$$

varying smoothly in S_0, S_1 and S_2, which satisfies the usual consistency condition: the choice of Floer datum on any boundary stratum should agree with the previously inductively chosen datum along any boundary stratum for which (it is possibly a product of moduli spaces for) we have already inductively picked data. Moreover, this choice should satisfy a series of additional requirements.

First, for $S_0 \in {}_k\overline{\check{\mathcal{R}}}{}^1_d$:

(5-141) At a boundary stratum of the form (5-121), the Floer datum for S_0 is equivalent to the one pulled back from ${}_{k-1}\overline{\check{\mathcal{R}}}{}^1_d$ via the forgetful map $\check{\pi}_i$.

Next, for $S_1 \in {}_k\overline{\mathcal{R}}{}^{S^1}_d$,

(5-142) On the codimension-one loci ${}_k\overline{\mathcal{R}}{}^{S^1_i}_d$, where p_{k+1} points at z_i, the Floer datum should agree with the pullback by τ_i of the existing Floer datum for the open–closed map.

(5-143) The Floer datum should be κ–equivariant, where κ is the map (5-129).

(5-144) At a boundary stratum of the form (5-125), the Floer datum for S_1 is conformally equivalent to the one pulled back from ${}_{k-1}\overline{\mathcal{R}}{}^{S^1}_d$ via the forgetful map $\pi_i^{S^1}$.

Finally, for $S_2 \in {}_k\overline{\hat{\mathcal{R}}}{}^1_d$:

(5-145) The choice of Floer datum on strata containing $\mathcal{R}^{d,f_i}$ components should be constant along fibers of the forgetful map $\mathcal{R}^{d,f_i} \to \mathcal{R}^{d-1}$.

(5-146) The Floer datum on the main component $(S_2)_0$ of $\bar{\pi}_f(S_2)$ should coincide with the Floer datum chosen on $(S_2)_0 \in {}_k\mathcal{R}^{1,\text{free}}_d \subset {}_k\mathcal{R}^{S^1}_d$.

(5-147) At a boundary stratum of the form (5-133), the Floer datum on the main component of S_2 is conformally equivalent to the one pulled back from ${}_k\overline{\mathcal{R}}{}^{S^1}_d$ via the forgetful map $\hat{\pi}_{\text{boundary}}$.

(5-148) At a boundary stratum of the form (5-134), the Floer datum for S_2 is conformally equivalent to the one pulled back from ${}_{k-1}\overline{\hat{\mathcal{R}}}{}^1_d$ via the forgetful map $\hat{\pi}_i$.

The above system of requirements can be split into three broad categories: the first type concerns the compatibility with forgetful maps of Floer data along the lower strata which were not previously constrained, the second type concerns the equivariance (under a free properly discontinuous action) of the Floer data on ${}_k\overline{\mathcal{R}}^{S^1}_d$ as well as the relationship between the Floer datum chosen here and the ones chosen on $\overline{\check{\mathcal{R}}}^1_d$ and ${}_k\overline{\hat{\mathcal{R}}}^1_d$.

Proposition 5.18 *A Floer datum for the cyclic open–closed map exists.*

Proof Since the choices of Floer data at each stage are contractible, this follows from the straightforward verification that, for a suitably chosen inductive order on strata, the conditions satisfied by the Floer data at various strata do not contradict each other. We use the following inductive order: first, say we have chosen a Floer datum for the A_∞ structure as in Section 3.3, along with a BV compatible Floer datum for the nonunital open–closed map following Section 5.3. In particular, we have chosen Floer data for the moduli spaces $\overline{\mathcal{R}}^{d,f_i}$ (per Section A.2), for ${}_0\overline{\check{\mathcal{R}}}_d$, for the auxiliary moduli space ${}_0\overline{\mathcal{R}}^{S^1}_d$, and (using the conditions above) we have induced a particular choice of Floer datum on ${}_0\overline{\hat{\mathcal{R}}}_d$. Next, inductively assuming that we have made all choices at level $k-1$ with $k>0$, we first choose Floer data for ${}_k\overline{\check{\mathcal{R}}}_d$ for each d, then ${}_k\overline{\mathcal{R}}^{S^1}_d$ for each d (by pulling back a choice of Floer datum on the quotient by κ in order to satisfy the equivariance condition), and finally note that a choice is fixed for ${}_k\overline{\hat{\mathcal{R}}}_d$ by the above constraints. □

Fixing a Floer datum for the cyclic open–closed map, we obtain, for any d–tuple of Lagrangians $L_0,\dots,L_{d-1}$, and asymptotic conditions

$$\begin{cases} \vec{x} = (x_d,\dots,x_1) \quad \text{with } x_i \in \chi(L_{i-1}, L_{i \bmod d}), \\ y_{\text{out}} \in \mathcal{O}, \end{cases} \tag{5-149}$$

Gromov–Floer compactified moduli spaces

$$ {}_k\overline{\check{\mathcal{R}}}^1_d(y_{\text{out}},\vec{x}), \tag{5-150}$$

$$ {}_k\overline{\mathcal{R}}^{S^1}_d(y_{\text{out}},\vec{x}), \tag{5-151}$$

$$ {}_k\overline{\hat{\mathcal{R}}}^1_d(y_{\text{out}},\vec{x}), \tag{5-152}$$

of maps into M from an arbitrary element S of the moduli spaces (5-107), (5-108) and (5-109) respectively (or rather from $\pi_f(S)$ in the case of (5-109)) satisfying Floer's equation using the Floer datum chosen for the given S as in (5-4), with asymptotics and Lagrangian boundary conditions as in (5-5), again with the modifications as in Remarks 5.1 or 5.2 for compact or wrapped Fukaya categories of Liouville manifolds. The virtual dimension of each component of these moduli spaces coincides (mod 2 or exactly, depending on whether we are $\mathbb{Z}/2$– or $\mathbb{Z}$–graded) with, respectively,

$$\deg(y_{\mathrm{out}}) - n + d - 1 - \sum_{i=1}^{d} \deg(x_i) + 2k \quad \text{for } {}_k\overline{\check{\mathcal{R}}}{}^1_d(y_{\mathrm{out}}, \vec{x}), \tag{5-153}$$

$$\deg(y_{\mathrm{out}}) - n + d - \sum_{i=1}^{d} \deg(x_i) + 2k \quad \text{for } {}_k\overline{\mathcal{R}}{}^{S^1}_d(y_{\mathrm{out}}, \vec{x}), \tag{5-154}$$

$$\deg(y_{\mathrm{out}}) - n + d - \sum_{i=1}^{d} \deg(x_i) + 2k \quad \text{for } {}_k\overline{\hat{\mathcal{R}}}{}^1_d(y_{\mathrm{out}}, \vec{x}). \tag{5-155}$$

By Assumption 3.10, for generic choices of Floer data, the Gromov–Floer compactifications of the components of virtual dimension ≤ 1 of (5-150)–(5-152) are compact manifolds-with-boundary of the expected dimension. For rigid elements u in the moduli spaces (5-150)–(5-152), which occur for asymptotics $(y, \vec{x})$ satisfying

$$(5\text{-}153) = 0, \quad (5\text{-}154) = 0 \quad \text{or} \quad (5\text{-}155) = 0,$$

respectively, the orientations (5-138)–(5-140) and [1, Lemma C.4] induce isomorphisms of orientation lines

$$({}_k\check{\mathcal{R}}^1_d)_u \colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_y, \tag{5-156}$$

$$({}_k\mathcal{R}^{S^1}_d)_u \colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_y, \tag{5-157}$$

$$({}_k\hat{\mathcal{R}}^1_d)_u \colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_y. \tag{5-158}$$

Summing the application of these isomorphisms over all rigid u (or "counting rigid elements") defines the $|o_{y_{\mathrm{out}}}|_{\boldsymbol{k}}$ component of three families of operations $\check{\mathcal{OC}}^k$, $\mathcal{OC}^{S^1,k}$ and $\hat{\mathcal{OC}}^k$, up to a sign twist specified below. Define

$$\check{\mathcal{OC}}^k([x_d], \dots, [x_1]) := \sum_{u \in {}_k\overline{\check{\mathcal{R}}}{}^1_d(y_{\mathrm{out}}; \vec{x}) \text{ rigid}} (-1)^{\check{\star}_d} ({}_k\check{\mathcal{R}}^1_d)_u([x_d], \dots, [x_1]), \tag{5-159}$$

$$\mathcal{OC}^{S^1,k}([x_d], \dots, [x_1]) := \sum_{u \in {}_k\overline{\mathcal{R}}{}^{S^1}_d(y_{\mathrm{out}}; \vec{x}) \text{ rigid}} (-1)^{\star^{S^1}_d} ({}_k\mathcal{R}^{S^1}_d)_u([x_d], \dots, [x_1]), \tag{5-160}$$

$$\hat{\mathcal{OC}}^k([x_d], \dots, [x_1]) := \sum_{u \in {}_k\overline{\hat{\mathcal{R}}}{}^1_d(y_{\mathrm{out}}; \vec{x}) \text{ rigid}} (-1)^{\hat{\star}_d} ({}_k\hat{\mathcal{R}}^1_d)_u([x_d], \dots, [x_1]), \tag{5-161}$$

where the signs are given by

$$\check{\star}_d = \deg(x_d) + \sum_i i \cdot \deg(x_i), \tag{5-162}$$

$$\star_d^{S^1} = \clubsuit_d \sum_{i=1}^{d} (i+1) \cdot \deg(x_i) + \deg(x_d) + d - 1 = \check{\star}_d + \maltese_d - 1, \tag{5-163}$$

$$\hat{\star}_d = \sum_i i \cdot \deg(x_i). \tag{5-164}$$

A codimension-one analysis of the moduli spaces (5-150) and (5-152) reveals:

Proposition 5.19 *The following equations hold for each $k \geq 0$:*

$$(-1)^n \sum_{i=0}^{k} \delta_i \check{\mathcal{OC}}^{k-i} = \mathcal{OC}^{S^1,k-1} + \check{\mathcal{OC}}^k b, \tag{5-165}$$

$$(-1)^n \sum_{i=0}^{k} \delta_i \hat{\mathcal{OC}}^{k-i} = \hat{\mathcal{OC}}^k b' + \check{\mathcal{OC}}^k (1-t). \tag{5-166}$$

Proof The boundary of the one-dimensional components of (5-150) are covered by the (rigid components of) the following types of strata:

- Spaces of maps with domain lying on the codimension-one boundary of the moduli space, ie in (5-118)–(5-121).
- Semistable breakings, namely those of the form

$${}_k\overline{\check{\mathcal{R}}}{}_d^1(y_1; \vec{x}) \times \overline{\mathcal{M}}(y_{\mathrm{out}}; y_1), \tag{5-167}$$

$$\overline{\mathcal{R}}^1(x; x_i) \times {}_k\overline{\check{\mathcal{R}}}{}_d^1(y_{\mathrm{out}}; \tilde{\vec{x}}), \tag{5-168}$$

  where again $\tilde{\vec{x}}$ denotes the collection of inputs $\vec{x}$ with x_i replaced with x.

All together, this implies, up to sign, that

$$(-1)^n \sum_{i=0}^{k} \delta_i \check{\mathcal{OC}}^{k-i} = \mathcal{OC}^{S^1,k-1} + \check{\mathcal{OC}}^k b + \sum_{i=1}^{k-1} \check{\mathcal{OC}}^{k,i,i+1}, \tag{5-169}$$

where $\check{\mathcal{OC}}^{k,i,i+1}$ is an operation corresponding with some sign twist to (5-121). Of course equation (5-169) is a shorthand for saying, for a tuple of d cyclically composable morphisms $x_d, \dots, x_1$ (recalling the abuse of notation $x_i := [x_i]$), that

$$
\begin{aligned}
(5\text{-}170)\quad & (-1)^n \sum_{i=0}^{k} \delta_i \check{\mathcal{OC}}_d^{k-i}(x_d, \dots, x_1) \\
&= \mathcal{OC}_d^{S^1,k-1}(x_d, \dots, x_1) + \sum_{i=1}^{k-1} \check{\mathcal{OC}}_d^{k,i,i+1}(x_d, \dots, x_1) \\
&\quad + \sum_{i,s} (-1)^{\maltese_1^s} \check{\mathcal{OC}}_{d-i+1}^{k}(x_d, \dots, x_{s+i+1}, \mu^i(x_{s+i}, \dots, x_{s+1}), x_s, \dots, x_1) \\
&\quad + \sum_{i,j} (-1)^{\sharp_j^i} \check{\mathcal{OC}}^{k}(\mu^{i+j+1}(x_i, \dots, x_1, x_d, \dots, x_{d-j}), x_{d-j-1}, \dots, x_{i+1}).
\end{aligned}
$$

We first note that in fact the operation $\check{\mathcal{OC}}^{k,i,i+1} = \sum_d \check{\mathcal{OC}}_d^{k,i,i+1}$ is zero, because by condition (5-141), the Floer datum chosen for elements S in (5-121) are constant along the one-dimensional fibers of $\check{\pi}_i$. Hence, elements of the moduli space with source in (5-121) are never rigid; see Lemma 4.11 for an analogous and more detailed explanation.

Thus, it suffices to verify that the signs coming from the codimension-one boundary are exactly those appearing in (5-169). We can safely ignore studying any signs for the vanishing operations such as $\hat{\mathcal{OC}}^{k,i,i+1}$. The remaining sign analysis is exactly as in Proposition 5.16; more precisely, note that the forgetful map $\check{F}_k \colon {}_k\check{\mathcal{R}}_d^1 \to {}_1\check{\mathcal{R}}_d^1$ which forgets $p_1, \dots, p_{k-1}$ has complex oriented fibers, and in particular (since the marked points p_i contribute complex domain orientations and do not introduce any new orientation lines) the sign computations sketched in Proposition 5.16 carry over for any stratum whose domain is pulled back from a boundary stratum of ${}_1\check{\mathcal{R}}_d^1$; in turn, as described in Proposition 5.16, the sign computations for ${}_1\check{\mathcal{R}}_d^1$ largely reduce to those for ${}_0\check{\mathcal{R}}_d^1$. This verifies (5-169).

Similarly, for the hat moduli space, an analysis of the boundary of one-dimensional moduli spaces of maps tells us, up to sign verification,

$$
(5\text{-}171)\quad (-1)^n \sum_{i=0}^{k} \delta_i \hat{\mathcal{OC}}^{k-i} = \hat{\mathcal{OC}}^k b' + \check{\mathcal{OC}}^k (1-t) + \hat{\mathcal{OC}}^{k,k,k+1} + \sum_{i=1}^{k-1} \hat{\mathcal{OC}}^{k,i,i+1},
$$

where $\hat{\mathcal{OC}}^{k,k,k+1}$ is an operation corresponding with some sign twist to (5-133), and $\hat{\mathcal{OC}}^{k,i,i+1}$ is an operation corresponding with some sign twist to (5-134). The conditions (5-147)–(5-148) similarly imply that $\hat{\mathcal{OC}}^{k,k,k+1}$ and $\hat{\mathcal{OC}}^{k,i,i+1}$ are zero, so it is not necessary to even establish what the signs for these terms are.

To verify signs for (5-171), we apply the principle discussed in the proof of Lemma 5.8, in which by treating the auxiliary boundary marked point z_f as possessing a "formal unit element asymptotic constraint e_+", therefore viewing $\hat{\mathcal{OC}}^k(x_d \otimes \cdots \otimes x_1) := \hat{\mathcal{OC}}^k(x_d, \dots, x_1)$ formally as $\hat{\mathcal{OC}}^k(e^+ \otimes x_d \otimes \cdots \otimes x_1)$, the signs for (5-171) applied to strings $(x_d \otimes \cdots \otimes x_1)$ of length d follow from the sign computations for $\check{\mathcal{OC}}$ applied to strings $(e^+ \otimes x_d \otimes \cdots \otimes x_1)$ of length $d+1$. This analysis applies to the term $\hat{\mathcal{OC}}^{k,k,k+1}$ as well, which is the hat version of $\mathcal{OC}^{S^1,k}$; however, the former operation happens to be zero because extra symmetries imply the moduli space controlling this operation is never rigid. □

Next, by decomposing the moduli space ${}_k\mathcal{R}_d^{S^1}$ into sectors, we can write the auxiliary operation $\mathcal{OC}^{S^1,j}$ in terms of $\hat{\mathcal{OC}}^j$ and Connes' B operator.

Proposition 5.20 *As chain-level operations*,

$$\mathcal{OC}^{S^1,k} = \hat{\mathcal{OC}}^k \circ B^{\mathrm{nu}}. \tag{5-172}$$

Proof The proof directly emulates Proposition 5.9, and as such we will give fewer details. We begin by defining, for $i \in \mathbb{Z}/d\mathbb{Z}$, operations

$$\hat{\mathcal{OC}}^k_{d,\tau_i} \tag{5-173}$$

associated to various "sectors" of the $k+1^{\text{st}}$ marked point p_{k+1} of ${}_k\mathcal{R}_d^{S^1}$. Once more, to gain better control of the geometry of these sectors in the compactification (when the sector size can shrink to zero), we pass to an alternative model for the compactification. Define

$${}_k\mathcal{R}^1_{d,\tau_i} \tag{5-174}$$

to be the abstract moduli space of discs with $d+1$ boundary punctures, $z_1, \dots, z_i$, z_f, $z_{i+1}, \dots, z_d$ arranged in counterclockwise order, one interior negative puncture z_{out} with asymptotic marker, and k additional interior auxiliary marked points $p_1, \dots, p_k$ which are *strictly radially ordered* with norms in $\left(0, \frac{1}{2}\right)$ for a representative fixing z_0 at 1 and z_{out} at 0, so

$$0 < |p_1| < \cdots < |p_k| < \tfrac{1}{2}. \tag{5-175}$$

Moreover, as before:

(5-176) The asymptotic marker on z_{out} points in the direction θ_1 (or towards z_f if $k = 0$).

There is a bijection

$$\tau_i\colon {}_k\mathcal{R}^1_{d,\tau_i} \to {}_k\widehat{\mathcal{R}}^1_d \tag{5-177}$$

given by cyclically permuting boundary labels, and in particular we also have an *auxiliary-rescaling map*, as in (5-137),

$$ {}_k\mathcal{R}^1_{d,\tau_i} \to {}_k\mathcal{R}_d^{S^1_{i,i+1}}, \tag{5-178}$$

which, for a representative with $|z_{\text{out}}| = 0$, adds a point p_{k+1} on the line between z_{out} and z_f with $|p_{k+1}| = \frac{1}{2}$ and deletes z_f. We choose orientations on ${}_k\mathcal{R}^1_{d,\tau_i}$ to be compatible with (5-178); more concretely, for a slice fixing the positions of z_{out} and z_d, consider the top form

$$r_1 \cdots r_k \, dz_1 \wedge dz_2 \wedge \cdots \wedge dz_{d-1} \wedge dz_d \wedge dz_f \wedge dr_1 \wedge d\theta_1 \wedge \cdots \wedge dr_k \wedge d\theta_k. \tag{5-179}$$

The compactification ${}_k\overline{\mathcal{R}}^1_{d,\tau_i}$ is inherited from the identification (5-177); the salient point is that we treat bubbled-off boundary strata containing the point z_f as coming from $\mathcal{R}^{d,f_i}$, the moduli space of discs with i^{th} marked point forgotten (where the i^{th} marked point is z_f), constructed in Section A.2.

We choose as a Floer datum for ${}_k\overline{\mathcal{R}}^1_{d,\tau_i}$ the pulled-back Floer datum from (5-177); it automatically then exists and is universal and consistent as desired. Moreover we have chosen orientations as in the case $k = 0$ so that the auxiliary rescaling map (5-178) is an oriented diffeomorphism extending to a map between compactifications.

Thus, for a given a Lagrangian labeling $\{L_0, \dots, L_{d-1}\}$ and compatible asymptotics $\{x_1, \dots, x_d; y_{\text{out}}\}$ we obtain a moduli space of maps satisfying Floer's equation with the chosen boundary and asymptotics,

$$ {}_k\overline{\mathcal{R}}^1_{d,\tau_i}(y_{\text{out}}; \vec{x}) := {}_k\overline{\widehat{\mathcal{R}}}{}^1_d(y_{\text{out}}; x_{i-1}, \dots, x_1, x_d, \dots, x_i), \tag{5-180}$$

which is (for components of virtual dimension ≤ 1) a manifold of dimension equal to the virtual dimension of the right-hand side, namely

$$\deg(y_{\text{out}}) - n + d - \sum_{j=1}^{d} \deg(x_j) + 2k,$$

with $\mathbb{Z}$–gradings or mod 2 if working with $\mathbb{Z}/2$–gradings. The isomorphisms of orientation lines

$$({}_k\mathcal{R}^1_{d,\tau_i})_u \colon o_{x_d} \otimes \cdots \otimes o_{x_1} \to o_{y_{\text{out}}} \tag{5-181}$$

induced by elements u of the zero-dimensional components of (5-180) define the $|o_{y_{\mathrm{out}}}|_{\boldsymbol{k}}$ component of the operation $\hat{\check{\mathcal{OC}}}{}^k_{d,\tau_i}$, up to the sign twist

$$\hat{\check{\mathcal{OC}}}{}^k_{d,\tau_i}([x_d],\dots,[x_1]) := \sum_{u \in {}_k\bar{\check{\mathcal{R}}}{}^1_{d,\tau_i}(y_{\mathrm{out}};\vec{x}) \text{ rigid}} (-1)^{\hat{\star}_d} ({}_k\mathcal{R}^1_{d,\tau_i})_u([x_d],\dots,[x_1]), \tag{5-182}$$

where $\star_d^{S^1} = \sum_{i=1}^d (i+1)\cdot\deg(x_i) + \deg(x_d) + d - 1$.

Now, exactly as in Lemma 5.12, there is a chain-level equality of signed operations

$$\mathcal{OC}^{S^1,k}_d = \sum_{i=0}^{d-1} \hat{\check{\mathcal{OC}}}{}^k_{d,\tau_i}. \tag{5-183}$$

We recall the geometric statement underlying this: the point is that by construction there is an oriented embedding

$$\coprod_i {}_k\mathcal{R}^1_{d,\tau_i} \xrightarrow{\coprod_i \pi_f^i} \coprod_i {}_k\mathcal{R}_d^{S^1_{i,i+1}} \hookrightarrow {}_k\mathcal{R}_d^{S^1}, \tag{5-184}$$

compatible with Floer data, covering all but a codimension-one locus in the target, and moreover all the sign twists defining the operations $\mathcal{OC}^k_{d,\tau_i}$ are chosen to be compatible with the sign twist in the operation $\mathcal{OC}^{S^1,k}$ — this uses the fact that the Floer data on ${}_k\mathcal{R}^{S^1_{i,i+1}}$ agrees with the data on ${}_k\hat{\mathcal{R}}^1_d$ via the cyclic permutation map κ^{-i} by (5-55). After perturbation, zero-dimensional solutions to Floer's equation can be chosen to come from the complement of any codimension-one locus in the source abstract moduli space, implying the equality (5-183).

Finally, all that remains is a sign analysis, whose conclusion is that

$$\hat{\check{\mathcal{OC}}}{}^k_{d,\tau_i} = \hat{\check{\mathcal{OC}}}{}^k_d \circ s^{\mathrm{nu}} \circ t^i, \tag{5-185}$$

where s^{nu} is the operation arising from changing a check term to a hat term with a sign twist (3-20). (The equality up to comparing signs is immediate, as the operations are constructed with identical Floer data and hence involve counts of identical moduli spaces.) The details of this sign comparison are exactly the same as in Lemma 5.13, including with signs, since when orienting the moduli of maps, the additional marked points $p_1, \dots p_k$ only contribute complex orientations to the moduli spaces of domains (and no additional orientation line terms). □

Proof of Proposition 5.17 This is an immediate corollary of the previous two propositions. □

We now collect all of this information to finish the proof of our main result.

Proof of Theorem 1.1 The premorphism $\widetilde{\mathcal{OC}} \in \mathrm{Rhom}^n_{S^1}(\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{F}), CF^*(M))$, written u–linearly as $\sum_i \mathcal{OC}^k u^k$, where the $\mathcal{OC}^k = \check{\mathcal{OC}}^k \oplus \hat{\mathcal{OC}}^k$ are as constructed above, satisfies $\partial\widetilde{\mathcal{OC}} = 0$ by Proposition 5.17. Hence $\widetilde{\mathcal{OC}}$ is closed, or an S^1–complex homomorphism, also known as an A_∞ $C_{-*}(S^1)$–module homomorphism; see Section 2.1. Note that $[\mathcal{OC}^0] = [\mathcal{OC}] = [\check{\mathcal{OC}}]$, where the first equality holds by definition and the second holds by Corollary 5.4. Hence $\widetilde{\mathcal{OC}}$ is an enhancement of $\check{\mathcal{OC}}$, as defined in Section 2.1. □

Proof of Corollary 1.5 This is an immediate consequence of Theorem 1.1 and the induced homotopy-invariance properties for equivariant homology groups discussed in Section 2, particularly Corollary 2.18 and Proposition 2.19. □

5.6 Variants of the cyclic open–closed map

5.6.1 Using singular (pseudo)cycles instead of Morse cycles Let M be Liouville or compact and admissible (in which case by our convention $\overline{M} = M$ and $\partial\overline{M} = \varnothing$),[18] and let us consider the version of $\widetilde{\mathcal{OC}}$ with target the relative cohomology $H^*(\overline{M}, \partial\overline{M})$ as in Section 4.1.2. Instead of using a C^2–small Hamiltonian to define the Floer complex computing $H^{*+n}(\overline{M}, \partial\overline{M})$ (which we only did for simultaneous compatibility with the symplectic cohomology case), we can pass to a geometric cycle model for the group, and then build a version of the map $\widetilde{\mathcal{OC}}$ with such a target, which simplifies many of the constructions in the previous section, in the sense that the codimension-one boundary strata of moduli spaces, and hence the equations satisfied by $\widetilde{\mathcal{OC}}$, are strictly a subset of the terms appearing above. As such, it will be sufficient to fix some notation for the relevant moduli spaces, and state the relevant simplified results.

We let

$$ {}_k\check{\mathcal{P}}^1_d, \tag{5-186}$$

$$ {}_k\mathcal{P}^{S^1}_d, \tag{5-187}$$

$$ {}_k\hat{\mathcal{P}}^1_d \tag{5-188}$$

denote copies of the abstract moduli spaces (5-107)– (5-109), where the interior puncture z_{out} is filled in and replaced by a marked point $\bar{z}_{\mathrm{out}}$, *without any asymptotic marker*. The compactifications of these moduli spaces are exactly as before, except that the

[18] Technically we should write $\mathrm{QH}^*(M)$ in the latter case, but additively $\mathrm{QH}^*(M) = H^*(M)$, and correspondingly no sphere bubbling occurs in the moduli spaces we define here, so there is no difference for the purposes of this discussion.

auxiliary points $p_1, \dots, p_k$ are now allowed to coincide with $\overline{z}_{\mathrm{out}}$, without breaking off an angle-decorated cylinder or element of $\mathcal{M}_r$ (in the language of Section 4.3). In other words, the real blow-up of Deligne–Mumford compactifications at z_{out} described in Section A.1, which was responsible for the boundary strata containing $\mathcal{M}_r$ factors, *no longer occurs*, but all other degenerations do occur. Correspondingly the codimension-one boundaries of compactified moduli spaces have all of the strata as before except for strata containing the $\mathcal{M}_r$ factors.

Inductively choose smoothly varying families Floer data as before on these moduli spaces of domains, satisfying all of the requirements and consistency conditions as before, except for any consistency conditions involving $\mathcal{M}_r$ moduli spaces, which no longer occur on the boundary. For a basis $\beta_1, \dots, \beta_s$ of smooth (pseudo)cycles in homology $H_*(M)$ whose Poincaré duals $[\beta_i^\vee]$ generate the cohomology $H^*(\overline{M}, \partial\overline{M})$, one obtains moduli spaces

$$ {}_k\check{\mathcal{P}}^1_d(\beta_i; \vec{x}), \tag{5-189}$$

$$ {}_k\mathcal{P}^{S^1}_d(\beta_i, \vec{x}), \tag{5-190}$$

$$ {}_k\hat{\mathcal{P}}^1_d(\beta_i, \vec{x}) \tag{5-191}$$

of moduli spaces of maps into M with source an arbitrary element of the relevant domain moduli space, satisfying Floer's equation as before, with Lagrangian boundary and asymptotics $\vec{x}$ as before, *with the additional point constraint that* $\overline{z}_{\mathrm{out}}$ *lie on the cycle* β_i. As before, standard methods ensure that zero- and one-dimensional moduli spaces are (for generic choices of perturbation data and/or β_i) transversely cut-out manifolds of the "right" dimension and boundary, which is all that we need.

Then, define the coefficient of $[\beta_i^\vee] \in H^*(\overline{M}, \partial\overline{M})$ in $\check{\mathcal{OC}}^k(x_d \otimes \cdots \otimes x_1)$ to be given by signed counts (with the same sign twists as before) of the moduli spaces (5-189); similarly for $\hat{\mathcal{OC}}^k$ and $\mathcal{OC}^{S^1,k}$ using the moduli spaces (5-191) and (5-190). A simplification of the arguments already given (in which the δ_k operations no longer occur, but every other part of the argument carries through) implies:

Proposition 5.21 *The premorphism*

$$\widetilde{\mathcal{OC}} = \sum_{i=0}^{\infty} \mathcal{OC}^k u^k \in \mathrm{Rhom}^n_{S^1}(\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{F}, \mathcal{F}), H^*(\overline{M}, \partial\overline{M}))$$

satisfies

$$\widetilde{\mathcal{OC}} \circ b_{\mathrm{eq}} = 0, \tag{5-192}$$

where $b_{\mathrm{eq}} = b^{\mathrm{nu}} + uB^{\mathrm{nu}}$. In other words, $\widetilde{\mathcal{OC}}$ is a homomorphism of S^1–complexes between $\mathrm{CH}_*^{\mathrm{nu}}(\mathcal{F},\mathcal{F})$ with its strict S^1–action and $H^*(\overline{M},\partial\overline{M})$ with its trivial S^1–action.

As usual this model of $\widetilde{\mathcal{OC}}$ again induces maps $\widetilde{\mathcal{OC}}^{+/-/\infty}$ between homotopy orbit complexes, homotopy fixed-point complexes etc; note that the relevant equivariant homology chain complexes are particularly simple for the latter $H^*(\overline{M},\partial\overline{M})$, seeing as there is no differential and trivial circle action; for instance,

$$H^*(\overline{M},\partial\overline{M})_{\mathrm{h}S^1} = \bigl(H^*(\overline{M},\partial\overline{M})(\!(u)\!)/uH^*(\overline{M},\partial\overline{M})[\![u]\!], \delta_{\mathrm{eq}} = 0\bigr).$$

5.6.2 Compact Lagrangians in noncompact manifolds Now let us explicitly restrict to the case of M a Liouville manifold, and denote by $\mathcal{F}\subset\mathcal{W}$ the full subcategory consisting of a finite collection of compact exact Lagrangian branes contained in the compact region $\overline{M}$. By Poincaré duality we may think of the map $\mathcal{OC}$ (and its cyclic analogue, $\widetilde{\mathcal{OC}}$) with target $H^*(\overline{M},\partial\overline{M})$ as a pairing $\mathrm{CH}_*^{\mathrm{nu}}(\mathcal{F},\mathcal{F})\otimes C^*(M)\to \boldsymbol{k}[n]$. In this case, there is a nontrivial refinement of this pairing to

$$\mathcal{OC}_{\mathrm{cpct}}\colon \mathrm{CH}_*(\mathcal{F})\otimes SC^*(M)\to\boldsymbol{k}[-n], \tag{5-193}$$

where $SC^*(M)$ is the *symplectic cohomology* cochain complex.

Remark 5.22 The refinement (5-193) relies on extra flexibility in Floer theory for compact Lagrangians compared to noncompact Lagrangians (compare Remarks 3.16 and 3.17), first alluded to in this form in [54]. This extra flexibility allows us to define operations without outputs — and in particular study a version of the open–closed map where the interior marked point and boundary marked points are all inputs — for instance by Poincaré dually treating some boundary inputs as outputs with "negative weight".

One way to implement such operations, using the type of Floer data discussed in Remark 3.17, is by allowing the *subclosed one-form* α_S used in Floer-theoretic perturbations to have complete freedom along boundary conditions corresponding to compact Lagrangians; in contrast, along possibly noncompact Lagrangian boundary conditions, α_S is required to vanish in order to appeal to the integrated maximum principle. In particular, if we allow α_S to be nonvanishing along boundary components, Stokes' theorem no longer implies that α_S being subclosed implies that the total "output" weights must be greater than the total "input" weights.

Remark 5.23 The existence of a map $SC^*(M) \to \mathrm{CH}_*(\mathcal{F})^\vee[-n]$ is well known. Namely, categories $\mathcal{C}$ with a *weak proper Calabi–Yau structure*[19] of dimension n come equipped with isomorphisms between the dual of Hochschild chains and Hochschild cochains $\mathrm{CH}_*(\mathcal{C})^\vee[-n] \simeq \mathrm{CH}^*(\mathcal{C})$, and the existence of a map $SC^*(M) \to \mathrm{CH}^*(\mathcal{F})$ was observed in [49].

The geometric moduli spaces used to establish our main result apply verbatim in this case, with the interior marked point changed to an input, and the ordering of the auxiliary marked points $p_1, \dots, p_k$ appearing in the cyclic open–closed map reversed. In this case, the operations associated to such moduli spaces imply:

Proposition 5.24 *Consider* $\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{F}) \otimes SC^*(M)$ *as an* S^1*–complex with its diagonal* S^1*–action* (see Lemma 2.11 in Section 2.1), *and* $\boldsymbol{k} = \underline{\boldsymbol{k}}^{\mathrm{triv}} \in S^1$–mod *with its trivial* S^1*–complex structure. The map from* $\mathrm{CH}_*(\mathcal{F}) \otimes SC^*(M)$ *to* $\boldsymbol{k}$ *can be enhanced to a homomorphism of* S^1*–complexes*

$$\widetilde{\mathcal{OC}}_{\mathrm{cpct}} \in \mathrm{Rhom}^n_{S^1}(\mathrm{CH}^{\mathrm{nu}}_*(\mathcal{F}) \otimes SC^*(M), \boldsymbol{k}).$$

For example, $\widetilde{\mathcal{OC}}_{\mathrm{cpct}}$ *satisfies* $\partial \widetilde{\mathcal{OC}}_{\mathrm{cpct}} = 0$. *In other words, in the notation of Section 2.3, there exists a map*

$$\widetilde{\mathcal{OC}}_{\mathrm{cpct,eq}} = \sum_{i=0}^{\infty} \mathcal{OC}_{\mathrm{cpct},i} u^i : \mathrm{CH}^{\mathrm{nu}}_*(\mathcal{F}) \otimes SC^*(M) \to \boldsymbol{k}[[u]]$$

of pure degree n, *with* $[\mathcal{OC}_{\mathrm{cpct},0}] = [\mathcal{OC}_{\mathrm{cpct}}]$, *such that*

$$\widetilde{\mathcal{OC}}_{\mathrm{cpct,eq}} \circ \bigl((-1)^{\deg(y)} b_{\mathrm{eq}}(\sigma) \otimes y + \sigma \otimes \delta^{SC}_{\mathrm{eq}}(y)\bigr) = 0.$$

To clarify the relevant moduli spaces used, we define the spaces

$$ {}_k\check{\mathcal{R}}^1_{d,\mathrm{cpct}}, \tag{5-194}$$

$$ {}_k\mathcal{R}^{S^1}_{d,\mathrm{cpct}}, \tag{5-195}$$

$$ {}_k\hat{\mathcal{R}}^1_{d,\mathrm{cpct}}, \tag{5-196}$$

to be copies of the abstract moduli spaces (5-107)–(5-109) where the interior puncture z_{out} is now a *positive* puncture (still equipped with an asymptotic marker), and all of the other inputs and auxiliary points are as before, except we've reversed the order

[19] Such as the Fukaya category of compact Lagrangians; see eg [52, (12j); 53, Proof of Proposition 5.1, Step 1; 60, Section 2.8].

of the labelings $p_1, \dots, p_k$ (for notational convenience), so the ordering constraints all now read as $0 < |p_k| < \cdots < |p_1| < \frac{1}{2}$. The compactified moduli spaces have boundary strata agreeing with the boundary strata of the compactified (5-107)–(5-109), except now the $\mathcal{M}_r$ cylinders break "above" the ${}_{k-r}\mathcal{R}^1_{d,\mathrm{cpct}}$ (equipped with $\check{}$, $\hat{}$ or S^1 decoration) discs instead of "below". The reversal of the ordering of auxiliary marked points is designed to be compatible with the ordering of the auxiliary marked points on the $\mathcal{M}_r$ moduli spaces when it breaks "above" (as in $\mathcal{M}_r$, the label numbers of the auxiliary marked points increase from top to bottom).

Equipping these moduli spaces with perturbation data satisfying the same consistency conditions and other requirements as before, and counting solutions with sign twists as before, defines the terms of the premorphism exactly as in the previous subsections, with identical analysis to show that, for instance, the operation corresponding to ${}_k\mathcal{R}^{S^1}_{d,\mathrm{cpct}}$ is the operation corresponding to ${}_{k-1}\hat{\mathcal{R}}^1_{d,\mathrm{cpct}}$ composed with Connes' B operator, the boundary strata in which $|p_i|$ and $|p_{i+1}|$ are coincident contributes trivially, and so on.

6 Calabi–Yau structures

6.1 The proper Calabi–Yau structure on the Fukaya category

Here we review the notion of a *proper Calabi–Yau structure*, following Kontsevich and Soibelman [37], and construct proper Calabi–Yau structures on Fukaya categories of compact Lagrangians in a compact admissible or Liouville manifold. A proper Calabi–Yau structure induces chain-level topological-field-theoretic operations on the Hochschild chain complex of the given category, controlled by the open moduli space of curves with marked points equipped with asymptotic markers, at least one of which is an input [14; 37]. Note that Costello's work [14] constructing field-theoretic operations has the (a priori stronger) requirement that the underlying A_∞ category be *cyclic*, but in characteristic zero any proper Calabi–Yau structure determines a unique quasi-isomorphism between the underlying A_∞ category and a cyclic A_∞ category [37, Theorem 10.7]; see Remark 1.11 for more discussion.

We say an A_∞ category $\mathcal{A}$ is *proper* (sometimes called *compact*) if its cohomological morphism spaces $H^*(\hom_{\mathcal{A}}(X, Y))$ have total finite rank over $\boldsymbol{k}$ for each X, Y. Recall that for any object $X \in \mathcal{A}$, there is an inclusion of chain complexes

$$\hom(X, X) \to \mathrm{CH}_*(\mathcal{A}),$$

inducing a map

$$[i]\colon H^*(\hom(X,X)) \to \mathrm{HH}_*(\mathcal{A}).$$

Definition 6.1 Let $\mathcal{A}$ be a proper category. A chain map $\mathrm{tr}\colon \mathrm{CH}_{*+n}(\mathcal{A}) \to \boldsymbol{k}$ is called a *weak proper Calabi–Yau structure*, or *nondegenerate trace* of dimension n if, for any two objects $X, Y \in \operatorname{ob}\mathcal{A}$, the composition

$$H^*(\hom_{\mathcal{A}}(X,Y)) \otimes H^{n-*}(\hom_{\mathcal{A}}(Y,X)) \xrightarrow{[\mu^2_{\mathcal{A}}]} H^n(\hom_{\mathcal{A}}(Y,Y)) \xrightarrow{[i]} \mathrm{HH}_n(\mathcal{A}) \xrightarrow{[\mathrm{tr}]} \boldsymbol{k} \tag{6-1}$$

is a perfect pairing; this nondegeneracy property only depends on the homology class [tr]. A chain map from the nonunital Hochschild complex $\mathrm{tr}\colon \mathrm{CH}^{\mathrm{nu}}_{*+n}(\mathcal{A}) \to \boldsymbol{k}$ is called a weak proper Calabi–Yau structure if composition with the inclusion $\mathrm{CH}_{*+n}(\mathcal{A}) \subset \mathrm{CH}^{\mathrm{nu}}_{*+n}(\mathcal{A})$ is a weak proper Calabi–Yau structure in the sense above.

Remark 6.2 In the symplectic literature, weak proper Calabi–Yau structures of dimension n are sometimes defined as bimodule quasi-isomorphisms $\mathcal{A}_\Delta \xrightarrow{\sim} \mathcal{A}^\vee[n]$, where $\mathcal{A}_\Delta$ denotes the *diagonal bimodule* and $\mathcal{A}^\vee$ the *linear dual diagonal bimodule*; see [52, (12j)] and Section 6.2 for brief conventions on A_∞–bimodules, see also [62]. To explain the relationship between this definition and the one above, which has sometimes been called a *weakly cyclic structure* or ∞*–inner product* [62; 60], note that for any compact A_∞ category $\mathcal{A}$, there are quasi-isomorphisms (with explicit chain-level models)

$$(\mathrm{CH}_*(\mathcal{A}))^\vee = \mathrm{CH}^*(\mathcal{A}, \mathcal{A}^\vee) \xleftarrow{\sim} \hom_{\mathcal{A}\text{–}\mathcal{A}}(\mathcal{A}_\Delta, \mathcal{A}^\vee), \tag{6-2}$$

where $\hom_{\mathcal{A}\text{–}\mathcal{A}}$ denotes morphisms in the category of A_∞–bimodules; see eg [50] or [24]. Under this correspondence, nondegenerate morphisms from $\mathrm{HH}_*(\mathcal{A}) \to \boldsymbol{k}$ as defined above correspond precisely (cohomologically) to weak Calabi–Yau structures, for instance, those bimodule morphisms from $\mathcal{A}_\Delta$ to $\mathcal{A}^\vee$ which are cohomology isomorphisms.

Remember that the Hochschild chain complex of an A_∞ category $\mathcal{A}$ comes equipped with a natural chain map to the (positive) cyclic homology chain complex, the *projection to homotopy orbits* (2-22),

$$\mathrm{pr}\colon \mathrm{CH}^{\mathrm{nu}}_*(\mathcal{A}) \to \mathrm{CC}^+_*(\mathcal{A}),$$

modeled on the chain level by the map that sends $\alpha \mapsto \alpha \cdot u^0$ for $\alpha \in \mathrm{CH}^{\mathrm{nu}}(\mathcal{A})$; compare with (2-35).

Definition 6.3 (cf Kontsevich and Soibelman [37]) A (*strong*) *proper Calabi–Yau structure of degree n* is a chain map

$$\widetilde{\mathrm{tr}}\colon \mathrm{CC}_*^+(\mathcal{A}) \to \boldsymbol{k}[-n] \tag{6-3}$$

from the (positive) cyclic homology chain complex of $\mathcal{A}$ to $\boldsymbol{k}$ of degree $-n$, such that the induced map $\mathrm{tr} = \widetilde{\mathrm{tr}} \circ \mathrm{pr}\colon \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{A}) \to \boldsymbol{k}[-n]$ — or equivalently the composition $\check{\mathrm{tr}}$ of tr with the inclusion $\mathrm{CH}_*(\mathcal{A}) \subset \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{A})$ — is a weak proper Calabi–Yau structure.

Via the model for cyclic chains given as

$$\mathrm{CC}_*^+(\mathcal{A}) := (\mathrm{CH}_*^{\mathrm{nu}}(\mathcal{A})(\!(u)\!)/u\mathrm{CH}_*^{\mathrm{nu}}(\mathcal{A})[\![u]\!], b + uB^{\mathrm{nu}}),$$

such an element $\widetilde{\mathrm{tr}}$ takes the form

$$\widetilde{\mathrm{tr}} := \sum_{i=0}^{\infty} \mathrm{tr}^k\, u^k, \tag{6-4}$$

where

$$\mathrm{tr}^k := (\check{\mathrm{tr}}^k \oplus \hat{\mathrm{tr}}^k)\colon \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{A}) \to \boldsymbol{k}[-n-2k]. \tag{6-5}$$

We now complete the proof of Theorem 1.12 described and sketched in Section 1: first, define the putative proper Calabi–Yau structure as the composition

$$\widetilde{\mathrm{tr}}\colon \mathrm{CC}_*^+(\mathcal{F}) \xrightarrow{\widetilde{\mathcal{OC}}^+} C^{*+n}(\overline{M}, \partial\overline{M}) \otimes_{\boldsymbol{k}} \boldsymbol{k}(\!(u)\!)/u\boldsymbol{k}[\![u]\!] \to \boldsymbol{k}, \tag{6-6}$$

where the last map (cohomologically) sends $\mathrm{PD}(\mathrm{pt})\cdot u^0 \in H^{2n}(\overline{M}, \partial\overline{M})$ to 1, and other elements to 0; ie it projects to the u^0 factor then integrates over $[M]$. Instead of using a C^2–small Hamiltonian to define the Floer complex computing $H^{*+n}(\overline{M}, \partial\overline{M})$, which we only did for simultaneous compatibility with the symplectic cohomology case, we can pass to a geometric cycle model for $\widetilde{\mathcal{OC}}^+$ (and therefore $\widetilde{\mathrm{tr}}$), which as described in Section 5.6.1 directly maps (on the chain level) to

$$H^{*+n}(\overline{M}, \partial\overline{M}) \otimes_{\boldsymbol{k}} \boldsymbol{k}(\!(u)\!)/u\boldsymbol{k}[\![u]\!].$$

With respect to this model, the map $\widetilde{\mathrm{tr}}$ involves counts of the moduli spaces described there, where the interior marked point $\bar{z}_{\mathrm{out}}$ is *unconstrained*, eg ${}_k\check{\mathcal{P}}_d^1([M]; \vec{x})$, ${}_k\hat{\mathcal{P}}_d^1([M]; \vec{x})$ and ${}_k\mathcal{P}_d^{S^1}([M]; \vec{x})$; see Figure 16.

The following well-known lemma verifies the nondegeneracy property of the map $\widetilde{\mathrm{tr}}$.

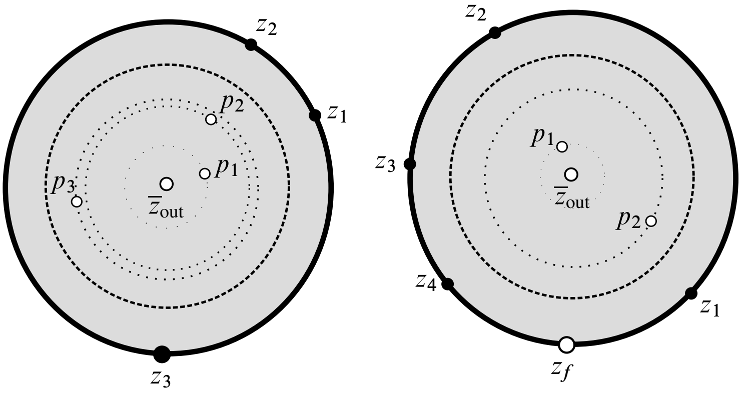


Figure 16: An image of representatives of moduli spaces ${}_3\check{\mathcal{P}}^1_3([M];\vec{x})$ and ${}_2\hat{\mathcal{P}}^1_4([M];\vec{x})$, which appear in the map $\widetilde{\mathrm{tr}}$.

Lemma 6.4 [52, (12j); 60, Lemma 2.4] *The corresponding morphism*

$$[\mathrm{tr}]\colon \mathrm{HH}_{*+n}(\mathcal{F}) \to \boldsymbol{k}$$

is a nondegenerate trace (*or weak proper Calabi–Yau structure*).

Sketch of proof This is an immediate consequence of Poincaré duality in Lagrangian Floer cohomology; see the cited references. As a brief sketch, note that $\check{\mathrm{tr}}^0 \circ \mu^2 = \check{\mathrm{tr}} \circ \mu^2\colon \hom(X,Y) \otimes \hom(Y,X) \to \boldsymbol{k}$ is chain homotopic (and hence equal in cohomology) to a chain map which counts holomorphic discs with an interior marked point satisfying an empty constraint, and two (positive) boundary asymptotics on p and q, with corresponding Lagrangian boundary on x and y. Via a further homotopy of Floer data, one can arrange that the generators of $\hom(X,Y)$ and $\hom(Y,X)$ are in bijection (for instance if one is built out of time-1 flowlines of H and one out of time-1 flowlines of $-H$), and the only such rigid discs are constant discs between p and the corresponding $p^\vee$. □

Proof of Theorem 1.12 The above discussion constructs $\widetilde{\mathrm{tr}}$ and Lemma 6.4 verifies nondegeneracy. □

6.2 The smooth Calabi–Yau structure on the Fukaya category

We give an overview of (a categorical version of) the notion of a (*strong*) *smooth Calabi–Yau structure*, and construct such smooth Calabi–Yau structures on wrapped or compact Fukaya categories under the "nondegeneracy" hypotheses of [24]. Smooth

Calabi-Yau structures were proposed by Kontsevich and Vlassopoulous [36] and later comprehensively studied by Kontsevich, Takeda and Vlassopoulous [39]. Other expositions appear, for instance, in [27] and [7]; in the latter work the terminology "left" is used instead of "smooth", and "right" instead of "proper". A smooth Calabi–Yau structure (analogously to the proper case) induces chain-level topological field theory operations on the Hochschild chain complex of the given category, controlled by the open moduli space of curves with marked points equipped with asymptotic markers, at least one of which is an output [39; 38].[20]

To state the relevant definitions, we make use of some of the theory of *A_∞–bimodules over a category $\mathcal{C}$*. We do so without much explanation, instead referring readers to existing references [50; 62; 24]. An A_∞–bimodule $\mathcal{P}$ over $\mathcal{C}$ is a bilinear A_∞ functor from $\mathcal{C}^{\mathrm{op}} \times \mathcal{C}$ to chain complexes, which is roughly the data of a chain complex $(\mathcal{P}(X,Y), \mu^{0|1|0})$ for every pair of objects $\mathcal{C}$, along with "higher multiplication maps"

$$\begin{aligned}\mu^{s|1|t}\colon \hom_{\mathcal{C}}(X_{s-1},X_s)\otimes\cdots\otimes\hom_{\mathcal{C}}(X_0,X_1)\otimes\mathcal{P}(X_0,Y_t)\\ \otimes\hom_{\mathcal{C}}(Y_{t-1},Y_t)\otimes\cdots\otimes\hom_{\mathcal{C}}(Y_0,Y_1)\to\mathcal{P}(X_s,Y_0)\end{aligned}$$

satisfying a generalization of the A_∞ equations. A_∞–bimodules over $\mathcal{C}$ form a dg category $\mathcal{C}$–mod–$\mathcal{C}$, with morphisms denoted by $\hom^*_{\mathcal{C}\text{–}\mathcal{C}}(\mathcal{P},\mathcal{Q})$. (For dg bimodules over a dg category, this chain complex corresponds to a particular chain model for the "derived morphism space" using the bar resolution.) The basic examples of bimodules we require are:

- The *diagonal bimodule* $\mathcal{C}_\Delta$, which associates to a pair of objects (K,L) the chain complex $\mathcal{C}_\Delta(K,L) := \hom_{\mathcal{C}}(L,K)$.
- For any pair of objects A, B, there is a *Yoneda bimodule* $\mathcal{Y}^l_A \otimes_{\boldsymbol{k}} \mathcal{Y}^r_B$, which associates to a pair of objects (K,L) the chain complex $\mathcal{Y}^l_A \otimes_{\boldsymbol{k}} \mathcal{Y}^r_B(K,L) := \hom_{\mathcal{C}}(A,K)\otimes\hom_{\mathcal{C}}(L,B)$.

Yoneda bimodules are the analogues of the free bimodule $A\otimes A^{\mathrm{op}}$ in the category of bimodules over an associative algebra A (which are the same as $A\otimes A^{\mathrm{op}}$–modules). Accordingly, we say a bimodule $\mathcal{P}$ is *perfect* if, in the category $\mathcal{C}$–mod–$\mathcal{C}$, it is split-generated by (ie isomorphic to a retract of a finite complex of) Yoneda bimodules. We say that a category $\mathcal{C}$ is (*homologically*) *smooth* if $\mathcal{C}_\Delta$ is a perfect $\mathcal{C}$–bimodule.

Recall for what follows that for any bimodule $\mathcal{P}$ there is a *cap product action*

$$\cap\colon \mathrm{HH}^*(\mathcal{C},\mathcal{P})\otimes\mathrm{HH}_*(\mathcal{C},\mathcal{C})\to\mathrm{HH}_*(\mathcal{C},\mathcal{P}), \tag{6-7}$$

[20] In contrast, note that in the proper case all operations should have at least one *input*.

and hence for any class $[\sigma] \in \mathrm{HH}_*(\mathcal{C},\mathcal{C})$ there is an induced map

$$[\cap\sigma]\colon \mathrm{HH}^*(\mathcal{C},\mathcal{P}) \to \mathrm{HH}_{*+\deg(\sigma)}(\mathcal{C},\mathcal{P}). \tag{6-8}$$

More generally, the cap products acts as $\mathrm{HH}^*(\mathcal{C},\mathcal{P}) \otimes \mathrm{HH}_*(\mathcal{C},\mathcal{Q}) \to \mathrm{HH}_*(\mathcal{C},\mathcal{P}\otimes_{\mathcal{C}}\mathcal{Q})$; here we are considering $\mathcal{Q}=\mathcal{C}_\Delta$, and then composing with the equivalence $\mathcal{P}\otimes_{\mathcal{C}}\mathcal{C}_\Delta \cong \mathcal{P}$. See for instance [24, Section 2.10] for explicit chain-level formulae in the variant case that $\mathcal{P}=\mathcal{C}_\Delta$, which can be straightforwardly adapted to the general case and then specialized to the case here.

Definition 6.5 Let $\mathcal{C}$ be a homologically smooth A_∞ category. A cycle $\sigma \in \mathrm{CH}_{-n}(\mathcal{C},\mathcal{C})$ is said to be a *weak smooth Calabi–Yau structure*, or a *nondegenerate cotrace* if, for any objects K, L, the operation of capping with σ induces a homological isomorphism

$$[\cap\sigma]\colon \mathrm{HH}^*(\mathcal{C},\mathcal{Y}^l_K\otimes_{\boldsymbol{k}}\mathcal{Y}^r_L) \xrightarrow{\cong} \mathrm{HH}_{*-n}(\mathcal{C},\mathcal{Y}^l_K\otimes_{\boldsymbol{k}}\mathcal{Y}^r_L) \simeq H^*(\hom_{\mathcal{C}}(K,L)). \tag{6-9}$$

(This nondegeneracy property only depends on the homology class $[\sigma]$.) A cycle in the nonunital Hochschild complex $\sigma \in \mathrm{CH}^{\mathrm{nu}}_{-n}(\mathcal{C})$ is said to be a weak smooth Calabi–Yau structure if again $[\sigma] \in H^*(\mathrm{CH}^{\mathrm{nu}}_{-n}(\mathcal{C})) \cong \mathrm{HH}_{-n}(\mathcal{C})$ is nondegenerate in the sense of (6-9).

Remark 6.6 The second isomorphism $\mathrm{HH}_{*-n}(\mathcal{C},\mathcal{Y}^l_K\otimes_{\boldsymbol{k}}\mathcal{Y}^r_L) \simeq H^*(\hom_{\mathcal{C}}(K,L))$ always holds for cohomologically unital categories, such as the Fukaya category; the content is in the first.

Remark 6.7 Continuing Remark 6.2, there is an alternative perspective on Definition 6.5 using bimodules. Namely, for any bimodule $\mathcal{P}$, there is a naturally associated *bimodule dual* $\mathcal{P}^!$, defined for a pair of objects (K,L) as the chain complex

$$\mathcal{P}^!(K,L) := \hom^*_{\mathcal{C}\text{-}\mathcal{C}}(\mathcal{P},\mathcal{Y}^l_K\otimes_{\boldsymbol{k}}\mathcal{Y}^r_L).$$

The higher bimodule structure is defined in [24, Definition 2.40]; for an A–bimodule B, it is an A_∞ analogue of defining $B^! := \mathrm{RHom}_{A\otimes A^{\mathrm{op}}}(B, A\otimes A^{\mathrm{op}})$, where RHom is taken with respect to the outer bimodule structure on $A\otimes A^{\mathrm{op}}$, and the bimodule structure on $B^!$ comes from the inner bimodule structure; see eg [28, Section 20.5].

We abbreviate $\mathcal{C}^! := \mathcal{C}^!_\Delta$ and call $\mathcal{C}^!$ the *inverse dualizing bimodule*, following [37]. (Observe that $H^*(\mathcal{C}^!(K,L)) \cong \mathrm{HH}^*(\mathcal{C},\mathcal{Y}^l_K\otimes_{\boldsymbol{k}}\mathcal{Y}^r_L)$.) For a homologically smooth category $\mathcal{C}$, one notes that there is a quasi-isomorphism $\mathrm{CH}_{*-n}(\mathcal{C}) \simeq \hom^*_{\mathcal{C}\text{-}\mathcal{C}}(\mathcal{C}^!_\Delta[n],\mathcal{C}_\Delta)$ (see [37, Remark 8.11] for the case of A_∞–algebras), where the equivalence associates to any element the bimodule morphism whose cohomology-level map is the cap product operation (6-9). Nondegenerate cotraces in $\mathrm{CH}_{-n}(\mathcal{C})$ then correspond precisely to bimodule quasi-isomorphisms $\mathcal{C}^![n] \xrightarrow{\sim} \mathcal{C}_\Delta$. Further discussion of these structures in the A_∞ categorical setting will appear as part of forthcoming work with Cohen [12].

Let $\iota\colon \mathrm{CC}_*^-(\mathcal{C}) \to \mathrm{CH}_*^{\mathrm{nu}}(\mathcal{C})$ denote the "inclusion of homotopy fixed points" chain map from (2-24); concretely, as described in (2-36), this is the chain map sending $\sum_{i=0}^\infty \alpha_i u^i \mapsto \alpha_0$.

Definition 6.8 Let $\mathcal{C}$ be a homologically smooth A_∞ category. A (*strong*) *smooth Calabi–Yau structure* is a cycle $\tilde{\sigma} \in \mathrm{CC}_{-n}^-(\mathcal{C})$ such that the corresponding element $\iota(\tilde{\sigma}) \in \mathrm{CH}_{-n}^{\mathrm{nu}}(\mathcal{C})$ is a weak smooth Calabi–Yau structure.

Using these definitions and the cyclic open–closed map, we will now restate and prove Theorem 1.16. We adopt the notation of wrapped Fukaya categories in the below result, using $\mathcal{W}$ and $SC^*(M)$ in place of $\mathcal{F}$ and $CF^*(M)$, with the understanding that for a compact symplectic manifold, these are the same.

Theorem 6.9 (Theorem 1.16 above) *Suppose a Liouville* (*or compact admissible symplectic*) *manifold is* ***nondegenerate*** *in the sense of* [24], *meaning that the map* $[\mathcal{OC}]\colon \mathrm{HH}_{*-n}(\mathcal{W}) \to SH^*(M)$ *hits* 1. *Then the Fukaya category* $\mathcal{W}$ *possesses a* (*cohomologically*) *canonical geometrically defined* ***smooth Calabi–Yau structure****.*

Proof In [24] it was proven, assuming nondegeneracy of M, that the map

$$[\mathcal{OC}]\colon \mathrm{HH}_{*-n}(\mathcal{W}) \to SH^*(M)$$

is an isomorphism, $\mathcal{W}$ is homologically smooth, and moreover that the preimage $[\sigma]$ of 1 gives a *weak smooth Calabi–Yau structure* in the sense described above; see [25; 27; 26] for a proof of some of these facts specifically tailored to the case of compact Lagrangians in compact symplectic manifolds. Let us briefly recall how the nondegeneracy condition (6-9) is proven (a fact which is left slightly implicit in [24]). First, a geometric morphism of bimodules $\mathcal{CY}\colon \mathcal{W}_\Delta \to \mathcal{W}^![n]$ is constructed and shown in [24, Theorem 1.3] to be a quasi-isomorphism under the given nondegeneracy hypotheses. Then, it is shown that capping with $[\sigma]$ is a one-sided inverse to the homological map $[\mathcal{CY}]$, and thus an isomorphism also, by the following argument. We establish that the following diagram is commutative (up to an overall sign of $(-1)^{n(n+1)/2}$); it can be thought of as coming from the compatibility of $\mathcal{OC}$ with module structures for Hochschild (co)homology with coefficients in $\mathcal{Y}^l_K \otimes_{\boldsymbol{k}} \mathcal{Y}^r_L$, and can be extracted from the holomorphic curve theory appearing in [24, Theorem 13.1]:

(6-10)
$$\begin{tikzcd}
\begin{array}{c} \mathrm{HH}_{*-n}(\mathcal{W},\mathcal{W}) \\ \otimes \\ H^*(\hom_{\mathcal{W}}(K,L)) \end{array} \arrow[r, "{(\mathrm{id},[\mathcal{CY}])}"] \arrow[d, "{([\mathcal{OC}],\mathrm{id})}"] & \begin{array}{c} \mathrm{HH}_{*-n}(\mathcal{W},\mathcal{W}) \\ \otimes \\ \mathrm{HH}^{*+n}(\mathcal{W},\mathcal{Y}^l_K \otimes_{\boldsymbol{k}} \mathcal{Y}^r_L) \end{array} \arrow[d, "\cap"] \\
\begin{array}{c} SH^*(M) \\ \otimes \\ H^*(\hom_{\mathcal{W}}(K,L)) \end{array} \arrow[r, "{[\mu^2(\mathcal{CO}_0(-),-)]}"] & \begin{array}{l} H^*(\hom_{\mathcal{W}}(K,L)) \\ \quad = \mathrm{HH}_*(\mathcal{W},\mathcal{Y}^l_K \otimes_{\boldsymbol{k}} \mathcal{Y}^r_L) \end{array}
\end{tikzcd}$$

Here $[\mathcal{CO}_0]$ is the length-zero part of the closed open map for the object L, mapping $SH^*(M)$ to $H^*(\hom_{\mathcal{W}}(L,L))$. Plugging $[\sigma]$ into $\mathrm{HH}_*(\mathcal{W},\mathcal{W})$ and noting that $[\mathcal{OC}]([\sigma]) = 1$ and $[\mu^2(\mathcal{CO}_0(1),-)] = [\mu^2]([e_L],-)$ is the identity map establishes, as desired, that $[\sigma \cap (\mathcal{CY}(y))] = [y]$.

To lift the weak smooth Calabi–Yau structure to a (strong) smooth Calabi–Yau structure, first we note that, because $[\mathcal{OC}]$ is an isomorphism, Corollary 1.5 implies that there is a commutative diagram of isomorphisms

(6-11)
$$\begin{tikzcd}
\mathrm{HC}^-_{*-n}(\mathcal{W}) \arrow[r, "{[\iota]}"] \arrow[d, "{[\widetilde{\mathcal{OC}}^-]}"'] & \mathrm{HH}_{*-n}(\mathcal{W}) \arrow[d, "{[\mathcal{OC}]}"] \\
H^*(SC^*(M)^{\mathrm{h}S^1}) \arrow[r, "{[\iota]}"] & SH^*(M)
\end{tikzcd}$$

where the horizontal maps ι are the "inclusion of homotopy fixed points" maps $\iota\colon P^{\mathrm{h}S^1} \to P$ defined for any S^1–complex P, sending $\sum_{i=0}^{\infty} \alpha_i u^i \mapsto \alpha_0$.

In Section 4.4, and specifically (4-68), it was shown that there is a canonical geometrically defined element $\tilde{1} \in H^*(SC^*(M)^{\mathrm{h}S^1})$ lifting the unit $1 \in SH^*(M)$; essentially this is because the map 1 is in the image of the map $H^*(M) \to SH^*(M)$, which on the chain level (as this map comes from "the inclusion of constant loops into the free loop space" and "constant loops are acted on by S^1 trivially") can be canonically lifted to a map $C^*(M) \to C^*(M)^{\mathrm{h}S^1} = C^*(M)[\![u]\!] \to SC^*(M)^{\mathrm{h}S^1}$.

Since $[\widetilde{\mathcal{OC}}^-]$ is an isomorphism, it follows that there is a unique (cohomological) element $[\tilde{\sigma}] \in \mathrm{HC}^-_{*-n}(\mathcal{W})$ hitting $\tilde{1}$ via $[\widetilde{\mathcal{OC}}]$. By (6-11), $[\iota]([\tilde{\sigma}])] = [\sigma]$, establishing that (any cycle representing) $[\tilde{\sigma}]$ is a smooth Calabi–Yau structure. □

Appendix Moduli spaces and operations

A.1 A real blow-up of Deligne–Mumford space

We review, in a special case, the compactifications of moduli spaces of surfaces where some interior marked points are equipped with asymptotic markers, which are a real blow-up of Deligne–Mumford moduli space as constructed in [34]. In particular, we show how boundary strata of the abstract compactifications in the sense of [34] can be identified with the specific models of the moduli spaces we use in Section 5. The appearance of the compactifications [34] in Floer theory is not new; see eg [58].

To begin, let

$$\mathcal{M}_{2,0} \tag{A-1}$$

denote the space of spheres with two marked points z_1, z_2 removed and asymptotic markers τ_1, τ_2 around the z_1 and z_2, modulo automorphism. Fixing the position of z_1 and z_2 and one of τ_1 or τ_2 gives a diffeomorphism

$$\mathcal{M}_{2,0} \cong S^1.$$

On an arbitrary representative in $\mathcal{M}_{2,0}$, we can think of the map to S^1 as coming from the *difference in angles* between τ_1 and τ_2 — after, say, parallel transporting one tangent space to the other along a geodesic path.

It is convenient to parametrize this difference by a point on the sphere itself, in the following manner (though this will break symmetry between z_1 and z_2). Let

$$\mathcal{M}_{2,1} \tag{A-2}$$

be the space of spheres with two marked points z_1, z_2 removed, an extra marked point p, and asymptotic markers τ_1, τ_2 around the z_1 and z_2, modulo automorphism, such that, for any representative with position of z_1, z_2 and p fixed, τ_2 is pointing towards p. The remaining freedom in τ_1 once more gives a diffeomorphism $\mathcal{M}_{2,1} \cong S^1$.

We can take a different representative for elements of $\mathcal{M}_{2,1}$: up to biholomorphism any element of (A-2) is equal to a cylinder sending z_1 to $+\infty$, z_2 to $-\infty$, with fixed asymptotic direction around $+\infty$ and an extra marked point p at fixed height freely varying around S^1, such that the asymptotic marker at $-\infty$ coincides with the S^1 coordinate of p. Thus, we obtain an identification

$$\mathcal{M}_{2,1} \cong \mathcal{M}_1, \tag{A-3}$$

where $\mathcal{M}_1$ is the space in Definition 4.7, with p_1 corresponding to p here.

Now let

$$ {}_k\mathcal{R}^1_d \tag{A-4}$$

denote the moduli space of discs $(S, z_1, \dots, z_d, z_{\mathrm{out}}, \tau_{z_{\mathrm{out}}}, p_1, \dots, p_k)$ with d boundary marked points $z_1, \dots, z_d$ arranged in counterclockwise order, an interior marked point with asymptotic marker $(z_{\mathrm{out}}, \tau_{z_{\mathrm{out}}})$, and interior marked points with no asymptotic markers $p_1, \dots, p_k$ satisfying two constraints to be described below, modulo automorphism. Up to automorphism, every equivalence class of the unconstrained moduli space of such $(S, z_1, \dots, z_d, z_{\mathrm{out}}, \tau_{z_{\mathrm{out}}}, p_1, \dots, p_k)$ admits a unique unit-disc representative with z_d fixed at 1 and z_{out} at 0; call this the (z_d, z_{out}) *standard representative*, or simply the standard representative. The positions of the asymptotic marker, remaining marked points, and interior marked points identify this unconstrained moduli space with an open subset of $S^1 \times \mathbb{R}^{2k} \times \mathbb{R}^d$. With respect to this identification, the space (A-4) consists of those discs satisfying the (open) "ordering constraint" on the positions of the interior marked points

$$\text{on the standard representative, } 0 < |p_1| < |p_2| < \cdots < |p_k| < \tfrac{1}{2}, \tag{A-5}$$

along with a (codimension-one) condition on the asymptotic marker,

$$\text{on the standard representative, } \tau_{z_{\mathrm{out}}} \text{ points at } p_1. \tag{A-6}$$

The condition (A-5), which cuts out a manifold with corners of the larger space in which the p_i are unconstrained, is technically convenient, as it reduces the types of bubbles that can occur with z_{out}. The compactification of interest, denoted by

$$ {}_k\overline{\mathcal{R}}^1_d, \tag{A-7}$$

differs from the Deligne–Mumford compactification in a couple of respects: firstly, we allow points p_i and p_{i+1} to be coincident without bubbling off (alternatively, we can Deligne–Mumford compactify and collapse the relevant strata). More interestingly, (A-7) is a real blow-up of the usual Deligne–Mumford compactification along any strata in which z_{out} and p_i points bubble off, as in [34]. We will proceed to describe the codimension-one boundary strata of (A-7) along with (after identification with the moduli spaces we introduce in this paper) the boundary chart gluing maps. Let $\Sigma = S_0 \cup_{z^+_{\mathrm{int}} = z^-_{\mathrm{int}}} S_1$ denote a nodal surface, where

- S_0 is a sphere containing interior marked points $(z_{\mathrm{out}}, \tau_{z_{\mathrm{out}}})$, $p_1, \dots, p_j$ and another marked point z^+_{int}, and

- S_1 is a disc with d boundary marked points $z_1, \dots, z_d$ and interior marked points z_{int}^-, $p_{j+1}, \dots, p_k$.

To occur as a possible degenerate limit of (A-4), the relevant points p_i on S_0 and S_1 must satisfy an ordering condition:

(A-8) For any S_0' which is biholomorphic to S_0, with z_{out} and z_{int}^+ at opposite poles, we have $0 < |p_1| < \cdots < |p_j| < |z_{\mathrm{int}}^+|$, where $|p|$ denotes the geodesic distance from z_{out} to p on S_0'.

(A-9) For the $(z_d, z_{\mathrm{int}}^-)$ standard representative of S_1, $0 < |p_{j+1}| < \cdots < |p_k| < \frac{1}{2}$.

Also:

(A-10) For S_0' as in (A-8), the asymptotic marker $\tau_{z_{\mathrm{out}}}$ should point (geodesically) towards p_1.

The relevant codimension-one stratum of (A-7) consists of all (automorphism classes of) such broken configurations $S_0 \cup_{z_{\mathrm{int}}^+ = z_{\mathrm{int}}^-} S_1$ as above, equipped additionally with a *gluing angle* at the node, which is a real positive line $\tau_{z_{\mathrm{int}}^+, z_{\mathrm{int}}^-}$ in $T_{z_{\mathrm{int}}^+} S_0 \otimes T_{z_{\mathrm{int}}^-} S_1$, or equivalently, a pair of asymptotic markers $(\tau_{z_{\mathrm{int}}^+}, \tau_{z_{\mathrm{int}}^-})$ around each of z_{int}^+ and z_{int}^-, modulo the diagonal S^1 rotation action. Note that the set of gluing angles (which is allowed to vary) is S^1, making this stratum codimension-1 (the corresponding stratum in Deligne–Mumford space does not have gluing angles, and hence has real codimension two). The gluing map takes, for a fixed pair of cylindrical ends around z_{int}^+ and z_{int}^- compatible with the pair of asymptotic markers in the sense of (5-3), the usual gluing with respect to the chosen cylindrical ends. Note first that for a given gluing parameter, if the cylindrical ends are chosen to simply rotate as $(\tau_{z_{\mathrm{int}}^+}, \tau_{z_{\mathrm{int}}^-})$ vary, the result of gluing after rotating $\tau_{z_{\mathrm{int}}^+}$ by θ_1 and $\tau_{z_{\mathrm{int}}^-}$ by θ_2 differs from the initial gluing by a rotation of the bottom component by $\theta_2 - \theta_1$. In particular, the glued surface indeed only depends on the gluing angle associated to $(\tau_{z_{\mathrm{int}}^+}, \tau_{z_{\mathrm{int}}^-})$, ie it is unchanged by simultaneously rotating $(\tau_{z_{\mathrm{int}}^+}, \tau_{z_{\mathrm{int}}^-})$.

We can recast this stratum by taking a slice of the quotient by the diagonal S^1–action appearing in the definition of gluing angle: First, note that z_{int}^- on S_1 possesses a canonical asymptotic marker $(\tau_{z_{\mathrm{int}}}^-)_{\mathrm{canon}}$, which (on the standard representative) points towards p_{j+1}; our convention is that $p_{s+1} = z_d$, so $\tau_{z_{\mathrm{int}}}^-$ points at z_d if $j = s$. Choosing the representative $(\tau_{z_{\mathrm{int}}^+}, \tau_{z_{\mathrm{int}}^-})$ of each gluing angle for which $\tau_{z_{\mathrm{int}}^-}$ is the canonical asymptotic marker $(\tau_{z_{\mathrm{int}}}^-)_{\mathrm{canon}}$, we see that the stratum described above can be identified with the space of broken configurations $S_0 \cup_{z_{\mathrm{int}}^+ = z_{\mathrm{int}}^-} S_1$ (up to automorphism) of the form:

- S_1 is as above (ie satisfies (A-9)) but is additionally equipped with $(\tau_{z_{\mathrm{int}}^-})_{\mathrm{canon}}$, ie $S_1 \in {}_{k-j}\mathcal{R}_1^d$.
- S_0 is equipped with interior marked points with asymptotic markers $(z_{\mathrm{out}}, \tau_{z_{\mathrm{out}}})$, $(z_{\mathrm{int}}^+, \tau_{z_{\mathrm{int}}^+})$ and additional marked points $p_1, \dots, p_j$ satisfying (A-8) and (A-10).

Just as in (A-3), the space of such S_0 up to biholomorphism is precisely $\mathcal{M}_j$ as in Definition 4.7, ie given any S_0, there is a one-dimensional space of biholomorphisms to a cylinder sending z_{int} and z_{out} to ∞ and $-\infty$ while fixing the angle of $\tau_{z_{\mathrm{int}}^+}$ to 1; any two such biholomorphisms differ by translation.

Thus, we have identified this stratum with

$$ {}_{k-j}\mathcal{R}_1^d \times \mathcal{M}_j, \tag{A-11}$$

which will be useful in defining the relevant pseudoholomorphic curve counts. From this perspective, the boundary chart gluing maps, defined with respect to the cylindrical ends (4-33) and (4-34) on $\mathcal{M}_j$ and with respect to a smoothly varying choice of cylindrical end over elements of ${}_{k-j}\mathcal{R}_1^d$ compatible with $(\tau_{z_{\mathrm{int}}^-})_{\mathrm{canon}}$, just as in (4-36), rotate the (standard representative of the) angle-decorated cylinder S_0 to match the angle of its top asymptotic marker with the angle of $(\tau_{z_{\mathrm{int}}^-})_{\mathrm{canon}}$, which coincides with the argument of p_{j+1} on the standard representative. In other words, if we denote by θ_i the angle of p_i in S_1 for $j+1 \le i \le k$ — with respect to any standard representative of S_1, with the usual convention that θ_{k+1} is the argument of z_d on the standard representative, so in particular θ_{j+1} is well defined even if $j = k$ — and denote by $\overline{\theta}_s$ the angle of p_s in S_0 for $1 \le s \le j$, the gluing of S_0 and S_1 for small gluing parameter has (on its standard representative) marked points $p_1, \dots, p_k$ with the angles

$$\begin{aligned}(\arg(p_1), \dots, \arg(p_k)) \\ = (\overline{\theta}_1 + \theta_{j+1}, \overline{\theta}_2 + \theta_{j+1}, \dots, \overline{\theta}_j + \theta_{j+1}, \theta_{j+1}, \theta_{j+2}, \dots, \theta_k).\end{aligned} \tag{A-12}$$

A.2 Operations with a forgotten marked point

We introduce auxiliary degenerate operations that will arise as the codimension-one boundary of the open–closed map and equivariant structure. This subsection is a very special case of the general discussion in [24].

Let $d \geq 2$ and $i \in \{1, \dots, d\}$. The *moduli space of discs with d marked points with i^{th} boundary point forgotten*,

$$\mathcal{R}^{d,f_i}, \tag{A-13}$$

is exactly the moduli space of discs $\mathcal{R}^d$, with i^{th} boundary marked point labeled as auxiliary.

The Deligne–Mumford compactification

$$\overline{\mathcal{R}}^{d,f_i} \tag{A-14}$$

is exactly the usual Deligne–Mumford compactification, along with the data of an *auxiliary label* at the relevant boundary marked point.

For $d > 2$, the *i–forgetful map*

$$\mathcal{F}_{d,i} \colon \mathcal{R}^{d,f_i} \to \mathcal{R}^{d-1} \tag{A-15}$$

associates to a surface S the surface obtained by putting the i^{th} point back in and forgetting it. This map admits an extension to the Deligne–Mumford compactification

$$\overline{\mathcal{F}}_{d,i} \colon \overline{\mathcal{R}}^{d,f_i} \to \overline{\mathcal{R}}^{d-1} \tag{A-16}$$

as follows: eliminate any nonmain components with only one nonauxiliary marked point p, and label the positive marked point below this component by p. We say that any component not eliminated is *f–stable* and any component eliminated is *f–semistable*. The above map is only well defined for $d > 2$. In the semistable case $d = 2$, the space $\mathcal{R}^{2,f_i}$ is a point so one can define an ad hoc map

$$\mathcal{F}_i^{\text{ss}} \colon \mathcal{R}^{2,f_i} \to \text{pt}, \tag{A-17}$$

which associates to a surface S the (unstable) strip $\Sigma_1 = (-\infty, \infty) \times [0, 1]$ as follows: take the unique representative of S which, after its three marked points are removed, is biholomorphic to the strip Σ_1 with an additional puncture $(0, 0)$. Then, forget/put back in the point $(0, 0)$.

Definition A.1 A *forgotten Floer datum* for a stable disc with i^{th} point auxiliary $S \in \overline{\mathcal{R}}^{d,f_i}$ consists, for every component T of S, of

- a Floer datum for T, if T does not contain the auxiliary point,
- a Floer datum for $\mathcal{F}_j(T)$, if T is f–stable and contains the auxiliary point as its j^{th} input,
- a Floer datum on $\mathcal{F}_i^{\text{ss}}(T)$ which is *translation invariant* if T is *f–semistable*.

By *translation invariant*, we mean the following: note that Σ_1 has a canonical $\mathbb{R}$–action given by linear translation in the s coordinate. We require H, J and the time-shifting map/weights to be invariant under this $\mathbb{R}$–action, and in particular they should only depend on $t \in [0,1]$ at most.

In particular, this Floer datum should only depend on the point $\bar{\mathcal{F}}_{d,i}(S)$.

Proposition A.2 *Let $i \in \{1,\dots,d\}$ with $d > 1$. Then the operation associated to $\overline{\mathcal{R}}^{d,f_i}$ is zero if $d > 2$, and the identity operation $I(\cdot)$ (up to a sign) when $d = 2$.*

Sketch Suppose first that $d > 2$, and let u be any solution to Floer's equation over the space $\mathcal{R}^{d,f_i}$ with domain S. Since the Floer data on S only depends on $\mathcal{F}_{d,i}(S)$, we see that maps from S' with $S' \in \mathcal{F}_{d,i}^{-1}(\mathcal{F}_{d,i}(S))$ also give solutions to Floer's equation with the same asymptotics. Moreover, the fibers of the map $\mathcal{F}_{d,i}$ are one-dimensional, implying that u cannot be rigid, and thus the associated operation is zero.

Now suppose that $d = 2$. Then the forgetful map associates to the single point $[S] \in \mathcal{R}^{2,f_i}$ the unstable strip with its translation-invariant Floer datum. Since nonconstant solutions can never be rigid — as, by translating, one can obtain other nonconstant solutions — it follows that the only solutions are constant ones, and the resulting operation is therefore the identity. □

References

[1] **M Abouzaid**, *A geometric criterion for generating the Fukaya category*, Publ. Math. Inst. Hautes Études Sci. 112 (2010) 191–240 MR Zbl

[2] **M Abouzaid**, **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Quantum cohomology and split generation in Lagrangian Floer theory*, in preparation

[3] **M Abouzaid**, **P Seidel**, *An open string analogue of Viterbo functoriality*, Geom. Topol. 14 (2010) 627–718 MR Zbl

[4] **P Albers**, **K Cieliebak**, **U Frauenfelder**, *Symplectic Tate homology*, Proc. Lond. Math. Soc. 112 (2016) 169–205 MR Zbl

[5] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR Zbl

[6] **F Bourgeois**, **A Oancea**, *S^1–equivariant symplectic homology and linearized contact homology*, Int. Math. Res. Not. 2017 (2017) 3849–3937 MR Zbl

[7] **C Brav**, **T Dyckerhoff**, *Relative Calabi–Yau structures*, Compos. Math. 155 (2019) 372–412 MR Zbl

[8] **D Burghelea**, *Cyclic homology and the algebraic K–theory of spaces, I*, from "Applications of algebraic K–theory to algebraic geometry and number theory, I, II" (S J Bloch, R K Dennis, E M Friedlander, M R Stein, editors), Contemp. Math. 55, Amer. Math. Soc., Providence, RI (1986) 89–115 MR Zbl

[9] **C-H Cho**, **S Lee**, *Potentials of homotopy cyclic A_∞–algebras*, Homology Homotopy Appl. 14 (2012) 203–220 MR Zbl

[10] **K Cieliebak**, **A Floer**, **H Hofer**, *Symplectic homology, II: A general construction*, Math. Z. 218 (1995) 103–122 MR Zbl

[11] **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

[12] **R Cohen**, **S Ganatra**, *Calabi–Yau categories, the Floer theory of the cotangent bundle, and the string topology of the base*, in preparation

[13] **A Connes**, *Noncommutative differential geometry*, Inst. Hautes Études Sci. Publ. Math. 62 (1985) 257–360 MR Zbl

[14] **K Costello**, *Topological conformal field theories and Calabi–Yau categories*, Adv. Math. 210 (2007) 165–214 MR Zbl

[15] **K Costello**, *The partition function of a topological field theory*, J. Topol. 2 (2009) 779–822 MR Zbl

[16] **V Dotsenko**, **S Shadrin**, **B Vallette**, *De Rham cohomology and homotopy Frobenius manifolds*, J. Eur. Math. Soc. 17 (2015) 535–547 MR Zbl

[17] **Y Félix**, **S Halperin**, **J-C Thomas**, *Differential graded algebras in topology*, from "Handbook of algebraic topology" (I M James, editor), North-Holland, Amsterdam (1995) 829–865 MR Zbl

[18] **A Floer**, *Witten's complex and infinite-dimensional Morse theory*, J. Differential Geom. 30 (1989) 207–221 MR Zbl

[19] **A Floer**, **H Hofer**, *Symplectic homology, I: Open sets in $\mathbb{C}^n$*, Math. Z. 215 (1994) 37–88 MR Zbl

[20] **K Fukaya**, *Cyclic symmetry and adic convergence in Lagrangian Floer theory*, Kyoto J. Math. 50 (2010) 521–590 MR Zbl

[21] **K Fukaya**, *Counting pseudo-holomorphic discs in Calabi–Yau* 3*–folds*, Tohoku Math. J. 63 (2011) 697–727 MR Zbl

[22] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Lagrangian intersection Floer theory: anomaly and obstruction, II*, AMS/IP Studies in Advanced Mathematics 46.2, Amer. Math. Soc., Providence, RI (2009) MR Zbl

[23] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Lagrangian Floer theory and mirror symmetry on compact toric manifolds*, Astérisque 376, Soc. Math. France, Paris (2016) MR Zbl

[24] **S Ganatra**, *Symplectic cohomology and duality for the wrapped Fukaya category*, preprint (2013) arXiv 1304.7312

[25] **S Ganatra**, *Automatically generating Fukaya categories and computing quantum cohomology*, preprint (2016) arXiv 1605.07702

[26] **S Ganatra**, **T Perutz**, **N Sheridan**, *The cyclic open–closed map and noncommutative Hodge structures*, in preparation

[27] **S Ganatra**, **T Perutz**, **N Sheridan**, *Mirror symmetry: from categories to curve counts*, preprint (2015) arXiv 1510.03839

[28] **V Ginzburg**, *Lectures on noncommutative geometry*, lecture notes (2005) arXiv math/0506603

[29] **D Kaledin**, *Non-commutative Hodge-to-de Rham degeneration via the method of Deligne–Illusie*, Pure Appl. Math. Q. 4 (2008) 785–875 MR Zbl

[30] **D Kaledin**, *Spectral sequences for cyclic homology*, from "Algebra, geometry, and physics in the 21st century" (D Auroux, L Katzarkov, T Pantev, Y Soibelman, Y Tschinkel, editors), Progr. Math. 324, Birkhäuser, Boston, MA (2017) 99–129 MR Zbl

[31] **C Kassel**, *Cyclic homology, comodules, and mixed complexes*, J. Algebra 107 (1987) 195–216 MR Zbl

[32] **B Keller**, *On the cyclic homology of exact categories*, J. Pure Appl. Algebra 136 (1999) 1–56 MR Zbl

[33] **B Keller**, *A–infinity algebras, modules and functor categories*, from "Trends in representation theory of algebras and related topics" (J A de la Peña, R Bautista, editors), Contemp. Math. 406, Amer. Math. Soc., Providence, RI (2006) 67–93 MR Zbl

[34] **T Kimura**, **J Stasheff**, **A A Voronov**, *On operad structures of moduli spaces and string theory*, Comm. Math. Phys. 171 (1995) 1–25 MR Zbl

[35] **M Kontsevich**, *XI Solomon Lefschetz memorial lecture series: Hodge structures in non-commutative geometry*, from "Non-commutative geometry in mathematics and physics" (G Dito, H García-Compeán, E Lupercio, F J Turrubiates, editors), Contemp. Math. 462, Amer. Math. Soc., Providence, RI (2008) 1–21 MR Zbl

[36] **M Kontsevich**, *Weak CY algebras*, conference talk (2013) Available at `https://tinyurl.com/Kontsevich-CY`

[37] **M Kontsevich**, **Y Soibelman**, *Notes on A_∞–algebras, A_∞–categories and non-commutative geometry*, from "Homological mirror symmetry" (A Kapustin, M Kreuzer, K-G Schlesinger, editors), Lecture Notes in Phys. 757, Springer (2009) 153–219 MR Zbl

[38] **M Kontsevich**, **A Takeda**, **Y Vlassopoulos**, *Pre-Calabi–Yau algebras and topological quantum field theories*, preprint (2021) arXiv 2112.14667

[39] **M Kontsevich**, **A Takeda**, **Y Vlassopoulos**, *Smooth Calabi–Yau structures and the noncommutative Legendre transform*, preprint (2023) arXiv 2301.01567

[40] **K Lefèvre-Hasegawa**, *Sur les A_∞ catégories*, PhD thesis, Université Paris 7 – Denis Diderot (2003) Available at `https://theses.hal.science/tel-00007761`

[41] **J-L Loday**, *Cyclic homology*, Grundl. Math. Wissen. 301, Springer (1992) MR Zbl

[42] **J-L Loday**, **D Quillen**, *Cyclic homology and the Lie algebra homology of matrices*, Comment. Math. Helv. 59 (1984) 569–591 MR Zbl

[43] **R McCarthy**, *The cyclic homology of an exact category*, J. Pure Appl. Algebra 93 (1994) 251–296 MR Zbl

[44] **J McCleary**, *A user's guide to spectral sequences*, 2nd edition, Cambridge Studies in Advanced Mathematics 58, Cambridge Univ. Press (2001) MR Zbl

[45] **L Menichi**, *String topology for spheres*, Comment. Math. Helv. 84 (2009) 135–157 MR Zbl

[46] **S Piunikhin**, **D Salamon**, **M Schwarz**, *Symplectic Floer–Donaldson theory and quantum cohomology*, from "Contact and symplectic geometry" (C B Thomas, editor), Publ. Newton Inst. 8, Cambridge Univ. Press (1996) 171–200 MR Zbl

[47] **A F Ritter**, *Topological quantum field theory structure on symplectic cohomology*, J. Topol. 6 (2013) 391–489 MR Zbl

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

[49] **P Seidel**, *Fukaya categories and deformations*, from "Proceedings of the International Congress of Mathematicians, II" (T Li, editor), Higher Ed., Beijing (2002) 351–360 MR Zbl

[50] **P Seidel**, *A_∞–subalgebras and natural transformations*, Homology Homotopy Appl. 10 (2008) 83–114 MR Zbl

[51] **P Seidel**, *A biased view of symplectic cohomology*, from "Current developments in mathematics" (B Mazur, T Mrowka, W Schmid, R Stanley, S-T Yau, editors), International, Somerville, MA (2008) 211–253 MR Zbl

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

[53] **P Seidel**, *Suspending Lefschetz fibrations, with an application to local mirror symmetry*, Comm. Math. Phys. 297 (2010) 515–528 MR Zbl

[54] **P Seidel**, *Categorical dynamics and symplectic topology*, lecture notes (2013) Available at `http://www-math.mit.edu/~seidel/texts/937-lecture-notes.pdf`

[55] **P Seidel**, *Disjoinable Lagrangian spheres and dilations*, Invent. Math. 197 (2014) 299–359 MR Zbl

[56] **P Seidel**, *Connections on equivariant Hamiltonian Floer cohomology*, Comment. Math. Helv. 93 (2018) 587–644 MR Zbl

[57] **P Seidel**, *Fukaya A_∞–structures associated to Lefschetz fibrations, VI*, preprint (2018) arXiv 1810.07119

[58] **P Seidel**, **J P Solomon**, *Symplectic cohomology and q–intersection numbers*, Geom. Funct. Anal. 22 (2012) 443–477 MR Zbl

[59] **N Sheridan**, *Homological mirror symmetry for Calabi–Yau hypersurfaces in projective space*, Invent. Math. 199 (2015) 1–186 MR Zbl

[60] **N Sheridan**, *On the Fukaya category of a Fano hypersurface in projective space*, Publ. Math. Inst. Hautes Études Sci. 124 (2016) 165–317 MR Zbl

[61] **N Sheridan**, *Formulae in noncommutative Hodge theory*, J. Homotopy Relat. Struct. 15 (2020) 249–299 MR Zbl

[62] **T Tradler**, *Infinity-inner-products on A–infinity-algebras*, J. Homotopy Relat. Struct. 3 (2008) 245–271 MR Zbl

[63] **B L Tsygan**, *Homology of matrix Lie algebras over rings and the Hochschild homology*, Uspekhi Mat. Nauk 38 (1983) 217–218 MR Zbl In Russian; translated in Russian Math. Surveys 38 (1983) 198–199

[64] **C Viterbo**, *Functors and computations in Floer homology with applications, I*, Geom. Funct. Anal. 9 (1999) 985–1033 MR Zbl

[65] **W-K Yeung**, *Pre-Calabi–Yau structures and moduli of representations*, preprint (2018) arXiv 1802.05398

[66] **J Zhao**, *Periodic symplectic cohomologies*, J. Symplectic Geom. 17 (2019) 1513–1578 MR Zbl

Department of Mathematics, University of Southern California
Los Angeles, CA, United States

sheel.ganatra@usc.edu

Proposed: Paul Seidel
Seconded: Ciprian Manolescu, Yakov Eliashberg

Received: 12 January 2020
Revised: 25 September 2021

Geometry & Topology **27**:9 (2023) 3585–3618
DOI: 10.2140/gt.2023.27.3585
Published: 5 December 2023

Congruences on K–theoretic Gromov–Witten invariants

JÉRÉMY GUÉRÉ

We study K–theoretic Gromov–Witten invariants of projective hypersurfaces using a virtual localization formula under finite group actions. In particular, it provides all K–theoretic Gromov–Witten invariants of the quintic threefold modulo 41, up to genus 19 and degree 40. As an illustration, we give an instance in genus one and degree one. Applying the same idea to a K–theoretic version of FJRW theory, we determine it modulo 205 for the quintic polynomial with minimal group and narrow insertions, in every genus.

14N35

0 Introduction

One of the first achievements of Gromov–Witten (GW) theory is the celebrated formula of Candelas, de la Ossa, Green and Parkes [4] computing genus-0 invariants of the quintic threefold in terms of a hypergeometric series solution of a Picard–Fuchs equation. It was a first instance of mirror symmetry and was proved by Givental [14] and Lian, Liu and Yau [28].

The K–theoretic version of GW theory, which we refer to as KGW theory, was constructed in Lee [25], and it is only recently that mirror symmetry in this context was developed by Givental in his series of preprints starting with [15]. It relates the KGW generating series to a q–hypergeometric function solution of a finite-difference equation.

Both GW and KGW theories rely on the notion of a perfect obstruction theory (see Behrend and Fantechi [2]), producing two fundamental objects on the moduli space $\overline{\mathcal{M}}_{g,n}(X,\beta)$ of stable maps to a given nonsingular variety X, namely the virtual cycle $[\overline{\mathcal{M}}_{g,n}(X,\beta)]^{\mathrm{vir}}$ living in the Chow ring of the moduli space, and the virtual structure sheaf $\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{g,n}(X,\beta)}$ living in its K–theory of coherent sheaves.

Given insertions $y_i \in \mathrm{CH}_*(X)$, or $Y_i \in K^0(X)$, and psi-classes $\psi_i \in \mathrm{CH}^1(\overline{\mathcal{M}}_{g,n})$, or psi-bundles $\Psi_i \in K^0(\overline{\mathcal{M}}_{g,n})$, we then form

$$a := \prod_{i=1}^{n} \mathrm{ev}^*(y_i)\cdot\psi_i^{d_i} \quad\text{and}\quad A := \bigotimes_{i=1}^{n} \mathrm{ev}^*(Y_i)\otimes\Psi_i^{\otimes d_i}, \quad \text{with } d_i \in \mathbb{Z},$$

using evaluation maps. It yields GW and KGW invariants

$$p_*(a\cdot[\overline{\mathcal{M}}_{g,n}(X,\beta)]^{\mathrm{vir}}) \in \mathbb{Q} \quad\text{and}\quad p_!(A\otimes\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{g,n}(X,\beta)}) \in \mathbb{Z},$$

where p is the projection map to a point along which we take pushforwards in Chow or in K–theory.[1] Each theory has an important feature: the virtual cycle is pure-dimensional, leading to a degree condition on the insertions for the GW invariant to be nonzero, and KGW invariants are all integers. Moreover, the two theories are related via a Hirzebruch–Riemann–Roch theorem (see Tonita [38]), saying that all KGW invariants of a nonsingular variety X can be reconstructed from the knowledge of all GW invariants of the DM stacks $[X/(\mathbb{Z}/M\mathbb{Z})]$ for all $M \in \mathbb{N}^*$; see Givental [16, Main Theorem].

Let T be a torus. When the variety X carries a nontrivial T–action, so does the moduli space of stable maps, and the virtual cycle and virtual structure sheaf are T–equivariant. One then benefits from the virtual localization formula of Graber and Pandharipande [17] to reduce the computation of invariants to the T–fixed locus, which greatly simplifies the calculation. Unfortunately, the automorphism group of a smooth projective hypersurface such as the quintic threefold is finite (except in the special cases of quadrics, elliptic curves and K3 surfaces), so that there is no nontrivial T–action.

Let G be a finite cyclic group. In this paper, we take advantage of the fact that the (virtual) localization formula holds with no change under finite group actions. Since projective hypersurfaces X admit such actions, we can apply it to the study of KGW

[1] In K–theory, it is also known as the Euler characteristic.

theory of X. However, we still have two difficulties. First, the G–fixed moduli space is in general quite involved and we cannot guarantee that it is smooth, so even after applying the virtual localization formula, we may not be able to finish the computation. Second, the (virtual) localization formula gives an answer in a localized ring. For instance, the G–equivariant K–theory of a point is isomorphic to the representation ring $R(G)$, which in the case of a finite cyclic group of order M yields $\mathbb{Z}[X]/(1-X^M)$. Instead of providing an answer in $R(G)$, the (virtual) localization formula only gives us the image in the localized complexified ring $R(G)_{\mathbb{C},\mathrm{loc}}$, where we invert a maximal ideal corresponding to a nonzero element in G. The issue with the localized ring is that the map $R(G) \to R(G)_{\mathbb{C},\mathrm{loc}}$ is in general not injective. In our example, we have $R(G)_{\mathbb{C},\mathrm{loc}} \simeq \mathbb{C}$ and the map sends X to a given primitive M^{th} root of unity, so that the nonzero polynomial $1 + X + \cdots + X^{M-1} \in R(G)$ is sent to $0 \in \mathbb{C}$. Notice that when the group is a torus T, the localization map $R(T) \to R(T)_{\mathbb{C},\mathrm{loc}}$ is injective; that is why we have no such issue in the previous paragraph.

We overcome the first difficulty by means of an "equivariant quantum Lefschetz theorem" that we developed for GW theory in [19, Section 2] and that we adapt to KGW theory and to finite group actions in Section 1; see Theorem 1.6. It compares the G–equivariant virtual structure sheaf of a hypersurface $X \subset \mathbb{P}^N$ to that of the ambient space $\mathbb{P}^N$, and then we use the T–action on the ambient space to apply the virtual localization formula. However, Theorem 1.6 requires that for every G–fixed stable map from a curve C to X, all stable components of C are contracted to a point in X. This condition could fail if the automorphism group of the curve is too big, leading us to impose restrictions on the genus of the curve and on the degree of the stable map.

The second difficulty is more serious. Indeed, we know the G–equivariant KGW invariant is of the form $a_0 + a_1 X + \cdots + a_{M-1} X^{M-1}$ for some integers a_i, and our goal would be the "nonequivariant" limit $a_0 + \cdots + a_{M-1}$, but we only have access to the complex number $a_0 + a_1\zeta + \cdots + a_{M-1}\zeta^{M-1}$, where ζ is a primitive M^{th} root of unity. Luckily, KGW are integers, so that when M is a prime number, we can sum all these complex numbers for primitive roots and obtain the KGW invariant modulo M.

As a conclusion, we seek automorphisms of X of prime order with isolated fixed points. For instance, the quintic threefold can be realized as the zero locus in $\mathbb{P}^4$ of the loop polynomial $x_0^4 x_1 + \cdots + x_4^4 x_0$ and the action

$$\zeta \cdot (x_0, \dots, x_4) = (\zeta x_0, \zeta^{-4} x_1, \zeta^{16} x_2, \zeta^{-64} x_3, \zeta^{256} x_4), \quad \text{where } \zeta := e^{2i\pi/41},$$

yields an automorphism of X of prime order 41, whose fixed points are coordinate points. As a result of Corollary 2.9, we obtain[2] all KGW invariants of the quintic threefold up to genus 19 and degree 40, modulo 41.

Remark 0.1 It happens that 41 is the biggest prime number p for which there exists an automorphism of order p for a smooth quintic hypersurface; see Oguiso and Yu [30].

In Proposition 2.13, we provide an instance of this calculation in genus one. Precisely, we compute

$$\chi\left(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{1,0}(X,1)}}{1-q\mathbb{E}^\vee}\right) \equiv (120q^2+180q+125)\frac{1-q^4-q^6}{(1-q^4)(1-q^6)} \in \mathbb{Z}/205\mathbb{Z}[\![q]\!],$$

where $\mathbb{E}$ denotes the Hodge bundle and q is a formal variable, so that the inverse of $1-q\mathbb{E}^\vee$ is defined as the geometric series in q of general term $\mathbb{E}^{-k}q^k$ (here $\mathbb{E}$ is a line bundle).

Interestingly, if we can find automorphisms of prime orders for infinitely many primes and if we can handle the respective localization formulas, then we are able to determine KGW invariants as integers instead of modulo a prime number. We apply this idea to elliptic curves. There are indeed p–torsion points for every prime number p, so that a translation by this point is an automorphism of order p. Furthermore, the localization formula is trivial since there are no fixed points. We deduce the vanishing of all KGW invariants of an elliptic curve, with homogeneous insertions.

Similarly to GW theory, Fan, Jarvis and Ruan [12; 11] developed a quantum singularity theory for Landau–Ginzburg orbifolds. It is known as the FJRW theory and an algebraic construction has been established by Polishchuk and Vaintrob in [33]. Precisely, they construct a matrix factorization over the moduli space of (W,G)–spin curves, where W is a nondegenerate quasihomogeneous polynomial and G is an admissible group of symmetries. We refer to Guéré [18] for details.

In Section 3, we explain how to construct a K–theoretic version of FJRW theory and we then pursue the same goal as for KGW theory: compute invariants by applying the localization theorem under finite group actions. We focus on the quintic polynomial

[2] It means we write the answer as a sum over dual graphs of contributions involving only KGW invariants of the point. This computation can be encoded into a computer. Note however that the calculation of KGW invariants of the point is nontrivial in genus more than two. As explained in Proposition 2.13, we use the computer for the calculation of an explicit example in genus one.

with group μ_5 for clarity of the exposition and we find all its K–theoretic FJRW invariants with narrow insertions in every genus and modulo 205; see Corollary 3.11. Here, we do not have restriction bounds on the genus of the curve.

In [18], we compute genus-0 FJRW invariants of chain polynomials using a characteristic class[3] $\mathfrak{c}_t \colon K^0(S) \to H^*(S)[\![t]\!]$, which we can define for a line bundle L over a smooth DM stack S as

$$\mathfrak{c}_t(L) = \mathrm{Ch}(\lambda_{-t} L^\vee)\,\mathrm{Td}(L) = c_1(L) \cdot \frac{e^{c_1(L)} - t}{e^{c_1(L)} - 1},$$

and then extend multiplicatively. Genus-0 FJRW invariants of a chain polynomial $x_1^{a_1} x_2 + \cdots + x_N^{a_N}$ are then equal to

$$c_{\mathrm{vir}} = \lim_{t \to 1} \prod_{j=1}^{N} \mathfrak{c}_{t_j}(-R\pi_*\mathcal{L}_j), \quad \text{with } g = 0, \tag{1}$$

where $t_j := t^{(-a_1)\cdots(-a_{j-1})}$ and $R\pi_*\mathcal{L}_j$ are the derived pushforwards of the universal line bundles over the moduli space of (W, G)–spin curves. It is remarkable that such a limit exists and the author has wondered since then whether other limits could exist, for instance when t tends to some root of unity. Interestingly, we prove[4] for the quintic loop polynomial with group μ_5 and narrow insertions that such a limit exists for all genus when t tends to a 41st root of unity ζ_{41}. It then converges to a $\mathbb{Z}/41\mathbb{Z}$–equivariant version of the FJRW virtual cycle, defined as follows. The two-periodic complex obtained from the Polishchuk–Vaintrob matrix factorization naturally decouples as a direct sum of 41 two-periodic complexes.[5] Each one of them provides a (virtual) cycle a_k, for $0 \leq k \leq 40$, and we define

$$c_{\mathrm{vir}}^{\mathbb{Z}/41\mathbb{Z}} := \sum_{k=0}^{40} a_k \zeta_{41}^k = \lim_{t \to \zeta_{41}} \prod_{j=1}^{N} \mathfrak{c}_{t_j}(-R\pi_*\mathcal{L}_j), \quad \text{with } g \geq 0.$$

We easily find similar results for other loop polynomials.

[3] Polishchuk and Vaintrob's definition of the virtual cycle in FJRW theory involves a Koszul complex of vector bundles; see [33]. The class $\mathfrak{c}_t$ then appears naturally from the definition of the Koszul complex, as it involves exterior powers of vector bundles. Note also that $\mathfrak{c}_1$ recovers the top Chern class of a vector bundle.

[4] The restriction to the group μ_5 is not necessary and we could consider other admissible groups. However, if one takes the maximal group μ_{1025}, then the action of $\mathbb{Z}/41\mathbb{Z}$ becomes trivial and gives no result.

[5] Precisely, the k^{th} two-periodic complex is the one containing vector bundles $\mathrm{Sym}^{k+41\cdot l} A_1^\vee$, with $l \in \mathbb{N}$, in the notation of [18].

As a conclusion, we mention a future line of research. Chiodo and Ruan [9] and Chiodo, Iritani and Ruan [8] studied the so-called genus-0 Landau–Ginzburg/Calabi–Yau (LG/CY) correspondence, which provides a striking relation between GW theory of a projective hypersurface and FJRW theory of the defining polynomial. Following Chiodo and Ruan [10], there is a similar correspondence in higher genus as well. Since we expect the LG/CY correspondence to hold in K–theory as well, it would be interesting to probe a K–theoretic version for the quintic threefold, up to genus 19, degree 40, and modulo 41.

Another question we may ask is: what information do we get on GW invariants of the quintic threefold up to genus 19? The quintic threefold X is special, its virtual cycle (with no markings) is 0–dimensional, so that a lot of its GW invariants vanish. In fact, they are all deduced from some rational numbers $n_{g,d} \in \mathbb{Q}$ for nonnegative integers g and d, corresponding to its GW invariants without markings. As a consequence, we expect some simplifications in the Hirzebruch–Riemann–Roch theorem of Tonita [38] and Givental [16], and to find formulas expressing KGW invariants of X in terms of the $n_{g,d}$. Moreover, it is proven in Fan and Lee [13], Guo, Janda and Ruan [20] and Chang, Guo and Li [5] that all values of $n_{g,d}$ are expressed in terms of low degrees, where $5d \leq 2g-2$. Up to genus 19, there are exactly 61 unknowns:

$$n_{4,1}, n_{5,1}, \dots, n_{18,6}, n_{19,7} \in \mathbb{Q}.$$

As we are able to compute all KGW invariants modulo 41 up to genus 19 and degree 40, we expect a lot of relations among these 61 unknowns. Moreover, KGW is not restricted by a degree condition on insertions, so we can also insert K–classes from $\mathbb{P}^4$, yielding indeed infinitely many relations among these 61 unknowns. Of course, we do not know yet how many of these relations are nontrivial. It would also be enlightening to express KGW invariants in terms of BPS numbers, which are integers as well; see [22] for a formula in genus zero.

Notation In this paper, we work over the complex numbers. We denote by $G_0(X)$ the Grothendieck group of coherent sheaves on a DM stack X and by $K^0(X)$ the Grothendieck ring of vector bundles on X. If a linear algebraic group G acts on X, then we denote by $G_0(G,X)$ and $K^0(G,X)$ the Grothendieck groups of G–equivariant coherent sheaves and vector bundles. They are identified when X is smooth, by Thomason [36]. When X is a point, then it equals the representation ring $R(G)$ of the group G. The G–fixed locus inside X is denoted by X^G. For an element $h \in G$, we denote by X^h the h–fixed locus. If V and W are G–equivariant vector bundles over X, then we denote by

$$\lambda^G_{-t}(V-W) = \sum_{k,l\geq 0} (-1)^k \Lambda^k V \otimes \mathrm{Sym}^l W\, t^{k+l} \in K^0(G,X)[\![t]\!]$$

the lambda-structure in K–theory. We extend multiplicatively the notation to any element $V \in K^0(G,X)$. When we forget the group action, we simply denote it by $\lambda_{-t}(V) \in K^0(X)[\![t]\!]$.

Let G be a diagonalizable group. The complexified representation ring $R(G)_{\mathbb{C}} := R(G)\otimes\mathbb{C}$ is identified with the coordinate ring $\mathcal{O}(G)$ of G. Hence, for every $h\in G$, there is a corresponding maximal ideal $\mathfrak{m}_h \subset R(G)_{\mathbb{C}}$. Let

$$\begin{aligned} G_0(G,X)_{\mathrm{loc}} &:= G_0(G,X)\otimes_{\mathbb{Z}} R(G)_{\mathbb{C},\mathfrak{m}_h},\\ K^0(G,X)_{\mathrm{loc}} &:= K^0(G,X)\otimes_{\mathbb{Z}} R(G)_{\mathbb{C},\mathfrak{m}_h} \end{aligned}$$

denote the localizations. Assume X is smooth and let $\iota\colon X^h \subset X$ be the inclusion of the h–fixed locus. The localization theorem says

$$A = \iota_!\left(\frac{\iota^* A}{\lambda^G_{-1}(N_\iota^\vee)}\right) \in K^0(G,X)_{\mathrm{loc}} \quad \text{for all } A\in K^0(G,X)_{\mathrm{loc}}; \tag{2}$$

see Thomason [37]. Note that λ^G_{-1} is the evaluation of the formula above at $t=1$. In general, it is not defined in $K^0(G,X)$ and it is only partially defined in $K^0(G,X)_{\mathrm{loc}}$. Precisely, for a vector bundle V, the term $\lambda^G_{-1}(V)$ is invertible if V has no G–fixed part. This is the case in equation (2).

Equation (2) is in particular true for finite groups G, even though the localization map $R(G)_{\mathbb{C}} \to R(G)_{\mathrm{loc}}$ is not injective in that case. Moreover, we can relax the smoothness condition on X. Indeed, if X is singular but carries a G–equivariant perfect obstruction theory $[E_{-1}\to E_0]$, then there is a G–equivariant virtual structure sheaf $\mathcal{O}_X^{\mathrm{vir},G} \in G_0(G,X)$; see Lee [25]. The obstruction theory pulls back to the G–fixed locus X^G and we denote by $N_\iota^{\mathrm{vir}} \in K^0(G,X^G)$ the K–theoretic class of the dual of its G–moving part. The G–fixed part gives a perfect obstruction theory on X^G and yields a virtual structure sheaf $\mathcal{O}^{\mathrm{vir}}_{X^G}$. Furthermore, we have the virtual localization formula

$$\mathcal{O}_X^{\mathrm{vir},G} = \iota_!\left(\frac{\mathcal{O}^{\mathrm{vir}}_{X^G}}{\lambda^G_{-1}(N_\iota^{\mathrm{vir}\vee})}\right) \in G_0(G,X)_{\mathrm{loc}}. \tag{3}$$

See Qu [34, Theorem 3.3] for the proof in the case where the group is a torus T, but the same proof holds word for word when we replace T by any diagonalizable group G. In particular, it applies to the moduli space $\overline{\mathcal{M}}(\mathcal{X})$ of stable maps to a smooth DM stack $\mathcal{X}$.

Here, we specify the genus, the degree and the number of markings as $\overline{\mathcal{M}}_{g,n}(\mathcal{X},\beta)$ when needed.

The letters GW stand for Gromov–Witten and KGW for K–theoretic Gromov–Witten.

1 Equivariant quantum Lefschetz theorem

This section is a generalization of [19, Section 2] to K–theory and to more general group actions. The main result is an "equivariant quantum Lefschetz" theorem which is of first importance in the next section.

1.1 Virtual localization formula

Let G be a linear algebraic group and $\mathcal{X}$ be a smooth DM stack equipped with a G–action. The moduli space $\overline{\mathcal{M}}(\mathcal{X})$ of stable maps to $\mathcal{X}$ carries a G–action, a G–equivariant perfect obstruction theory, and thus a G–equivariant virtual structure sheaf $\mathcal{O}^{\mathrm{vir},G}_{\overline{\mathcal{M}}(\mathcal{X})} \in G_0(G,\overline{\mathcal{M}}(\mathcal{X}))$.

Denote by $\iota\colon \overline{\mathcal{M}}(\mathcal{X})^G \hookrightarrow \overline{\mathcal{M}}(\mathcal{X})$ the embedding of the G–fixed locus. By definition, the virtual normal bundle $N_\iota^{\mathrm{vir}} \in K^0(G,\overline{\mathcal{M}}(\mathcal{X})^G)$ is the moving part of the pullback of the perfect obstruction theory to the fixed locus.[6] The virtual localization formula (3) states

$$\mathcal{O}^{\mathrm{vir},G}_{\overline{\mathcal{M}}(\mathcal{X})} = \iota_!\left(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{X})^G}}{\lambda^G_{-1}(N_\iota^{\mathrm{vir}\vee})}\right) \in G_0(G,\overline{\mathcal{M}}(\mathcal{X}))_{\mathrm{loc}}.$$

1.2 Enhancement of the group

Let $G \subset T$ be an embedding of linear algebraic groups and $\mathcal{X} \hookrightarrow \mathcal{P}$ be an embedding of smooth DM stacks equipped with a G–action. We assume that

- the G–fixed loci of $\mathcal{X}$ and of $\mathcal{P}$ are equal;
- for every G–fixed stable map to $\mathcal{P}$, all stable components of the source curve are sent to $\mathcal{P}^G$;
- $\mathcal{P}$ is equipped with a T–action extending the G–action;
- the normal bundle of $\mathcal{X} \hookrightarrow \mathcal{P}$ is the pullback of a T–equivariant vector bundle $\mathcal{N}$ over $\mathcal{P}$;
- $\mathcal{X}$ is the zero locus of a G–invariant section of the vector bundle $\mathcal{N}$; and

[6] The perfect obstruction theory admits a global presentation by a two-term complex of vector bundles, hence the virtual normal bundle is an element in K^0 rather than an element in G_0.

- the vector bundle $\mathcal{N}$ is convex up to two markings, ie for every stable map $f\colon \mathcal{C} \to \mathcal{P}$, where $\mathcal{C}$ is a smooth genus-0 orbifold curve with at most two markings, we have $H^1(\mathcal{C}, f^*\mathcal{N}) = 0$.

Let us first consider the G–fixed loci of the moduli spaces of stable maps and observe the fibered diagram

$$\begin{tikzcd}
\overline{\mathcal{M}}(\mathcal{X})^G \arrow[r, "j"] \arrow[d, "\iota"'] \arrow[dr, phantom, "\square"] & \overline{\mathcal{M}}(\mathcal{P})^G \arrow[d, "\tilde{\iota}"] \\
\overline{\mathcal{M}}(\mathcal{X}) \arrow[r, "\tilde{j}"'] & \overline{\mathcal{M}}(\mathcal{P})
\end{tikzcd}$$

Writing $i\colon \mathcal{X} \hookrightarrow \mathcal{P}$, we have a G–equivariant short exact sequence

$$0 \to T_{\mathcal{X}} \to i^*T_{\mathcal{P}} \to i^*\mathcal{N} \to 0$$

inducing a distinguished triangle

$$R\pi_* f^*T_{\mathcal{X}} \to R\pi_* f^*T_{\mathcal{P}} \to R\pi_* f^*\mathcal{N} \to (R\pi_* f^*T_{\mathcal{X}})[1]. \tag{4}$$

Note that their duals are parts of the perfect obstruction theories of $\overline{\mathcal{M}}(\mathcal{X})$ and of $\overline{\mathcal{M}}(\mathcal{P})$, the remaining parts being the perfect obstruction theory of the moduli space of stable curves itself.

The term $\mathcal{E} := R\pi_* f^*\mathcal{N}$, pulled back to $\overline{\mathcal{M}}(\mathcal{P})^G$, has a fixed and a moving part, that we denote respectively by $\mathcal{E}_{\text{fix}}$ and $\mathcal{E}_{\text{mov}}$.

Proposition 1.1 *The fixed part $\mathcal{E}_{\text{fix}}$ is a vector bundle over the fixed moduli space $\overline{\mathcal{M}}(\mathcal{P})^G$.*

Proof Let $f\colon \mathcal{C} \to \mathcal{P}$ be a stable map belonging to $\overline{\mathcal{M}}(\mathcal{P})^G$. We denote by $\rho\colon \mathcal{C} \to C$ the coarse map. It is enough to prove that

$$H^1(C, \rho_* f^*\mathcal{N})^{\text{fix}} = 0.$$

Take the normalization $\nu\colon C^\nu \to C$ of the curves at all their nodes. We have

$$C^\nu = \bigsqcup_{i\in I} C_i^{\text{fix}} \sqcup \bigsqcup_{j\in J} C_j^{\text{nf}},$$

where the superscripts refer respectively to fixed/nonfixed components of C^ν under the map f. In particular, nonfixed components are unstable curves, ie the projective line with one or two special points. By the normalization exact sequence, we obtain an exact sequence

$$\bigoplus_{\text{nodes}} H^0(\text{node}, f^*\mathcal{N}_{|\text{node}}) \to H^1(C, \rho_* f^*\mathcal{N}) \to H^1(C^\nu, \nu^*\rho_* f^*\mathcal{N}) \to 0,$$

with

$$H^1(C^\nu, \nu^*\rho_* f^*\mathcal{N}) = \bigoplus_{i\in I} H^1(C_i^{\text{fix}}, \nu^*\rho_* f^*\mathcal{N}) \oplus \bigoplus_{j\in J} H^1(C_j^{\text{nf}}, \nu^*\rho_* f^*\mathcal{N}).$$

Since the normal bundle has a nontrivial G–action once restricted to the fixed locus of $\mathcal{X}$ (or equivalently of $\mathcal{P}$), we have

$$H^0(\text{node}, f^*\mathcal{N}_{|\text{node}})^{\text{fix}} = 0 \quad \text{and} \quad H^1(C_i^{\text{fix}}, \nu^*\rho_* f^*\mathcal{N})^{\text{fix}} = 0.$$

Therefore, it remains to see the vanishing of H^1 for nonfixed unstable curves C_j^{nf}, for $j \in J$. The curve C_j^{nf} is isomorphic to $\mathbb{P}^1$ with either one or two markings, hence $H^1(C_j^{\text{nf}}, \nu^*\rho_* f^*\mathcal{N}) = 0$ by our assumption of convexity up to two markings. □

Denote by $\mathcal{O}^{\text{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G}$ the virtual structure sheaf obtained by the G–fixed part of the perfect obstruction theory $R\pi_* f^* T_{\mathcal{P}}$.

Proposition 1.2 *We have*

$$j_! \mathcal{O}^{\text{vir}}_{\overline{\mathcal{M}}(\mathcal{X})^G} = \lambda^G_{-1}(\mathcal{E}_{\text{fix}}^\vee) \otimes \mathcal{O}^{\text{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G} \in G_0(\overline{\mathcal{M}}(\mathcal{P})^G).$$

Furthermore, in the localized equivariant K–theoretic ring, we have

$$\frac{1}{\lambda^G_{-1}(N_\iota^{\text{vir}\,\vee})} = j^*\left(\frac{\lambda^G_{-1}(\mathcal{E}_{\text{mov}}{}^\vee)}{\lambda^G_{-1}(N_{\tilde{\iota}}^{\text{vir}\,\vee})}\right) \in K^0(G, \overline{\mathcal{M}}(\mathcal{P})^G)_{\text{loc}}.$$

Proof It follows from the standard proof using convexity; we recall here the main arguments.

The DM stack $\mathcal{X}$ is the zero locus of a G–invariant section of the vector bundle $\mathcal{N}$ over the ambient space $\mathcal{P}$. This section induces a map s from the moduli space of stable maps to $\mathcal{P}$ to the direct image cone $\pi_* f^*\mathcal{N}$; see [6, Definition 2.1]. Since the moduli space $\overline{\mathcal{M}}(\mathcal{P})^G$ is fixed by the action of G, it maps to the fixed part of the direct image cone, that is, the vector bundle $\mathcal{E}_{\text{fix}}$. Hence we have the fibered diagram

$$\begin{tikzcd}
\overline{\mathcal{M}}(\mathcal{X})^G \arrow[r, "j"] \arrow[d] \arrow[dr, phantom, "\square"] & \overline{\mathcal{M}}(\mathcal{P})^G \arrow[d, "s"] \\
\overline{\mathcal{M}}(\mathcal{P})^G \arrow[r, "0"'] & \mathcal{E}_{\text{fix}}
\end{tikzcd}$$

where the bottom map is the embedding as the zero section. The fixed part of the distinguished triangle (4) gives a compatibility datum of perfect obstruction theories for the fixed moduli spaces. Functoriality of the virtual structure sheaf gives

$$0^! \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G} = \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{X})^G};$$

see [34]. Applying the projection formula via the map j on both sides and using the Koszul resolution gives the first result. The second part of the statement follows from the moving part of the distinguished triangle (4). □

By the virtual localization formula, the G–equivariant virtual structure sheaf satisfies

$$\begin{aligned}\tilde{j}_! \mathcal{O}^{\mathrm{vir},G}_{\overline{\mathcal{M}}(\mathcal{X})} = \tilde{j}_! \iota_! \left(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{X})^G}}{\lambda^G_{-1}(N^{\mathrm{vir}\vee}_{\iota})} \right) &= \tilde{\iota}_! j_! \left(\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{X})^G} \otimes j^* \left(\frac{\lambda^G_{-1}(\mathcal{E}_{\mathrm{mov}}{}^\vee)}{\lambda^G_{-1}(N^{\mathrm{vir}\vee}_{\tilde{\iota}})} \right) \right) \\ &= \tilde{\iota}_! \left(\lambda^G_{-1}(\mathcal{E}_{\mathrm{fix}}{}^\vee) \otimes \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G} \otimes \frac{\lambda^G_{-1}(\mathcal{E}_{\mathrm{mov}}{}^\vee)}{\lambda^G_{-1}(N^{\mathrm{vir}\vee}_{\tilde{\iota}})} \right) \\ &= \tilde{\iota}_! \left(\lambda^G_{-1}(\mathcal{E}^\vee) \otimes \frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G}}{\lambda^G_{-1}(N^{\mathrm{vir}\vee}_{\tilde{\iota}})} \right),\end{aligned}$$

where equalities happen in $G_0(G, \overline{\mathcal{M}}(\mathcal{P}))_{\mathrm{loc}}$.

Remark 1.3 If it were defined, the right-hand side would equal

$$\lambda^G_{-1}(R\pi_* f^* \mathcal{N})^\vee \otimes \mathcal{O}^{\mathrm{vir},G}_{\overline{\mathcal{M}}(\mathcal{P})},$$

using the virtual localization formula, but it is not clear that the G–lambda class of $R\pi_* f^* \mathcal{N}$ is defined in $G_0(G, \overline{\mathcal{M}}(\mathcal{P}))_{\mathrm{loc}}$. However, we say that $\lambda^G_{-1}(R\pi_* f^* \mathcal{N})^\vee$ is defined after localization[7] to mean that its pullback to the fixed locus is defined.

Now, we aim to extend the right-hand side of the equality to the T action. The inclusion of groups $G \hookrightarrow T$ yields a morphism

$$\xi_* \colon G_0(T, \overline{\mathcal{M}}(\mathcal{P})) \to G_0(G, \overline{\mathcal{M}}(\mathcal{P})),$$

under which we get

$$\xi_*(\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})}) = \mathcal{O}^{\mathrm{vir},G}_{\overline{\mathcal{M}}(\mathcal{P})}.$$

Unfortunately, the map ξ_* is only partially defined when we localize equivariant parameters: the denominators could be nonzero in the T–localization but vanish in the G–localization. It is easier to work out this issue on the fixed locus of the moduli space.

[7] We find this definition for the formal quintic; see [27].

Let $\overline{\mathcal{M}}(\mathcal{P})^T \hookrightarrow \overline{\mathcal{M}}(\mathcal{P})$ denote the T–fixed locus of the moduli space. In particular, we have the inclusion $\hat{\iota}\colon \overline{\mathcal{M}}(\mathcal{P})^T \hookrightarrow \overline{\mathcal{M}}(\mathcal{P})^G$. We notice that the moduli space $\overline{\mathcal{M}}(\mathcal{P})^G$ is stable under the T–action from $\overline{\mathcal{M}}(\mathcal{P})$ and that the map $\hat{\iota}$ is T–equivariant. Moreover, we have a T–equivariant virtual structure sheaf

$$\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})^G} \in G_0(T,\overline{\mathcal{M}}(\mathcal{P})^G)$$

and the equality

$$\xi_*(\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})^G}) = \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G} \in G_0(G,\overline{\mathcal{M}}(\mathcal{P})^G).$$

By the virtual localization formula, we have

$$\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})^G} = \hat{\iota}_!\left(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\lambda^T_{-1}(N^{\mathrm{vir}\vee}_{\hat{\iota}})}\right) \in G_0(T,\overline{\mathcal{M}}(\mathcal{P})^G)_{\mathrm{loc}}.$$

Furthermore, in K–theory on the space $\overline{\mathcal{M}}(\mathcal{P})^T$ we have the equality

$$N^{\mathrm{vir}}_{\tilde{\iota}\circ\hat{\iota}} = \hat{\iota}^* N^{\mathrm{vir}}_{\tilde{\iota}} + N^{\mathrm{vir}}_{\hat{\iota}}. \tag{5}$$

Indeed, let $\mathcal{F}$ be the pullback of the perfect obstruction theory from $\overline{\mathcal{M}}(\mathcal{P})$ to $\overline{\mathcal{M}}(\mathcal{P})^T$. By definition, the virtual normal bundle $N^{\mathrm{vir}}_{\tilde{\iota}\circ\hat{\iota}}$ is the T–moving part $\mathcal{F}^{\mathrm{mov}}$, which decomposes as $\mathcal{F}^{\mathrm{mov}} = \mathcal{F}^{\mathrm{mov}}_{\mathrm{fix}} + \mathcal{F}^{\mathrm{mov}}_{\mathrm{mov}}$, where the subscript denotes the G–fixed/moving part. By definition, the virtual normal bundle $\hat{\iota}^* N^{\mathrm{vir}}_{\tilde{\iota}}$ is the G–moving part of $\mathcal{F}$, ie $\mathcal{F}^{\mathrm{mov}}_{\mathrm{mov}}$, since there is no G–moving T–fixed part in $\mathcal{F}$. The virtual normal bundle $N^{\mathrm{vir}}_{\hat{\iota}}$ identifies with $\mathcal{F}^{\mathrm{mov}}_{\mathrm{fix}}$.

Remark 1.4 The virtual normal bundle $N^{\mathrm{vir}}_{\tilde{\iota}}$ is defined on $\overline{\mathcal{M}}(\mathcal{P})^G$ and we have a well-defined equality

$$\xi_*(\lambda^T_{-1}(N^{\mathrm{vir}\vee}_{\tilde{\iota}})^{-1}) = \lambda^G_{-1}(N^{\mathrm{vir}\vee}_{\tilde{\iota}})^{-1} \in K^0(G,\overline{\mathcal{M}}(\mathcal{P})^G)_{\mathrm{loc}}.$$

We also have seen the G–decomposition $\mathcal{E} = \mathcal{E}_{\mathrm{fix}} + \mathcal{E}_{\mathrm{mov}}$ over $\overline{\mathcal{M}}(\mathcal{P})^G$ with $\mathcal{E}_{\mathrm{fix}}$ being a T–equivariant vector bundle. Indeed, the vector bundle $\mathcal{N}$ over $\mathcal{P}$ is T–equivariant, thus so are $\mathcal{E}$ and $\mathcal{E}_{\mathrm{fix}}$. As a consequence, the equality

$$\xi_*(\lambda^T_{-1}(\mathcal{E})) = \lambda^G_{-1}(\mathcal{E}) \in K^0(G,\overline{\mathcal{M}}(\mathcal{P})^G)_{\mathrm{loc}}$$

is well defined.

Proposition 1.5 *Consider the well-defined class*

$$C_T := \hat{\iota}^*(\lambda^T_{-1}(\mathcal{E}^\vee)) \otimes \frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\lambda^T_{-1}(N^{\mathrm{vir}\vee}_{\tilde{\iota}\circ\hat{\iota}})} \in G_0(T,\overline{\mathcal{M}}(\mathcal{P})^T)_{\mathrm{loc}}.$$

Then its pushforward under the inclusion $\widehat{\iota}$ equals

$$\widehat{\iota}_!(C_T)=\lambda_{-1}^T(\mathcal{E}^\vee)\otimes\frac{\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})^G}}{\lambda_{-1}^T(N_{\widetilde{\iota}}^{\mathrm{vir}\vee})}\in G_0(T,\overline{\mathcal{M}}(\mathcal{P})^G)_{\mathrm{loc}}.$$

In particular, we have

$$\xi_*(\widehat{\iota}_!(C_T))=\lambda_{-1}^G(\mathcal{E}^\vee)\otimes\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G}}{\lambda_{-1}^G(N_{\widetilde{\iota}}^{\mathrm{vir}\vee})}\in G_0(G,\overline{\mathcal{M}}(\mathcal{P})^G)_{\mathrm{loc}}.$$

Proof By the virtual localization above and equation (5), we have

$$\begin{aligned}\widehat{\iota}_!(C_T)&=\lambda_{-1}^T(\mathcal{E}^\vee)\otimes\widehat{\iota}_!\biggl(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\lambda_{-1}^T(N_{\widetilde{\iota}\circ\widehat{\iota}}^{\mathrm{vir}\vee})}\biggr)\\&=\lambda_{-1}^T(\mathcal{E}^\vee)\otimes\widehat{\iota}_!\biggl(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\widehat{\iota}^*(\lambda_{-1}^T(N_{\widetilde{\iota}}^{\mathrm{vir}\vee}))\otimes\lambda_{-1}^T(N_{\widehat{\iota}}^{\mathrm{vir}\vee})}\biggr)\\&=\frac{\lambda_{-1}^T(\mathcal{E}^\vee)}{\lambda_{-1}^T(N_{\widetilde{\iota}}^{\mathrm{vir}\vee})}\otimes\widehat{\iota}_!\biggl(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\lambda_{-1}^T(N_{\widehat{\iota}}^{\mathrm{vir}\vee})}\biggr)=\frac{\lambda_{-1}^T(\mathcal{E}^\vee)\otimes\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})^G}}{\lambda_{-1}^T(N_{\widetilde{\iota}}^{\mathrm{vir}\vee})}.\end{aligned}$$

The last sentence follows from the following property of ξ_*. Let Z be a DM stack with a T–action and take $A\in K^0(T,Z)_{\mathrm{loc}}$, $B\in G_0(T,Z)_{\mathrm{loc}}$, $a\in K^0(G,Z)_{\mathrm{loc}}$ and $b\in G_0(G,Z)_{\mathrm{loc}}$. If $\xi_*(A)=a$ and $\xi_*(B)=b$ are well-defined equalities, then $\xi_*(A\otimes B)$ is well defined and equals the localized class $a\otimes b$. □

The pushforward maps $\widetilde{\iota}_!$ and ξ_* commute when the latter is well defined. Precisely, the map $\widetilde{\iota}$ is T–equivariant and for any localized class $C\in G_0(T,\overline{\mathcal{M}}(\mathcal{P})^G)_{\mathrm{loc}}$ such that $\xi_*(C)$ is well defined in $G_0(G,\overline{\mathcal{M}}(\mathcal{P})^G)_{\mathrm{loc}}$, the localized class $\widetilde{\iota}_!(C)$ is well defined under ξ_* and we have

$$\widetilde{\iota}_!\xi_*(C)=\xi_*\widetilde{\iota}_!(C)\in G_0(G,\overline{\mathcal{M}}(\mathcal{P}))_{\mathrm{loc}}.$$

1.3 Equivariant quantum Lefschetz formula

Summarizing our discussion, we obtain the following.

Theorem 1.6 (equivariant quantum Lefschetz) *Let $\mathcal{X}\hookrightarrow\mathcal{P}$ be a G–equivariant embedding of smooth DM stacks satisfying assumptions listed at the beginning of this section. Then we have*

$$\widetilde{j}_!\mathcal{O}^{\mathrm{vir},G}_{\overline{\mathcal{M}}(\mathcal{X})}=\xi_*\bigl(\lambda_{-1}^T(R\pi_*f^*\mathcal{N})^\vee\otimes\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})}\bigr)\in G_0(G,\overline{\mathcal{M}}(\mathcal{P}))_{\mathrm{loc}},$$

where $\widetilde{j}$ is the embedding of moduli spaces and ξ_* is the specialization of T–equivariant parameters into G–equivariant parameters. Here, the T–equivariant lambda class $\lambda_{-1}^{T}(R\pi_* f^*\mathcal{N})^\vee$ is defined after localization; see Remark 1.3.

Proof Using previous equalities, we get

$$\widetilde{j}_!\mathcal{O}^{\mathrm{vir},G}_{\overline{\mathcal{M}}(\mathcal{X})}=\widetilde{\iota}_!\Biggl(\lambda_{-1}^{G}(\mathcal{E}^\vee)\otimes\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^G}}{\lambda_{-1}^{G}(N_{\widetilde{\iota}}^{\mathrm{vir}\,\vee})}\Biggr)=\xi_*\widetilde{\iota}_!\widehat{\iota}_!\Biggl(\widehat{\iota}^*(\lambda_{-1}^{T}(\mathcal{E}^\vee))\otimes\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\lambda_{-1}^{T}(N_{\widetilde{\iota}\circ\widehat{\iota}}^{\mathrm{vir}\,\vee})}\Biggr).$$

Following Remark 1.3, the meaning of “defined after localization” is precisely

$$\begin{aligned}\xi_*\bigl(\lambda_{-1}^{T}(R\pi_*f^*\mathcal{N})^\vee\otimes\mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}(\mathcal{P})}\bigr)&=\xi_*\Biggl(\lambda_{-1}^{T}(R\pi_*f^*\mathcal{N})^\vee\otimes\widetilde{\iota}_!\widehat{\iota}_!\Biggl(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\lambda_{-1}^{T}(N_{\widetilde{\iota}\circ\widehat{\iota}}^{\mathrm{vir}\,\vee})}\Biggr)\Biggr)\\&=\xi_*\widetilde{\iota}_!\widehat{\iota}_!\Biggl(\widehat{\iota}^*(\lambda_{-1}^{T}(\mathcal{E}^\vee))\otimes\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}(\mathcal{P})^T}}{\lambda_{-1}^{T}(N_{\widetilde{\iota}\circ\widehat{\iota}}^{\mathrm{vir}\,\vee})}\Biggr).\end{aligned}$$

□

2 K–theoretic Gromov–Witten theory

2.1 Automorphisms of loop hypersurfaces

Let X be a smooth degree-d hypersurface in $\mathbb{P}^N$. K–theoretic Gromov–Witten (KGW) theory is invariant under smooth deformations, so that we can choose any degree-d homogeneous polynomial P to define X, as long as it satisfies the Jacobian criterion for smoothness. Here, we will focus on loop polynomials, ie we take

$$X:=\{x_0^{d-1}x_1+\cdots+x_N^{d-1}x_0=0\}\subset\mathbb{P}^N.$$

Let $\overline{M}:=|1-(1-d)^{N+1}|/d$ and consider on $\mathbb{P}^N$ the $\mathbb{Z}/\overline{M}\mathbb{Z}$–action

$$\zeta\cdot(x_0,\dots,x_N)=(x_0,\zeta x_1,\zeta^{u_2}x_2,\dots,\zeta^{u_N}x_N),$$

where $u_0:=0$ and $u_{j+1}:=1-(d-1)u_j$. We have $(d-1)u_N\equiv 1$ modulo $\overline{M}$, so that the hypersurface X is $\mathbb{Z}/\overline{M}\mathbb{Z}$–invariant. Explicitly, we have

$$u_j=\sum_{l=0}^{j-1}(1-d)^l\quad\text{and}\quad\overline{M}=\sum_{l=0}^{N}(1-d)^l.$$

Notation 2.1 Let $\overline{M}_1:=\overline{M}$ and

$$\overline{M}_{j+1}:=\frac{\overline{M}_j}{\gcd(u_{j+1},\overline{M}_j)}\quad\text{for } 1\le j<N.$$

We write $M:=\overline{M}_N$ and $G:=\mathbb{Z}/M\mathbb{Z}$.

Proposition 2.2 *The group G acts on $\mathbb{P}^N$, leaves the hypersurface X invariant, and for all nonzero $g \in G$, the g–fixed locus in $\mathbb{P}^N$ consists of all coordinate points. Furthermore, assuming the Calabi–Yau condition $N+1=d$ and assuming d is a prime number, then we have $M=\overline{M}$. It holds in particular for the quintic hypersurface in $\mathbb{P}^4$.*

Proof By the construction of M, we see that every u_j with $1 \le j \le N$ is coprime with M. It implies that every pairwise difference $u_i - u_j$ is coprime with M. Indeed, let $0 \le i < j \le N$, then we have

$$u_j - u_i = (1-d)^i u_{j-i},$$

so that it is enough to prove that $d-1$ is coprime with M. If a nonzero integer p divides M and $d-1$, then it divides $\overline{M}$, and from its expression in terms of powers of $d-1$, we get $1 \equiv 0$ modulo p, so that $p=1$.

Therefore, we have, for all $0 \le i < j \le N$ and all $0 < k < M$,

$$\zeta^{ku_i} \neq \zeta^{ku_j}, \quad \text{where } \zeta = e^{2\mathrm{i}\pi/M},$$

and hence the statement about the fixed locus.

For the second statement, let $d = N+1$ be a prime number. Then we have $\overline{M} \equiv 0$ modulo d. Thus, if we have $ku_j \equiv 0$ modulo $\overline{M}$, then we have $ku_j \equiv 0$ modulo d. But $u_j \equiv j$ modulo d, so that $k=0$. As a consequence, every u_j is coprime with $\overline{M}$, and $\overline{M} = M$. □

Example 2.3 We realize the quintic hypersurface in $\mathbb{P}^4$ as

$$P = x_0^4 x_2 + \cdots + x_4^4 x_0.$$

Then the group is $G = (\mathbb{Z}/205\mathbb{Z})$, acting as

$$\zeta \cdot \underline{x} = (x_0, \zeta x_1, \zeta^{-3} x_2, \zeta^{13} x_3, \zeta^{-51} x_4).$$

2.2 Virtual localization formula

Gromov–Witten (GW) theory of $\mathbb{P}^N$ and its K–theoretic version is computed by the virtual localization formula under the natural action of the torus $T = (\mathbb{C}^*)^N$. Unfortunately, there are in general no nontrivial torus-actions preserving a smooth degree-d hypersurface X, leading to many difficulties in the computation of its GW and KGW invariants. Nevertheless, we have an action of the finite group G on X.

In cohomology or in Chow theory, the action of the finite group G is useless with respect to the localization formula. Indeed, we have, for example, $A_*^G(\mathrm{pt}) = \mathbb{C}$. On the other hand, we have $K^0(G,\mathrm{pt}) = R(G)$, the representation ring of the group G. Moreover, there exists in K–theory a (virtual) localization formula under finite group actions.

Unfortunately, the (virtual) localization formula does not give a result in $K^0(G,\mathrm{pt})$, but in a localized ring where we invert equivariant parameters. For instance, in the case of an abelian group $G = \mathbb{Z}/M\mathbb{Z}$, the representation ring (taken with complex coefficients) is

$$R(G)_{\mathbb{C}} \simeq \mathbb{C}[X]/(1-X^M),$$

and the multiplicative set we use for localization is generated by

$$\{1-X,\dots,1-X^{M-1}\}.$$

As a consequence, the localized ring is isomorphic to $\mathbb{C}$ and the map $R(G)_{\mathbb{C}} \to R(G)_{\mathbb{C},\mathrm{loc}}$ is not injective. Precisely, the map sends X to a primitive M^{th} root of unity ζ, so that for every prime divisor p of M, the polynomial

$$\sum_{k=0}^{p} X^{kM/p} \mapsto \sum_{k=0}^{p} \zeta^{kM/p} = 0.$$

In conclusion, the (virtual) localization formula successfully computes a G–equivariant K–class expressed using roots of unity, but we cannot extract the "nonequivariant" limit corresponding to the map

$$K^0(G,\mathrm{pt}) \simeq \mathbb{C}[X]/(1-X^M) \to \mathbb{C} \simeq K^0(\mathrm{pt}), \quad X \mapsto 1.$$

Nevertheless, we find a way to extract some information. Indeed, K–theoretic invariants have another important feature: they are integers. Therefore, when the order of the group is a prime number p, the defect of injectivity of the map $R(G) \to R(G)_{\mathbb{C},\mathrm{loc}}$ amounts to the uncertainty

$$1 + X + \cdots + X^{p-1},$$

which equals p in the nonequivariant limit $X \mapsto 1$. To conclude, we are left with the desired integer modulo p. Furthermore, if we have several finite actions of different prime orders, we can increase our knowledge about the result.

Let us go back to the degree-d hypersurface $X \subset \mathbb{P}^N$. The action of $G = \mathbb{Z}/M\mathbb{Z}$ on $\mathbb{P}^N$ leaving X invariant induces a G–action on the moduli spaces of stable maps to $\mathbb{P}^N$ and to X, so that their virtual structure sheaves are G–equivariant, namely

$$\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_g(\mathbb{P}^N,\beta)} \in G_0(G,\overline{\mathcal{M}}_g(\mathbb{P}^N,\beta)) \quad \text{and} \quad \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_g(X,\beta)} \in G_0(G,\overline{\mathcal{M}}_g(X,\beta)).$$

By the virtual localization formula, we then obtain

$$\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_g(X,\beta)} = \iota_!\biggl(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_g(X,\beta)^{\mathrm{fix}}}}{\lambda_{-1}(N_\iota^{\mathrm{vir}\vee})}\biggr) \in G_0(G,\overline{\mathcal{M}}_g(X,\beta))_{\mathrm{loc}},$$

where $\iota\colon \overline{\mathcal{M}}_g(X,\beta)^{\mathrm{fix}} \hookrightarrow \overline{\mathcal{M}}_g(X,\beta)$ denotes the G–fixed locus and N_ι^{vir} denotes the moving part of the perfect obstruction theory on the fixed locus. At last, we get

$$\chi(\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_g(X,\beta)}) = \chi\biggl(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_g(X,\beta)^{\mathrm{fix}}}}{\lambda_{-1}(N_\iota^{\mathrm{vir}\vee})}\biggr) \in \mathbb{C}.$$

The next step is to use Theorem 1.6 to relate this formula to a formula for $\mathbb{P}^N$, where an explicit localization formula is available via the torus action.

2.3 Fixed locus

We easily check all the conditions listed in Section 1.2 but the second:

- every stable component of a fixed stable map is contracted.

We are able to prove it under the following restrictions on genus of the source curve and degree of the map.

Proposition 2.4 *Let $G = \mathbb{Z}/M\mathbb{Z}$ act on a smooth projective variety X so that, for every nonzero element $h \in G$, the h–fixed locus X^h consists of isolated points in X. Let $f\colon C \to X$ be a stable map corresponding to a G–fixed point in the moduli space $\overline{\mathcal{M}}_{g,n}(X,\beta)$. We assume*

$$g < \tfrac{1}{2}(p-1) \quad \textit{and} \quad \beta < M, \tag{6}$$

where p is the greatest prime divisor of M. Then every stable component of the curve C is mapped to one of the G–fixed points in X.

Proof First, we claim that if $f\colon C \to X$ is a G–fixed stable map of positive degree, then the group G is a subgroup of the group $\mathrm{Aut}(C)$ of automorphisms of C. Indeed, let $\zeta \in G$ be a primitive element. Since the stable map is fixed, we can choose an automorphism $\phi_1 \in \mathrm{Aut}(C)$ of the curve C such that

$$\zeta \cdot f(x) = f(\phi_1(x)) \quad \text{for all } x \in C.$$

We then define $\phi_k := \phi_1^k$ for any $k \in \mathbb{N}$. Since the degree of the map f is positive and all but a finite number of points in X are not fixed by any element of G, we can choose a point $x \in C$ such that the points

$$\zeta^k \cdot f(x) = f(\phi_k(x))$$

are all distinct for $0 \leq k < M$. Since the automorphism ϕ_M is an automorphism of the stable map f, it is of finite order. Thus we can consider the smallest integer $K \in \mathbb{N}^*$ such that $\phi_K = \mathrm{id}$, so that we have

$$\zeta^K \cdot f(x) = f(\phi_K(x)) = f(x),$$

and the integer M divides the integer K. As a consequence, the map sending $0 \leq k < M$ to $\phi_{K/M}^k$ embeds the group G in $\mathrm{Aut}(C)$.

Secondly, let us assume C is a stable curve and is not contracted. Since G is a subgroup of $\mathrm{Aut}(C)$, the prime number p divides the order of $\mathrm{Aut}(C)$. By [29, Proposition 3.6], we get[8] $p \leq 2g + 1$, which is a contradiction.

Lastly, we consider the case where C is not a stable curve (and therefore the degree of the map is positive). Let Γ_f be a dual graph representing the stable map f, where we represent every stable component of C by a vertex and every unstable component by an edge. Furthermore, we add on the graph labels to keep track of the genus, number of markings and degree. It is clear that every automorphism ϕ_k of the curve C induces an automorphism of the dual graph Γ_f. Moreover, for each stable component D of C whose corresponding vertex is fixed in Γ_C, the restriction $f_{|D}\colon D \to X$ is fixed by the group G.

We aim to show that the set $V_{>0}$ of vertices with positive degree is empty. Assume it is not. Then, if the group G acts on $V_{>0}$ without fixed points, the total degree of the map is at least M, which is a contradiction. Therefore, there is at least one fixed point, ie there exists a stable component D of C such that the restriction $f_{|D}$ is G–fixed. As we have seen above, the stable component D is then contracted to a point, which contradicts the fact that its corresponding vertex is in $V_{>0}$. □

Remark 2.5 Proposition 2.4 also holds if the condition $g < \frac{1}{2}(p-1)$ is replaced by "for every stable curve of genus less than g, there is no automorphism of order equal to M".

2.4 Equivariant and congruent formulas

Let us apply Theorem 1.6 to our situation.

[8] The reference is only for $g \geq 2$. Nevertheless, it also holds for genus-0 stable curves, as their automorphism groups are all trivial. For genus-1 stable curves, the automorphism group is the largest when there is only one marking. In this case, we have three possibilities: $\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z}$ or $\mathbb{Z}/6\mathbb{Z}$. As a consequence, the prime number p can only be 2 or 3 and is indeed less than $2g + 1 = 3$.

Theorem 2.6 *Let g, n and β be nonnegative integers. Let X be a degree-d loop hypersurface in $\mathbb{P}^N$ and take a subgroup $H \subset \mathbb{Z}/M\mathbb{Z}$ of order q acting on X via the action*

$$k \cdot (x_0, \dots, x_N) = (x_0, \zeta^k x_1, \zeta^{k \cdot u_2} x_2, \dots, \zeta^{k \cdot u_N} x_N) \quad \text{for } k \in H \subset \mathbb{Z}/M\mathbb{Z}; \tag{7}$$

see Section 2.1. This action depends on the choice of a primitive q^{th} root of unity ζ. Moreover, we have the usual $T := (\mathbb{C}^*)^N$–action on $\mathbb{P}^N$ and we see that it extends the H–action via the embedding $\varphi\colon H \hookrightarrow T := (\mathbb{C}^*)^N$ sending k to $(\zeta^k, \zeta^{k \cdot u_2}, \dots, \zeta^{k \cdot u_N})$.

Assume the bounds

$$g < \tfrac{1}{2}(p-1) \quad \text{and} \quad \beta < q,$$

where p is the greatest prime divisor of q. Let $A := \bigotimes_{i=1}^n \Psi_i^{a_i} \otimes \mathrm{ev}^*(Y_i)$ denote some insertions of Psi-classes and K–classes $Y_i \in K^0(T, \mathbb{P}^N)$ coming from the ambient space. Then the corresponding H–equivariant K–theoretic GW invariant equals

$$\chi^T\big(\lambda_{-1}^T(R\pi_* f^*\mathcal{O}(d))^\vee \otimes A \otimes \mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}_{g,n}(\mathbb{P}^N,\beta)}\big) = \chi^H(A \otimes \mathcal{O}^{\mathrm{vir},H}_{\overline{\mathcal{M}}_{g,n}(X,\beta)}) \in \mathbb{C}.$$

Precisely, the class $\lambda_{-1}^T(R\pi_* f^*\mathcal{O}(d))^\vee$ is only defined after localization, so we first apply the virtual localization formula to the left-hand side, then we compute it in $K^0(T, \overline{\mathcal{M}}_{g,n})_{\mathrm{loc}} = K^0(\overline{\mathcal{M}}_{g,n}) \otimes \mathbb{C}(t_1, \dots, t_N)$ as rational fractions in the T–equivariant parameters, then we specialize them to $(t_1, \dots, t_N) = (\zeta, \zeta^{u_2}, \dots, \zeta^{u_N})$ using $\varphi\colon H \hookrightarrow T$ and obtain a well-defined K–class in $K^0(\overline{\mathcal{M}}_{g,n}) \otimes \mathbb{C}$. Eventually we take its Euler characteristic and land in $R(H)_{\mathbb{C},\mathrm{loc}} \simeq \mathbb{C}$, where the last isomorphism depends on the primitive q^{th} root of unity ζ.

Remark 2.7 The localization map $R(H) \to R(H)_{\mathbb{C},\mathrm{loc}}$ corresponds to the map $\mathbb{Z}[X]/(1 - X^q) \to \mathbb{C}$ sending the variable X to ζ.

Remark 2.8 In Theorem 2.6, it is important that for every nonzero element $h \in H$, the h–fixed locus consists of coordinate points in $\mathbb{P}^N$. It is guaranteed by Proposition 2.2 and the fact that $H \subset \mathbb{Z}/M\mathbb{Z}$.

Corollary 2.9 *We take the same notation and assumptions as in Theorem 2.6. We further assume that the order q of the group H is a prime number. For each $1 \le k < q$, denote by $B_k \in \mathbb{C}$ the result of Theorem 2.6 when $\zeta = e^{2\mathrm{i}k\pi/q}$. Then the K–theoretic GW invariant of X equals*

$$\chi(A \otimes \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{g,n}(X,\beta)}) \equiv -(B_1 + \dots + B_{q-1}) \in \mathbb{Z}/q\mathbb{Z}.$$

Proof The H–equivariant Euler characteristic $\chi^H(A \otimes \mathcal{O}^{\mathrm{vir},H}_{\overline{\mathcal{M}}_{g,n}(X,\beta)})$ lies in the representation ring $R(H) \simeq \mathbb{Z}[X]/(1-X^q)$, so there exist integers $a_0, \dots, a_{q-1}$ such that

$$\chi^H(A \otimes \mathcal{O}^{\mathrm{vir},H}_{\overline{\mathcal{M}}_{g,n}(X,\beta)}) = \sum_{l=0}^{q-1} a_l X^l \in \mathbb{Z}[X]/(1-X^q).$$

Our goal would be to compute

$$\chi(A \otimes \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{g,n}(X,\beta)}) = \sum_{l=0}^{q-1} a_l \in \mathbb{Z},$$

but Theorem 2.6 only gives us $\sum_{l=0}^{q-1} a_l \zeta^l \in \mathbb{C}$. However, since q is a prime number, we can apply Theorem 2.6 to every primitive q^{th} root of unity ζ^k for $1 \le k < q$. Summing the various results, we obtain

$$\sum_{k=1}^{q-1} \sum_{l=0}^{q-1} a_l \zeta^{kl} = q a_0 - \sum_{l=0}^{q-1} a_l \in \mathbb{Z},$$

leading to the congruence. □

Remark 2.10 Assume the order q of the group H is not a prime number and choose a nonzero element $h \in H$. Even when h is not a primitive element, we can apply Theorem 2.6 to the subgroup $\langle h \rangle$, but we then have the bounds

$$g < \tfrac{1}{2}(p-1) \quad \text{and} \quad \beta < \mathrm{ord}(h),$$

where $\mathrm{ord}(h)$ denotes the order of the element h, and p is its greatest prime divisor. In order to obtain the KGW invariant in H, we then need to sum all the results of Theorem 2.6 for all nonzero elements $h \in H$. Therefore, we have to restrict to the bounds

$$g < \tfrac{1}{2}(p-1) \quad \text{and} \quad \beta < p,$$

where p is the smallest prime divisor of q.

Example 2.11 For the quintic threefold of Example 2.3, the specialization of equivariant parameters corresponding to $G \hookrightarrow T$ is

$$(t_0, \dots, t_4) = (1, \zeta, \zeta^{-3}, \zeta^{13}, \zeta^{-51}), \quad \text{where } \zeta^{205} = 1.$$

Moreover, we have a subgroup $H := \mathbb{Z}/41\mathbb{Z} \subset \mathbb{Z}/205\mathbb{Z}$, so that by Corollary 2.9, we are able to compute all KGW invariants modulo 41 up to genus 19 and degree 40. Moreover, by Remark 2.10, we are able to compute all KGW invariants modulo 205 in genera 0 and 1 up to degree 4.

Remark 2.12 Another way to realize the quintic hypersurface in $\mathbb{P}^4$ is

$$X = \{x_0^5 + \cdots + x_4^5 = 0\} \subset \mathbb{P}^4.$$

Then the group is $(\mathbb{Z}/5\mathbb{Z})^4$, but to ensure that the g–fixed locus consists of isolated points for every element g of the group, we need to consider the subgroup $G = \mathbb{Z}/5\mathbb{Z}$, acting as

$$\zeta \cdot \underline{x} = (x_0, \zeta x_1, \zeta^2 x_2, \zeta^3 x_3, \zeta^4 x_4).$$

Furthermore, we observe that the G–fixed locus is empty. We then deduce that all KGW invariants in genera 0 and 1 and up to degree 4 vanish modulo 5.

2.5 Example of the quintic threefold

We illustrate Theorem 2.6 and Corollary 2.9 by a computation of the genus-one degree-one unmarked KGW invariant in the case of the quintic hypersurface in $\mathbb{P}^4$, modulo 205.

Proposition 2.13 *Let $X \subset \mathbb{P}^4$ be a smooth quintic hypersurface. We find that*

$$\chi\left(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{1,0}(X,1)}}{1 - q\mathbb{E}^\vee}\right) \equiv (120q^2 + 180q + 125)\frac{1 - q^4 - q^6}{(1-q^4)(1-q^6)} \in \mathbb{Z}/205\mathbb{Z}[\![q]\!].$$

In order to prove Proposition 2.13, we first write the general graph sum formula coming from torus localization and we then specialize to $(g, n, \beta) = (1, 0, 1)$.

Following the general scheme of Theorem 2.6, we compute the K–theoretic class

$$\chi^T\big(\lambda_{-1}^T(R\pi_* f^*\mathcal{O}(5))^\vee \otimes A \otimes \mathcal{O}^{\mathrm{vir},T}_{\overline{\mathcal{M}}_{g,n}(\mathbb{P}^4,\beta)}\big) \in K^0(\overline{\mathcal{M}}_{g,n}) \otimes \mathbb{C}(t_0, \dots, t_4).$$

It is done via the standard virtual localization formula of [17], lifted to K–theory, as a sum over dual graphs. Indeed, the class λ_{-1} is multiplicative in K–theory, just as the Euler class in cohomology, so that the whole proof of [17, Section 4] holds. Therefore, we take the same notation as in [17], to which we refer, for instance, for the description of graphs, except that we take the convention $t_j = e^{-\lambda_j}$ with respect to their T–weights.

Let Γ be a graph in the localization formula of $\mathbb{P}^4$. We denote by $\overline{\mathcal{M}}_\Gamma$ the associated moduli space of stable curves and by A_Γ the group of automorphisms coming from the graph Γ and from degrees of the edges, so that the corresponding fixed locus in $\overline{\mathcal{M}}_{g,n}(\mathbb{P}^4, \beta)$ is the quotient stack $[\overline{\mathcal{M}}_\Gamma/\mathrm{A}_\Gamma]$; see [17]. The contribution of the graph Γ to the localization formula is of the form

$$\chi\big([\overline{\mathcal{M}}_\Gamma/\mathrm{A}_\Gamma]; \mathrm{Contr(flags)} \cdot \mathrm{Contr(vertices)} \cdot \mathrm{Contr(edges)}\big),$$

where we have

$$\mathrm{Contr}(\text{flags}) = \frac{1}{1-\left(\dfrac{t_{i(F)}}{t_{j(F)}}\right)^{1/d_e}\Psi_F}\cdot\frac{\displaystyle\prod_{m\neq i(F)} 1-\frac{t_{i(F)}}{t_m}}{1-\dfrac{t_{i(F)}^5}{t_0^4t_1}},$$

$$\mathrm{Contr}(\text{vertices}) = \frac{1-\dfrac{t_{i(v)}^5}{t_0^4t_1}}{\displaystyle\sum_{k=0}^{g(v)}(-1)^k\left(\frac{t_{i(v)}^5}{t_0^4t_1}\right)^k\Lambda^k\mathbb{E}}\cdot\prod_{m\neq i(v)}\frac{\displaystyle\sum_{k=0}^{g(v)}(-1)^k\left(\frac{t_{i(v)}}{t_m}\right)^k\Lambda^k\mathbb{E}}{1-\dfrac{t_{i(v)}}{t_m}},$$

$$\begin{aligned}\mathrm{Contr}(\text{edges}) = \prod_{a=1}^{d_e}\left(2-\left(\frac{t_{j'}}{t_j}\right)^{a/d_e}-\left(\frac{t_j}{t_{j'}}\right)^{a/d_e}\right)^{-1}&\cdot\prod_{\substack{a+b=d_e\\k\neq j,j'}}\left(1-\frac{(t_j^at_{j'}^b)^{1/d_e}}{t_k}\right)^{-1}\\&\cdot\prod_{a+b=5d_e}\left(1-\frac{(t_j^at_{j'}^b)^{1/d_e}}{t_0^4t_1}\right),\end{aligned}$$

and where we write here the contribution of an edge linking the coordinate points p_j and $p_{j'}$. These formulae follow exactly from [17, Section 4], replacing the Euler class with the lambda class in K–theory.

Remark 2.14 In the contribution of vertices, we can rewrite the sum in terms of the lambda-structure as $\lambda_{-u}(\mathbb{E})$, with $u := t_{i(v)}^5/t_0^4t_1$.

Remark 2.15 All individual contributions are in $K^0(\overline{\mathcal{M}}_\Gamma)\otimes\mathbb{C}(t_0^{1/d},\dots,t_4^{1/d})_{d\in\mathbb{N}^*}$, but it is a consequence of the virtual localization formula that the final answer is in $K^0(\overline{\mathcal{M}}_\Gamma)\otimes\mathbb{C}(t_0,\dots,t_4)$.

Let us now specialize the formula to $(g,n,\beta)=(1,0,1)$. The graph Γ has only two vertices v_1 and v_2, of respective genera 1 and 0, and one degree-one edge in between. Moreover, as the vertex v_2 has valence one, it corresponds to a free point (not marked, not a node) rather than to a stable component of the curve. We denote by $0\leq i_1\neq i_2\leq 4$ the indices of the coordinate points p_{i_1} and p_{i_2} to which the vertices v_1 and v_2 are sent by the stable map. Note also that such a graph has no automorphisms and the moduli space $\overline{\mathcal{M}}_\Gamma$ is isomorphic to $\overline{\mathcal{M}}_{1,1}$. Furthermore, we recall that the Hodge bundle $\mathbb{E}$ over $\overline{\mathcal{M}}_{1,1}$ is identified with the cotangent line Ψ_1 at the marking. As a consequence, the virtual localization formula equals

$$\sum_{0\le i_1\ne i_2\le 4} \frac{\prod_{a+b=5}\left(1-\frac{t_{i_1}^a t_{i_2}^b}{t_0^4 t_1}\right)}{\left(2-\frac{t_{i_1}}{t_{i_2}}-\frac{t_{i_2}}{t_{i_1}}\right)\cdot \prod_{k\ne i_1,i_2}\left(1-\frac{t_{i_1}}{t_k}-\frac{t_{i_2}}{t_k}+\frac{t_{i_1}t_{i_2}}{t_k^2}\right)} \cdot \chi\left(\overline{\mathcal{M}}_{1,1};\frac{1}{1-q\mathbb{E}^\vee}\frac{\prod_{m\ne i_1,i_2}\left(1-\frac{t_{i_1}}{t_m}\Psi_1\right)}{1-\frac{t_{i_1}^5}{t_0^4t_1}\Psi_1}\right).$$

Once we specialize to $(t_0,\dots,t_4)=(1,\zeta,\zeta^{-3},\zeta^{13},\zeta^{-51})$, where ζ is any primitive root of unity of order 41, we notice that denominators never vanish, but the numerator could vanish; precisely,

$$1-\frac{t_{i_1}^a t_{i_2}^b}{t_0^4t_1}=0 \iff \begin{cases} i_2=i_1+1 \text{ and } (a,b)=(4,1), \text{ or} \\ i_1=i_2+1 \text{ and } (a,b)=(1,4), \end{cases}$$

with the cyclic convention on indices, ie $t_5:=t_0$. Moreover, we have

$$1-\frac{t_{i_1}}{t_{i_1+1}}\Psi_1 = 1-\frac{t_{i_1}^5}{t_0^4t_1}\Psi_1,$$

so that the specialization of the localization formula gives

$$\sum_{\substack{0\le i_1\ne i_2\le 4\\ i_2\ne i_1+1\\ i_1\ne i_2+1}} \left[\frac{\prod_{a+b=5}\left(1-\frac{t_{i_1}^a t_{i_2}^b}{t_0^4 t_1}\right)}{\left(2-\frac{t_{i_1}}{t_{i_2}}-\frac{t_{i_2}}{t_{i_1}}\right)\cdot \prod_{k\ne i_1,i_2}\left(1-\frac{t_{i_1}}{t_k}-\frac{t_{i_2}}{t_k}+\frac{t_{i_1}t_{i_2}}{t_k^2}\right)} \cdot\chi\left(\overline{\mathcal{M}}_{1,1};\frac{1}{1-q\mathbb{E}^\vee}\prod_{m\ne i_1,i_1+1,i_2}\left(1-\frac{t_{i_1}}{t_m}\Psi_1\right)\right)\right]_{(t_0,\dots,t_4)=(1,\zeta,\zeta^{-3},\zeta^{13},\zeta^{-51})}.$$

By [26, Proposition 2.9], we have

$$\chi\left(\overline{\mathcal{M}}_{1,1};\frac{1}{1-q\mathbb{E}^\vee}\frac{1}{1-q_1\Psi_1}\right) = \frac{(1-qq_1)(1-q^4-q^6-q_1^2q^6-q_1^2q^8-q_1^4q^8+q^2q_1^2+q^4q_1^4+q^6q_1^6+q^8q_1^8)}{(1-q^4)(1-q^6)(1-q_1^4)(1-q_1^6)}.$$

Hence, we get

$$\chi\Bigg(\overline{\mathcal{M}}_{1,1};\frac{\prod_{m\neq i_1,i_1+1,i_2}\Big(1-\frac{t_{i_1}}{t_m}\Psi_1\Big)}{1-q\mathbb{E}^\vee}\Bigg)=\frac{(1-q^4-q^6)}{(1-q^4)(1-q^6)}\prod_{m\neq i_1,i_1+1,i_2}\Big(1+\frac{t_{i_1}}{t_m}q\Big).$$

As a consequence, our formula simplifies as

$$\frac{1-q^4-q^6}{(1-q^4)(1-q^6)}\cdot\sum_{\substack{0\leq i_1\neq i_2\leq 4\\ i_2\neq i_1+1\\ i_1\neq i_2+1}}\left[\frac{\prod_{a+b=5}\Big(1-\frac{t_{i_1}^a t_{i_2}^b}{t_0^4 t_1}\Big)\prod_{m\neq i_1,i_1+1,i_2}\Big(1+\frac{t_{i_1}}{t_m}q\Big)}{\Big(2-\frac{t_{i_1}}{t_{i_2}}-\frac{t_{i_2}}{t_{i_1}}\Big)\cdot\prod_{k\neq i_1,i_2}\Big(1-\frac{t_{i_1}}{t_k}-\frac{t_{i_2}}{t_k}+\frac{t_{i_1}t_{i_2}}{t_k^2}\Big)}\right]_{(t_0,\dots,t_4)=(1,\zeta,\zeta^{-3},\zeta^{13},\zeta^{-51})}.$$

Finally, we must take the opposite of the sum of these expressions over all primitive roots ζ of order 41. First, we notice that the term inside the sum is a polynomial in q of degree at most two, so that it is enough to evaluate it at $q\in\{0,1,2\}$. Using Sagemath, we find

$$\begin{aligned}\chi\Bigg(\frac{\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{1,0}(X,1)}}{1-q\mathbb{E}^\vee}\Bigg)&\equiv(-85q^2+590q-80)\frac{1-q^4-q^6}{(1-q^4)(1-q^6)}\\&\equiv(38q^2+16q+2)\frac{1-q^4-q^6}{(1-q^4)(1-q^6)}\in\mathbb{Z}/41\mathbb{Z}[\![q]\!].\end{aligned}$$

Furthermore, using Remark 2.12, we obtain the result of Proposition 2.13.

2.6 Special case of elliptic curves

In this section, we use the ideas behind Corollary 2.9 to prove that KGW theory with homogeneous insertions of an elliptic curve is trivial.

Proposition 2.16 *Let E be an elliptic curve. Then for every genus g, degree β, number of markings n and insertions $A:=\bigotimes_{i=1}^n\Psi_i^{a_i}\otimes Y_i$, with $2g-2+n>0$ and $Y_i\in K^0(E)$ homogeneous K–classes, the corresponding KGW invariant vanishes*:

$$\chi(A\otimes\mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{g,n}(E,\beta)})=0.$$

Proof Let M be the largest possible order of an automorphism of a stable curve of genus g. Let p be any prime number larger than $M+1$ and $\beta+1$. Define $G:=\mathbb{Z}/p\mathbb{Z}$

and take a G–torsion point $x \in E$. Then the group G acts on the elliptic curve E by translation $y \mapsto y + x$, and for every nonzero element $h \in G$, the h–fixed locus is empty. By Remark 2.5 and Proposition 2.4, the G–fixed locus in the moduli space of stable maps $\overline{\mathcal{M}}_{g,n}(E, \beta)$ is empty. Therefore, by the localization formula, the G–equivariant KGW invariant vanishes, so that we get

$$\chi(A \otimes \mathcal{O}^{\mathrm{vir}}_{\overline{\mathcal{M}}_{g,n}(E,\beta)}) \equiv 0 \in \mathbb{Z}/p\mathbb{Z}$$

for the nonequivariant limit. Since it is true for infinitely many prime numbers p, we obtain the vanishing in $\mathbb{Z}$. □

Remark 2.17 Interestingly, KGW invariants are deduced from GW invariants via a Kawazaki–Riemann–Roch theorem; see [38; 16]. It would be instructive to compare Proposition 2.16 with GW theory of elliptic curves, which is nontrivial and described in [31; 32].

Remark 2.18 The same proof holds for abelian varieties. However, when the dimension of the abelian variety is greater than 2 and the degree-class β is nonzero, there is a trivial quotient of the obstruction theory, so that both GW and KGW theories are trivial. However, for degree-0 invariants, GW theory is nontrivial, but KGW theory is.

Remark 2.19 The main idea in the proof of Proposition 2.16 is to use congruence relations for infinitely many prime numbers. Indeed, if we were able to find, for a smooth DM stack X, automorphisms of prime orders for infinitely many primes and to compute the localization formulae, then we would be able to know all KGW invariants of X. Therefore, a necessary condition is that the automorphism group of X must be infinite. However, it is not sufficient. For instance, some K3 surfaces have infinitely many symmetries, but it was shown by [23] that the maximal order of a finite group acting faithfully on a K3 surface is 3840.

Remark 2.20 Here are a few remarks on finiteness of automorphism groups. For projective hypersurfaces (except quadrics, elliptic curves, and K3 surfaces), every automorphism is projective and the automorphism group is finite. All Batyrev Calabi–Yau (CY) 3–folds have finite automorphism groups [35]. Every projective variety of general type has finite automorphism group. CY varieties with Picard numbers 1 or 2 have finite automorphism groups. It is expected that most CY varieties with Picard number more than 4 have infinitely many automorphisms. In particular, it would be interesting to know whether the Schoen CY 3–fold has automorphisms of prime order for infinitely many primes and to study its KGW theory; see [21].

3 K–theoretic FJRW theory

Similarly to KGW theory, we aim in this section to compute the K–theoretic FJRW invariants of a Landau–Ginzburg (LG) orbifold modulo prime numbers. For simplicity of the exposition, we focus in this paper on the quintic polynomial with minimal group of symmetries. However, it is straightforward to apply the same ideas to an LG orbifold (W, H), where W is an invertible polynomial and H is an admissible group, as long as we only insert $\mathrm{Aut}(W)$–invariant states in the correlator. We refer to [18] for details.

3.1 Sketch of Polishchuk–Vaintrob construction

Let $W(x_1, \dots, x_5)$ be a quintic polynomial in five variables and let μ_5 act on $\mathbb{C}^5$ by multiplication by a fifth root of unity. The moduli space used in FJRW theory of (W, μ_5) is the moduli space $\mathcal{S}^{1/5}_{g,n}$, which parametrizes $(\mathcal{C}, \sigma_1, \dots, \sigma_n, \mathcal{L}, \phi)$. Precisely, the curve $\mathcal{C}$ is an orbifold genus-g stable curve with isotropy group μ_5 at the markings $\sigma_1, \dots, \sigma_n$ and at the nodes (and trivial everywhere else), $\mathcal{L}$ is a line bundle on $\mathcal{C}$, and $\phi\colon \mathcal{L} \to \omega_{\log} := \omega_{\mathcal{C}}(\sigma_1 + \dots + \sigma_n)$ is an isomorphism.

Let π be the projection from the universal curve to $\mathcal{S}^{1/5}_{g,n}$ and $\mathcal{L}$ be the universal line bundle. In [33], Polishchuk and Vaintrob constructed resolutions $R\pi_*(\mathcal{L}^{\oplus 5}) = [A \to B]$ by vector bundles over $\mathcal{S}^{1/5}_{g,n}$ such that there exists some morphism

$$\alpha\colon \mathrm{Sym}^4 A \to B^\vee$$

corresponding to the differentiation of the polynomial W; see [18] for details. Taking $p\colon X \to \mathcal{S}^{1/5}_{g,n}$ to be the total space of the vector bundle A, then the morphism α is interpreted as a global section of $p^* B^\vee$ over X, and the map $\beta\colon A \to B$ coming from the resolution is interpreted as a global section of $p^* B$. As a consequence, we obtain a Koszul matrix factorization

$$\mathbf{PV} := \{\alpha, \beta\} := (\Lambda^\bullet B^\vee, \alpha \wedge \cdot + \iota_\beta) \in D(X, \alpha(\beta))$$

of potential $\alpha(\beta)$ over the space X, and the support of this matrix factorization is exactly the moduli space $\mathcal{S}^{1/5}_{g,n}$.

The moduli space $\mathcal{S}^{1/5}_{g,n}$ has several components depending on the monodromies

$$\underline{\gamma} := (\gamma_1, \dots, \gamma_n) \in \mu_5^n$$

at the markings; we denote by $\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})$ the corresponding component. Assume all monodromies are nonzero; this is known as the *narrow condition*. Then the pairing $\alpha(\beta)$ is the zero function over X, and the matrix factorization $\mathbf{PV}$ becomes a two-periodic complex, exact off the moduli space $\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})$. Therefore, we can define the pushforward along the projection map p in the category of matrix factorizations, yielding

$$p_*(\mathbf{PV}) \in D(\mathcal{S}^{1/5}_{g,n}(\underline{\gamma}), 0) \simeq D^b(\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})),$$

where on the right we have the derived category of coherent sheaves.

Remark 3.1 If one allows trivial monodromies (ie one considers broad insertions), then the pairing $\alpha(\beta)$ does not vanish and we rather end with a functor

$$\begin{aligned}\Phi\colon D_\Gamma(\mathbb{A}^{\underline{\gamma}}, W_{\underline{\gamma}}) &\to D_{\overline{H}}(\mathcal{S}^{\mathrm{rig}}_{g,n}(\underline{\gamma}), 0) \simeq D([\mathcal{S}^{\mathrm{rig}}_{g,n}(\underline{\gamma})/\overline{H}]),\\ U &\mapsto p_*(\mathrm{ev}^*(U) \otimes \mathbf{PV}),\end{aligned}$$

where we need to consider rigidified moduli spaces; see [33] for details and notation.

In general, to any triangulated category $\mathcal{C}$ we associate a Grothendieck group $K_0(\mathcal{C})$ by taking the free abelian group generated by the objects of the category and then modding out the relation

$$[A] - [B] + [C] = 0$$

for every distinguished triangle $A \to B \to C$. Furthermore, any functor $f\colon \mathcal{C}_1 \to \mathcal{C}_2$ of triangulated dg categories induces a morphism of groups

$$f_!\colon K_0(\mathcal{C}_1) \to K_0(\mathcal{C}_2).$$

When the category is the derived category of coherent sheaves on a smooth DM stack, we recover the usual K–theory of the stack.

Remark 3.2 If we apply it to the functor of Remark 3.1, we get a morphism of groups

$$\Phi_! := K_0(\Phi)\colon K_0(D_\Gamma(\mathbb{A}^{\underline{\gamma}}, W_{\underline{\gamma}})) \to K^0([\mathcal{S}^{\mathrm{rig}}_{g,n}(\underline{\gamma})/\overline{H}]).$$

Definition 3.3 We define the K–theoretic class

$$\mathcal{O}^{\mathrm{vir}}_{\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})} := p_*(\mathbf{PV}) \in K^0(\mathcal{S}^{1/5}_{g,n}(\underline{\gamma}))$$

and we call it the virtual structure sheaf of the moduli space $\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})$.

Definition 3.4 Fix $d_1, \dots, d_n \in \mathbb{Z}$ and nontrivial monodromies $\gamma_1, \dots, \gamma_n \in \mu_5$. The K–theoretic FJRW invariant of the LG orbifold (W, μ_5) is

$$\chi(\Psi_1^{\otimes d_1} \otimes \cdots \otimes \Psi_n^{\otimes d_n} \otimes \mathcal{O}^{\mathrm{vir}}_{\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})}) \in \mathbb{Z},$$

where $\chi\colon K^0(\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})) \to K^0(\mathrm{pt}) = \mathbb{Z}$ is the Euler characteristic and the line bundle Ψ_i is the relative cotangent line at the i^{th} marked point.

Remark 3.5 A special feature of the LG orbifold (W, μ_5), and more generally when the polynomial has degree d and the group is μ_d, is that FJRW invariants do not depend on the polynomial W, as long as it is nondegenerate with the same weights and degree; see [12, Proposition 4.1.7]. The same result holds for the K–theoretic version, as it holds at the matrix factorization level. Hence, we can consider several choices for our quintic polynomial.

3.2 Invertible polynomials

In the context of mirror symmetry, a well-behaved class of polynomials has been introduced by Berglund and Hübsch [3]. We say that a polynomial is invertible when it is nondegenerate and has as many variables as monomials. According to Kreuzer and Skarke [24], every invertible polynomial is a (Thom–Sebastiani) sum of invertible polynomials, with disjoint sets of variables, of the following three types:

$$\begin{array}{lll} \textbf{Fermat} & x^{a+1}, & \\ \textbf{chain of length } \boldsymbol{c} & x_1^{a_1}x_2 + \cdots + x_{c-1}^{a_{c-1}}x_c + x_c^{a_c+1}, & \text{where } c \geq 2, \\ \textbf{loop of length } \boldsymbol{l} & x_1^{a_1}x_2 + \cdots + x_{l-1}^{a_{l-1}}x_l + x_l^{a_l}x_1, & \text{where } l \geq 2. \end{array} \tag{8}$$

In the case of the quintic polynomial, we have for example the following choices:

$$\begin{array}{lll} \textbf{Fermat} & x_1^5 + x_2^5 + x_3^5 + x_4^5 + x_5^5, & (\mu_5)^5, \\ \textbf{loop} & x_1^4x_2 + x_2^4x_3 + x_3^4x_4 + x_4^4x_5 + x_5^4x_1, & \mu_{1025}, \\ \textbf{chain} & x_1^4x_2 + x_2^4x_3 + x_3^4x_4 + x_4^4x_5 + x_5^5, & \mu_{1280}, \\ \textbf{2–loops} & x_1^4x_2 + x_2^4x_3 + x_3^4x_1 + x_4^4x_5 + x_5^4x_4, & \mu_{15} \times \mu_{65}, \\ \textbf{loop-Fermat} & x_1^4x_2 + x_2^4x_3 + x_3^4x_4 + x_4^4x_1 + x_5^5, & \mu_{255} \times \mu_5, \end{array} \tag{9}$$

where on the right we have written the group $\mathrm{Aut}(W)$ of diagonal matrices leaving the polynomial invariant.

Let W be an invertible quintic polynomial and $\mathrm{Aut}(W)$ be its maximal group of diagonal symmetries. Recall that the space X is the total space of the vector bundle A over the moduli space $\mathcal{S}_{g,n}^{1/5}(\underline{\gamma})$, and that we have

$$R\pi_*(\mathcal{L}^{\oplus 5}) = [A \to B].$$

Therefore, the vector bundles A and B are direct sums of five copies, which we write as

$$A = A_1 \oplus \cdots \oplus A_5 \quad \text{and} \quad B = B_1 \oplus \cdots \oplus B_5.$$

We then have a natural action of $\mathrm{Aut}(W)$ on the vector bundles A and B by rescaling the fibers. Precisely, the actions on X in the examples (9) are

$$\begin{array}{lll}
\textbf{Fermat} & (\zeta_1 x_1, \zeta_2 x_2, \zeta_3 x_3, \zeta_4 x_4, \zeta_5 x_5), & \zeta_j = e^{2\mathrm{i}\pi/5}, \\
\textbf{loop} & (\zeta x_1, \zeta^{-4} x_2, \zeta^{16} x_3, \zeta^{-64} x_4, \zeta^{256} x_5), & \zeta = e^{2\mathrm{i}\pi/1025}, \\
\textbf{chain} & (\zeta x_1, \zeta^{-4} x_2, \zeta^{16} x_3, \zeta^{-64} x_4, \zeta^{256} x_5), & \zeta = e^{2\mathrm{i}\pi/1280}, \\
\textbf{2–loops} & (\zeta_1 x_1, \zeta_1^{-4} x_2, \zeta_1^{16} x_3, \zeta_2 x_4, \zeta_2^{-4} x_5), & \zeta_1 = e^{2\mathrm{i}\pi/65}, \zeta_2 = e^{2\mathrm{i}\pi/15}, \\
\textbf{loop-Fermat} & (\zeta_1 x_1, \zeta_1^{-4} x_2, \zeta_1^{16} x_3, \zeta_1^{-64} x_4, \zeta_2 x_5), & \zeta_1 = e^{2\mathrm{i}\pi/255}, \zeta_2 = e^{2\mathrm{i}\pi/5}.
\end{array} \tag{10}$$

By construction, since the polynomial W is $\mathrm{Aut}(W)$–invariant, the matrix factorization $\mathbf{PV}$ is $\mathrm{Aut}(W)$–equivariant and so is the virtual structure sheaf.

However, we need to be careful when we compute the $\mathrm{Aut}(W)$–fixed locus. Indeed, the group of automorphisms of a (W, μ_5)–spin curve $(\mathcal{C}, \sigma_1, \dots, \sigma_n, \mathcal{L})$ fixing the coarse curve of $\mathcal{C}$ is $\mu_5 \times (\mu_5)^{\#\mathrm{nodes}}$, where the first factor rescales the line bundle $\mathcal{L}$, and the second factor acts only on the orbifold curve $\mathcal{C}$ — it is the so-called ghost automorphism; see [1, Proposition 7.1.1] and [7, Section 2.1.4]. As a consequence, we would rather consider the action of the group

$$G := \mathrm{Aut}(W)/\mu_5$$

on the space X over the moduli space $\mathcal{S}_{g,n}^{1/5}(\underline{\gamma})$. Here, the subgroup μ_5 is generated by $(e^{2\mathrm{i}\pi/5}, \dots, e^{2\mathrm{i}\pi/5})$ in the Fermat case, by $e^{2\mathrm{i}\pi/5}$ in the loop and chain cases, and by $(e^{2\mathrm{i}\pi/5}, e^{2\mathrm{i}\pi/5})$ in the 2–loops and loop-Fermat cases.

Unfortunately, we still have too many fixed points. For instance, in the Fermat example, the point

$$(\mathcal{C}, \sigma_1, \dots, \sigma_n, \mathcal{L}; x_1, 0, \dots, 0) \in X$$

is fixed. Another example is the chain polynomial, for which the point

$$(\mathcal{C}, \sigma_1, \dots, \sigma_n, \mathcal{L}; 0, \dots, 0, x_5) \in X$$

is fixed. In both cases, the G–fixed locus is noncompact. We easily check that, among all invertible quintic polynomials, the only cases where the G–fixed locus is compact are

- the loop polynomial with group $G = \mathrm{Aut}(W)/\mu_5 = \mu_{205}$,
- the 2–loops polynomial with group $G = (\mu_{65} \times \mu_{15})/\mu_5$.

Moreover, the G–fixed locus in the space X equals the base $\mathcal{S}_{g,n}^{1/5}(\underline{\gamma})$. We are therefore able to apply the (nonvirtual) localization formula on the smooth space X to get the next theorem.

Remark 3.6 It is more convenient to work with cyclic groups. Therefore, in the 2–loops polynomial case, we prefer to use $G = \mu_{195}$, where the G–action on X is

$$(\zeta^{15}x_1, \zeta^{-60}x_2, \zeta^{240}x_3, \zeta^{65}x_4, \zeta^{-260}x_5), \quad \text{where } \zeta = e^{2i\pi/195}.$$

Definition 3.7 Let $l \in \mathbb{Z}$. The Adams operation Ψ^l in K–theory is defined on a line bundle L over a space S as

$$\Psi^l(L) := L^{\otimes l},$$

and then extended as a ring homomorphism

$$\Psi^l \colon K^0(S) \to K^0(S).$$

Theorem 3.8 *Consider the two following situations*:

- *W is the loop polynomial, $G := \mu_{205}$, ζ a primitive 205^{th} root of unity, and $(a_1, \dots, a_5) = (1, -4, 16, -64, 256)$.*
- *W is the 2–loops polynomial, $G := \mu_{195}$, ζ is a primitive 195^{th} root of unity, and $(a_1, \dots, a_5) = (15, -60, 240, 65, -260)$.*

Let g and n be nonnegative integers in the stable range $2g-2+n>0$, and let $\underline{\gamma} \in \mu_5^n$ be nontrivial monodromies. Then the G–equivariant virtual structure sheaf equals

$$\mathcal{O}^{\mathrm{vir},G}_{\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})} = \exp\Bigg(\sum_{l \leq -1} \sum_{j=1}^{5} \frac{\zeta^{a_j \cdot l}}{l} \Psi^l(-R\pi_*\mathcal{L}) \Bigg) \in K^0(\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})) \otimes \mathbb{C}. \tag{11}$$

Proof In the G–equivariant K–theory of the space X, the matrix factorization equals

$$\mathbf{PV} = \lambda^G_{-1} p^* B^\vee \in K^0(G, X),$$

and by the localization formula we get

$$\mathbf{PV} = \iota_! \left(\frac{\lambda^G_{-1} B^\vee}{\lambda^G_{-1} A^\vee} \right) = \iota_!(\lambda^G_{-1}(B^\vee - A^\vee)) \in K^0(G, X)_{\mathrm{loc}}$$

in the localized ring, where $\iota \colon \mathcal{S}^{1/5}_{g,n}(\underline{\gamma}) \hookrightarrow X$ is the zero section. Taking the pushforward along the projection map p, we obtain the G–equivariant virtual structure sheaf

$$\mathcal{O}^{\mathrm{vir},G}_{\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})} = \lambda^G_{-1}(B^\vee - A^\vee) \in K^0(G, \mathcal{S}^{1/5}_{g,n}(\underline{\gamma}))_{\mathrm{loc}} \simeq K^0(\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})) \otimes \mathbb{C}.$$

If V is a vector bundle, we can express the λ–structure in terms of Adams operators via the formula

$$\lambda_{-p}(V^\vee) = \exp\Bigg(\sum_{l \leq -1} \frac{p^{-l}}{l} \Psi^l(V) \Bigg).$$

Moreover, if the action of a group G on the vector bundle V is by rescaling fibers with $\zeta \in G$, then

$$\lambda^G_{-1}(V^\vee) = \lambda_{-\zeta^{-1}}(V^\vee) = \exp\Bigl(\sum_{l\leq -1} \frac{\zeta^l}{l}\Psi^l(V)\Bigr).$$

In our situation, we find formula (11). □

Remark 3.9 In [18], we define the characteristic class $\mathfrak{c}_t \colon K^0(S) \to H^*(S)[\![t]\!]$ by

$$\mathfrak{c}_t(B-A) = \mathrm{Ch}(\lambda_{-t}(B^\vee - A^\vee))\,\mathrm{Td}(B-A),$$

and we then obtain the formula

$$\lim_{t\to 1} \prod_{i=1}^{5} \mathfrak{c}_{t_j}(-R\pi_*\mathcal{L}) = c_{\mathrm{vir}} \in H^*(\mathcal{S}^{1/5}_{0,n}(\underline{\gamma}))$$

for the FJRW virtual cycle of (W,μ_5), where $t_j := t^{a_j}$. This formula is only valid in genus 0 and we do not expect the left-hand side to converge in positive genus when $t \to 1$. However, by Theorem 3.8, we see that the formula converges for every genus when $t \to \zeta$.

In order to get congruences for the nonequivariant limit, we need to consider a subgroup of $\mathrm{Aut}(W)$ with prime order and whose fixed locus in X is compact. The only invertible polynomial for which it is possible is the loop polynomial, together with the subgroup μ_{41} acting on X as

$$(\zeta_{41}x_1, \zeta_{41}^{37}x_2, \zeta_{41}^{16}x_3, \zeta_{41}^{18}x_4, \zeta_{41}^{6}x_5), \quad \text{where } \zeta_{41} := e^{2i\pi/41}.$$

Remark 3.10 The prime decomposition of 205 is $5\cdot 41$, so we could also hope for a congruence modulo 5. However, the subgroup μ_5 acts trivially on X. Indeed, it acts as

$$(\zeta_5x_1, \zeta_5x_2, \zeta_5x_3, \zeta_5x_4, \zeta_5x_5), \quad \text{where } \zeta_5 := e^{2i\pi/5},$$

which is rescaled by the automorphism group of the (W,μ_5)–spin curve, so that the fixed locus is X. Nevertheless, from its definition using the quintic Fermat polynomial, we observe that the virtual structure sheaf decomposes into five identical summands, each one corresponding to the so-called 5–spin theory. It is then divisible by five in the K–theoretic ring with $\mathbb{Z}$ coefficients. As a consequence, all FJRW correlators of the quintic vanish modulo 5.

Corollary 3.11 *Let W be the loop polynomial and let*

$$(a_1,\dots,a_5) = (1,-4,16,-64,256).$$

For any nonnegative integers g and n in the stable range $2g-2+n>0$, nontrivial monodromies $\underline{\gamma} \in \mu_5^n$, and integers $d_1,\dots,d_n \in \mathbb{Z}$, the K–theoretic FJRW invariant *of* (W,μ_5),

$$\chi(\Psi_1^{\otimes d_1}\otimes\cdots\otimes\Psi_n^{\otimes d_n}\otimes\mathcal{O}^{\mathrm{vir}}_{\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})})\in\mathbb{Z},$$

equals

$$-\sum_{k=1}^{40}\chi\biggl(\Psi_1^{\otimes d_1}\otimes\cdots\otimes\Psi_n^{\otimes d_n}\otimes\exp\biggl(\sum_{l\leq -1}\sum_{j=1}^{5}\frac{e^{2\mathrm{i}\pi kl\cdot a_j/41}}{l}\Psi^l(-R\pi_*\mathcal{L})\biggr)\biggr)\in\mathbb{Z}/41\mathbb{Z}.$$

Using Remark 3.10, the correlator vanishes modulo 5, *and we can then compute it modulo* 205.

Remark 3.12 More generally, the K–class

$$B_{41}:=-\sum_{k=1}^{40}\exp\biggl(\sum_{l\leq -1}\sum_{j=1}^{5}\frac{e^{2\mathrm{i}\pi kl\cdot a_j/41}}{l}\Psi^l(-R\pi_*\mathcal{L})\biggr)\in K^0(G,\mathcal{S}^{1/5}_{g,n}(\underline{\gamma}))$$

lies in the K–theoretic ring with $\mathbb{Z}$–coefficients and we know there exists another K–class R in $K^0(G,\mathcal{S}^{1/5}_{g,n}(\underline{\gamma}))$ with $\mathbb{Z}$–coefficients such that

$$\mathcal{O}^{\mathrm{vir}}_{\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})}=B_{41}+41\cdot R\in K^0(G,\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})).$$

This yields the following formula for the FJRW virtual cycle of (W,μ_5):

$$c_{\mathrm{vir}}=\sum_{k=1}^{40}\biggl(\lim_{t\to\zeta_{41}^k}\prod_{i=1}^{5}\mathfrak{c}_{t_j}(-R\pi_*\mathcal{L})+41\cdot\mathrm{Ch}(R)\biggr)\mathrm{Td}(-R\pi_*\mathcal{L})^5\in H^*(\mathcal{S}^{1/5}_{g,n}(\underline{\gamma}))_{\mathbb{Q}}.$$

However, since the cohomology is taken with $\mathbb{Q}$–coefficients, we do not obtain congruence results on the virtual cycle. An idea would be to guess a formula for the K–class R as an integral linear combination of (natural) vector bundles over $\mathcal{S}^{1/5}_{g,n}(\underline{\gamma})$. Since the virtual cycle is pure-dimensional and the right-hand side of the formula above is most likely not pure-dimensional when we take a generic R, only special integral coefficients in this linear combination would work.

References

[1] **D Abramovich**, **A Corti**, **A Vistoli**, *Twisted bundles and admissible covers*, Comm. Algebra 31 (2003) 3547–3618 MR Zbl

[2] **K Behrend**, **B Fantechi**, *The intrinsic normal cone*, Invent. Math. 128 (1997) 45–88 MR Zbl

[3] **P Berglund**, **T Hübsch**, *A generalized construction of mirror manifolds*, Nuclear Phys. B 393 (1993) 377–391 MR Zbl

[4] **P Candelas**, **X C de la Ossa**, **P S Green**, **L Parkes**, *A pair of Calabi–Yau manifolds as an exactly soluble superconformal theory*, Nuclear Phys. B 359 (1991) 21–74 MR Zbl

[5] **H-L Chang**, **S Guo**, **J Li**, *BCOV's Feynman rule of quintic* 3*–folds*, preprint (2019) arXiv 1810.00394v2

[6] **H-L Chang**, **J Li**, *Gromov–Witten invariants of stable maps with fields*, Int. Math. Res. Not. 2012 (2012) 4163–4217 MR Zbl

[7] **A Chiodo**, **G Farkas**, *Singularities of the moduli space of level curves*, J. Eur. Math. Soc. 19 (2017) 603–658 MR Zbl

[8] **A Chiodo**, **H Iritani**, **Y Ruan**, *Landau–Ginzburg/Calabi–Yau correspondence, global mirror symmetry and Orlov equivalence*, Publ. Math. Inst. Hautes Études Sci. 119 (2014) 127–216 MR Zbl

[9] **A Chiodo**, **Y Ruan**, *Landau–Ginzburg/Calabi–Yau correspondence for quintic three-folds via symplectic transformations*, Invent. Math. 182 (2010) 117–165 MR Zbl

[10] **A Chiodo**, **Y Ruan**, *A global mirror symmetry framework for the Landau–Ginzburg/ Calabi–Yau correspondence*, Ann. Inst. Fourier (Grenoble) 61 (2011) 2803–2864 MR Zbl

[11] **H Fan**, **T J Jarvis**, **Y Ruan**, *The Witten equation and its virtual fundamental cycle*, preprint (2007) arXiv 0712.4025

[12] **H Fan**, **T Jarvis**, **Y Ruan**, *The Witten equation, mirror symmetry, and quantum singularity theory*, Ann. of Math. 178 (2013) 1–106 MR Zbl

[13] **H Fan**, **Y-P Lee**, *Towards a quantum Lefschetz hyperplane theorem in all genera*, Geom. Topol. 23 (2019) 493–512 MR Zbl

[14] **A Givental**, *A mirror theorem for toric complete intersections*, from "Topological field theory, primitive forms and related topics" (M Kashiwara, A Matsuo, K Saito, I Satake, editors), Progr. Math. 160, Birkhäuser, Boston, MA (1998) 141–175 MR Zbl

[15] **A B Givental**, *Permutation-equivariant quantum K–theory*, *I*: *Definitions. Elementary K–theory of* $\overline{\mathcal{M}}_{0,n}/S_n$, preprint (2015) arXiv 1508.02690

[16] **A B Givental**, *Permutation-equivariant quantum K–theory*, *IX*: *Quantum Hirzebruch–Riemann–Roch in all genera*, preprint (2017) arXiv 1709.03180

[17] **T Graber**, **R Pandharipande**, *Localization of virtual classes*, Invent. Math. 135 (1999) 487–518 MR Zbl

[18] **J Guéré**, *A Landau–Ginzburg mirror theorem without concavity*, Duke Math. J. 165 (2016) 2461–2527 MR Zbl

[19] **J Guéré**, *Hodge–Gromov–Witten theory*, preprint (2019) arXiv 1908.11409

[20] **S Guo**, **F Janda**, **Y Ruan**, *Structure of higher genus Gromov–Witten invariants of quintic* 3*–folds*, preprint (2018) arXiv 1812.11908

[21] **Y-H He**, *The Calabi–Yau landscape: from geometry, to physics, to machine learning*, Lecture Notes in Math. 2293, Springer (2021) MR Zbl

[22] **H Jockers**, **P Mayr**, *Quantum K–theory of Calabi–Yau manifolds*, J. High Energy Phys. (2019) 1–20 MR Zbl

[23] **S Kondō**, *The maximum order of finite groups of automorphisms of K*3 *surfaces*, Amer. J. Math. 121 (1999) 1245–1252 MR Zbl

[24] **M Kreuzer**, **H Skarke**, *On the classification of quasihomogeneous functions*, Comm. Math. Phys. 150 (1992) 137–147 MR Zbl

[25] **Y-P Lee**, *Quantum K–theory, I: Foundations*, Duke Math. J. 121 (2004) 389–424 MR Zbl

[26] **Y-P Lee**, **F Qu**, *Euler characteristics of universal cotangent line bundles on $\overline{\mathcal{M}}_{1,n}$*, Proc. Amer. Math. Soc. 142 (2014) 429–440 MR Zbl

[27] **H Lho**, **R Pandharipande**, *Holomorphic anomaly equations for the formal quintic*, Peking Math. J. 2 (2019) 1–40 MR Zbl

[28] **B H Lian**, **K Liu**, **S-T Yau**, *Mirror principle, I*, Asian J. Math. 1 (1997) 729–763 MR Zbl

[29] **K Liu**, **H Xu**, *Intersection numbers and automorphisms of stable curves*, Michigan Math. J. 58 (2009) 385–400 MR Zbl

[30] **K Oguiso**, **X Yu**, *Automorphism groups of smooth quintic threefolds*, Asian J. Math. 23 (2019) 201–256 MR Zbl

[31] **A Okounkov**, **R Pandharipande**, *Gromov–Witten theory, Hurwitz theory, and completed cycles*, Ann. of Math. 163 (2006) 517–560 MR Zbl

[32] **A Okounkov**, **R Pandharipande**, *Virasoro constraints for target curves*, Invent. Math. 163 (2006) 47–108 MR Zbl

[33] **A Polishchuk**, **A Vaintrob**, *Matrix factorizations and cohomological field theories*, J. Reine Angew. Math. 714 (2016) 1–122 MR Zbl

[34] **F Qu**, *Virtual pullbacks in K–theory*, Ann. Inst. Fourier (Grenoble) 68 (2018) 1609–1641 MR Zbl

[35] **M F Tehrani**, *Automorphism group of Batyrev Calabi–Yau threefolds*, Manuscripta Math. 146 (2015) 299–306 MR Zbl

[36] **R W Thomason**, *Equivariant resolution, linearization, and Hilbert's fourteenth problem over arbitrary base schemes*, Adv. in Math. 65 (1987) 16–34 MR Zbl

[37] **R W Thomason**, *Une formule de Lefschetz en K–théorie équivariante algébrique*, Duke Math. J. 68 (1992) 447–462 MR Zbl

[38] **V Tonita**, *A virtual Kawasaki–Riemann–Roch formula*, Pacific J. Math. 268 (2014) 249–255 MR Zbl

Institut Fourier, CNRS, Université de Grenoble Alpes
Grenoble, France

jeremy.guere@univ-grenoble-alpes.fr

Proposed: Jim Bryan
Seconded: Dan Abramovich, Richard P Thomas

Received: 16 April 2021
Revised: 7 January 2022

Geometry & Topology **27**:9 (2023) 3619–3698
DOI: 10.2140/gt.2023.27.3619
Published: 5 December 2023

Moduli of spherical tori with one conical point

ALEXANDRE EREMENKO
GABRIELE MONDELLO
DMITRI PANOV

We determine the topology of the moduli space $\mathcal{MS}_{1,1}(\vartheta)$ of surfaces of genus one with a Riemannian metric of constant curvature 1 and one conical point of angle $2\pi\vartheta$. In particular, for $\vartheta \in (2m-1, 2m+1)$ nonodd, $\mathcal{MS}_{1,1}(\vartheta)$ is connected, has orbifold Euler characteristic $-\frac{1}{12}m^2$, and its topology depends on the integer $m > 0$ only. For $\vartheta = 2m+1$ odd, $\mathcal{MS}_{1,1}(\vartheta)$ has $\lceil \frac{1}{6}m(m+1) \rceil$ connected components. For $\vartheta = 2m$ even, $\mathcal{MS}_{1,1}(\vartheta)$ has a natural complex structure and it is biholomorphic to $\mathbb{H}^2/G_m$ for a certain subgroup G_m of $\mathrm{SL}(2,\mathbb{Z})$ of index m^2, which is nonnormal for $m > 1$.

58D27; 32Q20, 52B70

1 Introduction and main results

A spherical metric on a surface S with *conical points* at the points $\boldsymbol{x} = \{x_1, \dots, x_n\} \in S$ is a Riemannian metric of curvature 1 on $\dot{S} := S \setminus \boldsymbol{x}$ such that a neighborhood of x_j is isometric to a cone with a conical angle $2\pi\vartheta_j > 0$.

Let us immediately specify what we mean by the moduli space $\mathcal{MS}_{g,n}(\vartheta)$ of spherical surfaces. As a set, $\mathcal{MS}_{g,n}(\vartheta)$ parametrizes compact connected oriented surfaces of genus g with a spherical metric that has conical angles $(2\pi\vartheta_1,\dots,2\pi\vartheta_n)$ at marked points $x_1,\dots,x_n$. Two surfaces correspond to the same point of the space if there is a marked isometry from one to the other. In order to define a topology on $\mathcal{MS}_{g,n}(\vartheta)$, we consider the bi-Lipschitz distance between marked surfaces; see Gromov [15]. Such a distance defines a metric, and the corresponding topology on $\mathcal{MS}_{g,n}(\vartheta)$ is called the *Lipschitz topology*; its properties are discussed in Section 6.

As a spherical metric defines a conformal structure on the surface, we have the *forgetful map* $F\colon \mathcal{MS}_{g,n}(\theta)\to\mathcal{M}_{g,n}$, where $\mathcal{M}_{g,n}$ is the moduli space of conformal structures on $(S,\boldsymbol{x})$.

Since a neighborhood of a smooth point on S is isometric to an open set on the sphere equipped with the standard spherical metric, by an analytic continuation we obtain an orientation-preserving locally isometric *developing map* $f:\dot{S}\to\mathbb{S}^2$. Strictly speaking, the developing map is defined on the universal cover of $\dot{S}$ but it is sometimes convenient to think of it as a multivalued function on $\dot{S}$.

The developing map defines a representation of the fundamental group of $\dot{S}$ to the group SO(3) of rotations of the unit sphere $\mathbb{S}^2$. The image of this representation is called the *monodromy group*.

Our goal is to provide an explicit description of the moduli space $\mathcal{MS}_{1,1}(\vartheta)$ of spherical tori with one conical point.

Spherical tori with one conical point were also studied by Chai, Lin and Wang [2], Chen and Lin [6], Chen, Kuo and Lin [5], Eremenko [10], Eremenko and Gabrielov [11] and Lin and Wang [19; 20].

1.1 Main results

Our main results consist of Theorems A–F, which are stated in the next three subsections.

1.1.1 ϑ not an odd integer

Theorem A (topology of $\mathcal{MS}_{1,1}(\vartheta)$ for ϑ not odd) *Take $\vartheta\in(1,\infty)$ that is not an odd integer and set $m=\lfloor\frac{1}{2}(\vartheta+1)\rfloor$. The moduli space $\mathcal{MS}_{1,1}(\vartheta)$ of spherical tori with a conical point of angle $2\pi\vartheta$ is a connected orientable 2–dimensional orbifold of finite type with the following properties*:

(i) *As a surface,* $\mathcal{MS}_{1,1}(\vartheta)$ *has genus* $\lfloor\frac{1}{12}(m^2-6m+12)\rfloor$ *and* m *punctures.*

(ii) $\mathcal{MS}_{1,1}(\vartheta)$ *has orbifold Euler characteristic* $\chi(\mathcal{MS}_{1,1}(\vartheta))=-\frac{1}{12}m^2$*. Moreover, it has at most one orbifold point of order* 4 *and at most one orbifold point of order* 6*. All the other points are orbifold points of order* 2*.*

(iii) $\mathcal{MS}_{1,1}(\vartheta)$ *has one orbifold point of order* 6 *if and only if* $d_1(\vartheta, 6\mathbb{Z})>1$*.*

(iv) $\mathcal{MS}_{1,1}(\vartheta)$ *has one orbifold point of order* 4 *if and only if* $d_1(\vartheta, 4\mathbb{Z})>1$*.*

Note that for $\vartheta=2m$ this theorem gives a positive answer to the question of Chai, Lin and Wang [2, Question 4.6.6(a)] as to whether $\mathcal{MS}_{1,1}(2m)$ is connected.

We refer to Cooper, Hodgson and Kerckhoff [7] for a general treatment of orbifolds. In fact we adopt a slightly more general definition of orbifolds that includes the case in which all points can have orbifold order greater than 1. The definition of orbifold Euler characteristic is given on page 29 of [7]. This is consistent with the definition used, for example, in Harer and Zagier [16]. A few properties of the orbifold Euler characteristic are listed in Remark 4.7.

Note that, in [13], with Gabrielov we used a different convention and we endowed our moduli spaces with an orbifold structure for which the order of each point is half the number of automorphisms of the corresponding object. Thus, the orbifold Euler characteristics computed in [13] are twice the ones that would be obtained following the convention here.

Remark 1.1 (orbifold structure and isometric involution) For ϑ not odd, spherical metrics in $\mathcal{MS}_{1,1}(\vartheta)$ are invariant under the unique conformal involution σ of tori (see Proposition 2.17). Thus every such spherical torus is a double cover of a spherical surface of genus 0 with conical points of angles $(\pi\vartheta,\pi,\pi,\pi)$, and so the moduli space $\mathcal{MS}_{1,1}(\vartheta)$ is homeomorphic to $\mathcal{MS}_{0,4}\big(\frac{1}{2}\vartheta,\frac{1}{2},\frac{1}{2},\frac{1}{2}\big)/\mathrm{S}_3$ as a topological space. On the other hand, the orbifold order of a point in $\mathcal{MS}_{1,1}(\vartheta)$ exactly corresponds to the number of (orientation-preserving) self-isometries of the corresponding spherical torus. This explains why every point of $\mathcal{MS}_{1,1}(\vartheta)$ has even orbifold order, as stated in Theorem A. Thus $\mathcal{MS}_{1,1}(\vartheta)$ is not isomorphic to the orbifold quotient $\mathcal{MS}_{0,4}\big(\frac{1}{2}\vartheta,\frac{1}{2},\frac{1}{2},\frac{1}{2}\big)/\mathrm{S}_3$.

An important geometric input on which Theorem A hinges is the notion of *balanced spherical triangles*; Theorem B describes the relation between spherical tori and balanced triangles.

Definition 1.2 (spherical polygons) A *spherical polygon P* with angles $\pi(\vartheta_1,\dots,\vartheta_n)$ is a closed disk equipped with a Riemannian metric of constant curvature 1, with n distinguished boundary points $x_1,\dots,x_n$ which are called *vertices*, and such that the arcs between the adjacent vertices are geodesics forming an interior angle $\pi\vartheta_i$ at the i^{th} vertex. Two polygons are *isometric* if there is an isometry between them that preserves the labeling.

Spherical polygons with two or three vertices are called *digons* or *triangles*.[1]

Definition 1.3 (balanced triangles) A spherical triangle Δ with angles $\pi(\vartheta_1,\vartheta_2,\vartheta_3)$ is called *balanced* if the numbers ϑ_1, ϑ_2 and ϑ_3 satisfy the three triangle inequalities. If the triangle inequalities are satisfied strictly, we call the triangle *strictly balanced.* If, for some permutation (i,j,k) of $(1,2,3)$, we have $\vartheta_i=\vartheta_j+\vartheta_k$, we call the triangle *semibalanced.* If $\vartheta_i>\vartheta_j+\vartheta_k$ for some i, we call the triangle *unbalanced.*

Semibalanced triangles are called *marginal* in Eremenko and Gabrielov [12] and [13].

Whenever a spherical triangle is realized as a subset of a surface, we will induce on it the orientation of the surface. We will say that two oriented spherical surfaces (or polygons) are *conformally isometric* (or *congruent*) if there is an orientation-preserving isometry from one surface (or polygon) to the other.

Terminology (integral angles) Throughout the paper angles will be measured in radians. Nevertheless, an angle $2\pi\vartheta$ at a conical point of a spherical surface is called *integral* if $\vartheta\in\mathbb{Z}_{>0}$; similarly, an angle $\pi\vartheta$ at a vertex of a spherical polygon is called *integral* if $\vartheta\in\mathbb{Z}_{>0}$.

Now we describe a construction that will be omnipresent:

Construction 1.4 To each spherical triangle Δ with vertices x_1, x_2 and x_3 one can associate a spherical torus $T(\Delta)$ with one conical point by taking a conformally isometric triangle Δ' with vertices x_1', x_2' and x_3' and isometrically identifying each side x_ix_j with the side $x_j'x_i'$ (in such a way that x_i is identified to x_j' and x_j is identified to x_i') for $i,j\in\{1,2,3\}$. The angle at the conical point of $T(\Delta)$, which corresponds to the vertices of the triangles, is twice the sum of the angles of Δ. If Δ is endowed with an orientation, then $T(\Delta)$ canonically inherits an orientation.

[1] We note that spherical triangles in the sense of our definition are sometimes called *Schwarz–Klein triangles* to distinguish them from triangles understood as broken geodesic lines on the sphere; see for instance [12].

To state the next result we need two more notions. Let T be a spherical torus with one conical point. An isometric orientation-reversing involution on T will be called a *rectangular involution* if its set of fixed points consists of two connected components. By a *geodesic loop* γ based at a conical point x we mean a loop based at x which is geodesic in $\dot{T} = T \setminus \{x\}$ and which passes through x only at its endpoints.

Theorem B (canonical decomposition of a spherical torus for nonodd ϑ) *Let (T, x) be a spherical torus with one conical point of angle $2\pi\vartheta$ such that $\vartheta \in (1, \infty) \setminus (2\mathbb{Z}+1)$.*

(i) *If T does not have a rectangular involution, then there exists a unique (up to a reordering) triple of geodesic loops $(\gamma_1, \gamma_2, \gamma_3)$ based at x that cuts T into two congruent strictly balanced spherical triangles.*

(ii) *If T has a rectangular involution, there exist exactly two (unordered) triples of geodesic loops such that each of them cuts T into two congruent balanced triangles. Moreover, such triangles are semibalanced. These two triples are exchanged by the rectangular involution.*

We recall that, by Mondello and Panov [23, Section 4], the Voronoi graph associated to a spherical surface with n conical points decomposes such a surface into the union of n topological disks with one conical point each. Indeed, the role of this Voronoi graph is analogous to the role of the critical graph of a Jenkins–Strebel differential (a procedure that allows one to build a spherical surface out of a Jenkins–Strebel differential is described by Song, Cheng, Li and Xu [26]).

In order to prove Theorem B, we note that the complement of the Voronoi graph of the spherical torus (T, x) is one disk, and that this disk can be further split into two congruent triangles using the conformal involution of the torus. As a consequence of Theorem B, to each spherical torus T one can associate an essentially unique balanced spherical triangle $\Delta(T)$. Such uniqueness will permit us to reduce the description of the moduli space $\mathcal{MS}_{1,1}(\vartheta)$ to that of the moduli space of balanced triangles of area $\pi(\vartheta - 1)$.

1.1.2 ϑ an odd integer

The case when ϑ is an odd integer is quite different, as not all spherical metrics are invariant under the unique (nontrivial) conformal involution σ of the tori. We begin by stating our result for metrics that are σ–invariant:

Theorem C (topology of $\mathcal{MS}_{1,1}(2m+1)^\sigma$) *Fix an integer $m \geq 1$ and consider the moduli space $\mathcal{MS}_{1,1}(2m+1)^\sigma$ of tori with a σ–invariant spherical metric of area $4m\pi$.*

(a) *As a topological space, $\mathcal{MS}_{1,1}(2m+1)^\sigma$ is homeomorphic to the disjoint union of $\left\lceil \frac{1}{6}m(m+1) \right\rceil$ 2–dimensional open disks.*

(b) $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ *is naturally endowed with the structure of a* 2*–dimensional orbifold with* $\left\lceil \frac{1}{6}m(m+1)\right\rceil$ *connected components, which can be described as follows*:

(b-i) *If* $m \not\equiv 1 \pmod 3$, *then all components are isomorphic to the quotient* $\mathcal{D}$ *of* $\mathring{\Delta}^2 = \{y \in \mathbb{R}^3_+ \mid y_1+y_2+y_3 = 2\pi\}$ *by the trivial* $\mathbb{Z}_2$*–action. Hence, every point of* $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ *has orbifold order* 2.

(b-ii) *If* $m \equiv 1 \pmod 3$, *then one component is isomorphic to the quotient* $\mathcal{D}'$ *of* $\mathring{\Delta}^2$ *by* $\mathbb{Z}_2 \times \mathrm{A}_3$, *where* $\mathbb{Z}_2$ *acts trivially and* A_3 *acts by cyclically permuting the coordinates of* $\mathring{\Delta}^2$, *and all the other components are isomorphic to* $\mathcal{D}$. *Hence, one point of* $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ *has orbifold order* 6 *and all the other points have order* 2.

Remark 1.5 Similarly to Remark 1.1, as a topological space $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ is homeomorphic to $\mathcal{MS}_{0,4}\big(m+\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\big)/\mathrm{S}_3$ (though they are not isomorphic as orbifolds). Thus, Theorem C has a connection with the results of Chai, Lin and Wang [2], Chen and Lin [6], Eremenko [10], Eremenko and Gabrielov [11] and Lin and Wang [20].

The following description of the moduli space of tori with metrics that are not necessarily σ–invariant will be deduced from Theorem C:

Theorem D (topology of $\mathcal{MS}_{1,1}(2m+1)$) *For each positive integer* m, *the moduli space* $\mathcal{MS}_{1,1}(2m+1)$ *is a* 3*–dimensional orbifold with* $\left\lceil \frac{1}{6}m(m+1)\right\rceil$ *connected components.*

(i) *If* $m \not\equiv 1 \pmod 3$, *then all components of* $\mathcal{MS}_{1,1}(2m+1)$ *are isomorphic to the quotient* $\mathcal{M}$ *of* $\mathring{\Delta}^2 \times \mathbb{R}$ *by the involution* $(y,t) \mapsto (y,-t)$.

(ii) *If* $m \equiv 1 \pmod 3$, *then one component of* $\mathcal{MS}_{1,1}(2m+1)$ *is isomorphic to the quotient* $\mathcal{M}'$ *of* $\mathring{\Delta}^2 \times \mathbb{R}$ *by* $\mathbb{Z}_2 \times \mathrm{A}_3$, *where* $\mathbb{Z}_2$ *acts via the involution* $(y,t) \mapsto (y,-t)$ *and the alternating group* A_3 *acts by cyclically permuting the coordinates of* $\mathring{\Delta}^2$. *All the other components are isomorphic to* $\mathcal{M}$.

The locus $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ *of* σ*–invariant metrics correspond to* $t=0$.

In order to understand what happens for spherical metrics that are not necessarily σ–invariant, we recall:

Definition 1.6 (coaxiality) A monodromy is *coaxial* if and only if it is contained inside a one-parameter subgroup $\mathrm{SO}(3,\mathbb{R})$. A spherical surface is called *coaxial* if its monodromy is.

Note that every spherical metric with nontrivial coaxial monodromy on a surface belongs to a 1–parameter family of metrics that induce the same $\mathbb{CP}^1$–structure; we will say that metrics in the same 1–parameter family are *projectively equivalent*.

In the present case, a spherical metric on a torus T with one conical point of angle $2\pi\vartheta$ has nontrivial monodromy. Moreover, the monodromy is coaxial if and only if ϑ is odd. This fact is proven in [2, Theorem 5.2] and can be also deduced by combining the observations of Li, Song and Xu [18, page 8] with Chen, Wang, Wu and Xu [4, Proposition 1.4]. We reprove this statement using an argument based on monodromy considerations (see Corollary A.2).

The above discussion shows that every spherical surface in $\mathcal{MS}_{1,1}(2m+1)$ belongs to a 1–parameter family of projectively equivalent metrics, which thus traces a copy of $\mathbb{R}$ inside $\mathcal{MS}_{1,1}(2m+1)$. Moreover, in every family there exists exactly one metric which is σ–invariant (see Proposition 2.17). For this reason $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ is isomorphic to the moduli space $\mathcal{MS}_{1,1}(2m+1)/\mathrm{proj}$ of projective classes of spherical tori of area $4m\pi$, and so $\mathcal{MS}_{1,1}(2m+1)$ is 3–dimensional.

Another major difference from the nonodd case concerns the forgetful map: for ϑ nonodd, the forgetful map $\mathcal{MS}_{1,1}(\vartheta) \to \mathcal{M}_{1,1}$ is proper (see Mondello and Panov [23]) and surjective, whereas this is not so for odd ϑ; see Lin and Wang [19]. The boundary of $\mathcal{MS}_{1,1}(2m+1)/\mathrm{proj}$ inside the space of $\mathbb{CP}^1$–structures describes interesting real-analytic curves (see [2]) that are investigated in the sequel to this paper [13].

Theorems C and D will rely on the following result, which links moduli spaces of tori to moduli spaces of balanced triangles with integral angles:

Theorem E (canonical decomposition of a spherical torus with odd ϑ) *Fix a spherical torus with one conical point of angle $2\pi(2m+1)$. In the same projective class there exists a unique spherical torus (T,x) that admits an isometric orientation-preserving involution. Also, there exists a unique collection of three geodesic loops $(\gamma_1,\gamma_2,\gamma_3)$ based at x that cuts T into two congruent balanced spherical triangles Δ and Δ' with integral angles $\pi(m_1,m_2,m_3)$.*

1.1.3 ϑ an even integer Our final main result concerns moduli spaces $\mathcal{MS}_{1,1}(2m)$ where m is a positive integer. It is known (see Chai, Lin and Wang [2] and Eremenko and Tarasov [14]) that these moduli spaces have a natural holomorphic structure with respect to which they are compact Riemann surfaces with punctures. This is the unique conformal structure which makes the forgetful map to $\mathcal{M}_{1,1}$ holomorphic. With this structure $\mathcal{MS}_{1,1}(2m)$ is an algebraic curve.

Theorem F ($\mathcal{MS}_{1,1}(2m)$ is a Belyi curve) *For each integer $m > 0$ there exists a subgroup $G_m < \mathrm{SL}(2,\mathbb{Z})$ of index m^2 such that the orbifold $\mathcal{MS}_{1,1}(2m)$ is biholomorphic to the quotient $\mathbb{H}^2/G_m$. Such G_m is nonnormal for $m > 1$. Moreover, the points in $\mathbb{H}^2/G_m$ that project to the geodesic ray $[i,\infty)$ in the modular curve $\mathbb{H}^2/\mathrm{SL}(2,\mathbb{Z})$ correspond to tori T such that the triangle $\Delta(T)$ has one integral angle.*

1.2 Analytic representation of spherical metrics

Let (T,x) be a spherical torus with a conical singularity at x of angle $2\pi\vartheta$. The pullback of the spherical metric via the universal cover $\mathbb{C} = \widetilde{T} \to T$ has area element $e^u|dz|^2$. Then the function u satisfies the nonlinear PDE

$$\Delta u + 2e^u = 2\pi(\vartheta - 1)\delta_\Lambda, \tag{1}$$

where δ_Λ is the sum of delta functions over the lattice Λ and T is biholomorphic to $\mathbb{C}/\Lambda$. So our results describe the moduli spaces of pairs (Λ,u), where u is a Λ–periodic solution of (1).

Equation (1) is the simplest representative of the class of "mean field equations", which have important applications in physics; see Tarantello [27].

The general solution of (1) can be expressed in terms of a function $f\colon \mathbb{C} \to \mathbb{CP}^1$, the developing map, which is related to the conformal factor u by

$$u = \log \frac{4|f'|^2}{(1+|f|^2)^2}.$$

The developing map $f = w_1/w_2$ is the ratio of two linearly independent solutions w_1 and w_2 of the Lamé equation

$$w'' = \Big(\frac{\vartheta^2-1}{4}\wp - c\Big)w, \tag{2}$$

where $\wp$ is the Weierstrass function of the lattice Λ and $c \in \mathbb{C}$ is an accessory parameter. So our results can be also interpreted as a description of the moduli space of projective structures on tori whose monodromies are subgroups of $\mathrm{SO}(3,\mathbb{R})$.

Most of the known results on spherical tori are formulated in terms of (1) and (2). For example, it is proved in Chen and Lin [3] that when ϑ is not an odd integer, then the Leray–Schauder degree of the nonlinear operator in (1) equals $\lfloor \frac{1}{2}(\vartheta+1)\rfloor$. An especially well-studied case is the classical Lamé equation (2) where ϑ is an integer; see [2; 13]. Solutions of (2) with odd integer ϑ are special functions of mathematical physics; see Maier [21] and Whittaker and Watson [29].

1.3 The idea of the proof of Theorem A

Here we give a brief summary of the proof of Theorem A, since various parts of it stretch through the whole paper. Fix $\vartheta > 1$ not odd and consider spherical tori with a conical point of angle $2\pi\vartheta$, and area $2\pi(\vartheta - 1)$.

- By Proposition 2.17(i), on every torus the unique nontrivial conformal involution is an isometry.

- Every spherical torus is obtained by gluing two isometric copies of a spherical balanced triangle with labeled vertices in an essentially unique way (Theorem B, proven in Section 2.4). This result has a clear refinement for tori with a 2–marking (namely, a labeling of its 2–torsion points); see Construction 4.5.

- The doubled space $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ of balanced triangles of area $\pi(\vartheta - 1)$ is the double of the space $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ of balanced triangles of area $\pi(\vartheta - 1)$ and it describes oriented balanced triangles up to some identifications that only involve semibalanced triangles (Definition 3.21).

- The space $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is an orientable connected surface with boundary, and its topology is completely determined (see Proposition 3.20) and so is the topology of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$; see Proposition 3.22.

- As a topological space, the space $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ of isomorphism classes of 2–marked tori is homeomorphic to $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$; see Theorem 6.5.

- As an orbifold, $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is isomorphic to the quotient of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ by the trivial $\mathbb{Z}_2$–action. This allows us to determine the topology and the orbifold Euler characteristic of $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$; see Theorem 4.8.

- The map $\mathcal{MS}^{(2)}_{1,1}(\vartheta) \to \mathcal{MS}_{1,1}(\vartheta)$ that forgets the 2–marking is an unramified orbifold S_3–cover, where S_3 acts on $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ by permuting the 2–markings (see Remark 6.28). This allows us to describe the points in $\mathcal{MS}_{1,1}(\vartheta)$ of orbifold order greater than 2 (Proposition 4.4) and to determine the topology and the orbifold Euler characteristic of $\mathcal{MS}_{1,1}(\vartheta)$; see Theorem A, proved towards the end of Section 4.1.

1.4 Content of the paper

The relation between spherical tori with one conical point and balanced spherical triangles is established in Section 2, which culminates in the proof of Theorem B. The section contains a careful analysis of the Voronoi graph of a torus and of the action of the unique nontrivial conformal involution σ on its spherical metric.

In Section 3 we describe the topology of the space of balanced triangles of area $\pi(\vartheta-1)$ and of its double, separately considering the cases ϑ nonodd and ϑ odd. Here we visualize the space of spherical triangles with assigned area, which is a manifold, by looking at its image (which we call a *carpet*) through the angle map Θ. The balanced carpet will turn out to be a useful tool in computing the topological invariants of the space of balanced triangles.

In Section 4 we describe the topology of the moduli spaces of spherical tori with one conical point, endowed with the Lipschitz metric (which we study in Section 6). For ϑ nonodd, we first establish a homeomorphism between the doubled space of balanced triangles and the topological space of 2–marked tori using tools from Section 6. Then we prove Theorem A. For ϑ odd, we first prove Theorem E using results from Sections 2 and 3, which immediately allows us to prove part (a) of Theorem C. Then we endow our moduli space of σ–invariant spherical tori with a 2–dimensional orbifold structure and prove part (b) of Theorem C. Finally, using one-parameter projective deformations of σ–invariant spherical metrics, we put a 3–dimensional orbifold structure on the moduli space of (not necessarily σ–invariant) tori and prove Theorem D.

In Section 5 we analyze the moduli space of tori with ϑ even, and prove Theorem F by identifying it to a Hurwitz space of covers of $\mathbb{CP}^1$ branched at three points. This permits us to exhibit this moduli space as a Belyi curve and to characterize tori that sit on the 1–dimensional skeleton of its dessin.

Section 6 deals with properties of the Lipschitz metric on moduli spaces of spherical surfaces with conical points with area bounded from above. The main result of the section is Theorem 6.3 on properness of the inverse of the systole function. Then the treatment is specialized to tori with one conical point of angle $2\pi\vartheta$ with ϑ nonodd (or with ϑ odd and a σ–invariant metric). The section culminates with establishing the homeomorphism between the space of 2–marked tori and the doubled space of balanced triangles, needed in Section 4. A last remark explains how to use such results to endow our moduli spaces with an orbifold structure.

In the short appendix we prove a general $\mathrm{SU}(2)$–lifting theorem for the monodromy of a spherical surface, and we apply it to the cases of ϑ odd and ϑ even to explain their special features.

Acknowledgments We would like to thank Andrei Gabrielov, Alexander Lytchak, Anton Petrunin and Dimitri Zvonkine for useful discussions. We also thank the referee for carefully reading the paper and for numerous suggestions and remarks.

Eremenko was supported by NSF grant DMS-1665115. Mondello was partially supported by INdAM and by grant PRIN 2017 *Moduli and Lie theory*. Panov was supported by EPSRC grant EP/S035788/1.

2 Voronoi diagrams and the proof of Theorem B

In this section we will study the Voronoi graphs of spherical tori (T, x, ϑ) with one conical point and prove Theorem B.

2.1 Properties of Voronoi graphs, functions and domains

In this subsection we recall the definition of a Voronoi graph [23, Section 4] and apply it to spherical tori with one conical point.

Definition 2.1 (Voronoi function and Voronoi graph) Let S be a surface with a spherical metric and conical points $\boldsymbol{x}$. The *Voronoi function* $\mathcal{V}_S \colon S \to \mathbb{R}$ is defined as $\mathcal{V}_S(p) := d(p, \boldsymbol{x})$. The *Voronoi graph* $\Gamma(S)$ is the locus of points $p \in \dot{S}$ at which the distance $d(p, \boldsymbol{x})$ is realized by two or more geodesic arcs joining p to $\boldsymbol{x}$. We will simply write $\Gamma = \Gamma(S)$ when no ambiguity is possible. The *Voronoi domains* of S are connected components of the complement $S \setminus \Gamma(S)$. Each Voronoi domain D_i contains a unique conical point x_i and this point is the closest conical point to all the points in the domain.

Various properties of Voronoi functions, graphs and domains of spherical surfaces were proven in [23, Section 4], and Proposition 2.3 lists some of the facts needed here. To formulate the last two properties we need one more definition:

Definition 2.2 (convex star-shaped polygons) Let D be a disk with a spherical metric, containing a unique conical point $x \in D$, such that its boundary is composed of a collection of geodesic segments. We say that D is a *convex and star-shaped polygon* if any two neighboring sides of D meet under an interior angle smaller than π and for any point $p \in D$ there is a unique geodesic segment that joins x with p.

Proposition 2.3 (basic properties of the Voronoi function and graph) *Let S be a spherical surface of genus g with conical points $x_1, \dots, x_n$.*

(i) *The Voronoi function is bounded from above by π, namely $\mathcal{V}_S < \pi$.*

(ii) *The Voronoi graph $\Gamma(S)$ is a graph with geodesic edges embedded in S and contains at most $-3\chi(\dot{S}) = 6g - 6 + 3n$ edges.*

(iii) *The valence of each vertex of $\Gamma(S)$ is at least three. For any point $p \in \Gamma(S)$ its valence coincides with the multiplicity μ_p, ie there exist exactly μ_p geodesic segments in S of length $\mathcal{V}_S(p)$ that join p with conical points of S.*

(iv) *The metric completion of each Voronoi domain*[2] *is a convex and star-shaped polygon with a unique conical point in its interior.*

(v) *Let γ be an open edge of $\Gamma(S)$. Let D_i and D_j be two Voronoi domains adjacent to γ. Let $\Delta \subset D_i$ and $\Delta' \subset D_j$ be the two triangles with one vertex x_i or x_j, respectively, and opposite side γ. Then Δ and Δ' are anticonformally isometric by an isometry fixing γ.*

Proof (i) This is proven in [23, Lemma 4.2].

(ii) This is proven in [23, Lemma 4.5 and Corollary 4.7].

(iii) The valence of vertices is at least three by [23, Corollary 4.7]. The valence of a point on $\Gamma(S)$ coincides with its multiplicity by [23, Lemma 4.5].

(iv) The convexity is proven in [23, Lemma 4.8]. The fact that each domain is star-shaped follows from the fact that each point p in it can be joined, by a unique geodesic segment of length $\mathcal{V}_S(p)$, with the conical point. Such a segment varies continuously with p, since $\mathcal{V}_S(p) < \pi$.

(v) To find the isometry between Δ and Δ' just notice that by definition each point $p \in \gamma$ can be joined by two geodesics of the same length with x_i and x_j. Also these two geodesics intersect γ under the same angle. The isometry between the triangles is obtained by the map exchanging each pair of such geodesics. □

Example 2.4 (Voronoi graph in a sphere with three conical points) Let S be a sphere with three conical points. It follows from Proposition 2.3(ii) that the Voronoi graph $\Gamma(S)$ is either a trefoil graph, an eight graph or an eyeglasses graph; see Figure 1. Indeed, $\Gamma(S)$ splits S into three disks, and it has at most three edges.

The next definition and remark explain how to define Voronoi functions and graphs for spherical polygons, mimicking Definition 2.1.

Definition 2.5 (Voronoi function and graph of a polygon) Let P be a spherical polygon with vertices $\boldsymbol{x}$. Then the Voronoi function $\mathcal{V}_P\colon P \to \mathbb{R}$ is defined as

[2] The metric completion can differ from the closure of the domain inside S; see the rightmost example in Figure 2.

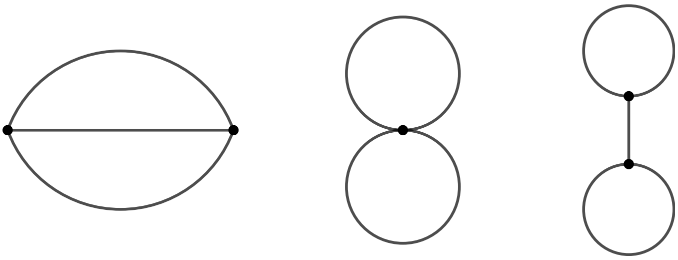


Figure 1: Voronoi graphs on a sphere with three conical points. From left to right: the trefoil, the eight graph and the eyeglass graph.

$\mathcal{V}_P(p) := d(p, \boldsymbol{x})$. The Voronoi graph $\Gamma(P)$ of P consists of points p of two types: first, the points for which there exist at least two geodesic segments of length $d(p, \boldsymbol{x})$ that join p with $\boldsymbol{x}$, and second, the points p on ∂P for which the closest vertex of P does not lie on the edge to which p belongs.

Remark 2.6 (doubling a polygon: Voronoi function and graph) To each spherical polygon P one can associate a sphere $S(P)$ with conical singularities by *doubling*[3] P across its boundary. Such a sphere has an anticonformal isometry that exchanges P and its isometric copy P', and fixes their boundary. It is easy to see that the function $\mathcal{V}_{S(P)}$ restricts to $\mathcal{V}_P$ on $P \subset S$ and to $\mathcal{V}_{P'}$ on $P' \subset S$. One can also check that the Voronoi graph $\Gamma(S(P))$ is the union $\Gamma(P) \cup \Gamma(P')$. As a result, the statements of Proposition 2.3 have their analogues for spherical polygons.

The following lemma gives an efficient criterion permitting one to verify whether a given geodesic graph on a spherical surface is in fact its Voronoi graph:

Lemma 2.7 (Voronoi graph criterion) *Let S be a spherical surface of genus g with conical points $x_1, \dots, x_n$ and let $\Gamma'(S) \subset S$ be a finite graph with geodesic edges embedded in S. Then $\Gamma'(S) = \Gamma(S)$ if and only if the following two conditions hold*:

(a) *$S \setminus \Gamma'(S)$ is a union of disks whose metric completions are convex and star-shaped polygons each with a unique conical point in its interior.*

(b) *For each point $p \in \Gamma'(S)$ all geodesic segments that join p with some conical point of S and intersect $\Gamma'(S)$ only at p have the same length.*

Proof Since by Proposition 2.3 the graph $\Gamma(S)$ satisfies the conditions (a) and (b), we only need to prove the "only if" direction.

[3] Given a topological space X and a closed subset A, the *doubling of X along A* is obtained from $X \times \{0, 1\}$ by making the identification $(a, 0) \sim (a, 1)$ for every $a \in A$.

For each conical point x_i let D_i be the Voronoi domain of x_i (namely the connected component of $S \setminus \Gamma(S)$ that contains x_i), and let D'_i be the component of $S \setminus \Gamma'(S)$ containing x_i. Let's assume, for contradiction, that there is a point $p \in D_i$ that is not contained in D'_i. By definition of D_i there is a unique geodesic segment $\gamma(p)$ of length $\mathcal{V}_S(p)$ that joins p with x_i. Denote by $\gamma'(p)$ the connected component of the intersection $\gamma(p) \cap D'_i$ that contains x_i and let $p' \notin D'_i$ be the point in its closure. Clearly p' belongs to $\Gamma'(S)$. By (a) each component of $S \setminus \Gamma'(S)$ is star-shaped, so using (b) we get a second (different from $\gamma'(p)$) geodesic segment of length $\mathcal{V}_S(p')$ that joins p' with a conical point. Hence $p' \in \Gamma(S)$, which contradicts the fact that p' is in D_i.

We proved that $D_i \subset D'_i$ for each i. It follows that $D_i = D'_i$, hence $\Gamma'(S) = \Gamma(S)$. □

Lemma 2.8 (Voronoi graphs of a sphere with three conical points) *Let S be a sphere with three conical points x_i of conical angles $2\pi\vartheta_i$.*

(i) $\Gamma(S)$ *is a trefoil if and only if ϑ_1, ϑ_2 and ϑ_3 satisfy the triangle inequality strictly.*

(ii) $\Gamma(S)$ *is an eight graph if and only if $\vartheta_i = \vartheta_j + \vartheta_k$ for some permutation (i, j, k) of $\{1, 2, 3\}$.*

(iii) $\Gamma(S)$ *is an eyeglasses graph if and only if $\vartheta_i > \vartheta_j + \vartheta_k$ for some permutation (i, j, k) of $\{1, 2, 3\}$.*

(iv) *In cases* (i) *and* (ii) *the vertices of $\Gamma(S)$ are equidistant from x_1, x_2 and x_3. In case* (iii) *the vertices of $\Gamma(S)$ are not equidistant from x_1, x_2 and x_3.*

Proof It is enough to prove the "only if" parts of claims (i), (ii) and (iii); the cases are mutually exclusive and so the "if" part will follow.

For the proof of the "only if" part, all three cases are treated in a similar way. Let us consider, for example, the case when $\Gamma(S)$ is a trefoil graph. Let's show that in this case the ϑ_i satisfy the triangle inequality strictly. Denote the two vertices of $\Gamma(S)$ by A and B. The three edges of $\Gamma(S)$ cut S into three Voronoi disks, each of which contains one conical point. Let us denote these three segments of $\Gamma(S)$ by γ_1, γ_2 and γ_3, as shown in the leftmost picture in Figure 2. Let us join each of the x_i with the vertices A and B by geodesics x_iA and x_iB of lengths $\mathcal{V}_S(A)$ and $\mathcal{V}_S(B)$, respectively. These geodesic segments are depicted in gray.

Consider now the spherical quadrilaterals Ax_3Bx_1, Ax_1Bx_2 and Ax_2Bx_3 into which the gray geodesics cut S. It follows from Proposition 2.3(v) for $i, j \in \{1, 2, 3\}$ that the

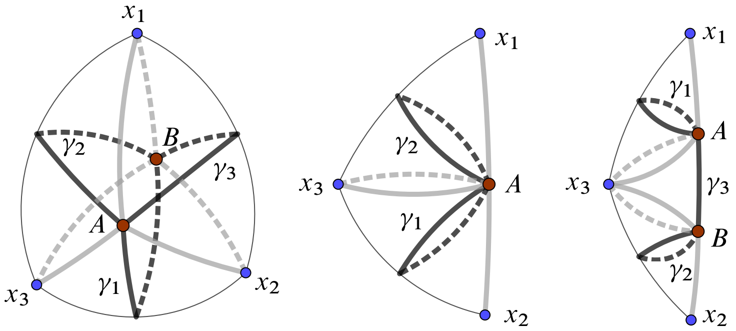


Figure 2: Three types of spheres.

angles of Ax_iBx_j at x_i and x_j are equal. This implies that ϑ_1, ϑ_2 and ϑ_3 satisfy the triangle inequality strictly.

(ii)–(iii) One treats the cases when $\Gamma(S)$ is an eight graph or an eyeglasses graph in a similar way; the corresponding pictures are shown in Figure 2.

(iv) This is clear from the way $\Gamma(S)$ is embedded in S; see Figure 2. In particular, if $\Gamma(S)$ is an eyeglasses graph, $d(A,x_1) = d(A,x_3) < d(A,x_2)$ and $d(B,x_2) = d(B,x_3) < d(B,x_1)$. □

2.2 The circumcenters of balanced triangles

It is well known that the circumcenter of a Euclidean triangle Δ is contained in Δ if and only if Δ is not obtuse. Moreover, in the case when Δ is right-angled, the circumcenter is the midpoint of the hypotenuse. It is also a classical fact that the circumcenter of a Euclidean triangle is the point of intersection of the axes[4] of its sides. The next theorem is a generalization of the above statements to spherical triangles. By an *involutive triangle* we mean a triangle that admits an anticonformal isometric involution that fixes one vertex and exchanges the other two.[5]

Theorem 2.9 (circumcenters of balanced triangles) *Let Δ be a spherical triangle with vertices x_1, x_2 and x_3.*

(i) *The triangle Δ contains a point O equidistant from x_1, x_2 and x_3 if and only if Δ is balanced.*

[4] The *axis* of a segment is the perpendicular through the midpoint of such segment.

[5] Note that every Euclidean or hyperbolic isosceles triangle admits an isometric involution exchanging the equal sides. This is not the case for spherical triangles; for example the triangle with angles $\frac{5}{2}\pi$, $\frac{13}{2}\pi$ and $\frac{9}{2}\pi$ is equilateral but clearly has no symmetries.

(ii) *The point O (equidistant from x_1, x_3 and x_3) is in the interior of Δ if and only if Δ is strictly balanced. The point O is the midpoint of a side of Δ if and only if Δ is semibalanced.*

(iii) *If Δ is strictly balanced, then the geodesic segments Ox_1, Ox_2 and Ox_3 cut Δ into three involutive triangles.*

(iv) *Suppose that Δ is semibalanced and the angle $\angle x_i = \pi\vartheta_i$ is the largest one. Then O is the midpoint of the side opposite to x_i, and x_iO cuts Δ into two involutive triangles.*

To prove this theorem we need the following lemma.

Lemma 2.10 (some isosceles triangles are involutive triangles) *Let Δ be a spherical triangle with vertices q_1, q_2 and q_3 and denote by $|q_iq_j|$ the length of the side q_iq_j. Suppose that $|q_1q_2| = |q_1q_3| < \pi$ and $\angle q_1 < 2\pi$. Then there is an isometric reflection τ of Δ that fixes q_1 and exchanges q_2 with q_3. In particular, $\angle q_2 = \angle q_3$. Moreover, τ pointwise fixes a geodesic segment that joins q_1 with the midpoint of q_2q_3 and splits Δ into two isometric triangles. Furthermore, $|q_2q_3| < 2\pi$.*

Proof First, let $\angle q_1 = \pi$. In this case Δ can be isometrically identified with a digon so that q_1 is identified with the midpoint of one of its sides. Since each digon has an isometric reflection fixing the midpoints of both sides, the lemma holds.

From now on we assume that $\angle q_1 \neq \pi$. Consider the unique spherical triangle $\Delta' \subset \mathbb{S}^2$ with vertices q'_1, q'_2 and q'_3 such that $|q'_1q'_2| = |q'_1q'_3| = |q_1q_2|$, $\angle q'_1 = \angle q_1$, and $\mathrm{Area}(\Delta') < 2\pi$. We will show that Δ' admits an isometric embedding into Δ that sends q'_i to q_i. This will prove the lemma since this implies that Δ is isometric to a triangle obtained by gluing a digon to the side $q'_2q'_3$ of Δ'. And such a triangle clearly has an isometric reflection τ. This will also prove that $|q_2q_3| < 2\pi$, since $|q'_2q'_3| < 2\pi$ and either $|q_2q_3| = |q'_2q'_3|$ or $|q_2q_3| + |q'_2q'_3| = 2\pi$.

To prove the existence of the embedding, denote by $\iota\colon \Delta \to \mathbb{S}^2$ the developing map of triangle Δ. We may assume that $\iota(q_i) = q'_i$, $\iota(q_1q_2) = q'_1q'_2$ and $\iota(q_1q_3) = q'_1q'_3$. Note that ι sends q_2q_3 to the unique[6] geodesic circle that contains $\iota(q_2)$ and $\iota(q_3)$. Hence, it is not hard to see that the preimages of Δ' in Δ form a union of some number of isometric copies of Δ'. One of them, containing sides q_1q_2 and q_1q_3 of Δ, is the embedding we are looking for. □

[6] This circle is unique since $\angle q_1 \neq \pi$, and also it intersects the segments $q'_1q'_2$ and $q'_1q'_3$ only at the points q'_2 and q'_3.

Remark 2.11 This lemma is sharp in the sense that neither of the two conditions $|q_1q_2| = |q_1q_3| < \pi$ and $\angle q_1 < 2\pi$ can be dropped.

Proof of Theorem 2.9 (i) Let $S(\Delta)$ be the sphere obtained by doubling Δ across its boundary, ie by gluing Δ with the triangle Δ' that is anticonformally isometric to Δ. Then, by Remark 2.6, the graph $\Gamma(S(\Delta))$ is the union of $\Gamma(\Delta)$ with $\Gamma(\Delta')$.

Suppose first that Δ contains a point O equidistant from all the x_i. Then, since the restriction of $\mathcal{V}_{S(\Delta)}$ to Δ equals $\mathcal{V}_\Delta$, we see that O is equidistant from x_i on S as well. So, by Proposition 2.3(iii), the point O corresponds to a vertex of $\Gamma(S(\Delta))$ of multiplicity at least 3. Furthermore, by Lemma 2.8(iv), we conclude that $\Gamma(S)$ is either a trefoil or an eight graph. Hence, again by Lemma 2.8, the triangle Δ is balanced.

Suppose now that Δ is balanced, ie ϑ_1, ϑ_2 and ϑ_3 satisfy the triangle inequality. Then, by Lemma 2.8(i)–(ii), the graph $\Gamma(S(\Delta))$ is a trefoil or a eight graph, and so by Lemma 2.8(iv) there is a point O in S equidistant from all x_i. It follows that Δ contains such a point as well.

(ii) We first prove the "only if" direction. Suppose that O is in the interior of Δ. Then $\Gamma(S(\Delta))$ has two vertices of valence 3. So according to (i), $\Gamma(S(\Delta))$ is a trefoil. Hence, Δ is strictly balanced by Lemma 2.8(i).

Suppose that O is on the boundary of Δ. Without loss of generality assume that O is on the side of Δ opposite to x_1. For $i = 1, 2, 3$ let γ_i be the geodesic segment of length $\mathcal{V}_\Delta(O)$ that joins O with x_i. Let γ_i' be the image of γ_i in $\Delta' \subset S(\Delta)$ under the anticonformal involution. Since the multiplicity of O in $\Gamma(S)$ is at most 4, we conclude that $\gamma_2 = \gamma_2'$ and $\gamma_3 = \gamma_3'$. Hence, O is the midpoint of the side x_2x_3.

To prove the "if" direction, one needs to apply Lemma 2.8(iv). Indeed, if Δ is strictly balanced, $\Gamma(S(\Delta))$ has two vertices of multiplicity 3 and one of them lies in Δ. If Δ is semibalanced, $\Gamma(S(\Delta))$ has one vertex and it has to lie on the boundary of Δ.

(iii) Since Δ is strictly balanced, by (ii) there is a point O in the interior of Δ equidistant from points x_1, x_2 and x_3. Since $\mathcal{V}_\Delta(O) < \pi$, we have $|Ox_1| = |Ox_1| = |Ox_3| < \pi$. Hence all three isosceles triangles x_iOx_j are involutive triangles by Lemma 2.10.

(iv) This proof is identical to the proof of (iii). □

Remark 2.12 Theorem 2.9 can be used to construct the Voronoi graph $\Gamma(\Delta)$ of a balanced triangle Δ with vertices x_1, x_2 and x_3. Indeed, according to this theorem, the geodesic segments Ox_i cut Δ into three or two involutive triangles, and, using a

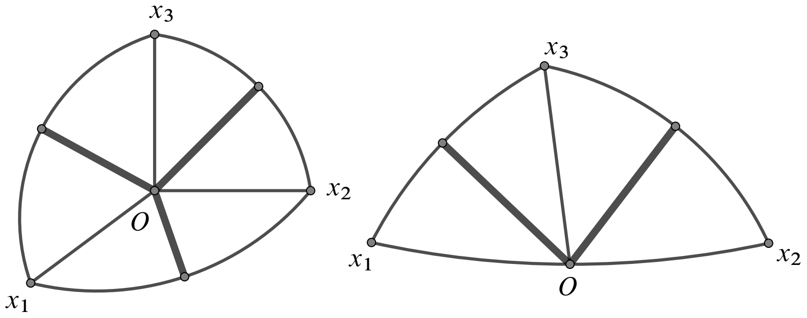


Figure 3: Voronoi graphs of balanced triangles.

variation of Lemma 2.7, one can show that $\Gamma(\Delta)$ is the union of symmetry axes of these triangles; see Figure 3.

We will see that some results we are interested in about balanced triangles indeed concern the following class of triangles.

Definition 2.13 (short-sided triangles) A spherical triangle *short-sided* if all its sides have length $l_i < 2\pi$. In this case, we set $\bar{l}_i := \min(l_i, 2\pi - l_i)$.

Theorem 2.9 has two simple corollaries:

Corollary 2.14 (balanced triangles are short-sided) *Let Δ be a balanced triangle with vertices x_1, x_2 and x_3. Then Δ is short-sided, ie $|x_i x_j| < 2\pi$.*

Proof Let us treat the case when Δ is strictly balanced. The semibalanced case is similar. By Theorem 2.9(iii), the triangle Δ can be cut into three involutive triangles $x_i O x_j$, where $\angle O < 2\pi$ and $|Ox_i| = |Ox_j| < \pi$. Applying Lemma 2.10 to the triangle $x_i O x_j$, we conclude that $|x_i x_j| < 2\pi$. □

Corollary 2.15 (short geodesic in a balanced triangle) *Let Δ be a balanced triangle with vertices x_1, x_2 and x_3. Suppose that $\{i, j, k\} = \{1, 2, 3\}$, ordered so that the value $\bar{l}_k = \min(|x_i x_j|, 2\pi - |x_i x_j|)$ is minimal. Then there is a geodesic segment γ_Δ in Δ that joins x_i with x_j and is such that $\ell(\gamma_\Delta) = \bar{l}_k \le \frac{2}{3}\pi$, which in fact realizes the minimum distance between distinct vertices.*

Proof Let us again treat the case when Δ is strictly balanced. Let $x_i O x_j$ be three involutive triangles into which Δ is cut. Consider the developing map $\iota\colon \Delta \to \mathbb{S}^2$. Then, for each $\{i, j, k\} = \{1, 2, 3\}$, the value $\bar{l}_k$ is equal to the distance between $\iota(x_i)$ and $\iota(x_j)$ on $\mathbb{S}^2$, and so $d(x_i, x_j) \ge d(\iota(x_i), \iota(x_j)) = \bar{l}_k$. For this reason, it is not hard to see that the minimum of the value $\bar{l}_k$ is attained for the triangle $x_i O x_j$ for which the angle at O is the minimal one. In particular, in such a triangle the angle at O is at most $\frac{2}{3}\pi$. It follows that there is a geodesic segment γ_Δ in such a triangle $x_i O x_j$ of length less than $\frac{2}{3}\pi$ that joins x_i and x_j. Since it cuts out of $x_i O x_j$ a digon with one side $x_i x_j$, we conclude that $\ell(\gamma_\Delta) = \bar{l}_k = d(x_i, x_j)$. □

2.3 Isometric conformal involutions on tori

In this short section we prove the following useful proposition:

Lemma 2.16 (invariance of projective structures on one-pointed tori) *Let (T, x) be a flat one-pointed torus and let σ be its unique nontrivial conformal involution. Then every projective structure on T whose Schwarzian derivative has at worst a double pole at x is invariant under σ.*

Proof We represent our torus T as $\mathbb{C}/\Lambda$, where Λ is a lattice in $\mathbb{C}$, and suppose that x corresponds to the lattice points. We also endow T with the corresponding projective structure.

The involution σ pulls back to the map $z \mapsto -z$ on $\widetilde{T} = \mathbb{C}$. The Schwarzian derivative (see for example [25]) of a projective structure is a quadratic differential on the torus T. By hypothesis, it has at worst a double pole at x. The vector space of such quadratic differentials is 2–dimensional, generated by the constants and the Weierstrass elliptic function. Hence, all its elements are invariant under the involution σ, and so are all solutions of the associated Schwarz equations. As a consequence, all such projective structures are σ–invariant. □

Proposition 2.17 (spherical metrics and conformal involution) *Let σ be the unique conformal involution of a spherical torus T that fixes the unique conical point x.*

(i) *If $\vartheta \notin 2\mathbb{Z}+1$, then σ is an isometry.*

(ii) *If $\vartheta \in 2\mathbb{Z}+1$, then each projective equivalence class of spherical metrics is parametrized by a copy of $\mathbb{R}$, on which σ acts as an orientation-reversing diffeomorphism. Thus, σ is an isometry for a unique spherical metric in its projective equivalence class.*

Proof Consider the projective structure associated to a spherical metric on (T, x). By Lemma 2.16, such projective structure is σ–invariant.

(i) Every spherical metric is noncoaxial by Corollary A.2, and so in each projective equivalence class there is at most one spherical metric. Hence, this metric must be invariant under σ.

(ii) Fix a spherical metric h in $\mathcal{MS}_{1,1}(2m+1)$. Let $\tilde{\dot{T}}$ be the universal cover of $\dot{T}$ and let $\widehat{T}$ be its completion. Denote by $\hat{x}_i$ the points in $\partial\widehat{T} = \widehat{T} \setminus \tilde{\dot{T}}$ which project to $x \in T$. Pick a developing map ι for h, which in fact extends to $\hat{\iota}\colon \widehat{T} \to \mathbb{S}^2 \cong \mathbb{CP}^1$, and let ρ be the associated monodromy representation.

By Corollary A.2, the monodromy ρ is coaxial but nontrivial. Fix an element α of $\pi_1(T)$ such that $\rho(\alpha) = e^X \neq I$ with $X \in \mathfrak{su}_2$. Up to conjugation, we can assume that $\infty \in \mathbb{CP}^1$ is the attracting point and $0 \in \mathbb{CP}^1$ is the repelling point for $\rho(\alpha)' := e^{iX}$. The orbits of the group (e^{tX}) on $\mathbb{CP}^1 \setminus \{0, \infty\}$ will be called "parallels" and the unique geodetic orbit will be called the "equator".

First, we claim that $\hat{\iota}(\hat{x}_i) \neq 0, \infty$ for all $\hat{x}_i \in \partial\widehat{T}$, and they all sit on the same parallel. In fact, the holomorphic vector field $z(\partial/\partial z)$ on $\mathbb{CP}^1$ is invariant for the monodromy, and so its pullback descends to a nonzero holomorphic vector field V on $\dot{T}$, possibly with a pole in x. If $\hat{\iota}(\hat{x}_i) \in \{0, \infty\}$, then V would have a zero at x, contradicting $\chi(T) = 0$. The second assertion is clear, since $\hat{\iota}(\partial\widehat{T})$ is an orbit for the action of the monodromy.

Second, note that all spherical metrics $(h_t)_{t\in\mathbb{R}}$ projectively equivalent to h have developing maps $e^t\iota$ and monodromy representation ρ. Thus, up to replacing h by some h_{t_0}, we can assume that $\hat{\iota}(\partial\widehat{T})$ is contained inside the equator.

The function $d\colon \mathbb{CP}^1 \to [0, \pi]$ that measures the distance from the repelling point of $\rho(\alpha)'$ is invariant for the monodromy action, and so its pullback via ι_t to $\widehat{T}$ descends to a function $d_t\colon T \to [0, \pi]$. We observe that t can be recovered from $d_t(x)$ via $e^t = \frac{1}{2}\tan(d_t(x))$.

Now, $(\rho\circ\sigma)(\alpha) = \rho(\alpha)^{-1} = e^{-X}$. Thus, when considering the developing map $(e^t\iota)\circ\sigma$ with monodromy representation $\rho\circ\sigma$, the attracting point of $(\rho\circ\sigma)(\alpha)'$ is 0 and the repelling point is ∞. It follows that the distance of $(e^t\hat{\iota})\circ\sigma(x) = e^t\hat{\iota}(x)$ from the repelling point ∞ is $\pi - d_t(x)$. Hence, $(e^t\iota)\circ\sigma$ is a developing map for h_{-t}. It follows that σ acts on the family of metrics $(h_t)_{t\in\mathbb{R}}$ by sending h_t to h_{-t}, and so fixing the unique metric h_0 whose developing map sends $\partial\widehat{T}$ to the equator. It follows that σ acts on (T, x, h_t) as an isometry if and only if $t = 0$. □

Proposition 2.17(ii) was also proved in [2, Theorem 5.2]; see also [11, Theorem 1].

2.4 Proof of Theorem B

The goal of this section is to prove Theorem B and to make preparations for the proof of Theorem C. Throughout the section we will mainly consider the class of tori that have a conformal isometric involution. By Proposition 2.17, we know that such an involution exists automatically in the case when the conical angle is not $2\pi(2m+1)$. We start with the following simple lemma:

Lemma 2.18 (points of Γ fixed by a conformal isometric involution) *Let S be a spherical surface with conical points $\boldsymbol{x}$ that admits an isometric conformal involution σ. Let p be a point in $\dot{S} = S \setminus \boldsymbol{x}$ fixed by σ. Then p belongs to $\Gamma(S)$, its multiplicity μ_p is even, and there exist exactly $\frac{1}{2}\mu_p$ geodesic segments or loops*[7] *of lengths $2\mathcal{V}_S(p) < 2\pi$ based at $\boldsymbol{x}$ and passing through p. The point p cuts each such geodesic segment into two halves of equal length.*

Proof Consider any geodesic segment γ of length $\mathcal{V}_S(p)$ that joins p with one of the conical points. Since $\sigma(\gamma) \neq \gamma$ we see that p belongs to $\Gamma(S)$. If p is not a vertex of $\Gamma(S)$, then γ and $\sigma(\gamma)$ are the only two geodesic segments of length $\mathcal{V}_S(p)$ that join p with $\boldsymbol{x}$. Clearly, since σ is a conformal involution the union $\gamma \cup \sigma(\gamma)$ is a geodesic segment or loop based at $\boldsymbol{x}$. Its length is less than 2π by Proposition 2.3(i).

The case when p is a vertex of $\Gamma(S)$ is similar. Since σ is a conformal involution and it sends $\Gamma(S)$ to $\Gamma(S)$ we see that the valence of p in Γ_S is even. By Proposition 2.3(iii) the number μ_p of geodesic segments of length $\mathcal{V}_S(p)$ that join p with $\boldsymbol{x}$ is equal to this valence. Clearly, altogether these μ_p segments form $\frac{1}{2}\mu_p$ geodesic segments (or loops) of length $2\mathcal{V}_S(p)$, all of which have midpoint p. □

Now we concentrate on the case of spherical tori with one conical point. It will be convenient for us to recall first the construction of *hexagonal* and *square* flat tori.

Example 2.19 (hexagonal and square flat tori) Let T_6 and T_4 be the flat tori obtained by identifying opposite sides of a regular flat hexagon and a square, respectively. Denote by $\Gamma_6 \subset T_6$ and $\Gamma_4 \subset T_4$ the graphs formed by the images of the polygons' boundaries. Then it is easy to check that Γ_6 and Γ_4 are Voronoi graphs in T_6 and T_4 with respect to the images of the centers of the polygons.

Lemma 2.20 (Voronoi graph of a spherical torus) *Let T be a spherical torus with one conical point and let Γ be its Voronoi graph. Then Γ is either a trefoil or an eight*

[7] We always assume that a geodesic loop or segment can intersect $\boldsymbol{x}$ only at its endpoints.

graph. In the first case the pair (T, Γ) is homeomorphic to the pair (T_6, Γ_6). In the second case it is homeomorphic to the pair (T_4, Γ_4).

Proof By [23, Corollary 4.7] the Voronoi graph Γ has at most three edges and two vertices. Since the complement to the Voronoi graph is a disk, the graph has at least two edges.

Suppose first that Γ has three edges. By [23, Corollary 4.7] the vertices of Γ have multiplicity at least 3, so Γ is a trivalent graph with two vertices, ie a trefoil or an eyeglasses graph. Note that the punctured torus $\dot{T}$ is homeomorphic to a thickening $\mathrm{Th}(\Gamma)$ of Γ, and such $\mathrm{Th}(\Gamma)$ is uniquely determined by choosing a cyclic ordering of the half-edges incident at each vertex of Γ. Now, up to isomorphism, such a cyclic ordering is unique for the eyeglass graph, and its thickening is homeomorphic to a three-punctured sphere. Hence Γ must be a trefoil.

It is easy to see that $\mathrm{Th}(\Gamma)$ can be endowed with a metric such that, if we cut along Γ, we obtain a flat regular hexagon with its center removed. If $\widehat{\mathrm{Th}}(\Gamma)$ is the completion of $\mathrm{Th}(\Gamma)$ obtained by adding one point, then (T_6, Γ_6) is homeomorphic to $(\widehat{\mathrm{Th}}(\Gamma), \Gamma)$, which in turn is homeomorphic to (T, Γ).

The case when Γ has two edges is similar. □

The following is the main proposition on which the proof of Theorem B relies:

Proposition 2.21 (from tori to balanced triangles) *Let (T, x) be a spherical torus with one conical point x and suppose that T has a nontrivial isometric conformal involution σ. Let $\Gamma(T)$ be the Voronoi graph of T.*

(i) *Suppose $\Gamma(T)$ is a trefoil. Then σ permutes the two vertices of $\Gamma(T)$ and fixes the midpoints p_1, p_2 and p_3 of the three edges of $\Gamma(T)$. Moreover, there exist exactly three σ–invariant simple geodesic loops γ_1, γ_2 and γ_3 based at x such that γ_i intersects $\Gamma(T)$ orthogonally at p_i. These geodesic loops cut the torus into the union of two congruent strictly balanced triangles that are exchanged by σ.*

(ii) *Suppose $\Gamma(T)$ is an eight graph with the vertex A. Then σ fixes the vertex and the midpoints p_1 and p_2 of the two edges of $\Gamma(T)$. Moreover there exist four σ–invariant simple geodesic loops γ_1, γ_2, η_1 and η_2 based at x and uniquely characterized by the following properties: each geodesic γ_i intersects $\Gamma(T)$ orthogonally at p_i, and each geodesic η_i passes through A and has length $2d(A, x)$. Moreover, for $i = 1, 2$, the triple of loops $(\gamma_1, \gamma_2, \eta_i)$ cuts T into the union of two congruent semibalanced triangles that are exchanged by σ.*

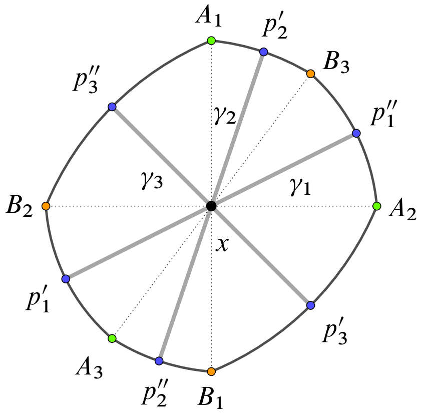


Figure 4: The trefoil case.

(iii) *T has a rectangular involution if and only if its Voronoi graph is an eight graph. For a torus T with a rectangular involution, the triangles into which γ_1, γ_2 and η_1 cut T are reflections of the triangles into which γ_1, γ_2 and η_2 cut T.*

Proof (i) Since σ is an isometry of T it sends $\Gamma(T)$ to itself. Let's denote the vertices of $\Gamma(T)$ by A and B. Since their valence is 3 and σ is a conformal isometric involution, σ can fix neither A nor B. Indeed, since σ is of order 2, if σ fixed A then it would fix at least one half-edge outgoing from A, and so it would be the identity. Hence σ permutes A and B, which implies in particular that A and B are at the same distance from x.

Next, since σ is an orientation-preserving involution and $\Gamma(T)$ is a trefoil, from simple topological considerations it follows that σ sends each edge γ_i of $\Gamma(T)$ into itself. It follows that the midpoints of the edges p_1, p_2 and p_3 are fixed by σ.

Let us now cut T along $\Gamma(T)$ and consider the completion $\overline{D}$ of the obtained open disk. Clearly $\overline{D}$ is a spherical hexagon with the conical point x in its interior. Moreover, σ induces an isometric involution on $\overline{D}$ without fixed points on $\partial\overline{D}$. It follows that σ sends each vertex of $\overline{D}$ to the opposite one.

Next, let's denote the vertices of $\overline{D}$ by A_1, B_2, A_3, B_1, A_2 and B_3, as is shown in Figure 4. Here all the points A_i correspond to A and B_i to B when we reassemble T from the disk. In a similar way we mark midpoints of the sides of $\overline{D}$ by p_i' and p_i''.

According to Lemma 2.18, for each i there is a geodesic loop γ_i of length $2d(p_i, x)$ based at x for which p_i is the midpoint. Let us show that γ_1, γ_2 and γ_3 cut T into two equal strictly balanced triangles whose vertices are identified to the point x.

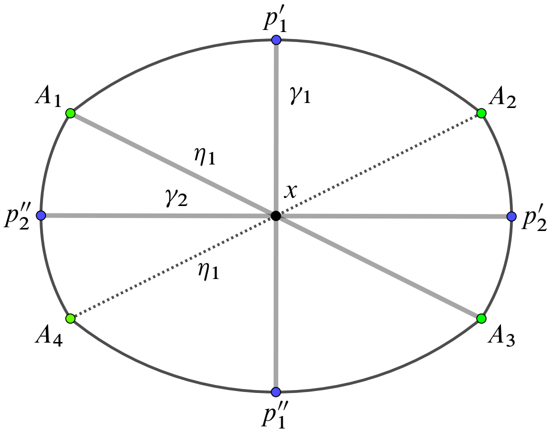


Figure 5: The eight graph case.

Indeed, the first triangle, which we will call Δ_A, is assembled from three quadrilaterals $A_1 p_3'' x p_2'$, $A_2 p_1'' x p_3'$ and $A_3 p_2'' x p_1'$. The second triangle Δ_B is assembled from the remaining three quadrilaterals. Clearly $\sigma(\Delta_A) = \Delta_B$, so these two triangles are congruent.

Finally, Δ_A is strictly balanced according to Theorem 2.9(i). Indeed the point A lies in the interior of Δ_A and is at distance $d(A, x)$ from all the vertices of Δ_A.

(ii) Let us now consider the case when $\Gamma(T)$ is an eight graph with a vertex labeled by A. Clearly, A is fixed by σ since this is the unique point of $\Gamma(T)$ of valence 4.

As before, we see that the midpoints p_1 and p_2 of the two edges of $\Gamma(T)$ are fixed by σ, and this gives us two σ–invariant geodesic loops γ_1 and γ_2. To construct η_1 and η_2 we apply Lemma 2.18 to the point A.

Now let us cut T along the Voronoi graph $\Gamma(T)$ and consider the completion $\overline{D}$ of the obtained open disk. Clearly this disk is a quadrilateral with one conical point in the interior. Let us mark the vertices of this quadrilateral and the midpoints of its edges as shown in Figure 5.

As before, the loops γ_1, γ_2 and η_1 cut T into two congruent triangles, exchanged by σ. To show that these triangles are semibalanced consider one of these triangles obtained as a union of two triangles $A_1 x p_2''$ and $A_3 x p_1''$ and the quadrilateral $x p_1' A_2 p_2'$. To assemble this triangle one has to identify the pairs of sides $(A_1 p_2'', A_2 p_2')$ and $(A_2 p_1', A_3 p_1'')$. The resulting triangle is semibalanced by Theorem 2.9(ii).

(iii) Suppose first that $\Gamma(T)$ is an eight graph. Then we are in the setting of case (ii) of this proposition. Let us construct an involution τ_1 of $\overline{D}$ that pointwise fixes γ_1. We define τ_1 so that $\tau_1(A_1) = A_2$ and $\tau_1(A_3) = A_4$. Then in order show that τ_1 extends to $\overline{D}$ it is enough to show that the triangle A_1xA_4 is isometric to A_2xA_3 and that the geodesic γ_1 is the axis of symmetry of both triangles A_1xA_2 and A_3xA_4. The former statement follows from Proposition 2.3(v). To prove the latter statement, note again that A_1xA_2 is isometric to A_4xA_3 by Proposition 2.3(v) and then compose this isometry with σ. This induces the desired reflections on both triangles A_1xA_2 and A_4xA_3. The involution τ_2 fixing γ_2 is constructed in the same way.

Suppose now that T has a rectangular involution τ. Let us show that $\Gamma(T)$ is an eight graph. Since τ is a rectangular involution, its fixed locus is a union of two disjoint geodesic loops. One of these loops passes through x while the other one, say ξ, is a simple smooth closed geodesic. For any point $p \in \xi$ there exist at least two length-minimizing geodesic segments that join it with x (they are exchanged by τ). It follows that ξ lies in $\Gamma(T)$. And since a trefoil graph can't contain a smooth simple closed geodesic, we conclude that $\Gamma(T)$ is an eight graph. □

Later we will need the following, which is a part of the proof of Proposition 2.21:

Remark 2.22 Suppose we are in case (ii) of Proposition 2.21. Consider the four sectors into which geodesic loops η_1 and η_2 cut a neighborhood of x. Then, for each $i = 1, 2$, the geodesic loop γ_i bisects two of these sectors.

The final preparatory proposition of this subsection is the converse to Proposition 2.21:

Proposition 2.23 (from balanced triangles to tori) *Let Δ be a balanced triangle and let Δ' be a triangle congruent to it. Let $T(\Delta)$ be the torus obtained by identifying the sides of Δ and Δ' through orientation-reversing isometries.*

(i) *The Voronoi graph $\Gamma(T(\Delta))$ coincides with the union in $T(\Delta)$ of $\Gamma(\Delta)$ and $\Gamma(\Delta')$.*

(ii) *If Δ is strictly balanced then the Voronoi graph $\Gamma(T(\Delta))$ has two vertices. Moreover, the images of the three sides of Δ in $T(\Delta)$ coincide with three canonical geodesic loops γ_1, γ_2 and γ_3 on $T(\Delta)$ constructed in Proposition 2.21*(i).

(iii) *If Δ is semibalanced then $\Gamma(T(\Delta))$ has one vertex. Moreover, the images of the three sides of Δ in $T(\Delta)$ coincide with three canonical geodesic loops γ_1, γ_2 and η_i on $T(\Delta)$ constructed in Proposition 2.21*(ii). *Here the side of Δ opposite to the largest angle of Δ corresponds to η_i.*

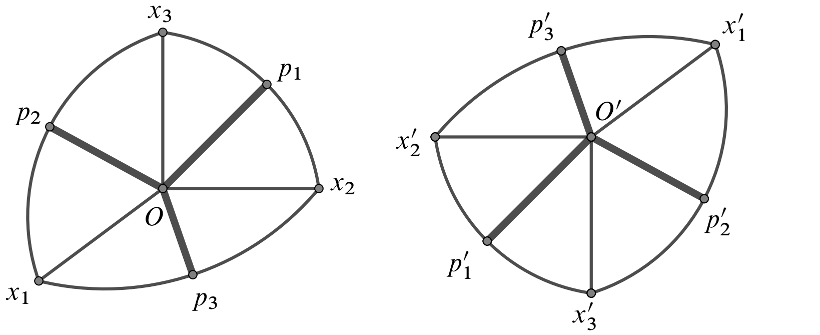


Figure 6: Two isomorphic triangles Δ and Δ'.

Proof (i) Assume first that Δ is strictly balanced. Let $\check{\Gamma}$ be the graph obtained as the union $\Gamma(\Delta) \cup \Gamma(\Delta')$. In order to prove that $\check{\Gamma} = \Gamma(T(\Delta))$, it is enough to show that $\check{\Gamma}$ satisfies properties (a) and (b) of Lemma 2.7.

Recall that by Theorem 2.9(ii) there is a point O in the interior of Δ that is equidistant from the points x_i. Denote by p_i and p'_i the midpoints of sides opposite to x_i and x'_i, as in Figure 6. Then, by Remark 2.12, $\Gamma(\Delta)$ is the union of the segments Op_i and $\Gamma(\Delta')$ is the union of the segments Op'_i. It follows that $T(\Delta) \setminus \check{\Gamma}$ is convex and star-shaped with respect to x, which means that property (a) of Lemma 2.7 holds. As for property (b), it holds since $\Gamma(\Delta)$ and $\Gamma(\Delta')$ are Vornoi graphs of Δ and Δ'.

The case when Δ is semibalanced is treated in the same way, so we omit it.

(ii) Since Δ is strictly balanced, it follows from (i) that $\Gamma(T(\Delta))$ has two vertices. Now, it follows from (i) that for any permutation $\{i, j, k\}$ the side $x_i x_j \subset T(\Delta)$ intersects an edge of $\Gamma(T(\Delta))$ at its midpoint and it is orthogonal to it at this point. Hence, by Proposition 2.21(ii), each geodesic $x_i x_j$ coincides with the geodesic loop γ_k.

(iii) The proof of this result is similar to case (ii) and we omit it. □

Remark 2.24 Proposition 2.23 does not hold for any unbalanced triangle. Indeed, if Δ is unbalanced one can still construct a torus $T(\Delta)$ from Δ and its copy of Δ'. However, the union of the Voronoi graphs of Δ and Δ' will be an eyeglasses graph in $T(\Delta)$. Such a graph can never be the Voronoi graph of a torus with one conical point.

Proof of Theorem B Let T be a spherical torus with one conical point of angle $2\pi\vartheta$ with $\vartheta \notin 2\mathbb{Z}+1$. By Proposition 2.17, there exists a conformal isometric involution σ on T. Hence we can apply Proposition 2.21. In particular, by Proposition 2.21(iii), the torus T has a rectangular involution if and only if $\Gamma(T)$ is an eight graph.

(i) The Voronoi graph $\Gamma(T)$ of T is a trefoil, and we get a collection of three geodesics γ_1, γ_2 and γ_3 that cut T into two congruent strictly balanced triangles. Such a collection of geodesics is unique on T by Proposition 2.23.

(ii) The Voronoi graph $\Gamma(T)$ is an eight graph, and by Proposition 2.21 we get two triples of geodesics, $(\gamma_1, \gamma_2, \eta_1)$ and $(\gamma_1, \gamma_2, \eta_2)$, both cutting T into two congruent semibalanced triangles. Again, it follows from Proposition 2.23 that these two triples are the only ones that cut T into two isometric balanced triangle, and they are exchanged by the rectangular involution. □

3 Balanced spherical triangles

The main goal of this section is to describe the space of balanced spherical triangles with assigned area. To do this, we recall in Section 3.1 several theorems describing the inequalities satisfied by the angles of spherical triangles. We also give an explicit constructions of such triangles. Section 3.2 is mainly expository. It recalls the results from [12] that the space $\mathcal{MT}$ of all (unoriented) spherical triangles has the structure of a 3–dimensional real-analytic manifold. From this we deduce that the space of balanced triangles of a fixed noneven area is a smooth-bordered surface. In Section 3.3 we describe a natural cell decomposition of the space $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ of all balanced triangles of fixed area $\pi(\vartheta-1)$ with $\vartheta \notin 2\mathbb{Z}+1$.

3.1 The shape of spherical triangles

We start this section by recalling the classifications [9] of spherical triangles. In fact, such triangles are in one-to-one correspondence with spheres with a spherical metric with three conical points, provided we exclude spheres and triangles with all integral angles. Indeed, for each S^2 with a spherical metric and three conical points that are not all integral, there is a unique isometric anticonformal involution τ such that S^2/τ is a spherical triangle. Conversely, for each spherical triangle Δ we can take the sphere $S(\Delta)$ formed by gluing together two copies of Δ.

It will be useful to introduce the following notation:

Notation Let $\mathbb{Z}^3_e$ be the subset of $\mathbb{Z}^3$ consisting of triples (n_1, n_2, n_3) with $n_1+n_2+n_3$ even. By d_1 we denote the ℓ_1–distance in $\mathbb{R}^3$ defined by $d_1(\boldsymbol{v}, \boldsymbol{w}) = \sum_i |v_i - w_i|$. If a spherical triangle has angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$, then we call $(\vartheta_1, \vartheta_2, \vartheta_3) \in \mathbb{R}^3$ its associated *angle vector*.

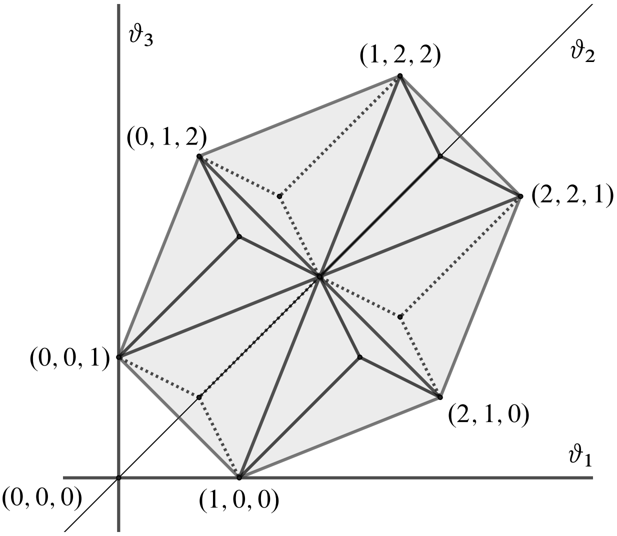


Figure 7: Angle vectors of spherical triangles.

We collect the results into three subsections, depending on the number of integral angles, recalling that there cannot be a triangle with exactly two integral angles.

3.1.1 Triangle with no integral angle

The first result we want to recall from [9] is the following:

Theorem 3.1 (triangles with nonintegral angles [9]) *Suppose ϑ_1, ϑ_2 and ϑ_3 are positive and not integers. A spherical triangle with angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$ exists if and only if*

$$d_1((\vartheta_1, \vartheta_2, \vartheta_3), \mathbb{Z}^3_e) > 1. \tag{3}$$

Moreover, such a triangle is unique, when it exists.

The unique triangle with three nonintegral angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$ will be denoted by $\Delta(\vartheta_1, \vartheta_2, \vartheta_3)$.

Remark 3.2 Let us decipher (3). Note that the subset $d_1((\vartheta_1, \vartheta_2, \vartheta_3), \mathbb{Z}^3_e) \le 1 \subset \mathbb{R}^3$ is a union of octahedra of diameter 2 centered at points of $\mathbb{Z}^3_e$. The complement to this set is a disjoint union of open tetrahedra, each contained in a unit cube with integer vertices. This collection of tetrahedra is invariant under translations of $\mathbb{R}^3$ by elements of $\mathbb{Z}^3_e$. Theorem 3.1 states that if a point $(\vartheta_1, \vartheta_2, \vartheta_3) \in \mathbb{R}^3_{>0}$ lies in one of these tetrahedra, the corresponding spherical triangle exists and is unique. Figure 7 depicts the union of six such tetrahedra in the octant $\mathbb{R}^3_{>0}$.

Section 3.1.2 of [22] contains an explicit construction of balanced spherical triangles. In fact, this was used previously by Klein [17].

3.1.2 Triangles with one integral angle

The second result we wish to recall from [9] is the following:

Theorem 3.3 (triangles with one integral angle [9]) *If ϑ_1 is an integer and ϑ_2 and ϑ_3 are not integers, then a spherical triangle with angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$ exists if and only if at least one of the following conditions is satisfied*:

(a) *$|\vartheta_2 - \vartheta_3|$ is an integer n of opposite parity from ϑ_1 and $n = |\vartheta_2 - \vartheta_3| \le \vartheta_1 - 1$.*

(b) *$\vartheta_2 + \vartheta_3$ is an integer n of opposite parity from ϑ_1 and $n = \vartheta_2 + \vartheta_3 \le \vartheta_1 - 1$.*

Moreover, when the ϑ_i satisfy (a) or (b) there is a one-parameter family of triangles with angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$ that is parametrized by the length $|x_1x_2|$ (or $|x_1x_3|$).

It is obvious that triangles satisfying the hypotheses of Theorem 3.3(b) are never balanced.

Remark 3.4 It is easy to see that, in the case when a triple $(\vartheta_1, \vartheta_2, \vartheta_3)$ of positive numbers satisfies the triangle inequality and the integrality constraints of Theorem 3.3(a), there are integers $n_1, n_2, n_3 \ge 0$ and a number $\theta \in (0, 1)$ such that $\vartheta_1 = n_2 + n_3 + 1$, $\vartheta_2 = n_1 + n_3 + \theta$ and $\vartheta_3 = n_1 + n_2 + \theta$.

Finally, we give a full description of balanced triangles with exactly one integral angle:

Proposition 3.5 (balanced triangles with one integral angle) *Let Δ be a balanced spherical triangle with vertices x_1, x_2 and x_3 and angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$, where ϑ_1 is an integer while ϑ_2 and ϑ_3 are not integers. Let n_1, n_2, n_3 and θ be as in Remark 3.4. Then the following hold*:

(i) $|x_2x_3| = \pi$.

(ii) *There is a unique pair of geodesic segments $\gamma_{12}, \gamma_{13} \subset \Delta$, with $|\gamma_{12}| + |\gamma_{13}| = \pi$, that cut Δ into three domains. The first is a digon with angles πn_3 bounded by the sides x_1x_2 and γ_{13}. The second is a digon with angles πn_2 bounded by the sides x_1x_3 and γ_{13}. The third is a triangle with sides γ_{12}, γ_{13} and x_2x_3, and angles $\pi(\theta + n_1, \theta + n_1, 1)$ opposite to the sides.*

(iii) *All balanced triangles with angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$ are parametrized by the interval $(0, \pi)$, where one can choose either $|x_1x_2|$ or $2\pi - |x_1x_2|$ as a parameter, depending on whether n_3 is even or odd.*

Proof (i) Since Δ is balanced, by Corollary 2.14 we have $|x_1x_2|, |x_2x_3|, |x_3x_1| < 2\pi$. Consider the developing map $\iota\colon \Delta \to \mathbb{S}^2$. Since ϑ_1 is integer, the images $\iota(x_1x_2)$ and $\iota(x_1x_3)$ belong to one great circle C in $\mathbb{S}^2$. At the same time, since the angle ϑ_2 is not an integer, the image $\iota(x_2x_3)$ does not belong to C. This means that $\iota(x_2)$ and $\iota(x_3)$ are opposite points on $\mathbb{S}^2$, and so $|x_2x_3| = \pi$.

(ii) Since $|x_2x_3| = \pi$ by part (i), there exists a maximal digon embedded in Δ, with one edge equal to x_2x_3. The other edge of such a digon must pass through x_1 by maximality, and so it is the concatenation of two geodesics, γ_{12} from x_1 to x_2 and γ_{13} from x_1 to x_3, that form an angle π at x_1. It is easy to see that these are the geodesics we are looking for. The uniqueness of γ_{12} and γ_{13} follows, because n_1 and θ are uniquely determined.

(iii) This follows from part (ii). □

The next lemma is a partial converse to Proposition 3.5(i).

Lemma 3.6 (balanced triangles with one edge of length π) *Let Δ be a balanced spherical triangle with vertices x_1, x_2 and x_3 and angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$. Suppose that $|x_2x_3| = \pi$. Then ϑ_1 is an integer.*

Proof Consider the developing map $\iota\colon \Delta \to \mathbb{S}^2$. Since $|x_ix_j| < 2\pi$ by Corollary 2.14, we see that $\iota(x_i) \neq \iota(x_j)$ for $i \neq j$. In order to show that ϑ_1 is an integer, it is enough to prove that both images $\iota(x_1x_2)$ and $\iota(x_1x_3)$ lie on the same great circle. But this is clear, since the points $\iota(x_2)$ and $\iota(x_3)$ are opposite on $\mathbb{S}^2$, while $\iota(x_1)$ is different from both points. □

The last lemma concerns semibalanced triangles.

Lemma 3.7 (semibalanced triangles with one integral angle) *Suppose Δ is a semibalanced triangle with angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$.*

(i) *If ϑ_i is an integer, then $\vartheta_1 + \vartheta_2 + \vartheta_3$ is an even integer $2m$ and ϑ_j and ϑ_k are half-integers.*

(ii) *If $\vartheta_1 + \vartheta_2 + \vartheta_3 = 2m$, then one, ϑ_i, is an integer and the other two, ϑ_j and ϑ_k, are half-integers.*

Proof Without loss of generality, we can assume that $\vartheta_1 = \vartheta_2 + \vartheta_3$. So certainly $\vartheta_1 + \vartheta_2 + \vartheta_3$ cannot be an odd integer. It follows from [9, Theorem 2] that ϑ_1, ϑ_2 and ϑ_3 cannot be three integers.

(i) Note that ϑ_2 cannot be an integer, because the relation $\vartheta_1 - \vartheta_3 = \vartheta_2$ would violate Theorem 3.3(a). Similarly, ϑ_3 cannot be an integer. Hence ϑ_1 is an integer, and so Theorem 3.3(a) implies that ϑ_2 and ϑ_3 are half-integers.

(ii) Our hypotheses imply that $\vartheta_1 = m$ is an integer. By (i), we obtain that ϑ_2 and ϑ_3 are half-integers. □

3.1.3 Triangles with three integral angles

We begin by giving a description of all triangles with integral angles.

Proposition 3.8 (triangles with three integral angles) *For any spherical triangle Δ with integral angles $\pi(m_1, m_2, m_3)$:*

(i) *There exists a unique triple (n_1, n_2, n_3) of nonnegative integers such that $m_1 = n_2 + n_3 + 1$, $m_2 = n_3 + n_1 + 1$ and $m_3 = n_1 + n_2 + 1$. Moreover, there exist unique geodesic segments $\gamma_{12}, \gamma_{23}, \gamma_{13} \subset \Delta$, with $|\gamma_{12}| + |\gamma_{23}| + |\gamma_{13}| = 2\pi$, that join points x_i and cut Δ into the following four domains:*

- *the central disk Δ_0 isometric to a half-sphere and bounded by segments γ_{12}, γ_{23} and γ_{13};*
- *digons B_1, B_2 and B_3, where each B_i is bounded by segments γ_{jk} and $x_j x_k$ and has angle πn_i.*

(ii) *The space of triangles with angles $\pi(n_1, n_2, n_3)$ can be identified with the set of triples of positive numbers (l_{12}, l_{13}, l_{23}) satisfying $l_{12} + l_{23} + l_{13} = 2\pi$ (where the l_{ij} are interpreted as the lengths of the sides of Δ_0).*

(iii) *All sides of Δ are shorter than 2π. Moreover, there is at most one side of length π.*

Proof (i) Consider the developing map: $\iota\colon \Delta \to \mathbb{S}^2$. Since all the angles of Δ are integral, all its sides are sent to one great circle on $\mathbb{S}^2$. The full preimage of this circle cuts Δ into a collection of hemispheres. It is easy to see that only one of these hemispheres contains all three conical points; this is the disk Δ_0 in Δ. The conical points cut the boundary of the disk into three geodesic segments, γ_{12}, γ_{23} and γ_{13}. The complement of Δ_0 in Δ is the union of the three digons B_1, B_2 and B_3.

(ii) It is clear from (i) that Δ is uniquely defined by the three lengths $l_{ij} = |\gamma_{ij}|$ as well as n_1, n_2 and n_3. Conversely, for each positive triple l_{ij} with $l_{12} + l_{23} + l_{13} = 2\pi$ and each integer triple (n_1, n_2, n_3), one constructs a unique spherical triangle.

(iii) Since $|\gamma_{12}| + |\gamma_{23}| + |\gamma_{31}| = 2\pi$, all the γ_{ij} are shorter than 2π. If $n_k = 0$, then $x_i x_j = \gamma_{ij}$. If $n_k > 0$, then $x_i x_j$ bounds a digon B_k with angles πn_k. In both cases, $x_i x_j$ has length $|\gamma_{ij}|$ (if n_k is even) or $2\pi - |\gamma_{ij}|$ (if n_k is odd). Thus, $|x_i x_j| < 2\pi$.

Moreover, suppose that one of the sides x_ix_j, say x_2x_3, has length π. It follows that $|\gamma_{23}| = \pi$ and so $|\gamma_{12}|, |\gamma_{13}| < \pi$. As a consequence, x_1x_2 and x_1x_3 have length different from π. □

Remark 3.9 (existence of balanced triangles with integral angles) If (m_1, m_2, m_3) is a triple of positive integers that satisfies the triangle inequality, then there exist integers $n_1, n_2, n_3 \geq 0$ such that $m_i = 1 + n_j + n_k$ for $\{i, j, k\} = \{1, 2, 3\}$. Then the construction described in Proposition 3.8(i) shows that there exists a balanced spherical triangle with angles $\pi(m_1, m_2, m_3)$.

We thus obtain a characterization of such triangles (see also [9; 12]):

Corollary 3.10 (balanced triangles of area $2m\pi$) *Let Δ be a triangle.*

(i) *If Δ has integral angles $\pi(m_1, m_2, m_3)$, then Δ is strictly balanced and it has area $2m\pi$ with $m = \frac{1}{2}(m_1 + m_2 + m_3 - 1) \in \mathbb{Z}$.*

(ii) *If Δ has area $2m\pi$ for some integer $m > 0$ and it is balanced, then Δ has integral angles $\pi(m_1, m_2, m_3)$, with $m_1 + m_2 + m_3 = 2m + 1$.*

Proof (i) By Proposition 3.8, the central disk Δ_0 has angles $\pi(1, 1, 1)$ and so it is strictly balanced. Since Δ is obtained from Δ_0 by gluing digons along its edges, Δ is strictly balanced. The second claim is a consequence of [9, Theorem 2].

(ii) Suppose that Δ has angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$. Since $\text{Area}(\Delta) = \pi(\vartheta_1 + \vartheta_2 + \vartheta_3 - 1)$, we see that $\vartheta_1 + \vartheta_2 + \vartheta_3 = 2m + 1$. It follows easily that $d_1((\vartheta_1, \vartheta_2, \vartheta_3), \mathbb{Z}^3_e) = 1$. Hence, from Theorem 3.1, we conclude that at least one of the ϑ_i, say ϑ_1, is an integer.

Assume, for contradiction, that ϑ_2 and ϑ_3 are not integers, and so we are in the setting of Theorem 3.3. The possibility (b) can't hold because Δ is balanced. Assume that possibility (a) holds, in which case $\vartheta_2 - \vartheta_3$ is an integer, and $\vartheta_1 + \vartheta_2 - \vartheta_3$ is odd. But then, since $\vartheta_1 + \vartheta_2 + \vartheta_3$ is also odd, we see that ϑ_3 is an integer. This is a contradiction.

We conclude that all the ϑ_i are integers. □

3.1.4 Final considerations

The last statement of the section can be derived in many ways. Here we obtain it as a consequence of Theorems 3.1 and 3.3 and Proposition 3.8:

Corollary 3.11 (triangles are determined the side lengths and angles) *Let Δ be a spherical triangle with angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$, and let l_i be the length of the side opposite to the vertex x_i. Then Δ is uniquely determined by the ϑ_i and l_i.*

Proof If none of the ϑ_i is an integer, then Δ is uniquely determined by $(\vartheta_1, \vartheta_2, \vartheta_3)$ by Theorem 3.1.

If ϑ_1 is an integer while ϑ_2 and ϑ_3 are not integers, then the triangle Δ is uniquely determined by the angles ϑ_i and the length l_3 by Theorem 3.3.

If ϑ_1, ϑ_2 and ϑ_3 are integers, then it follows from Proposition 3.8 that all triangles with angles ϑ_i are uniquely determined by the lengths of their sides. □

3.2 The space of spherical triangles and its coordinates

Let us denote by $\mathcal{MT}$ be space of all (unoriented) spherical triangles with vertices labeled by x_1, x_2 and x_3, up to isometries that preserve the labeling. This space has a natural topology induced by the Lipschitz distance (see Section 6). We will denote by ϑ_1, ϑ_2, ϑ_3, l_1, l_2 and l_3 the functions on $\mathcal{MT}$ defined by requiring that $\pi\vartheta_i(\Delta)$ is the angle of the spherical triangle Δ at x_i and $l_i(\Delta)$ is the length of the side of Δ opposite to x_i.

By Corollary 3.11, the map $\Psi\colon \mathcal{MT} \to \mathbb{R}^6$ that associates to each triangle its angles and side lengths is one-to-one onto its image. Moreover:

Theorem 3.12 (space of spherical triangles [12, Theorem 1.2]) *Let $\mathcal{MT}$ be the space of spherical triangles. The image $\Psi(\mathcal{MT}) \subset \mathbb{R}^6$ is a smooth connected orientable real-analytic 3–dimensional submanifold of $\mathbb{R}^6$.*

This theorem says that the space $\mathcal{MT}$ has the structure of a smooth connected analytic manifold, and moreover at each point $\Delta \in \mathcal{MT}$ one can choose three functions among the ϑ_i and l_i as local analytic coordinates. It also follows from Theorem 3.12 that formulas of spherical trigonometry, that are usually stated for convex spherical triangles, hold for all spherical triangles. In particular, for any permutation (i, j, k) of $(1, 2, 3)$ and any $\Delta \in \mathcal{MT}$, the following cosine formula for lengths holds:[8]

$$\cos l_i \sin(\pi\vartheta_j)\sin(\pi\vartheta_k) = \cos(\pi\vartheta_i) + \cos(\pi\vartheta_j)\cos(\pi\vartheta_k). \tag{4}$$

Lemma 3.13 (some coordinates on the space $\mathcal{MT}$) *Consider the functions ϑ_1, ϑ_2 and ϑ_3 on $\mathcal{MT}$.*

(i) *The functions ϑ_1, ϑ_2 and ϑ_3 form global analytic coordinates on the (open dense) subset of $\mathcal{MT}$ consisting of triangles with nonintegral angles.*

(ii) *Suppose $\Delta \in \mathcal{MT}$ is short-sided and the angle sum $\vartheta_1(\Delta) + \vartheta_2(\Delta) + \vartheta_3(\Delta)$ is not an odd integer. Then the function $\vartheta_1 + \vartheta_2 + \vartheta_3$ has nonzero differential at Δ.*

[8] Indeed, an analytic function vanishing on an open subset of an irreducible analytic variety vanishes identically.

Proof (i) Consider the projection map from $\Psi(\mathcal{MT})$ to the angle space $\mathbb{R}^3$. According to Theorem 3.1, this map is one-to-one over the subset of $(\vartheta_1, \vartheta_2, \vartheta_3)$ in $\mathbb{R}^3_{>0}$ that satisfy (3). We need to show that this projection is in fact a diffeomorphism over this set. However, using the cosine formula (4) and the fact that none of the ϑ_i are integers, we see that the lengths l_i depend analytically on the ϑ_i.

(ii) As mentioned just before Section 3.1.1, there cannot be a spherical triangle with exactly two integral angles. Moreover, Proposition 3.8(i) implies that Δ cannot have three integral angles if $\vartheta_1(\Delta) + \vartheta_2(\Delta) + \vartheta_3(\Delta)$ is not an odd integer. Thus Δ can have at most one integral angle.

If all the ϑ_i are not integers, the statement follows immediately from (i). Suppose finally that exactly one of the ϑ_i, say ϑ_1, is an integer. Then, since Δ is short-sided, using exactly the same reasoning as in the proof of Proposition 3.5(i), we deduce that $l_i = \pi$. Now, for any $\theta > 0$, we can glue the digon with two sides of length π and angles $\pi\theta$ to the side x_2x_3 of Δ. The family of triangles thus constructed, which depends on θ, determines a straight segment in $\Psi(\mathcal{MT})$ starting from $\Psi(\Delta)$, and the linear function $\vartheta_1 + \vartheta_2 + \vartheta_3$ restricted to this segment has nonzero derivative. □

Definition 3.14 (spaces of triangles with assigned area) For any $\vartheta > 1$ we denote by $\mathcal{MT}(\vartheta) \subset \mathcal{MT}$ the surface consisting of triangles with $\vartheta_1 + \vartheta_2 + \vartheta_3 = \vartheta$. We denote by $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ and $\mathcal{MT}_{\mathrm{sh}}(\vartheta)$ the subsets of balanced and short-sided triangles, respectively.

The following statement is a corollary of Theorem 3.12 and Lemma 3.13:

Corollary 3.15 (space of balanced triangles with assigned area) *For any $\vartheta > 1$, the set $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is a nonsingular real-analytic orientable bordered submanifold of the manifold $\mathcal{MT}$ of all spherical triangles. The boundary of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ consists of semibalanced triangles.*

Proof Suppose first that $\vartheta_1 + \vartheta_2 + \vartheta_3 = 2m + 1$. Balanced spherical triangles of area $2m\pi$ are classified in Corollary 3.10 and Proposition 3.8. They have integral angles, and each connected component forms an open Euclidean triangle in $\mathbb{R}^6$. Clearly such a subset of $\mathcal{MT} \subset \mathbb{R}^6$ is a smooth submanifold.

Assume now that $\vartheta = \vartheta_1 + \vartheta_2 + \vartheta_3$ is not an odd integer. Clearly $\mathcal{MT}_{\mathrm{sh}}$ is an open subset of $\mathcal{MT}$, and so we deduce from Lemma 3.13(ii) that $\mathcal{MT}_{\mathrm{sh}}(\vartheta)$ is an open smooth

2–dimensional submanifold of $\mathcal{MT}$. The set $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is contained in $\mathcal{MT}_{\mathrm{sh}}(\vartheta)$ and its boundary is composed of semibalanced triangles. We need to show that such triangles form a smooth curve in $\mathcal{MT}_{\mathrm{sh}}(\vartheta)$.

Let $\Delta \in \mathcal{MT}_{\mathrm{sh}}(\vartheta)$ be a semibalanced triangle, say $\vartheta_1 = \vartheta_2 + \vartheta_3$. If ϑ_1, ϑ_2 and ϑ_3 are not integers, from Lemma 3.13(i) it follows immediately that the curve $\vartheta_1 - \vartheta_2 - \vartheta_3 = 0$ is smooth in a neighborhood of Δ. Suppose that one of the ϑ_i is an integer. Then we are in the setting of Lemma 3.7. In particular, by Lemma 3.7(i), $\vartheta_1 + \vartheta_2 + \vartheta_3 = 2m$. But then, applying Lemma 3.7(ii), all semibalanced triangles in $\mathcal{MT}_{\mathrm{bal}}(2m)$ have one integral and two half-integral angles. Such triangles are governed by Proposition 3.5, and their image under the map Ψ forms a collection of straight segments in $\mathbb{R}^6$. It follows that semibalanced triangles form a smooth curve in $\mathcal{MT}_{\mathrm{sh}}(2m)$.

Finally, let's show that $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is orientable. This is clear if ϑ is an odd integer, because a disjoint union of open triangles is orientable. If ϑ is not an odd integer, it suffices to show that $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ can be co-oriented, since $\mathcal{MT}$ is orientable. A co-orientation can indeed be chosen since the function $\vartheta_1 + \vartheta_2 + \vartheta_3 = \vartheta$ has nonzero differential along the surface $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ by Lemma 3.13(ii). □

3.3 Balanced spherical triangles of fixed area

The goal of this section is to describe the topology of the moduli space $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ of balanced triangles with marked vertices of fixed area $\pi(\vartheta - 1)$, where $\vartheta > 1$. To better visualize the structure of this space, we introduce the following object:

Definition 3.16 (angle carpet) Take $\vartheta > 1$ such that $\vartheta \notin 2\mathbb{Z} + 1$. The *angle carpet* is the subset of the plane $\Pi(\vartheta) := \{(\vartheta_1, \vartheta_2, \vartheta_3) \in \mathbb{R}^3_{>0} \mid \vartheta_1 + \vartheta_2 + \vartheta_3 = \vartheta\}$ consisting of points such that there exists a spherical triangle with angles $\pi(\vartheta_1, \vartheta_2, \vartheta_3)$, and is denoted by $\mathrm{Crp}(\vartheta)$. Points in $\mathrm{Crp}(\vartheta)$ with one integral coordinate are called *nodes*. The *balanced angle carpet* is the subset $\mathrm{Crp}_{\mathrm{bal}}(\vartheta) := \mathrm{Crp}(\vartheta) \cap \mathrm{Bal}(\vartheta)$, where $\mathrm{Bal}(\vartheta) = \{(\vartheta_1, \vartheta_2, \vartheta_3) \mid \vartheta_i \le \vartheta_j + \vartheta_k\}$. A node in $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ is *internal* if it does not lie on $\partial\,\mathrm{Bal}(\vartheta)$.

Now we separately treat the cases ϑ not odd and ϑ odd.

3.3.1 Case ϑ not odd

Throughout the section, assume $\vartheta \notin 2\mathbb{Z} + 1$. We will denote by $\mathcal{MT}^{\mathbb{Z}}_{\mathrm{bal}}(\vartheta)$ the subset of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ consisting of triangles with at least one integral angle. By Proposition 3.5, this subset is a disjoint union of smooth open intervals in

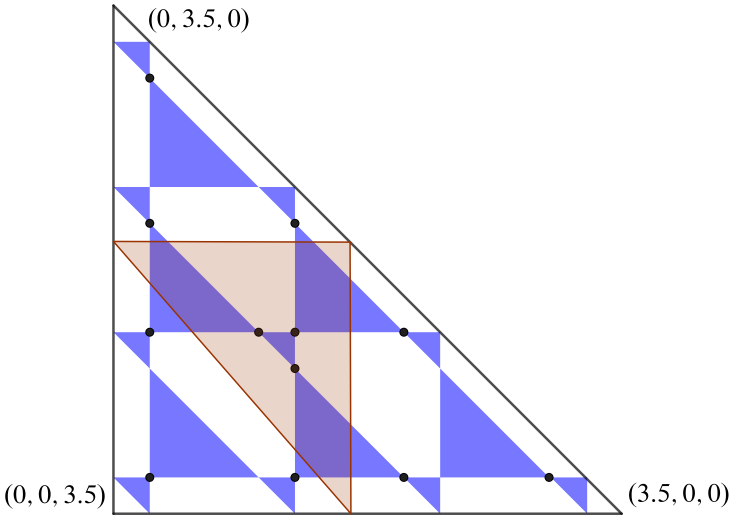


Figure 8: The angle carpet $\mathrm{Crp}(\frac{7}{2})$, composed of 16 open triangles and 12 nodes. The shaded area represents points in $\mathrm{Bal}(\vartheta)$.

$\mathcal{MT}_{\mathrm{bal}}(\vartheta)$. We will see that it cuts $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ into a union of topological disks. This decomposition is very well reflected in the structure of the associated balanced carpet, as we will see below.

The carpet $\mathrm{Crp}(\vartheta)$ is composed of a disjoint union of open triangles with a subset of their vertices (the nodes). In order to better visualize such carpets, we will often identify $\mathrm{Crp}(\vartheta)$ with its projection to the horizontal $(\vartheta_1, \vartheta_2)$–plane. Figure 8 shows the projection of $\mathrm{Crp}(3.5)$. It is a union of 16 disjoint open triangles (singled out by inequality (3) of Theorem 3.1) and a subset of 12 nodes (governed by condition (a) of Theorem 3.3) marked as black dots. Figure 9 depicts the projection of balanced angle carpets for five different values of ϑ.

The following lemma is a consequence of Theorems 3.1 and 3.3:

Lemma 3.17 (description of the angle carpets) *Take $\vartheta \in (1, \infty) \setminus \{2\mathbb{Z} + 1\}$ and set $m = \lfloor \frac{1}{2}(\vartheta + 1) \rfloor$.*

(i) *$\mathrm{Crp}(\vartheta)$ is the union of $4m^2$ open triangles with $3m^2$ nodes $(\vartheta_1, \vartheta_2, \vartheta_3)$ such that the unique integer coordinate ϑ_i of a node satisfies $\vartheta_i \geq |\vartheta_j - \vartheta_k| + 2l + 1$ for some integer $l \leq 0$.*

(ii) *All points $(\vartheta_1, \vartheta_2, \vartheta_3) \in \mathrm{Bal}(\vartheta)$ with one positive integer coordinate are nodes in $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$. Hence, the balanced carpet $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ is a connected set.*

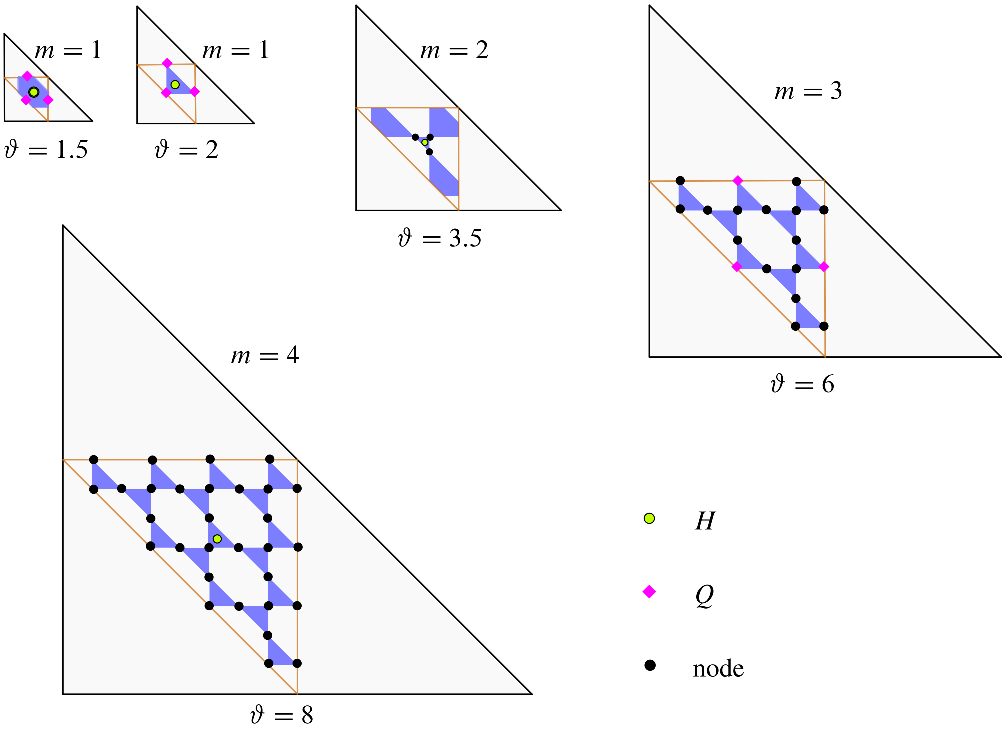


Figure 9: Balanced carpets for $\vartheta = 1.5, 2, 3.5, 6, 8$.

(iii) *The balanced carpet* $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ *intersects* E *open triangles and it contains* N *internal nodes*, *where*

$$E = \begin{cases} m^2 & \text{if } \vartheta \le 2m, \\ m^2 + 3m & \text{if } \vartheta > 2m, \end{cases} \qquad N = \begin{cases} \frac{3}{2}m(m-1) & \text{if } \vartheta \le 2m, \\ \frac{3}{2}m(m+1) & \text{if } \vartheta > 2m. \end{cases}$$

Hence $E - N = -\frac{1}{2}m(m-3)$.

(iv) *There exists a point in* $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ *with noninteger coordinates at which* $\vartheta_2 = \vartheta_3$.

Proof (i) Let us split the carpet into two subsets. The first subset consists of points such that none of the coordinates ϑ_i are integers, and the second subset is where one of the coordinates ϑ_i is an integer.

It is clear that the first subset is the union of open triangles given by intersecting the plane $\vartheta_1 + \vartheta_2 + \vartheta_3 = \vartheta$ with the open tetrahedra that are given by inequality (3) of Theorem 3.1. Since this plane does not pass through any vertex of the tetrahedra for ϑ nonodd, it follows that the number of triangles only depends on m, and so we can compute it for $\vartheta = 2m$. Look at the projection of $\mathrm{Crp}(2m)$ inside the $(\vartheta_1, \vartheta_2)$–plane and enumerate the open triangles as follows: to points of type $\left(0, l + \frac{1}{2}\right)$ with $l \in \{0, 1, \ldots, 2m-1\}$ we can associate a unique triangle, and to points of type $\left(n, l+\frac{1}{2}\right)$ with $n \in \{1, \ldots, 2m-1\}$ and $l \in \{0, \ldots, 2m-n-1\}$ we can associate two triangles. The number of such triangles is thus $4m^2$.

The second subset is governed by Theorem 3.3. Since $\vartheta_1 + \vartheta_2 + \vartheta_3 = \vartheta$ is not an odd integer, only the nodes that satisfy condition (a) of Theorem 3.3 lie in $\mathrm{Crp}(\vartheta)$. Again it's enough to count the nodes for $\vartheta = 2m$. Suppose first that ϑ_1 is an integer. We must have $|2\vartheta_2 + \vartheta_1 - 2m| = |\vartheta_2 - \vartheta_3| = \vartheta_1 - 1 - 2l$ for some integer l. If $\vartheta_1 \in \{1, 2, \ldots, m\}$, then $\vartheta_2 \in \frac{1}{2} + \{m - \vartheta_1, \ldots, m-1\}$ and so we have $\frac{1}{2}m(m+1)$ nodes. If $\vartheta_1 \in \{m+1, \ldots, 2m-1\}$, then $\vartheta_2 \in \frac{1}{2} + \{0, \ldots, 2m-1-\vartheta_1\}$ and so we have $\frac{1}{2}m(m-1)$ nodes. Thus, we have m^2 nodes with integral ϑ_1, and we conclude that we have $3m^2$ nodes in total.

(ii) Again, it is enough to consider the case where $\vartheta = 2m$. In the balanced carpet, $\vartheta_i \leq m$ for all i and so the first claim follows from the above enumeration of the nodes. Hence, $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ is connected.

(iii) Let us first consider N. For $\vartheta = 2m$ the enumeration in part (i) shows that $N = \frac{3}{2}m(m-1)$. If $\vartheta < 2m$, then N does not change. If $\vartheta > 2m$, then $N = \frac{3}{2}m(m-1) + 3m$, and the extra $3m$ is exactly the number of nodes sitting in $\partial\,\mathrm{Bal}(2m)$.

As for E, the enumeration in (i) for $\vartheta = 2m$ shows that $4E = 4m^2$, and so $E = m^2$. For $\vartheta < 2m$, the value of E does not change. For $\vartheta > 2m$, there $3m$ extra triangles intersected by $\mathrm{Bal}(\vartheta)$, which is exactly the number of nodes sitting in $\partial\,\mathrm{Bal}(2m)$.

(iv) The point with $\vartheta_1 = \frac{1}{4}(c+3)$ and $\vartheta_2 = \vartheta_3 = m - \frac{3}{8}(1-c)$ belongs to the interior of $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ and it is not a node. □

In order to understand the topology of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$, we consider the natural projection map $\Theta \colon \mathcal{MT}_{\mathrm{bal}}(\vartheta) \to \mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ that sends Δ to $(\vartheta_1(\Delta), \vartheta_2(\Delta), \vartheta_3(\Delta))$.

Analysis of the map Θ By Lemma 3.17, the balanced carpet $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ consists of E polygons $\{P_l\}$, bounded by some *semibalanced edges* that sit in $\partial\,\mathrm{Bal}(\vartheta)$ and some nodes. Note that we are considering P_l as closed subsets of $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$; in fact, P_l is not a closed subset of the plane $\Pi(\vartheta)$ as it misses the edges sitting on the lines $\vartheta_i = a + \frac{1}{2}(c+1)$ with $i \in \{1, 2, 3\}$ and $a \in \{0, 1, \ldots, m-1\}$. Such edges will be called *ideal edges*. In Figure 10 the polygon P_l on the right has two nodes, one semibalanced edge and three ideal edges. (Note that a node can be semibalanced too.)

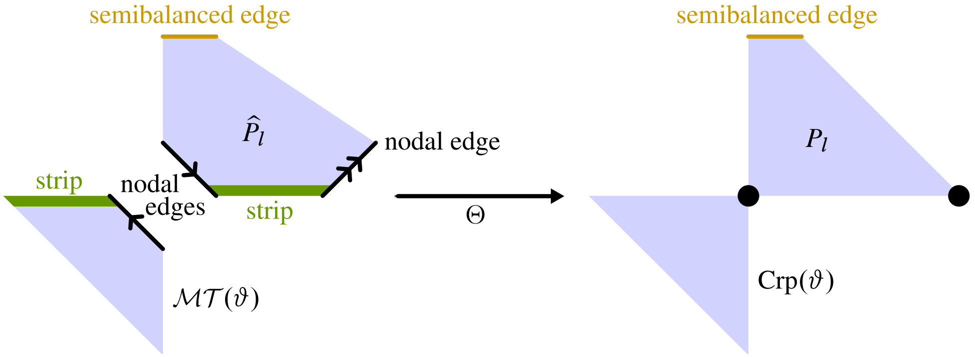


Figure 10: The map Θ. Unmarked edges are ideal edges.

For each polygon P_l, the real blow-up $\widehat{P}_l$ of P_l at its nodes is obtained from P_l by replacing each node by an open interval (*nodal edge*). The natural projection $\widehat{P}_l \to P_l$ contracts each nodal edge to the corresponding node. (Note that a nodal edge can also be semibalanced.) For every l we can fix a realization of $\widehat{P}_l$ inside $\mathbb{R}^2$ as the union of an open convex polygon with some of its open edges (nodal edges and semibalanced edges). Again, such a $\widehat{P}_l$ is not a closed subset of $\mathbb{R}^2$, as it misses the edges corresponding to the ideal edges of P_l. Such missing edges will be referred to as the ideal edges of $\widehat{P}_l$. In Figure 10 the polygon $\widehat{P}_l$ has two nodal edges, one semibalanced edge and three ideal edges.

We recall that $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is a surface by Corollary 3.15 and its boundary consists of semibalanced triangles, and that the map Θ contracts each open interval in $\mathcal{MT}^{\mathbb{Z}}_{\mathrm{bal}}(\vartheta)$ to a node by Proposition 3.5 and it is a homeomorphism elsewhere by Lemma 3.13(i).

It is easy then to see that $\Theta^{-1}(P_l \setminus \{\mathrm{nodes}\})$ is homeomorphic to $\widehat{P}_l \setminus \{\mathrm{nodes}\}$. Suppose now that two distinct polygons P_l and P_h intersect in a node $\bar{\boldsymbol{\vartheta}}$. The preimage $\Theta^{-1}(\bar{\boldsymbol{\vartheta}})$ is an open segment and $\Theta^{-1}(P_l \cup P_h)$ is homeomorphic to the space obtained from $\widehat{P}_l \sqcup \widehat{P}_h$ by identifying the nodal edges that correspond to $\bar{\boldsymbol{\vartheta}}$.

To understand this identification, choose an orientation of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ in a neighborhood of $\Theta^{-1}(\bar{\boldsymbol{\vartheta}})$ and an orientation of the plane $\Pi(\vartheta)$, so that P_l and $\widehat{P}_l$ inherit an orientation from $\Pi(\vartheta)$, and each nodal edge of $\widehat{P}_l$ inherits an orientation from $\widehat{P}_l$. Together with Corollary 3.15, the last paragraph of the proof of [12, Proposition 4.7] shows that Θ is orientation-preserving on one of the two polygons P_l or P_h and orientation-reversing on the other. Hence, the two nodal edges corresponding to $\bar{\boldsymbol{\vartheta}}$ are identified through a map that preserves their orientation; we can also prescribe that such an identification is a homothety in the chosen realizations of $\widehat{P}_l$ and $\widehat{P}_h$.

Part of the above analysis can be rephrased:

Lemma 3.18 *The space $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is homeomorphic to the real blow-up of* $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ *at its nodes.*

A further step in describing the topology of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is to study its ends:

Construction 3.19 (the strips $\mathcal{S}_{i,a}(\vartheta)$) As remarked above, every ideal edge of P_l has equation $\vartheta_i = a + \frac{1}{2}(c+1)$ for some $a \in \{0, \dots, m-1\}$ and $i \in \{1, 2, 3\}$. Viewing $\widehat{P}_l$ inside $\mathbb{R}^2$, an open thickening of the corresponding ideal edge intersects $\widehat{P}_l$ in a region $\mathcal{S}^l_{i,a}(\vartheta)$ homeomorphic to $[0, 1] \times \mathbb{R}$, where $\{0, 1\} \times \mathbb{R}$ corresponds to portions of nodal or semibalanced segments. In every $\widehat{P}_l$, such thickenings can be chosen so that the corresponding regions are disjoint and their ends $\{0\} \times \mathbb{R}$ and $\{1\} \times \mathbb{R}$ cover $\frac{1}{4}$ of the corresponding nodal or semibalanced segment. The complement inside $\widehat{P}_l$ of such strips is clearly compact. (One example of the region $\mathcal{S}^l_{i,a}(\vartheta)$ is illustrated in Figure 10 on the left: it is the darker thickening of the horizontal ideal edge of $\widehat{P}_l$.)

It follows that, for fixed $i \in \{1, 2, 3\}$ and $a \in \{0, 1, \dots, m-1\}$, the regions $\{\mathcal{S}^l_{i,a}(\vartheta)\}$ glue to give a strip $\mathcal{S}_{i,a}(\vartheta)$ homeomorphic to $[0, 1] \times \mathbb{R}$, with $\{0, 1\} \times \mathbb{R}$ corresponding to semibalanced triangles. Thus there are $3m$ disjoint such strips, each one associated to a pair (i, a).

We are now ready to completely determine the topology of the space $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$:

Proposition 3.20 (topology of the space of balanced triangles with assigned area) *Suppose that $\vartheta = 2m + c$ where $c \in (-1, 1)$.*

(i) *$\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is a connected orientable smooth-bordered surface of finite type whose boundary is the set of semibalanced triangles.*

(ii) *The boundary of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is a union of $3m$ disjoint open intervals.*

(iii) *The surface $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ has $3m$ ends, namely the strips $\mathcal{S}_{i,a}(\vartheta)$. Each strip corresponds in* $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ *to a line $\vartheta_i = a + \frac{1}{2}(c+1)$ for some $a \in \{0, 1, \dots, m-1\}$ and $i \in \{1, 2, 3\}$. Moreover, each $\mathcal{S}_{i,a}(\vartheta)$ is homeomorphic to* $[0, 1] \times \mathbb{R}$ *and* $\{0, 1\} \times \mathbb{R}$ *corresponds to semibalanced triangles.*

(iv) *The Euler characteristic of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is $\chi(\mathcal{MT}_{\mathrm{bal}}(\vartheta)) = -\frac{1}{2}m(m-3)$.*

Proof (i) Thanks to Corollary 3.15 we only need to prove that $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is connected and of finite type. Since the balanced carpet $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ is connected by Lemma 3.17(ii) and consists of finitely many nodes and polygons, both claims follow from Lemma 3.18.

(ii) It will be enough to show that the set of semibalanced triangles with angles ϑ_1, ϑ_2 and ϑ_3, satisfying $\vartheta_1 = \vartheta_2 + \vartheta_3$ and $\vartheta_1 + \vartheta_2 + \vartheta_3 = \vartheta$, is a union of m open intervals. In case $c = 0$ these m intervals correspond to m types of triangles with angles $\pi\bigl(m, \frac{1}{2} + l, \frac{1}{2} + m - l - 1\bigr)$ where $l \in [0, m-1]$ is an integer number. In case $c \neq 0$ these intervals correspond to the intersection of the line $\vartheta_1 = \vartheta_2 + \vartheta_3$ with m open triangles of the carpet $\mathrm{Crp}(\vartheta)$.

(iii) This follows from Construction 3.19.

(iv) The internal part of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is an orientable surface without boundary, and so the Euler characteristic of its cohomology with compact support coincides with its Euler characteristic by Poincaré duality. Decompose the interior of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ into a finite union of open 1–cells $\mathcal{MT}^{\mathbb{Z}}_{\mathrm{bal}}(\vartheta)$ (corresponding to internal nodes in the balanced carpet) and open 2–cells (corresponding to the intersection of $\mathrm{Bal}(\vartheta)$ with open triangles in the carpet). By Lemma 3.17(iii), the space $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is a union of E open 2–cells and N open 1–cells. Thus, its Euler characteristic is $E - N = -\frac{1}{2}m(m-3)$. □

Let us now consider balanced triangles (with labeled vertices, as usual) endowed with an orientation. We stress that the orientation and the labeling of the vertices are unrelated. Let $\mathcal{MT}^{+}_{\mathrm{bal}}(\vartheta)$ be the set of oriented balanced triangles of area $\pi(\vartheta - 1)$ in which the vertices are labeled anticlockwise, and let $\mathcal{MT}^{-}_{\mathrm{bal}}(\vartheta)$ be the analogous space in which the vertices are labeled clockwise. Both sets can be given the topology induced by the identification with $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$. The space of oriented balanced triangles is then $\mathcal{MT}^{+}_{\mathrm{bal}}(\vartheta) \sqcup \mathcal{MT}^{-}_{\mathrm{bal}}(\vartheta)$.

Definition 3.21 (doubled space of balanced triangles) The *doubled space of balanced triangles* of area $\pi(\vartheta - 1)$ is the space $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ obtained from $\mathcal{MT}^{+}_{\mathrm{bal}}(\vartheta) \sqcup \mathcal{MT}^{-}_{\mathrm{bal}}(\vartheta)$ by identifying an oriented semibalanced triangle Δ to the triangle obtained from Δ by reversing its orientation.

It follows that $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ is homeomorphic to the double of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$.

Proposition 3.22 (the doubled space of balanced triangles of assigned area) *Let $\vartheta > 1$ be a nonodd real number and let $m = \bigl\lfloor \frac{1}{2}(\vartheta + 1) \bigr\rfloor$.*

(i) *$\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ is a connected orientable surface of finite type, without boundary.*

(ii) *$\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ has Euler characteristic $-m^2$, genus $\frac{1}{2}(m-1)(m-2)$, and $3m$ punctures.*

(iii) *The action of* S_3 *by relabeling the vertices of the triangles consists of orientation-preserving homeomorphisms of* $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$.

(iv) *The action of* S_3 *on the set of punctures of* $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ *has* m *orbits of length* 3.

Proof (i) This is a consequence Proposition 3.20(i), since $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ is the double of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$.

(ii) Since $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ is an orientable surface without boundary, the Euler characteristic agrees with the Euler characteristic with compact support. By Proposition 3.20(ii), $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ has boundary consisting of $3m$ open segments. Hence, $\chi(\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)) = 2\chi(\mathcal{MT}_{\mathrm{bal}}(\vartheta)) - 3m = -m(m-3) - 3m = -m^2$.

By Proposition 3.20(iii), each end of $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ is associated to a strip $\mathcal{S}^i_a(\vartheta)$ with $a \in \{0, 1, \dots, m\}$ and $i \in \{1, 2, 3\}$, and it is homeomorphic to $[0, 1] \times \mathbb{R}$, so it doubles to punctured disk $S^1 \times \mathbb{R}$ inside $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$, which will be denoted by $\mathcal{E}^i_a(\vartheta)$. Hence, we obtain $3m$ punctures. The genus of $g(\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)) = 1 - \frac{3}{2}m - \frac{1}{2}\chi(\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta))$ is then easily computed.

(iii) Choose an arbitrary orientation of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$. We want to show that every transposition $(i\ j) \in \mathrm{S}_3$ acts on $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ through an orientation-preserving homeomorphism. Consider, for instance, the transposition $(2\,3)$, that sends a triangle in $\mathcal{MT}^{+}_{\mathrm{bal}}(\vartheta)$ with nonintegral angles $(\vartheta_1, \vartheta_2, \vartheta_3)$ to the triangle in $\mathcal{MT}^{-}_{\mathrm{bal}}(\vartheta)$ with nonintegral angles $(\vartheta_1, \vartheta_3, \vartheta_2)$. Since $\mathcal{MT}^{+}_{\mathrm{bal}}(\vartheta)$ and $\mathcal{MT}^{-}_{\mathrm{bal}}(\vartheta)$ have opposite orientations when viewed as subsets of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$, it is enough to show that $(2\,3)$ acts on $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ by reversing its orientation.

By Lemma 3.17(iv), there exists a point in $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ with noninteger coordinates $(\vartheta_1, \vartheta_2, \vartheta_2)$, and so a corresponding balanced triangle Δ in $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$. It is clear that the transformation $(\vartheta_1, \vartheta_2, \vartheta_3) \mapsto (\vartheta_1, \vartheta_3, \vartheta_2)$ of $\mathrm{Crp}_{\mathrm{bal}}(\vartheta)$ reverses the orientation at $(\vartheta_1, \vartheta_2, \vartheta_2)$. Hence, (23) acts on $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ by reversing its orientation.

(iv) Each orbit of the S_3–action on the ends $\mathcal{E}^i_a(\vartheta)$ is of type $\{\mathcal{E}^1_a(\vartheta), \mathcal{E}^2_a(\vartheta), \mathcal{E}^3_a(\vartheta)\}$. Since $a \in \{0, 1, \dots, m-1\}$, there are m orbits of length 3. □

3.3.2 Case ϑ odd

The case where $\vartheta = 2m + 1$ for some integer $m \geq 0$ is much easier to handle.

Lemma 3.23 (description of the balanced carpet) *The balanced carpet* $\mathrm{Crp}_{\mathrm{bal}}(2m+1)$ *consists of* $\frac{1}{2}m(m + 1)$ *internal nodes.*

Proof Triangles in $\mathcal{MT}_{\mathrm{bal}}(2m+1)$ have area $2m\pi$ by the Gauss–Bonnet theorem. By Corollary 3.10 and Remark 3.9, the balanced carpet $\mathrm{Crp}_{\mathrm{bal}}(2m+1)$ consists just of triples $(\vartheta_1,\vartheta_2,\vartheta_3)\in\mathbb{Z}^3$ such that $\vartheta_1+\vartheta_2+\vartheta_3=2m+1$ and $1\le\vartheta_i\le m$ for all i. It is easy to see that such points are $\frac{1}{2}m(m+1)$ internal nodes. □

This easily leads to the description of the moduli space of balanced triangles:

Proposition 3.24 (topology of the space of balanced triangles) $\mathcal{MT}_{\mathrm{bal}}(2m+1)$ *is diffeomorphic to the disjoint union of* $\frac{1}{2}m(m+1)$ *copies of the open* 2*–simplex* $\mathring{\Delta}^2$.

Proof Fix $(\vartheta_1,\vartheta_2,\vartheta_3)\in\mathrm{Crp}_{\mathrm{bal}}(2m+1)$. By Proposition 3.8, the locus of triangles Δ in $\mathcal{MT}_{\mathrm{bal}}(2m+1)$ with $\vartheta_i(\Delta)=\vartheta_i$ for $i=1,2,3$ is real-analytically diffeomorphic to the set of triples $(l_1,l_2,l_3)\in(0,2\pi)^3$ such that $l_1+l_2+l_3=2\pi$, which is clearly homothetic to $\mathring{\Delta}^2$. The conclusion then follows from Lemma 3.23. □

Let $\mathrm{Crp}^{\pm}_{\mathrm{bal}}(2m+1)$ be the disjoint union of two copies of $\mathrm{Crp}_{\mathrm{bal}}(2m+1)$. Namely its elements are of type $(\boldsymbol{\vartheta},\epsilon)$, where $\boldsymbol{\vartheta}\in\mathrm{Crp}^{\pm}_{\mathrm{bal}}(2m+1)$ and $\epsilon=\pm1$. We denote by $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ the doubled space of spherical triangles of area $2m\pi$ and by $\Theta^{\pm}\colon\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)\to\mathrm{Crp}^{\pm}_{\mathrm{bal}}(2m+1)$ the map that sends an oriented triangle Δ to $(\boldsymbol{\vartheta}(\Delta),\epsilon(\Delta))$, where $\epsilon(\Delta)=1$ if the vertices of Δ are numbered anticlockwise, and $\epsilon(\Delta)=-1$ otherwise.

Proposition 3.25 (topology of the doubled space of balanced triangles) *The space* $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ *is diffeomorphic to* $\mathrm{Crp}^{\pm}_{\mathrm{bal}}(2m+1)\times\mathring{\Delta}^2$, *namely to the disjoint union of* $m(m+1)$ *open* 2*–simplices. The permutation group* S_3 *that relabels the vertices of a triangle in* $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ *acts on an element* $(\boldsymbol{\vartheta},\epsilon,\boldsymbol{y})$ *of* $\mathrm{Crp}^{\pm}_{\mathrm{bal}}(2m+1)\times\mathring{\Delta}^2$ *by permuting the coordinates of* $\boldsymbol{\vartheta}$ *and* $\boldsymbol{y}$, *and through its sign on* ϵ.

Proof The first claim relies on Proposition 3.24. The others are straightforward. □

4 Moduli spaces of spherical tori

The goal of this section is to describe the topology of the moduli space $\mathcal{MS}_{1,1}(\vartheta)$ and so to prove Theorem A (case ϑ nonodd) and Theorems C and D (case ϑ odd).

We recall that, by *isomorphism* between two spherical tori, we mean an orientation-preserving isometry. We refer to Section 6 for the definition of Lipschitz distance and topology on $\mathcal{MS}_{1,1}(\vartheta)$ and $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ needed below.

The object of our interest is the following:

Definition 4.1 ($\mathcal{MS}_{1,1}(\vartheta)$ as a topological space) The space $\mathcal{MS}_{1,1}(\vartheta)$ is the set of isomorphism classes of spherical tori with one conical point of angle $2\pi\vartheta$, endowed with the Lipschitz topology.

In order to prove Theorem A it will be convenient to introduce the notion of 2–marking:

Definition 4.2 (2–marking) A *2–marking* of a spherical torus T with one conical point x is a labeling of its nontrivial 2–torsion points or, equivalently, an isomorphism $H_1(T;\mathbb{Z}_2) \cong (\mathbb{Z}_2)^2$.

There is a bijective correspondence between isomorphisms $\mu\colon (\mathbb{Z}_2)^2 \to H_1(T;\mathbb{Z}_2)$ and orderings of the three nontrivial elements of $H_1(T;\mathbb{Z}_2)$, sending μ to the triple $(\mu(e_1),\mu(e_2),\mu(e_1+e_2))$. In fact, the action of $\mathrm{SL}(2,\mathbb{Z}_2)$ on 2–markings corresponds to the S_3–action that permutes the orderings. If the torus T has a spherical metric with conical point x, the nontrivial conformal involution σ fixes x and its three nontrivial 2–torsion points. The above ordering is then equivalent to the labeling of these three points. In this case, an isomorphism between two 2–marked spherical tori is an orientation-preserving isometry compatible with the 2–markings.

Definition 4.3 ($\mathcal{MS}_{1,1}^{(2)}$ as a topological space) The space $\mathcal{MS}_{1,1}^{(2)}(\vartheta)$ is the set isomorphisms classes of 2–marked spherical tori with one conical point of angle $2\pi\vartheta$, endowed with the Lipschitz topology.

In Remark 6.28 we show that $\mathcal{MS}_{1,1}(\vartheta)$ and $\mathcal{MS}_{1,1}^{(2)}(\vartheta)$ can be given the structure of orbifolds in such a way that the map $\mathcal{MS}_{1,1}^{(2)}(\vartheta) \to \mathcal{MS}_{1,1}(\vartheta)$ that forgets the 2–marking is a Galois cover with group S_3 (which is unramified in the orbifold sense).

4.1 The case when ϑ is not an odd integer

Because of the relevance for the orbifold structure of the moduli spaces we are interested in, we first classify all possible automorphisms of spherical tori with one conical point:

Proposition 4.4 (automorphisms group of a spherical torus (ϑ nonodd)) *Suppose that $\vartheta \notin 2\mathbb{Z}+1$. For any spherical torus (T,x) of area $2\pi(\vartheta-1)$, the group of automorphisms G_T is isomorphic to $\mathbb{Z}_2$, $\mathbb{Z}_4$ or $\mathbb{Z}_6$.*

(i) *A torus with automorphism group $\mathbb{Z}_6$ exists if and only if $d_1(\vartheta,6\mathbb{Z})>1$.*

(ii) *A torus with automorphism group $\mathbb{Z}_4$ exists if and only if $d_1(\vartheta,4\mathbb{Z})>1$.*

(iii) *For each ϑ, there can be at most one torus with automorphism group $\mathbb{Z}_4$ and one torus with automorphism group $\mathbb{Z}_6$.*

(iv) *The subgroup of G_T of automorphisms that fix the 2–torsion points of T is isomorphic to $\mathbb{Z}_2$ and generated by the conformal involution.*

In Figure 9 we have highlighted with Q or H the triples $\Theta(\Delta)$ such that $T(\Delta)$ has automorphism group isomorphic to $\mathbb{Z}_4$ or $\mathbb{Z}_6$, respectively.

Proof Recall that, by Proposition 2.17, each torus has an automorphism of order 2, namely the conformal involution. Clearly this involution fixes the 2–torsion points of the torus. This implies (iv) and it proves that $|G_T|$ is even.

To bound the automorphism group we note that the action of G_T fixes x and preserves the conformal structure on T. Hence, when $|G_T| > 2$, the torus T is biholomorphic to either $T_4 = \mathbb{C}/(\mathbb{Z} \oplus \zeta_4\mathbb{Z})$ or $T_6 = \mathbb{C}/(\mathbb{Z} \oplus \zeta_6\mathbb{Z})$, where $\zeta_k = \exp(2\pi i/k)$, and its automorphism group is isomorphic to $\mathbb{Z}_4$ (generated by the multiplication by ζ_4) or to $\mathbb{Z}_6$ (generated by the multiplication by ζ_6), respectively.

Let us now prove the existence part of (i) and (ii).

(i) Suppose that $d_1(\vartheta, 6\mathbb{Z}) > 1$. According to Theorem 3.1, this condition is equivalent to the existence of a spherical triangle Δ with angles $\frac{1}{3}\pi\vartheta$. Such a triangle has a rotational $\mathbb{Z}_3$–symmetry. It follows that the torus $T(\Delta)$ has an automorphism of order 6.

(ii) Suppose that $d_1(\vartheta, 4\mathbb{Z}) > 1$. According to Theorem 3.1, this condition is equivalent to the existence of a spherical triangle Δ with angles $\pi\left(\frac{1}{2}\vartheta, \frac{1}{4}\vartheta, \frac{1}{4}\vartheta\right)$. This triangle has a reflection, ie an anticonformal isometry that exchanges two vertices of angles $\frac{1}{4}\pi\vartheta$. Gluing two copies of Δ along the edge that faces the angle $\frac{1}{2}\pi\vartheta$, we obtain a quadrilateral with four edges of the same length and four angles $\frac{1}{2}\pi\vartheta$. It is easy to see that such a quadrilateral has a rotational $\mathbb{Z}_4$–symmetry, and so $T(\Delta)$ has an order-4 automorphism.

Now let (T, x, ϑ) be any spherical torus with $|G_T| > 2$, and let us show that it has to be one of the two tori constructed above. Consider two cases.

First, suppose that the Voronoi graph $\Gamma(T)$ is a trefoil. In this case, by Proposition 2.21 and Theorem B, there is a unique collection of three geodesic loops $(\gamma_1, \gamma_2, \gamma_3)$ based at x that cuts T into two isometric strictly balanced triangles Δ and Δ'. This collection is sent by G_T to itself, and so $|G_T|$ is divisible by three; hence $|G_T| = 6$. It is easy to see then that the subgroup $\mathbb{Z}_3 \subset G_T$ sends Δ to itself and permutes its vertices. So Δ has angles $\frac{1}{3}\pi\vartheta$ and so we are in case (i). Since $\frac{1}{3}\vartheta$ cannot be integer, this also proves the uniqueness of a torus with automorphism group $\mathbb{Z}_6$.

Suppose now that the Voronoi graph $\Gamma(T)$ is an eight graph. Again by Proposition 2.21 and Theorem B, there is a canonical collection of four geodesic loops γ_1, γ_2, η_1 and η_2. Since G_T sends the pair (η_1, η_2) to itself, we see that geodesics η_1 and η_2 cut a neighborhood of x into four sectors of angles $\frac{1}{2}\pi\vartheta$. The same holds for the pair of loops γ_1 and γ_2. Since, by Remark 2.22, each γ_i bisects two sectors formed by η_1 and η_2, we see that, taken together, the geodesics γ_1, γ_2, η_1 and η_2 cut a neighborhood of x into eight sectors of angles $\frac{1}{4}\pi\vartheta$. Hence, γ_1, γ_2 and η_1 cut Δ into two semibalanced triangles with angles $\pi\left(\frac{1}{2}\vartheta, \frac{1}{4}\vartheta, \frac{1}{4}\vartheta\right)$, and so we are in case (ii). The uniqueness of a torus with automorphism group $\mathbb{Z}_4$ follows from the uniqueness of an isosceles triangle with angles $\pi\left(\frac{1}{2}\vartheta, \frac{1}{4}\vartheta, \frac{1}{4}\vartheta\right)$. □

We recall in more detail the construction mentioned in the introduction:

Construction 4.5 Consider the maps of sets

$$\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta) \underset{\Delta^{(2)}}{\overset{T^{(2)}}{\rightleftarrows}} \mathcal{MS}^{(2)}_{1,1}(\vartheta).$$

The map $T^{(2)}$ is defined by sending an oriented triangle Δ to the torus $T(\Delta)$, where we mark by p_i the midpoint of the side opposite to the vertex x_i of Δ.

As for $\Delta^{(2)}$, we proceed as follows. Let $(T, x, \boldsymbol{p})$ be a torus with its order-2 points marked by p_1, p_2 and p_3.

Suppose first that T does not have a rectangular involution. By Theorem B, there is a unique collection of three geodesics loops γ_i that cuts T into two congruent strictly balanced triangles Δ and Δ'. We enumerate the geodesics so that each p_i is the midpoint of γ_i. Next, we label the vertices of Δ by x_1, x_2 and x_3 so that x_i is opposite to γ_i. Hence, we associate to T a unique strictly balanced triangle with enumerated vertices. If the vertices of Δ go in anticlockwise order, we associate to Δ the corresponding point in the interior of $\mathcal{MT}^{+}_{\mathrm{bal}}(\vartheta)$, otherwise we associate to Δ a point in the interior of $\mathcal{MT}^{-}_{\mathrm{bal}}(\vartheta)$.

Suppose now that T has a rectangular involution. Then, by Theorem B, the torus T can be cut into two isomorphic semibalanced triangles in two different ways. At the same time, the rectangular involution sends one pair to the other by reversing the orientation and fixing the labeling of the vertices. This means that the two points associated to T in the boundaries of $\mathcal{MT}^{+}_{\mathrm{bal}}(\vartheta)$ and $\mathcal{MT}^{-}_{\mathrm{bal}}(\vartheta)$ are identified in $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$.

At this point we have the tools to prove the following preliminary fact:

Lemma 4.6 ($T^{(2)}$ is bijective) *The map $T^{(2)}\colon \mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta) \to \mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is a bijection and $\Delta^{(2)}$ is its inverse.*

Proof It is very easy to see that $T^{(2)} \circ \Delta^{(2)}$ is the identity of $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$. Conversely, $\Delta^{(2)} \circ T^{(2)}$ is the identify of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ by Theorem B. □

Remark 4.7 (orbifold Euler characteristic) We recall from the introduction that we are using the definition of orbifold Euler characteristic given by [7, page 29]. We are particularly interested in two properties enjoyed by the orbifold Euler characteristic:

(a) If $\mathcal{Y} \to \mathcal{Z}$ is an orbifold cover of degree d, then $\chi(\mathcal{Y}) = d \cdot \chi(\mathcal{Z})$.

(b) If $\mathcal{Y}$ is a connected orientable 2–dimensional orbifold with underlying topological space Y, then

$$\chi(\mathcal{Y}) = \frac{1}{\mathrm{ord}(\mathcal{Y})}\chi(Y) - \sum_{y}\Big(\frac{1}{\mathrm{ord}(\mathcal{Y})} - \frac{1}{\mathrm{ord}(y)}\Big),$$

where $\mathrm{ord}(\mathcal{Y})$ is the orbifold order of a general point of Y, $\mathrm{ord}(y)$ is the orbifold order of $y \in Y$, and the sum ranges over points $y \in Y$ that have orbifold order strictly greater than $\mathrm{ord}(\mathcal{Y})$.

Since we only compute χ for 2–dimensional connected orientable orbifolds, property (b) could even be taken as a definition.

The main ingredient for the proof of Theorem A is to show that the map $T^{(2)}$ is a homeomorphism, so, as a topological space, $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is a surface. As a consequence, we can endow $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ with an orbifold structure (as done in Remark 6.28) in such a way that every point has orbifold order 2, which is consistent with Proposition 4.4(iv).

Theorem 4.8 (moduli space of spherical tori with 2–marking) *Let $\vartheta > 1$ be a real number such that $\vartheta \notin 2\mathbb{Z}+1$ and let $m = \lfloor \frac{1}{2}(\vartheta+1) \rfloor$. As a topological space, $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ has the following properties:*

(i) *The map $T\colon \mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta) \to \mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is a homeomorphism, and so $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is a connected orientable surface of finite type without boundary.*

(ii) *It has genus $\frac{1}{2}(m-1)(m-2)$ and $3m$ punctures.*

(iii) *The group S_3 that permutes the 2–torsion points of a torus acts on $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ by orientation-preserving homeomorphisms.*

(iv) *The action of S_3 on the set of punctures of $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ has m orbits of length 3.*

As an orbifold, $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is isomorphic to the quotient of its underlying topological space by the trivial $\mathbb{Z}_2$–action, and its orbifold Euler characteristic is $-\frac{1}{2}m^2$.

Proof The map $T^{(2)}$ is bijective by Lemma 4.6, and in fact a homeomorphism by Theorem 6.5. Hence, (i)–(iv) follow from Proposition 3.22(i)–(iv). The orbifold structure was described just above the statement of the theorem: the involution σ is the only nontrivial automorphism of a point in $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ by Proposition 4.4(iv), and it acts trivially on $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$. Hence, $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is isomorphic to the quotient of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ by the trivial $\mathbb{Z}_2$–action. As a consequence, the orbifold Euler characteristic satisfies $\chi(\mathcal{MS}^{(2)}_{1,1}(\vartheta)) = \frac{1}{2}\chi(\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta))$. □

As above, we can endow $\mathcal{MS}_{1,1}(\vartheta)$ with an orbifold structure as in Remark 6.28, in such a way that the orbifold order of a point in $\mathcal{MS}_{1,1}(\vartheta)$ agrees with the number of automorphisms of the corresponding spherical torus.

Proof of Theorem A By Remark 6.28, the map $\mathcal{MS}^{(2)}_{1,1}(\vartheta) \to \mathcal{MS}_{1,1}(\vartheta)$ that forgets the 2–marking is an unramified S_3–cover of orbifolds. Hence, $\mathcal{MS}_{1,1}(\vartheta)$ is a smooth connected 2–dimensional orbifold of finite type by Theorem 4.8(i), and orientability follows from Proposition 4.4.

(ii)–(iv) Clearly $\chi(\mathcal{MS}_{1,1}(\vartheta)) = \chi(\mathcal{MS}^{(2)}_{1,1}(\vartheta))/|\mathrm{S}_3| = -\frac{1}{12}m^2$ by Theorem 4.8. Also, (iii)–(iv) and the remaining claim of (ii) are established in Proposition 4.4.

(i) The space $\mathcal{MS}_{1,1}(\vartheta)$ has m punctures by Theorem 4.8(ii) and (iv). Moreover, its (nonorbifold) Euler characteristic is $2\bigl(\frac{1}{12} - m^2 + \epsilon\bigr)$, where $\epsilon \in \bigl\{0, \frac{1}{4}, \frac{1}{3}, \frac{7}{12} = \frac{1}{4} + \frac{1}{3}\bigr\}$. Indeed, a point of order 4 in $\mathcal{MS}_{1,1}(\vartheta)$ contributes to ϵ with $\frac{1}{4} = \frac{1}{2} - \frac{1}{4}$ and a point of order 6 contributes with $\frac{1}{3} = \frac{1}{2} - \frac{1}{6}$. Hence, the genus of $\mathcal{MS}_{1,1}(\vartheta)$ is $1 - \frac{1}{2}\bigl(m + 2\bigl(-\frac{1}{12}m^2 + \epsilon\bigr)\bigr) = \bigl\lfloor \frac{1}{6}(m^2 - 6m + 12)\bigr\rfloor$. □

Let us finish this subsection with a simple corollary of Theorem 4.8. As a topological space, we denote by $\overline{\mathcal{MS}}^{(2)}_{1,1}(\vartheta)$ the unique smooth compactification of the surface $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ obtained by filling in the $3m$ punctures. As above, we endow $\overline{\mathcal{MS}}^{(2)}_{1,1}(\vartheta)$ with the orbifold structure given by taking the quotient of its underlying topological space by the trivial $\mathbb{Z}_2$–action.

Corollary 4.9 (a cell decomposition of $\overline{\mathcal{MS}}^{(2)}_{1,1}(\vartheta)$) *Suppose that $\vartheta = 2m + c$, where $c \in (-1, 1)$. As a topological space, $\overline{\mathcal{MS}}^{(2)}_{1,1}(\vartheta)$ has the following properties*:

(i) *It is a compact connected orientable surface of genus $\frac{1}{2}(m-1)(m-2)$.*

(ii) *It has a natural structure of a CW complex*, *where*

- *its* 0*–cells are the* $3m$ *added points*;
- *its* 1*–cells are formed by tori* T *such that* $\Delta(T)$ *is ether a semibalanced triangle*, *or a triangle with one integral angle*;
- *its* 2*–cells are the complement of the union of the* 0*–cells and* 1*–cells.*

Moreover, *for* $c \leq 0$, *the cell decomposition is a triangulation into* $2m^2$ *triangles.*

Proof Let us comment on the last claim, since the other claims are rather immediate after Theorem 4.8. Recall that in the proof of Proposition 3.20(iv), for $c \leq 0$, we constructed a decomposition of $\mathcal{MT}_{\text{bal}}(\vartheta)$ into the union of $\frac{3}{2}m(m+1)$ 1-cells and m^2 2-cells. One can check that each of theses m^2 cells has exactly three 1–cells in its boundary. Hence, we get a triangulation of the topological space $\overline{\mathcal{MS}}_{1,1}^{(2)}(\vartheta)$.

Note, however, that for $c > 0$ the total number of 2–cells is $2m^2 + 6m$, and the additional $6m$ cells are digons rather than triangles. □

4.2 The case when ϑ is an odd integer

In this subsection we prove Theorems C and D. Our first step will be to prove Theorem E, from which part (a) of Theorem C is easily obtained.

Proof of Theorem E According to Proposition 2.17, there is a unique curvature-1 metric on T with angle $2\pi(2m+1)$ in a given projective equivalence class, which is invariant under the conformal involution σ of T. Hence we can apply Proposition 2.21 to T endowed with such a σ–invariant metric. According to this proposition, there exist three geodesic loops based at the conical point x that cut T into two isometric balanced triangles Δ and Δ'. By the Gauss–Bonnet formula $\text{Area}(\Delta) = 2\pi m$, and so we can apply Corollary 3.10 to obtain that Δ is a balanced triangle with angles $2\pi(m_1, m_2, m_3)$ where $m_1 + m_2 + m_3 = 2m + 1$. □

This result directly allows us to describe $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ as a topological space:

Proof of Theorem C(a) As in the proof of Theorem E, we can associate to each torus with a σ–invariant metric a unique oriented balanced spherical triangle with integral angles and unmarked vertices. Clearly an orientation on a triangle is equivalent to a numbering of its vertices up to cyclic permutations, and this correspondence determines a bijective map

$$T \colon \mathcal{MT}_{\text{bal}}(2m+1)/\mathrm{A}_3 \to \mathcal{MS}_{1,1}(2m+1)^{\sigma},$$

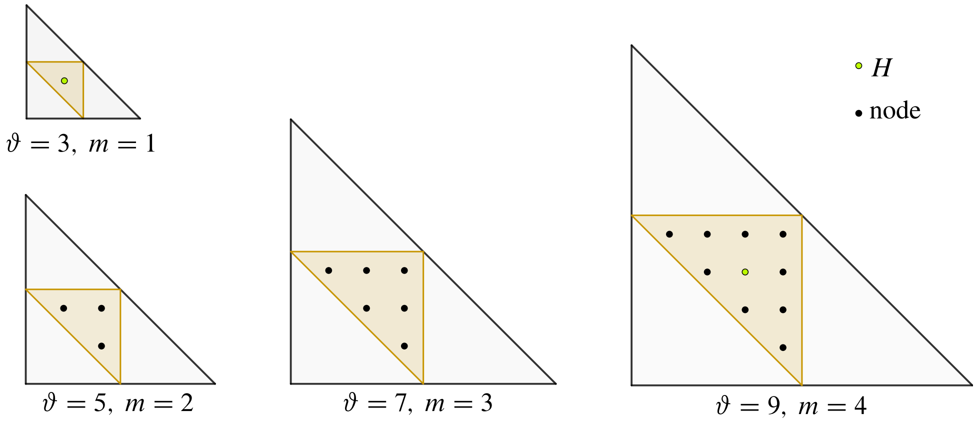


Figure 11: Angle carpets for $\vartheta = 2m+1$ an odd integer.

where the alternating group A_3 acts by relabeling the vertices of the triangle. Arguments entirely analogous to the ones used in Theorem 6.5(ii) show that T is continuous and proper, and hence a homeomorphism of topological spaces.

By Proposition 3.24, the space $\mathcal{MT}_{\text{bal}}(2m+1)$ is homeomorphic to the disjoint union of $\frac{1}{2}m(m+1)$ copies of the open standard simplex $\mathring{\Delta}^2$. Each component represents triangles of angles $\pi(m_1, m_2, m_3)$ with $m_1+m_2+m_3 = 2m+1$, where (m_1, m_2, m_3) is a triple of positive integers that satisfy the three triangle inequalities (see Figure 11).

Consider two cases:

(i) Suppose that $m \not\equiv 1 \pmod 3$. In this case, the integer $2m+1$ is not divisible by 3 and so neither of the spherical triangles in $\mathcal{MT}_{\text{bal}}(2m+1)$ has all equal angles. It follows that the action of A_3 does not send any component to itself. So the number of components of $\mathcal{MS}_{1,1}(2m+1)^\sigma$ is $\frac{1}{6}m(m+1)$ and each one is homeomorphic to the open 2–disk $\mathring{\Delta}^2$.

(ii) Suppose that $m \equiv 1 \pmod 3$. Then the component corresponding to triangles with angles $m_1 = m_2 = m_3 = \frac{1}{3}(2m+1)$ is the only one that is sent to itself. It contains a unique point fixed by A_3, namely the equilateral spherical triangle, and the quotient of this component by A_3 is homeomorphic to an open 2–disk. All the other $\frac{1}{2}(m(m+1)-2)$ components of $\mathcal{MT}_{\text{bal}}(2m+1)$ are nontrivially permuted by A_3; hence, they give $\frac{1}{6}(m(m+1)-2)$ components of $\mathcal{MS}_{1,1}(2m+1)^\sigma$ homeomorphic to $\mathring{\Delta}^2$. Therefore the total number of connected components of $\mathcal{MS}_{1,1}(2m+1)^\sigma$ is $\frac{1}{6}(m(m+1)+4)$. □

The rest of the subsection is devoted to a careful analysis of the orbifold structures on our moduli spaces, the proof of part (b) of Theorem C and the proof of Theorem D.

4.2.1 Voronoi graphs and decorations The orbifold structure on our moduli spaces is defined in Remark 6.28, but a more explicit interpretation of the structure for moduli spaces of tori of area $4m\pi$ relies on the notion of decoration.

We begin with a simple lemma:

Lemma 4.10 (Voronoi graphs of tori of area $4m\pi$) *The Voronoi graph $\Gamma(T)$ of a spherical torus T of area $4m\pi$ has two vertices and three edges of lengths $(2m_i+1)\pi$ for integers $m_i \geq 0$. The two vertices are exchanged by the conformal involution σ. Also, projectively equivalent spherical metrics on a torus have the same Voronoi graph.*

Proof Consider first the case $m=1$. A spherical triangle Δ_0 with vertices x_1, x_2 and x_3 of angles (π,π,π) is isometric to a hemisphere, and its circumcenter O is at distance $\frac{1}{2}\pi$ from the boundary of the hemisphere. So the rotations of the hemisphere that take x_i to x_j fix O. A torus T_0 with a σ–invariant metric h of area 4π is isometric to $T(\Delta_0)$ and so it has three edges and two vertices. Since σ fixes the Voronoi graph $\Gamma(T_0)$ and pointwise fixes the conical point and the midpoints of the three edges of $\Gamma(T_0)$, it does not fix any other point. In particular, σ exchanges the two vertices of $\Gamma(T_0)$. Moreover, the vertices of $\Gamma(T_0)$ are at distance $\frac{1}{2}\pi$ from $\partial\Delta_0$, and so the edges of $\Gamma(T_0)$ have length π. It follows that a (multivalued) developing map for T_0 sends the vertices of $\Gamma(T_0)$ to the two fixed points O and O' for the monodromy, and the edges of $\Gamma(T_0)$ to meridians running between O and O'. Note that another spherical metric on T_0 projectively equivalent to h is obtained by postcomposing the developing map of h by a Möbius transformation that fixes O and O'. Since such transformations preserve the meridians between O and O', the two metrics have the same Voronoi graph.

Suppose now $m>1$. By Theorem E and Proposition 3.8, a torus T with σ–invariant metric of area $4m\pi$ is obtained from a torus $T_0=T(\Delta_0)$ of area 4π as above by gluing a sphere S_i with two conical points of angles $2m\pi$ at distance $|x_jx_k|$ along the geodesic segment x_jx_k of T_0. The conclusion then follows from the analysis of the case $m=1$. □

In order to make the role of the conformal involution σ in Constructions 4.14 and 4.16 below more transparent, we will need:

Definition 4.11 (decorations on strictly balanced tori) A *decoration* v of a spherical torus (T,x) is a vertex v of its Voronoi graph $\Gamma(T)$.

The main reason for introducing decorations relies on the following fact:

Lemma 4.12 (rigidity of 2–marked decorated spherical tori) *Decorated* 2*–marked spherical tori of area* $4m\pi$ *have nontrivial automorphisms.*

Proof Being an isometry, an automorphism is in particular biholomorphic. It is a classical fact that the only nontrivial biholomorphism of a 2–marked conformal torus (T,x) is the involution σ. By Lemma 4.10, the Voronoi graph $\Gamma(T)$ has two vertices and they are exchanged by σ. □

As a consequence, we obtain the following modular interpretation of $\mathcal{MS}_{1,1}^{(2)}(2m+1)$ as a topological space:

Remark 4.13 The topological space $\mathcal{MS}_{1,1}^{(2)}(2m+1)$ is the moduli space of decorated 2–marked spherical tori of area $4m\pi$.

By Lemma 4.10, σ induces an action σ^* on $\mathcal{MS}_{1,1}^{(2)}(2m+1)$ by sending $(T,\boldsymbol{p},v,h)$ to $(T,\boldsymbol{p},v,\sigma^*h)$. Since $\sigma\colon(T,\boldsymbol{p},v,\sigma^*h)\to(T,\boldsymbol{p},\sigma(v),h)$ is an isomorphism, we also have $\sigma^*(T,\boldsymbol{p},v,h)=(T,\boldsymbol{p},\sigma(v),h)$.

4.2.2 Moduli spaces of σ–invariant spherical metrics of area $4m\pi$

Similarly to Section 4.1, we first discuss the space of decorated 2–marked tori:

Construction 4.14 (tori with σ–invariant metrics) If σ is the unique (nontrivial) conformal involution of a conformal torus, denote by $\mathcal{MS}_{1,1}^{(2)}(2m+1)^\sigma$ the set of 2–marked decorated tori $(T,x,\boldsymbol{p})$ with a σ–invariant spherical metric of angle $(2m+1)2\pi$ at x. We recall that triangles in $\mathcal{MT}_{\mathrm{bal}}^{\pm}(2m+1)$ have area $2m\pi$ and integral angles and they are strictly balanced. We then define the maps

$$\mathcal{MT}_{\mathrm{bal}}^{\pm}(2m+1)\underset{\Delta^{(2)}}{\overset{T^{(2)}}{\rightleftarrows}}\mathcal{MS}_{1,1}^{(2)}(2m+1)^\sigma$$

as in Construction 4.5. In particular, $T^{(2)}$ sends an oriented triangle Δ to the 2–marked torus $T(\Delta)$ obtained as the union of Δ and Δ', with the decoration given by the vertex of $\Gamma(T(\Delta))$ that sits inside Δ.

We easily have the following preliminary result:

Theorem 4.15 (moduli space of 2–marked σ–invariant tori of area $4m\pi$) *For* $m>0$ *an integer, the space* $\mathcal{MS}_{1,1}^{(2)}(2m+1)^\sigma$ *of decorated* 2*–marked tori with a* σ*–invariant spherical metric has the following properties*:

(i) *The map* $T^{(2)}\colon \mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1) \to \mathcal{MS}^{(2)}_{1,1}(2m+1)^{\sigma}$ *is a homeomorphism, with inverse* $\Delta^{(2)}$*.*

(ii) $\mathcal{MS}^{(2)}_{1,1}(2m+1)^{\sigma}$ *is a disjoint union of* $m(m+1)$ *open* 2*–disks* $\mathring{\Delta}^2$*.*

(iii) S_3 *acts on* $\mathcal{MS}^{(2)}_{1,1}(2m+1)^{\sigma}$ *by permuting its components. If* $m \not\equiv 1 \pmod 3$*, then all orbits have length* 6*. If* $m \equiv 1 \pmod 3$*, then one orbit has length* 2 *and all the others have length* 6*.*

(iv) *The action of* σ^* *on the topological space* $\mathcal{MS}^{(2)}_{1,1}(2m+1)^{\sigma}$ *is trivial.*

As an orbifold, the moduli space of 2*–marked tori with a* σ*–invariant spherical metric is isomorphic to the quotient of* $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ *by the trivial* $\mathbb{Z}_2$*–action.*

Proof (i) It is very easy to see that $T^{(2)} \circ \Delta^{(2)}$ is the identity on $\mathcal{MS}^{(2)}_{1,1}(2m+1)^{\sigma}$. Conversely $\Delta^{(2)} \circ T^{(2)}$ is the identify on $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ by Theorem E. Hence $T^{(2)}$ is bijective. Moreover, $T^{(2)}$ is a homeomorphism by Theorem 6.5.

(ii)–(iii) These follow from Propositions 3.25 and 3.24.

(iv) This is clear, since σ is an isomorphism between the 2–marked decorated spherical tori $(T, \boldsymbol{p}, v, h)$ and $(T, \boldsymbol{p}, v, \sigma^* h)$.

In view of Remark 6.28, the final claim follows from (iv). □

Now we discuss the moduli space $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ of σ–invariant spherical tori:

Proof of Theorem C(b) Recall $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ is endowed with a 2–dimensional orbifold structure by Remark 6.28. By Theorem 4.15(i), the space $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ is isomorphic to the moduli space of decorated 2–marked tori with a σ–invariant metric. Fix such a torus. Then 2–markings are permuted by S_3 and the decorations are exchanged by σ. Hence, the moduli space $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ is isomorphic (as an orbifold) to the quotient of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ by $\mathrm{S}_3 \times \langle 1, \sigma^* \rangle$. By Proposition 3.25, this quotient can be identified to $\mathcal{MT}_{\mathrm{bal}}(2m+1)/\mathrm{A}_3 \times \mathbb{Z}_2$, where the alternating group A_3 acts by cyclically relabeling the vertices of the triangles and $\mathbb{Z}_2$ acts trivially by Theorem 4.15(iv). By Proposition 3.24, the space $\mathcal{MT}_{\mathrm{bal}}(2m+1)$ consists of $\frac{1}{2}m(m+1)$ connected components and is diffeomorphic to $\mathrm{Crp}_{\mathrm{bal}}(2m+1) \times \mathring{\Delta}^2$.

Consider two cases.

(b-i) Suppose $2m+1$ is not divisible by 3. In this case, neither of the spherical triangles in $\mathcal{MT}_{\mathrm{bal}}(2m+1)$ have all equal angles, so the action of A_3 does not send any component to itself. So the number of components of $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$ is $\frac{1}{6}m(m+1)$

and each one is homeomorphic to the quotient $\mathcal{D}$ of $\mathring{\Delta}^2$ by the trivial $\mathbb{Z}_2$–action, and so all points have orbifold order 2.

(b-ii) Suppose $2m+1$ is divisible by 3. Then the component corresponding to triangles with angles $m_1=m_2=m_3=\frac{1}{3}(2m+1)$ is the only one that is sent to itself. It contains a unique point fixed by A_3, namely the equilateral spherical triangle. This point gives rise to an orbifold point of order 6 on $\mathcal{MS}_{1,1}(2m+1)^{\sigma}$, which belongs to a component homeomorphic to the quotient $\mathcal{D}'$ of $\mathring{\Delta}^2$ by $\mathbb{Z}_2\times\mathrm{A}_3$, where $\mathbb{Z}_2$ acts trivially. All the other $\frac{1}{2}m(m+1)-1$ components are nontrivially permuted by A_3, and they are all homeomorphic to $\mathcal{D}$. Hence, there are$\left\lceil\frac{1}{6}m(m+1)\right\rceil$ connected components, and all points except the equilateral spherical triangle have orbifold order 2. □

4.2.3 Moduli spaces of spherical metrics of area $4m\pi$

In order to treat spherical metrics that are not σ–invariant, we need a further construction.

Construction 4.16 Given a point $O\in\mathbb{S}^2$, let $R\in\mathfrak{su}(2)$ be the unique element with $\mathrm{tr}(R^2)=-\frac{1}{2}$ that generates anticlockwise rotations of $\mathbb{S}^2$ at O.

We view the topological space $\mathcal{MS}^{(2)}_{1,1}(2m+1)$ as a moduli space of decorated, 2–marked tori and we define the pair of maps

$$\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)\times\mathbb{R}\underset{\nu}{\overset{\Xi}{\rightleftarrows}}\mathcal{MS}^{(2)}_{1,1}(2m+1)$$

as follows.

In order to define Ξ, let Δ be an oriented triangle in $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$ and fix a developing map ι for Δ that sends its circumcenter v to $O\in\mathbb{S}^2$. Extend ι to the universal cover of the torus $T(\Delta)$, which has a σ–invariant metric h, and is given a 2–marking as in Construction 4.5. For every $t\in\mathbb{R}$, the map $e^{itR}\circ\iota\colon\widetilde{T}\to\mathbb{S}^2$ has the same equivariance of ι, and so the pullback of the metric of $\mathbb{S}^2$ via such a map descends to a spherical metric h_t on T. We then define $\Xi(T,x,\boldsymbol{p},v,h,t):=(T,x,\boldsymbol{p},v,h_t)$.

In order to define ν, consider a 2–marked decorated spherical torus $(T,x,\boldsymbol{p},v,\hat{h})$, whose metric $\hat{h}$ is not necessarily invariant under the conformal involution σ. Its developing map $\iota\colon\widetilde{T}\to\mathbb{S}^2$ has monodromy contained in a 1–parameter subgroup that fixes $O=\iota(\tilde{v})$, where $\tilde{v}$ is a lift of v, and a maximal circle E. Note that points in $e^{itR}E$ sit at constant distance $\arctan(2e^{-t})$ from O and that the distance from O corresponds to the distance function $d_v\colon T\to[0,\pi]$ from the vertex v. Thus we also have the function $t=-\log\tan\left(\frac{1}{2}d_v\right)\colon T\to[-\infty,\infty]$. We remark that a developing map of $\sigma^*(T,x,\boldsymbol{p},v,\hat{h})$ can be obtained by postcomposing ι with an isometry of $\mathbb{S}^2$ that exchanges O with $-O$. Hence, $t\circ\sigma^*=-t$. It follows that $\hat{h}$ is σ–invariant if and

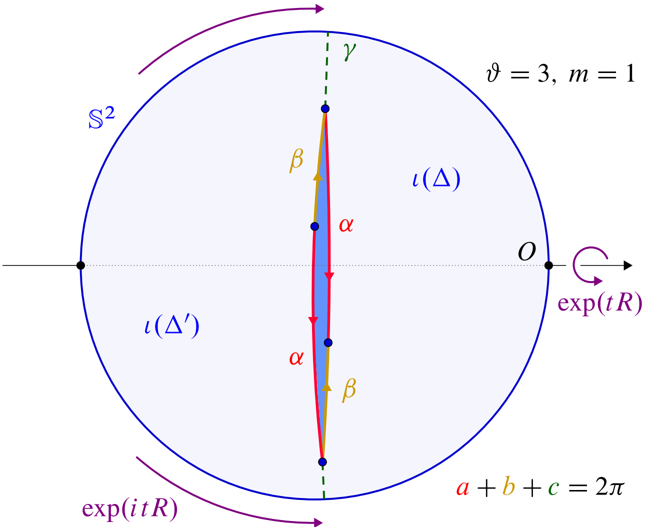


Figure 12: The developing map ι for $T(\Delta) \setminus (\alpha \cup \beta)$, where Δ is a triangle with $\vartheta = 3$ and edges α, β and γ of lengths a, b and c. The two congruent triangles Δ and Δ' are mapped to antipodal hemispheres, and their edges are mapped to the separating equator.

only if $t(x) = 0$, namely $\iota(\tilde{x}) \in E$ for any lift $\tilde{x}$ of x. It is easy to see that the modified developing map $e^{-it(x)R} \circ \iota$ has the same invariance as ι and sends $\tilde{x}$ to E. Hence, the round metric on $\mathbb{S}^2$ pulls back and descends to a σ–invariant metric h on T. We define $\nu(T, x, \boldsymbol{p}, v, \hat{h}) := \Delta^{(2)}(T, x, \boldsymbol{p}, v, h, t(x))$.

Before proceeding, we need a very simple lemma:

Lemma 4.17 (Lipschitz constant of projective transformations) *For every $t \in \mathbb{R}$, the transformation e^{itR} of $\mathbb{S}^2$ has* (bi)*Lipschitz constant* $\cosh(t)$. *Moreover, along the maximal circle E it has Lipschitz constant* $1/\cosh(t)$.

Proof If O is the origin of $\mathbb{C}$ and $2|dz|/(1 + |z|^2)$ is the spherical line element, then the transformation e^{itR} can be written as $z \mapsto e^{-t}z$. Through the map e^{itR} the metric decreases the most at $E = \{|z| = 1\}$, where the Lipschitz constant is exactly $1/\cosh(t)$. □

The first fact about Construction 4.16 is the following:

Proposition 4.18 (the homeomorphism Ξ) *The map Ξ is a homeomorphism and ν is its inverse.*

Proof It is routine to check that the maps Ξ and ν are set-theoretic inverses of each other. Note that the restriction of Ξ to $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)\times\{0\}$ is a homeomorphism by Theorem 4.15(i). Hence, the continuity of Ξ follows from Lemma 4.17.

To show that Ξ is proper, consider a diverging sequence in $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)\times\mathbb{R}$, which we can assume to be contained in a fixed connected component. By Proposition 3.25, an element of this component can be identified by a quadruple (l_1,l_2,l_3,t) with $0<l_i<2\pi$ and $l_1+l_2+l_3=2\pi$. A sequence of quadruples diverges if and only if some $\bar{l}_i\to 0$ or if $|t|\to\infty$ (up to subsequences). Since the systole of the triangle corresponding to (l_1,l_2,l_3) is $\min\{\bar{l}_i\}$ by Lemma 6.24, the systole of the torus $\Xi(l_1,l_2,l_3,t)$ is at most $\min\{\bar{l}_i\}/\log\cosh(t)\to 0$ by Lemma 4.17. It follows that Ξ sends diverging sequences to diverging sequences by Theorem 6.3. □

Since $\mathcal{MS}^{(2)}_{1,1}(2m+1)$ is a manifold by Proposition 4.18, it can be endowed with an orbifold structure as in Remark 6.28. We then have the following preliminary result:

Theorem 4.19 (moduli space of 2–marked tori of area $4m\pi$) *For $m>0$ an integer, the moduli space $\mathcal{MS}^{(2)}_{1,1}(2m+1)$ of 2–marked tori with spherical metric of area $4m\pi$ has the following properties*:

(i) *As an orbifold, it is isomorphic to the quotient of $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)\times\mathbb{R}$ by the action of the involution σ^* that flips the sign of the $\mathbb{R}$ factor. Hence it consists of $m(m+1)$ components isomorphic to $\mathring{\Delta}^2\times(\mathbb{R}/\{\pm 1\})$.*

(ii) *The locus in $\mathcal{MS}^{(2)}_{1,1}(2m+1)$ of metrics that are invariant under the conformal involution σ corresponds to $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)\times\{0\}$.*

(iii) *The group S_3 that permutes the 2–torsion points of the torus acts trivially on $\mathbb{R}$ and as in Proposition 3.25 on $\mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1)$.*

Proof (i) The action of σ is described in Construction 4.16. The claim follows from Theorem 4.15(i) and Proposition 4.18.

(ii) This is also clear by Construction 4.16.

(iii) This follows by noting that relabeling the 2–torsion points does not affect the decoration. □

Proof of Theorem D The forgetful map $\mathcal{MS}^{(2)}_{1,1}(2m+1)\to\mathcal{MS}_{1,1}(2m+1)$ is an unramified S_3–cover of orbifolds. By Theorem 4.19 such a quotient can be identified to

$(\mathcal{MT}_{\mathrm{bal}}(2m+1)\times\mathbb{R})/(\mathrm{A}_3\times\mathbb{Z}_2)$, where $\mathbb{Z}_2$ acts by flipping the sign of the $\mathbb{R}$ factor, and the alternating group A_3 acts by cyclically relabeling the vertices of the triangles.

The rest of argument is entirely analogous to the one used in the proof of Theorem C. □

5 $\mathcal{MS}_{1,1}(2m)$ and $\mathcal{MS}_{1,1}^{(2)}(2m)$ as Belyi curves

The goal of this section is to identify the moduli spaces $\mathcal{MS}_{1,1}(2m)$ and $\mathcal{MS}_{1,1}^{(2)}(2m)$ with Belyi curves, and to relate their cell decompositions constructed in Corollary 4.9 with the corresponding dessins. We recall [2, Section 2; 14] that these two spaces have a *canonical complex structure*. This structure is the unique one with respect to which the forgetful maps to $\mathcal{M}_{1,1}$ and $\mathcal{M}_{1,1}^{(2)}$ are holomorphic. We also recall that the compactification $\overline{\mathcal{MS}}_{1,1}^{(2)}(2m)$, obtained from $\mathcal{MS}_{1,1}^{(2)}(2m)$ by filling in the $3m$ punctures, has orbifold structure that makes it isomorphic to the quotient of its underlying topological space (which is in fact a Riemann surface) by the trivial $\mathbb{Z}_2$–action. The respective forgetful maps extend to the smooth compactifications of all the four orbifolds.

The following definition slightly differs from the usual definition of a dessin d'enfant, though it is very similar in spirit.

Definition 5.1 (Belyi functions and dessins) A *Belyi function* is a holomorphic map $\psi\colon S\to\mathbb{CP}^1$ from a compact Riemann surface S to the complex projective line $\mathbb{CP}^1$, ramified only over points 0, 1 and ∞. The *dessin* associated to ψ is the 3–partite graph embedded in S obtained as the preimage of the real line $\mathbb{RP}^1\subset\mathbb{CP}^1$ under ψ.

The dessin of ψ can also be seen as the 1–skeleton of the triangulation of S whose open cells are the preimages through ψ of the two open disks into which $\mathbb{RP}^1$ cuts $\mathbb{CP}^1$.

The main result of this section concerns the underlying Riemann surface $\overline{\mathcal{MS}}_{1,1}^{(2)}(2m)$:

Theorem 5.2 (the topological space $\overline{\mathcal{MS}}_{1,1}^{(2)}(2m)$ as a Belyi curve) *Let m be a positive integer. Then there is a holomorphic Belyi map $\psi_{\mathrm{Bel}}\colon\overline{\mathcal{MS}}_{1,1}^{(2)}(2m)\to\mathbb{CP}^1$ of degree m^2 from the Riemann surface underlying $\overline{\mathcal{MS}}_{1,1}^{(2)}(2m)$ with the following properties*:

(i) *The preimage of $\mathbb{CP}^1\setminus\{0,1,\infty\}$ under ψ_{Bel} coincides with the Riemann surface $\mathcal{MS}_{1,1}^{(2)}(2m)$.*

(ii) *The cycle type of ramification of ψ_{Bel} over points $\{0,1,\infty\}$ is $(1,3,\dots,2m-1)$.*

(iii) *The dessin of ψ_{Bel} is composed of tori T such that the triangle $\Delta(T)$ has one integral angle. In particular, the triangulation given by this dessin is the one described in Corollary 4.9.*

Definition 5.3 (Klein group and Klein sphere) The *Klein group* K_4 is the subgroup of diagonal matrices in $\mathrm{SO}(3,\mathbb{R})$. The *Klein sphere* S_{Kl} is the sphere with three conical points (y_1,y_2,y_3) of angles (π,π,π), obtained by taking the quotient of the unit sphere $\mathbb{S}^2$ by the action of $K_4\cong\mathbb{Z}_2\oplus\mathbb{Z}_2$. We denote by $S_{\mathrm{Kl}}(\mathbb{R})$ the circle in S_{Kl} which is invariant under the unique antiholomorphic isometric involution of S_{Kl}.

Using the conformal structure on S_{Kl} given by the spherical metric, we can view S_{Kl} as $\mathbb{CP}^1$, where $y_1=0$, $y_2=1$ and $y_3=\infty$, and $S_{\mathrm{Kl}}(\mathbb{R})$ as $\mathbb{RP}^1$.

Remark 5.4 (Klein sphere as a doubled triangle) The Klein sphere S_{Kl} can also be obtained by doubling the spherical triangle Δ with angles (π,π,π) across its boundary. This way $\partial\Delta$ corresponds to the circle $S_{\mathrm{Kl}}(\mathbb{R})$ in S_{Kl}. Recall that, in the triangle with three angles π, each vertex is at distance exactly $\frac{1}{2}\pi$ from each point of the opposite side. For this reason, points of $S_{\mathrm{Kl}}(\mathbb{R})$ are exactly the points on S_{Kl} that are at distance $\frac{1}{2}\pi$ from one conical point.

The key result to parametrize spherical tori using a Hurwitz space is the following:

Proposition 5.5 (tori of area $(2m-1)\pi$ cover the Klein sphere) *Let (T,x) be a spherical torus with a conical point of angle $4\pi m$ and with points of order 2 marked by p_1, p_2 and p_3. There exists a unique branched cover map $\varphi_{\mathrm{Kl}}\colon T\to S_{\mathrm{Kl}}$ of degree $4m-2$ which is a local isometry outside of branching points, and such that $\varphi_{\mathrm{Kl}}(p_i)=y_i$. Moreover, $\varphi_{\mathrm{Kl}}(x)\neq y_i$ for $i=1,2,3$.*

Proof We first construct the map and then count its degree. Recall [2, Proposition 1.5.1] that the image of the monodromy map $\rho\colon\pi_1(T,x)\to\mathrm{SO}(3,\mathbb{R})$ is the Klein group (see also Corollary A.3). Consider the developing map $\iota\colon\widetilde{T}\to\mathbb{S}^2$ from the universal cover $\widetilde{T}$ of T. This map is equivariant with respect to the action of $\pi_1(T,x)$ on $\widetilde{T}$ by deck transformation and on $\mathbb{S}^2$ by the monodromy representation. Hence, by taking the quotient, we get a map $\varphi_{\mathrm{Kl}}\colon T\to S_{\mathrm{Kl}}\cong\mathbb{S}^2/K_4$.

We now prove that the constructed map φ_{Kl} sends points p_i to the three distinct orbifold points of S_{Kl}. This will permit us to label these three points so that $\varphi_{\mathrm{Kl}}(p_i)=y_i$. In order to do this, consider the order-two automorphisms σ of T and denote by S the quotient T/σ. The surface S is a sphere with three conical points of angle π that are the images of the points p_i, and one conical point of angle $2\pi m$. Let us take a lift $\tilde{x} \in \widetilde{T}$ of x and let $\tilde{\sigma}$ be the lift of σ to $\widetilde{T}$ that fixes $\tilde{x}$. Since the conical angle at x is an even multiple of 2π, the maps ι and $\iota \circ \tilde{\sigma}$ coincide in a neighborhood of $\tilde{x}$. It follows that ι is $\tilde{\sigma}$–invariant and the map φ_{Kl} descends to a map $\varphi'_{\mathrm{Kl}}\colon S \to S_{\mathrm{Kl}}$. Now, by construction, the map φ'_{Kl} is a local isometry outside of ramification points. This implies that all three conical points of angle π on S are sent by φ'_{Kl} to conical points of angle π on S_{Kl}. Finally, to see that the images of the three conical points are distinct, we use the fact that the monodromy of S is generated by three loops winding simply around these points, and that it is isomorphic to K_4. Hence, we proved that points $\varphi_{\mathrm{Kl}}(p_i)$ in S_{Kl} are the three distinct conical points of S_{Kl}, and so we can label each $\varphi_{\mathrm{Kl}}(p_i)$ by y_i. This finishes the construction of the map. Its uniqueness is clear.

To prove that $\deg(\varphi_{\mathrm{Kl}}) = 4m-2$, we use the fact that φ_{Kl} is a local isometry outside of branching points, so $\deg(\varphi_{\mathrm{Kl}}) = \mathrm{Area}(T)/\mathrm{Area}(S_{\mathrm{Kl}}) = 2\pi(2m-1)/\pi = 4m-2$. Finally, if φ_{Kl} mapped x to some y_i, its local degree at x would be $(4m\pi)/\pi = 4m > \deg(\varphi_{\mathrm{Kl}})$. This contradiction proves the last claim. □

Corollary 5.6 (moduli space of 2–marked tori as a Hurwitz space) *As a differentiable orbifold, the moduli space $\mathcal{MS}_{1,1}^{(2)}(2m)$ is isomorphic to the Hurwitz space $\mathcal{H}_m$ of connected degree $4m-2$ covers, ramified over points 0, 1 and ∞ with cyclic type $(2,\dots,2)$, and over $\lambda \neq 0, 1, \infty$ with cyclic type $(1,\dots,1,2m)$.*

Proof To construct $\mathcal{MS}_{1,1}^{(2)}(2m) \to \mathcal{H}_m$ we use Proposition 5.5, which associates to each spherical torus (T,x) a 2–marking of the branched cover $\varphi_{\mathrm{Kl}}\colon T \to S_{\mathrm{Kl}}$. Using the conformal structure on S_{Kl} given by the spherical metric, we view it as $\mathbb{CP}^1$, where $y_1 = 0$, $y_2 = 1$ and $y_3 = \infty$. By Proposition 5.5, we know that $\lambda = \varphi_{\mathrm{Kl}}(x) \neq 0, 1, \infty$. To find the cyclic type of ramification over points $(0, 1, \infty, \lambda)$, we recall that the map φ_{Kl} is a local isometry outside of the branching locus, and so for each preimage of the points 0, 1 and ∞ the map has branching of order 2. Finally, there is only one conical point in the preimage of λ, hence the cyclic type over λ is $(1,\dots,1,2m)$.

To define the inverse map $\mathcal{H}_m \to \mathcal{MS}_{1,1}^{(2)}(2m)$, for each ramified cover $T \to \mathbb{CP}^1 \cong S_{\mathrm{Kl}}$ with the prescribed cyclic type we pull back the spherical metric of S_{Kl} to T. By Proposition 5.5, the 2–torsion points of T are mapped to y_1, y_2 and y_3, and we call p_i the unique 2–torsion point of T that is sent to y_i. □

In view of Corollary 5.6, we can give $\mathcal{MS}_{1,1}^{(2)}(2m)$ the unique structure of a complex-analytic orbifold that makes the isomorphism $\mathcal{MS}_{1,1}^{(2)}(2m) \cong \mathcal{H}_m$ complex-analytic.

Now, there are two interesting holomorphic maps. The first map $F\colon \mathcal{H}_w \to \mathcal{M}_{1,1}$ sends a cover $(T,x) \to (\mathbb{CP}^1, \lambda)$ to the isomorphism class of (T,x), and so it has finite fibers. Since F can be interpreted as the forgetful map $\mathcal{MS}^{(2)}_{1,1}(2m) \to \mathcal{M}_{1,1}$, which is proper and surjective (see [23]), the map F is a finite (possibly branched) holomorphic cover. The second map $\psi_{\mathrm{Bel}}\colon \mathcal{H}_w \to \mathbb{CP}^1 \setminus \{0,1,\infty\}$ sends a $(4m-2)$–cover branched over 0, 1, ∞ and λ with cyclic types (2^{2m-1}), (2^{2m-1}), (2^{2m-1}) and $(2m, 1^{2m-2})$ to λ. Since the cyclic types are fixed, ψ_{Bel} is a finite unramified cover. In view of the complex isomorphism between $\mathcal{H}_w$ and $\mathcal{MS}^{(2)}_{1,1}(2m)$, we have proven:

Corollary 5.7 ($\mathcal{MS}^{(2)}_{1,1}(2m)$ covers the 3–punctured sphere) *The map*

$$\psi_{\mathrm{Bel}}\colon \mathcal{MS}^{(2)}_{1,1}(2m) \to \mathbb{CP}^1 \setminus \{0,1,\infty\}$$

is a finite unramified holomorphic cover.

We need one last lemma:

Lemma 5.8 (dessin of ψ_{Bel}) *A torus T in the topological space $\mathcal{MS}^{(2)}_{1,1}(2m)$ belongs to the dessin of ψ_{Bel} if and only if the balanced triangle $\Delta(T)$ has one integral angle.*

Proof Let us prove the "if" direction. Suppose that Δ has an integral angle. Then it has one side of length π. This means that, for some i, the distance on T from x to p_i is $\frac{1}{2}\pi$. This means that the distance on S_{Kl} between y_i and $\varphi_{\mathrm{Kl}}(x)$ is $\frac{1}{2}\pi$. Using Remark 5.4, we deduce that $\varphi_{\mathrm{Kl}}(x)$ belongs to $S_{\mathrm{Kl}}(\mathbb{R})$. By the definition of the dessin of ψ_{Bel}, we see that T belongs to the dessin.

Let us now prove the "only if" direction. Suppose that $\varphi_{\mathrm{Kl}}(x)$ belongs to $S_{\mathrm{Kl}}(\mathbb{R})$. For example, assume $\varphi_{\mathrm{Kl}}(x) \in y_1y_2$. Let γ_3 be the geodesic loop on T based at x whose midpoint is p_3. Since half of this geodesic is projected by φ_{Kl} to the segment that joins y_3 with the segment y_1y_2, we see that $|\gamma_3| = \pi$. From Lemma 3.6, it follows that the angle of Δ opposite to γ_3 is integral. □

Proof of Theorem 5.2 (i) The ramified cover is the extension of the cover constructed in Corollary 5.7 to the compactified spaces.

(ii) Recall $\mathcal{MS}^{(2)}_{1,1}(2m)$ is glued from two copies of $\mathcal{MT}_{\mathrm{bal}}(2m)$ and that $\mathcal{MT}_{\mathrm{bal}}(2m)$ is obtained by gluing m^2 polygons $\widehat{P}_l$ as in Figure 10 (see also the case $\vartheta = 6$ in Figure 9). Let us call each connected component of $\mathbb{CP}^1 \setminus \mathbb{RP}^1$ a "hemisphere" and the intersection of a neighborhood of p with a closed hemisphere a "half-neighborhood" of a point $p \in \mathbb{RP}^1$. Recall now, from Construction 3.19, that the ends of $\mathcal{MS}^{(2)}_{1,1}(2m)$ are described by the strips $\mathcal{S}_{i,a}(2m)$ with $i = 1,2,3$ and $0 \le a \le m-1$. It is easy to see that the "length" of the strip $\mathcal{S}_{i,a}(2m)$, namely the number of regions $\mathcal{S}^{l}_{i,a}(2m)$ such a strip is made of, is exactly $2(2a+1)$.

By Lemma 5.8, the finite unramified cover ψ_{Bel} maps the interior of each $\widehat{P}_l$ onto a hemisphere, and the three nodal edges of $\widehat{P}_l$ are mapped to $\mathbb{RP}^1 \setminus \{0,1,\infty\}$. It follows that, up to labeling the coordinates, ψ_{Bel} maps each region $\mathcal{S}^{l}_{1,a}(2m)$ to a half-neighborhood of 0. Hence, ψ_{Bel} maps a strip $\mathcal{S}_{1,a}(2m)$ of length $2(2a+1)$ onto a (punctured) neighborhood of 0 with degree $2a+1$. It follows that the cycle type ramification of ψ_{Bel} over 0 is $(1,3,5,\dots,2m-1)$. Analogous considerations hold for the cycle type ramification over 1 and over ∞.

(iii) This is proven in Lemma 5.8. □

Proof of Theorem F To prove this result we will realize $\mathcal{MS}_{1,1}(2m)$ as an unramified orbifold cover of the modular curve $\mathbb{H}^2/\mathrm{SL}(2,\mathbb{Z})$. Recall that in Theorem 5.2 we constructed the unramified covering map ψ_{Bel} of degree m^2 from the topological space $\mathcal{MS}^{(2)}_{1,1}(2m)$ to $\mathbb{CP}^1 \setminus \{0,1,\infty\}$. Note that the quotient of $\mathbb{CP}^1 \setminus \{0,1,\infty\}$ by the trivial $\mathbb{Z}_2$–action is an orbifold isomorphic to $\mathbb{H}^2/\Gamma(2)$, where

$$\Gamma(2) = \{A \in \mathrm{SL}(2,\mathbb{Z}) \mid A \equiv I \pmod 2\}.$$

So the above cover can be promoted to an unramified cover of orbifolds $\mathcal{MS}^{(2)}_{1,1}(2m) \to \mathbb{H}^2/\Gamma(2)$ of the same degree.

The symmetric group S_3 acts on $\mathcal{MS}^{(2)}_{1,1}(2m)$ by relabeling the 2–torsion points of the tori, and it acts on $\mathbb{H}^2/\Gamma(2)$ through the isomorphism $\mathrm{S}_3 \cong \mathrm{SL}(2,\mathbb{Z})/\Gamma(2)$.

Since $\mathcal{MS}^{(2)}_{1,1}(2m)/\mathrm{S}_3 = \mathcal{MS}_{1,1}(2m)$ as orbifolds, the covering map then descends to an unramified orbifold covering $\mathcal{MS}_{1,1}(2m) \to \mathbb{H}^2/\mathrm{SL}(2,\mathbb{Z})$ of degree m^2. Note that the cycle type ramification of this cover at infinity is $(1,3,\dots,2m-1)$ by Theorem 5.2(ii). It follows that, for $m > 1$, such a cover is not Galois and so G_m is not a normal subgroup.

The last claim follows from Theorem 5.2(iii), noting that the real locus $\mathbb{RP}^1 \setminus \{0,1,\infty\}$ inside $\mathbb{CP}^1 \setminus \{0,1,\infty\} \cong \mathbb{H}^2/\Gamma(2)$ descends to $\{[it] \mid t \geq 1\}$ inside $\mathbb{H}^2/\mathrm{SL}(2,\mathbb{Z})$. □

6 Lipschitz topology on $\mathcal{MS}_{g,n}$

In this section we define a natural topology on the set of spherical surfaces with conical singularities and establish some of its basic properties. We choose the approach using Lipschitz distance, described, for example in [15, Example on page 71].

We first recall the definition of the Lipschitz distance between two marked metric spaces:

Definition 6.1 Let $(X, x_1, \dots, x_n; d_X)$ and $(Y, y_1, \dots, y_n; d_Y)$ be two metric spaces with distinct marked points x_i and y_i. The *Lipschitz distance* between them is defined by

$$d_{\mathcal{L}}((X, \boldsymbol{x}), (Y, \boldsymbol{y})) = \inf_f \log \max\{\mathrm{dil}(f), \mathrm{dil}(f^{-1})\},$$

where

$$\mathrm{dil}(f) = \sup_{p_1 \neq p_2 \in X} \frac{d_Y(f(p_1), f(p_2))}{d_X(p_1, p_2)}$$

and the infimum runs over bi-Lipschitz homeomorphisms between X and Y that send each x_i to y_i. The value $\max\{\mathrm{dil}(f), \mathrm{dil}(f^{-1})\}$ is called the *bi-Lipschitz constant* of the map f.

Furthermore, we say that a map $f : X \to Y$ is a *bi-Lipschitz embedding* with constant $c \geq 1$ if, for any two points x_1, x_2, we have

$$c^{-1} \cdot d_Y(f(x_1), f(x_2)) \leq d_X(x_1, x_2) \leq c \cdot d_Y(f(x_1), f(x_2)).$$

We will denote by $\mathcal{MS}_{g,n}$ the space of genus-g surfaces with n marked conical points up to a marked isometry. By $\mathcal{MS}_{g,n}(\leq A)$ we denote the subspace of surfaces with area bounded by $A > 0$. To state the main two results of this section we recall the notion of the systole of a spherical surface:

Definition 6.2 (systole) The *systole* $\mathrm{sys}(S)$ of a spherical surface S is the half length of the shortest geodesic segment or geodesic loop on S whose endpoints are conical points of S.

The *systole* $\mathrm{sys}(P)$ of a spherical polygon P is the minimum of half-distances between all vertices of P and the distances between a vertex of P with the unions of edges not adjacent to the vertex. Such a systole is clearly equal to the systole of the sphere obtained by doubling P along its boundary.

Let $\mathcal{MS}^{\geq s}_{g,n}(\leq A)$ be the subspace of $\mathcal{MS}_{g,n}(\leq A)$ of surfaces with systole at least s.

Theorem 6.3 *$\mathcal{MS}_{g,n}$ is a complete metric space with respect to Lipschitz distance. The function $\mathrm{sys}(S)^{-1}$ is proper on $\mathcal{MS}_{g,n}(\leq A)$ in the Lipschitz topology.*

Let us denote by $\mathcal{MP}_n$ the space of all spherical polygons with n cyclically labeled vertices up to isometries that preserve the labeling. We have the following similar result:

Corollary 6.4 *The space $\mathcal{MP}_n$ of spherical polygons with n vertices is complete with respect to Lipschitz distance. For any positive $A > 0$, the function* $\mathrm{sys}^{-1}(P)$ *is proper on the subset $\mathcal{MP}_n$ of polygons with area at most A.*

To prove Theorem 6.3, we show that surfaces from $\mathcal{MS}^{\geq s}_{g,n}(\leq A)$ admit triangulations into a finite number of *relatively large* triangles. This is done in Theorem 6.23, which itself relies on Delaunay triangulations, constructed in Proposition 6.15. The proof of Corollary 6.4 is similar.

As an application of Theorem 6.3 and Corollary 6.4, we get a result on the topology of the space $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ of 2–marked tori induced by Lipschitz metric $\mathcal{L}$:

Theorem 6.5 (i) *Suppose ϑ is not odd. Then $T^{(2)}\colon \mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta) \to (\mathcal{MS}^{(2)}_{1,1}(\vartheta), \mathcal{L})$ is a homeomorphism of surfaces.*

(ii) *Let m be a positive integer. Then $T^{(2)}\colon \mathcal{MT}^{\pm}_{\mathrm{bal}}(2m+1) \to (\mathcal{MS}^{(2)}_{1,1}(2m+1)^{\sigma}, \mathcal{L})$ is a homeomorphism of surfaces.*

Recall that the bijective map $T^{(2)}$ was defined in Construction 4.5, whereas the Lipschitz distance between two 2–marked tori is measured among maps that preserve 2–marking.

6.1 Lipschitz metric and its basic properties

Here we collect basic results concerning the Lipschitz metric, with an emphasis on spherical surfaces.

Lemma 6.6 *Lipschitz distance defines a metric on the space $\mathcal{MS}_{g,n}$ of spherical surfaces of genus g with n conical points.*

Proof Let $(S, \boldsymbol{x}, h)$ and $(S', \boldsymbol{x}', h')$ be genus-g spherical surfaces with n conical points. Let's show that $d_{\mathcal{L}}(S, S') < \infty$, ie that there is a bi-Lipschitz map $\varphi\colon (S, \boldsymbol{x}) \to (S', \boldsymbol{x}')$. By definition, every point x_i has a contractible neighborhood U_i with polar coordinates (r_i, ϕ_i) on which $h = dr_i^2 + \vartheta_i^2 r_i^2 d\phi_i^2$, and similarly for the points x_i'. Pick a small $\varepsilon > 0$ such that the subsets $U_i(\varepsilon) = \{r_i \leq \varepsilon\} \subset U_i$ and $U_i'(\varepsilon) = \{r_i' \leq \varepsilon\} \subset U_i'$ are compact. Define a map $\varphi_i : U_i(\varepsilon) \to U_i'(\varepsilon)$ such that it is the identity in polar coordinates. Manifestly, φ_i has bi-Lipschitz constant $\max\{\vartheta_i'/\vartheta_i,\ \vartheta_i/\vartheta_i'\}$, and it is a diffeomorphism away from x_i. Moreover, it can be extended to a homeomorphism $\varphi\colon (S, \boldsymbol{x}) \to (S', \boldsymbol{x}')$ that is a diffeomorphism from $\dot{S}$ to $\dot{S}'$. Such a map is clearly bi-Lipschitz.

Note that $d_{\mathcal{L}}(S, S') = 0$ if and only if S and S' are isometric by [1, Theorem 7.2.4].

All the other properties of the metric are obvious. □

Definition 6.7 The *Lipschitz topology* on the moduli space $\mathcal{MS}_{g,n}$ of spherical surfaces is the topology induced by the Lipschitz metric.

The next lemma explains how differences in the values of conical angles of two surfaces affects the Lipschitz distance between them.

Lemma 6.8 (continuity of angle functions) *Let U and U' be neighborhoods of conical points x and x' with conical angles ϑ and ϑ'. Suppose $f\colon U \to U'$ is a bi-Lipschitz homeomorphism. Then*

$$\max\{\mathrm{dil}(f), \mathrm{dil}(f^{-1})\} \geq \max\Big(\frac{\vartheta}{\vartheta'}, \frac{\vartheta'}{\vartheta}\Big)^{1/2}. \tag{5}$$

In particular, functions $\vartheta_i\colon \mathcal{MS}_{g,n} \to \mathbb{R}_+$ are continuous for the Lipschitz topology.

Proof After scaling by a large constant and passing to the limit, we can assume that the metrics on U and U' are flat; moreover both U and U' are flat cones with conical angles $2\pi\vartheta$ and $2\pi\vartheta'$, respectively. Note that as a result, the limit quantity $\max\{\mathrm{dil}(f), \mathrm{dil}(f^{-1})\}$ can only decrease. Replacing f by f^{-1} if necessary, we can assume that $\vartheta \leq \vartheta'$.

Let us now reason by contradiction. Assume that (5) is not satisfied. Consider the radius-1 circle S^1 centered at x on U. Since $\mathrm{dil}(f^{-1}) < (\vartheta'/\vartheta)^{1/2}$, the image $f(S^1)$ lies at distance c from x', where $c > (\vartheta/\vartheta')^{1/2}$. Hence, $l(f(S^1)) \geq 2\pi c\vartheta'$. At the same time,

$$\mathrm{dil}(f) \geq \frac{l(f(S^1))}{l(S^1)} = \frac{l(f(S^1))}{2\pi\vartheta} \geq \frac{2\pi c\vartheta'}{2\pi\vartheta} > \Big(\frac{\vartheta'}{\vartheta}\Big)^{1/2}.$$

This contradicts our assumption. □

Lemma 6.9 (continuity of systole function) *Let (S,h) and (S',h') be spherical surfaces from $\mathcal{MS}_{g,n}$ such that $d_{\mathcal{L}}(S,S') \leq d$. Then*

$$e^{-d}\,\mathrm{sys}(S,h) \leq \mathrm{sys}(S',h') \leq e^{d}\,\mathrm{sys}(S,h).$$

In particular, the function sys: $\mathcal{MS}_{g,n} \to \mathbb{R}_+$ *is continuous for the Lipschitz topology.*

Proof Let S be a spherical surface with conical points $x_1,\dots,x_n$. According to [23], $\mathrm{sys}(S)$ is equal to the minimum of half-distances between conical points and half-lengths of all (rectifiable) simple loops based at some conical point x_i contained in $\dot{S} \cup x_i$ and noncontractible in $\dot{S} \cup x_i$. Any bi-Lipschitz homeomorphism f from S to S' that sends conical points x_i of S to the corresponding points x_i' of S' also sends rectifiable loops based at x_i to rectifiable loops based at x_i'. By definition, for any

$\varepsilon > 0$, there exists a homeomorphism $f_\varepsilon \colon S \to S'$ with bi-Lipschitz constant $e^{d+\varepsilon}$. This clearly explains the above inequalities. □

6.2 Injectivity radius

Here we prove Proposition 6.11, which gives an estimate on the injectivity radius of points on spherical surfaces in terms of the value of the Voronoi function and the systole of the surface.

Definition 6.10 Let S be a spherical surface and $y \in \dot{S}$ be a nonconical point. The *injectivity radius* $\mathrm{inj}(y)$ is the supremum of r such that S contains an isometric copy of a spherical disk of radius r embedded in S and centered at y.

For a conical point $x_i \in S$, the injectivity radius is defined to be the minimum of all distances from x_i to other conical points and half lengths of geodesic loops based at x_i.

Proposition 6.11 *Let S be a spherical surface with conical angles $2\pi(\vartheta_1, \dots, \vartheta_n)$. Then, for any $y \in \dot{S}$,*

$$\mathrm{inj}(y) \geq \min(\mathrm{sys}(S), \mathcal{V}_S(y), \min_i \vartheta_i \mathcal{V}_S(y)). \tag{6}$$

Moreover:

(i) *If $\mathrm{inj}(y) < \mathcal{V}_S(y)$, then there exists a closed geodesic loop $\gamma \subset \dot{S}$ of length $2\,\mathrm{inj}(y)$ based at y. Also, $l(\gamma) = 2\,\mathrm{inj}(y) < \pi$.*

(ii) *If $\mathcal{V}_S(y) > \frac{1}{2}\pi$ then $\mathrm{inj}(y) = \mathcal{V}_S(y)$.*

(iii) *If $\mathrm{inj}(y) < \mathcal{V}_S(y)$ and so $\mathcal{V}_S(y) \leq \frac{1}{2}\pi$, then at least one of the following holds:*

 (a) $\mathrm{inj}(y) > \mathrm{sys}(S)$.

 (b) *There exists i such that $\vartheta_i < \frac{1}{2}$ and $\mathrm{inj}(y) > \min_i \vartheta_i \mathcal{V}_S(y)$.*

We will need one lemma to prove this result:

Lemma 6.12 *Let D be a spherical disk with one conical point x in its interior. Suppose that the boundary γ of D satisfies $\ell(\gamma) < 2\pi$ and γ is a geodesic loop with a unique nonsmooth point y. Then there is an orientation-reversing isometric involution τ on D.*

Proof Note first that the angle at x is not an integer, otherwise the univalent developing map from D to $\mathbb{S}^2$ would send γ onto a great circle. Consider the sphere S obtained from D by doubling along γ, and denote by τ_γ the corresponding isometric involution. Since not all conical angles of S are integers, there exists a unique anticonformal isometry τ of S fixing its conical points. Clearly τ commutes with τ_γ, and so τ leaves $\gamma \subset S$ invariant. Hence τ induces the desired involution on $D \subset S$. □

Proof of Proposition 6.11 Since clearly $\mathrm{inj}(y) \le \mathcal{V}_S(y)$, (6) immediately follows from (iii), so we only need to prove (i)–(iii).

(i) Since $\mathrm{inj}(y) < \mathcal{V}_S(y)$, the existence of a geodesic loop of length $2\,\mathrm{inj}(y)$, based at y is straightforward. Indeed, the midpoint of such a loop is a point at distance $\mathrm{inj}(y)$ from y, where the disk centered at y of radius $\mathrm{inj}(y)$ touches itself. One can check that $l(\gamma) \le \pi$, since otherwise there would be points close to the midpoint of γ that could be joined with y by two distinct geodesic segments of length less than $\mathrm{inj}(y)$. To see that $\mathrm{inj}(y) < \frac{1}{2}\pi$ we note that, in case $\mathrm{inj}(y) = \frac{1}{2}\pi$, the boundary of the open disk centered at y of radius $\frac{1}{2}\pi$ is a closed geodesic to which the disk is adjacent twice. This means that S is a standard $\mathbb{RP}^2$, which is impossible since S is orientable.

(ii) Assume $\mathcal{V}_S(y) > \frac{1}{2}\pi$ and suppose, for contradiction, that $\mathrm{inj}(y) < \mathcal{V}_S(y)$. Let γ be a geodesic constructed in (i). Let $2\pi\theta$ and $2\pi(1-\theta)$ be the angles into which γ cuts the neighborhood of y, and assume, without loss of generality, that $\theta \le \frac{1}{2}$.

Take a point $O \in \mathbb{S}^2$ and consider a spherical kite OP_1QP_2 in $\mathbb{S}^2$ with $\angle O = 2\pi\theta$, $\angle P_1 = \angle P_2 = \frac{1}{2}\pi$ and $l([OP_1]) = l([OP_2]) = \frac{1}{2}l(\gamma)$. Since $\theta \le \frac{1}{2}$ and $l([OP_1]) \le \frac{1}{2}\pi$, one can check that $l([OQ]) \le \frac{1}{2}\pi$. In particular, the kite lies in the interior of a disk $\mathbb{D}_r$ centered at O for any $r \in \left(\frac{1}{2}\pi, \mathcal{V}_S(y)\right)$. Since $\mathcal{V}_S(y) > r$, there exists a locally isometric immersion $\iota\colon \mathbb{D}_r \to \dot{S}$ such that $\iota(O) = y$. By precomposing ι with a rotation, we can arrange so that ι sends the sides OP_1 and OP_2 to γ, and $\iota(P_1) = \iota(P_2)$ is the midpoint of γ. It is clear then that the segments P_1Q and P_2Q are sent by ι to the same geodesic segment in $\dot{S}$. It follows that ι is not a locally isometric immersion in any neighborhood of Q. This is a contraction.

(iii) Since $\mathrm{inj}(y) < \mathcal{V}_S(y)$, by (i) there is a simple geodesic loop γ on $\dot{S}$ based at y of length $2\,\mathrm{inj}(y) < \pi$. We will consider separately two possibilities, depending on whether γ is *essential* (it doesn't bound on $\dot{S}$ a disk with at most one puncture) on $\dot{S}$.

If γ is essential on $\dot{S}$, it follows from [23] that $\mathrm{inj}(y) = \frac{1}{2}l(\gamma) > \mathrm{sys}(S)$, and so we are in case (a).

Let's assume now that γ is nonessential on $\dot{S}$. Then γ encircles on S a disk D with at most one conical point in its interior. Since $l(\gamma) < \pi$ by (i), the disk D should contain exactly one conical point, which we denote by x_i. Denote by $2\pi\theta$ the angle that γ forms at y in D.

Suppose first that $\theta \ge \frac{1}{2}$. In this case γ forms a convex boundary of the surface $S \setminus D$. Thanks to this, using exactly the same method as in [23, Corollary 3.11], one proves that $l(\gamma) > 2\,\mathrm{sys}(S)$, and we are in case (a).

Suppose now $\theta < \frac{1}{2}$. Since $\ell(\gamma) < \pi$, we can apply Lemma 6.12 to D to get its isometric involution τ. This involution fixes the midpoint p of γ, and fixes two geodesic segments yx_i and px_i that cut D into two isometric right-angled spherical triangles. Let yp be one of two halves of γ. The segments yx_i, px_i and yp border a triangle $x_i yp$ in D with $\angle x_i = \pi\vartheta_i$, $\angle y = \frac{1}{2}\pi\theta$ and $\angle p = \frac{1}{2}\pi$. Since the side yp of the triangle is shorter than π and two adjacent angles are less than π, the triangle is convex. Since $|yx_i| > |yp|$, we have $\theta_i < \frac{1}{2}$. Applying the sine rule to the triangle $x_i yp$ we get $\sin(|yp|) = \sin(\pi\vartheta_i)\sin(|x_i y|)$. Hence

$$\operatorname{inj}(y) = |yp| > \sin(\pi\vartheta_i)\sin(|x_i y|) > 2\vartheta_i \sin(\mathcal{V}_S(y)) > \frac{4}{\pi}\vartheta_i \mathcal{V}_S(y),$$

which proves that we are in case (b). □

6.3 Equivalence of Lipschitz and analytic topologies on $\mathcal{MT}$

In this section we prove that Lipschitz distance between triangles induces the same topology on $\mathcal{MT}$ as the topology induced by the embedding in $\mathbb{R}^6$ described in Theorem 3.12.

Definition 6.13 The *relative Lipschitz distance* $d_{\overline{\mathcal{L}}}$ (or $\overline{\mathcal{L}}$–distance) between two spherical triangles is the infimum of $\log\max(\operatorname{dil}(f), \operatorname{dil}(f^{-1}))$ over all the marked bi-Lipschitz homeomorphisms $f\colon \Delta_1 \to \Delta_2$ that restrict to a homothety on each edge of Δ_1.

The $\overline{\mathcal{L}}$–distance defines a metric on the space $\mathcal{MT}$ of spherical triangles, which we call the $\overline{\mathcal{L}}$–metric. We have the following natural statement.

Proposition 6.14 *The topologies defined on $\mathcal{MT}$ by the $\mathcal{L}$– and $\overline{\mathcal{L}}$–metrics coincide with the analytic topology given by the angle–side length embedding $\Psi\colon \mathcal{MT} \to \mathbb{R}^6$.*

Proof Note that the side lengths of Δ are clearly continuous functions in both the $\overline{\mathcal{L}}$ and $\mathcal{L}$ topologies. The angles of Δ are continuous in these topologies thanks to Lemma 6.8, applied to the double of Δ. Furthermore the $\overline{\mathcal{L}}$–distance is greater than or equal to the $\mathcal{L}$–distance. Hence, the $\overline{\mathcal{L}}$–topology is finer than the $\mathcal{L}$–topology, which is finer than the analytic topology. For this reason, we only need to show that, for any spherical triangle Δ and a sequence of triangles Δ_i converging to Δ in $\mathbb{R}^6$ (ie in the analytic topology), we have $\lim d_{\overline{\mathcal{L}}}(\Delta_i, \Delta) = 0$. This claim can be proven by exhibiting explicit bi-Lipschitz maps between spherical triangles. We will only treat the case when Δ is short-sided, since this is the only case needed for our purposes.

Following [12, Lemma 4.1], denote by U the open subset of $\mathcal{MT}$ consisting of triangles with angles $\pi\vartheta_i$, where $\vartheta_i < 2$. This subset consists of spherical triangles that admit an isometric embedding into $\mathbb{S}^2$. In particular, U lies in $\mathcal{MT}_{\mathrm{sh}}$, the space of all short-sided triangles. We first prove that the $\overline{\mathcal{L}}$–topology coincides with the analytic topology on U.

For two spherical triangles $\Delta = x_1x_2x_3$ and $\Delta' = x_1'x_2'x_3'$ embedded into $\mathbb{S}^2$ with incenters I_Δ and $I_{\Delta'}$, respectively, define the *incentric* map $\Phi\colon \Delta \to \Delta'$ as the unique map satisfying:

- $\Phi(x_i) = x_i'$, $\Phi(I_\Delta) = \Phi(I_{\Delta'})$.
- Φ is a homothety on each edge x_ix_j.
- For any point $p \in \partial\Delta$, Φ sends the geodesic segment $pI_{\Delta'}$ to a geodesic segment and restricts to a homothety on it.

Suppose now we have a sequence of embedded triangles $\Delta_i \in U$ whose angles and side lengths converge to those of $\Delta \in U$. Then it is not hard to see that the bi-Lipschitz constant of the incentric maps $\Phi\colon \Delta_i \to \Delta$ tends to 1. Hence Δ_i converges to Δ in the $\overline{\mathcal{L}}$–topology as well. This proves the statement for U.

Let us denote by $U_{klm} \subset \mathcal{MT}_{\mathrm{sh}}$ the subspace of triangles which can be obtained from an embedded triangle Δ by repeated gluing of $k-1$, $l-1$ and $m-1$ hemispheres to the sides x_1x_2, x_2x_3 and x_3x_1, respectively, of Δ. From [12, Theorem 4.7 and Lemma 5.2] it follows that the sets U_{klm} give an open cover of $\mathcal{MT}_{\mathrm{sh}}$. At the same time, the incentric map Φ between any two triangles Δ and Δ' from U can be naturally extended to a map $\widetilde{\Phi}\colon \tilde{\Delta} \to \tilde{\Delta}'$ between triangles with attached hemispheres. Namely, a radius of each hemisphere is sent isometrically to a radius and the restriction of $\widetilde{\Phi}$ to both sides of each hemisphere are homotheties. Since the Lipschitz constants of Φ and $\widetilde{\Phi}$ clearly coincide, the statement about the topologies is proven for each U_{klm}, and so for the whole space $\mathcal{MT}_{\mathrm{sh}}$. □

6.4 Delaunay triangulations

We now turn to triangulations of spherical surfaces into convex spherical triangles. We will not require the triangulation to induce the structure of a simplicial complex on the surface. In particular, a triangle can be adjacent to a vertex up to three times, and to an edge up to two times.

The first result is a variation of the famous Delaunay triangulations of the plane [8] (see also [24, Section 14] for a modern exposition).

Proposition 6.15 (Delaunay triangulations) *Let S be a spherical surface with conical points $x_1, \dots, x_n$, some of which might have angle 2π. Suppose that the Voronoi function $\mathcal{V}_S$ is bounded by $\frac{1}{2}\pi$. Then there exists a triangulation of S into convex spherical triangles with the following "empty circle" property: for each triangle $x_i x_j x_k$ of the triangulation, there exists a vertex $v \in \Gamma(S)$ at equal distance r from x_i, x_j and x_k such that $d(x_l, v) \geq r$ for all $l \in \{1, \dots, n\}$.*

The proof will follow the proof by Thurston of a similar result [28, Proposition 3.1] concerning triangulations of surfaces with flat metric and conical singularities. We will need the following elementary lemma:

Lemma 6.16 *Let $D, D' \subset \mathbb{S}^2$ be two disks of radius less than $\frac{1}{2}\pi$. Let $x_1, x_2 \in \partial D$ and $x_1', x_2' \in \partial D'$ be four distinct points. Suppose x_1 and x_2 don't lie in the interior of D', and x_1' and x_2' don't lie in the interior of D. Then the geodesic segments $x_1 x_2 \subset D$ and $x_1' x_2' \subset D'$ are disjoint in $\mathbb{S}^2$.*

Proof If D and D' are disjoint, there is nothing to prove. Suppose D and D' intersect, and let y_1 and y_2 be the two points of intersection of the boundary circles ∂D and $\partial D'$. Let γ be the unique great circle on $\mathbb{S}^2$ passing through y_1 and y_2. It is now easy to see that the complements $D \setminus D'$ and $D' \setminus D$ lie in different hemispheres of $\mathbb{S}^2$ with respect to γ. It follows that the segments $x_1 x_2$ and $x_1' x_2'$ also lie in different hemispheres, and so they can intersect only in their endpoints. However, the points x_i and x_i' are distinct, so $x_1 x_2$ and $x_1' x_2'$ are disjoint. □

Proof of Proposition 6.15 The proof closely follows the proof of [28, Proposition 3.1]. Let $\Gamma(S)$ be the Voronoi graph of S. Let us first explain how to associate to each edge $e \subset \Gamma(S)$ a *dual* geodesic segment $\check{e}$ with conical endpoints.

Let $p \in \Gamma(S)$ be a point in the interior of an edge $e \subset \Gamma(S)$, and set $r = \mathcal{V}_S(p)$. Then there exists a locally isometric immersion $\iota_p \colon \mathbb{D}_r \to S$, from a radius $r < \frac{1}{2}\pi$ spherical disk, that sends the center of $\mathbb{D}_r$ to p. Exactly two of the boundary points of $\mathbb{D}_r$, say y and z, are sent to two conical points x_i and x_j of S. Denote by $\check{e}$ the image $\iota_p(yz)$. It is easy to see that the segment $\check{e}$ is independent of the choice of $p \in e$.

Let us now deduce from Lemma 6.16 that, for any two edges $e, e' \subset \Gamma_S$, their dual edges $\check{e}$ and $\check{e}'$ do not intersect in their interior points. This is similar the proof of [28, Proposition 3.1]. Let D and D' be the disks immersed in S that correspond to e and e'. Assume, for contradiction, that $\check{e}$ and $\check{e}'$ intersect in their interior point p. Consider the (multivalued) developing map $\iota \colon S \to \mathbb{S}^2$. The images of D and D' under this map are embedded disks, and the images of $\check{e}$ and $\check{e}'$ are chords of these disks, intersecting in $\iota(p)$. This contradicts Lemma 6.16. Indeed, the endpoints of $\check{e}$ are conical points that belong to $\partial D \setminus D'$, and the endpoints of $\check{e}'$ are conical points that belong to $\partial D' \setminus D$. Hence, Lemma 6.16 is applicable to the 4–tuple $\iota(D, \check{e}, D', \check{e}')$.

Next, we associate to each vertex v of $\Gamma(S)$ a convex polygon embedded in S whose edges $\check{e}_1, \dots, \check{e}_k$ are dual to the half-edges of $\Gamma(S)$ adjacent to v. To do so, consider the immersion $\iota_v \colon \mathbb{D}_r \to S$ of a disk of radius $r = \mathcal{V}_S(v)$ that sends the center of $\mathbb{D}_r$ to v. There will be exactly k points, say $y_1, \dots, y_k$, on $\partial\mathbb{D}_r$ whose images in S are conical points. Let P_v be the convex hull of the points y_i in $\mathbb{D}_r$. Then the map ι_v is an embedding on the interior $\mathring{P}_v$ of the polygon P_v; it may identify some vertices and it may identify an edge to at most one other edge of P_v.

Our last observation is that the union of the $\iota_v(\mathring{P}_v)$ over all vertices v of $\Gamma(S)$ coincides with the complement in S of the union of edges $\check{e}$. Indeed, since the edges $\check{e}$ can only intersect at endpoints, each $\iota_v(\mathring{P}_v)$ is a connected component of the complement of edges $\check{e}$. At the same time, each edge $\check{e}$ is adjacent to one or two open polygons $\iota_v(\mathring{P}_v)$ corresponding to the vertices of the edge e dual to $\check{e}$. It follows that polygons $\iota_v(P_v)$ cover the whole S.

Finally, if some of convex polygons $\iota_v(P_v)$ are not triangles, we subdivide them by diagonals into a collection of triangles. This gives the desired triangulation of S, where for each triangle $x_i x_j x_k$, the point v is the corresponding vertex of the Voronoi graph. □

Remark 6.17 Let Δ be a triangle from a Delaunay triangulation with vertices x_i, x_j and x_k, and let v be the corresponding vertex of $\Gamma(S)$. Then the circumscribed radius of Δ is equal to $\mathcal{V}_S(v) = d(v, x_i)$.

6.4.1 Compact subsets of $\mathcal{MS}_{g,n}(\leq A)$ In this subsection we prove Proposition 6.22, which singles out a class of compact subsets of $\mathcal{MS}_{g,n}(\leq A)$ consisting of surfaces that admit triangulations into triangles of *bounded shapes*.

Definition 6.18 ((l, r)–bounded triangles and surfaces) Fix constants $l \in (0, \pi)$ and $r \in \left(0, \frac{1}{2}\pi\right)$. We say that a convex spherical triangle is *(l, r)–bounded* if all its sides have length at least l and its circumscribed circle has radius at most r. A spherical surface is *(l, r)-bounded* if it admits a triangulation into (l, r)–bounded spherical triangles.

We will denote by $\mathcal{MT}_{l,r}$ the subset of $\mathcal{MT}$ consisting of (l, r)–bounded triangles.

Remark 6.19 (compactness of $\mathcal{MT}_{l,r}$) The set $\mathcal{MT}_{l,r}$ is compact in the analytic topology of $\mathcal{MT}$. Indeed, let $\Delta_i \subset \mathbb{S}^2$ be a sequence of convex triangles from $\mathcal{MT}_{l,r}$

with vertices (x_1^i, x_2^i, x_3^i). Passing to a subsequence, we can assume that the sequences of vertices converge to x_1, x_2 and x_3. We have $|x_i x_j| \geq l$, and the circle on $\mathbb{S}^2$ containing x_1, x_2 and x_3 has radius at most r. Hence $x_1x_2x_3$ is a triangle from $\mathcal{MT}_{l,r}$.

Definition 6.20 (space of (l,r)–triangulated surfaces) Let μ be a combinatorial type of triangulations of a genus-g surface with n marked points such that the marked points are vertices of the triangulation. Denote by $Y_{l,r}^{\mu}(\leq A)$ the set of all spherical surfaces of area at most A with a chosen triangulation of type μ consisting of (l,r)–bounded triangles. The $\overline{\mathcal{L}}$–distance between two triangulated surfaces from $Y_{l,r}^{\mu}(\leq A)$ is the Lipschitz distance with respect to all the maps that send the triangulation to the triangulation and restrict to homotheties on the edges.

We recall that, given a compact surface S of genus g with n marked points $\boldsymbol{x}$, there always exists a triangulation of S whose set of vertices contains $\boldsymbol{x}$ as in Definition 6.20. Indeed, it is possible to pick a point $b \in S$ and $2g$ loops $\{\gamma_j\}$ based at b such that no γ_j passes through $\boldsymbol{x}$ and $S \setminus \bigcup_j \gamma_j$ is a topological disk. This shows that $(S, \boldsymbol{x})$ can be obtained from a $2g$–gon P with n marked points $\boldsymbol{x}'$ in its interior via pairwise identification of its edges. Thus, every triangulation of P whose vertices include $\boldsymbol{x}'$ descends to a triangulation of S whose vertices include $\boldsymbol{x}$. The existence of such a triangulation of P is obvious.

Lemma 6.21 *The set $Y_{l,r}^{\mu}(\leq A)$ is compact in the $\overline{\mathcal{L}}$–metric.*

Proof From Remark 6.19 and Proposition 6.14 it follows that the subset $\mathcal{MT}_{l,r} \subset \mathcal{MT}$ of (l,r)–bounded triangles is compact in the $\overline{\mathcal{L}}$ metric. At the same time, $Y_{l,r}^{\mu}(\leq A)$ can be identified with a closed subset of the set of $(\mathcal{MT}_{l,r})^{|\mu|}$, where $|\mu|$ is the number of triangles in μ. □

Proposition 6.22 ($\mathcal{L}$–compactness of (l,r)–bounded surfaces) *Fix $A > 0$, $l \in (0,\pi)$ and $r \in \left(0, \frac{1}{2}\pi\right)$. Then the subset $X_{l,r}(\leq A)$ of $\mathcal{MS}_{g,n}(\leq A)$ consisting of (l,r)–bounded surfaces is compact in the Lipschitz topology. The analogous statement holds for $\mathcal{MP}_n(\leq A)$.*

Proof Since the area of an (l,r)–bounded triangle is bounded from below, there exists only a finite number of combinatorial triangulations μ of surfaces from $\mathcal{MS}_{g,n}(\leq A)$. Note that, for each μ, the natural map $Y_{l,r}^{\mu}(\leq A) \to \mathcal{MS}_{g,n}(\leq A)$, that forgets the triangulation is continuous since it contracts the metric. Hence $X_{l,r}(\leq A)$ is a finite union of images of compact sets under continuous maps. □

6.5 Properness of the function $\mathrm{sys}(S)^{-1}$ on $\mathcal{MS}_{g,n}(\le A)$

In this section we deduce Theorem 6.3 and Corollary 6.4 from the following result.

Theorem 6.23 (bounded Delaunay triangulations) *For any* $s>0$:

(i) *Any spherical surface from* $\mathcal{MS}_{g,n}^{\ge s}$ *can be triangulated into* $\left(\frac{1}{2}s,\frac{1}{4}\pi\right)$*–bounded spherical triangles.*

(ii) *Any spherical polygon* P *with* $\mathrm{sys}(P)\ge s$ *can be triangulated into* $\left(f(s),\frac{1}{4}\pi\right)$*–bounded spherical triangles, where* f *is a positive and continuous function.*

Proof of Theorem 6.3 We start with the properness of sys^{-1}. Since $\mathrm{sys}\colon\mathcal{MS}_{g,n}\to\mathbb{R}_+$ is continuous by Lemma 6.9, the subset $\mathcal{MS}_{g,n}^{\ge s}(\le A)$ is closed inside $\mathcal{MS}_{g,n}(\le A)$. Furthermore, $\mathcal{MS}_{g,n}^{\ge s}(\le A)$ is contained in the subset $X_{s/2,\pi/4}(\le A)$ of $\mathcal{MS}_{g,n}(\le A)$ consisting of $\left(\frac{1}{2}s,\frac{1}{4}\pi\right)$–bounded surfaces by Theorem 6.23(i). Since $X_{s/2,\pi/4}(\le A)$ is compact by Proposition 6.22, it follows that $\mathcal{MS}_{g,n}^{\ge s}(\le A)$ is compact too, and so the restriction of sys^{-1} to $\mathcal{MS}_{g,n}(\le A)$ is proper.

For the completeness of $\mathcal{MS}_{g,n}$, it is enough to show that, for every $r>0$ and spherical surface S in $\mathcal{MS}_{g,n}$, the closed ball $\bar{B}(S,r)=\{S'\in\mathcal{MS}_{g,n}\mid d_{\mathcal{L}}(S,S')\le r\}$ is compact. By Lemma 6.9, $\bar{B}(S,r)$ is contained in $\mathcal{MS}_{g,n}^{\ge s}(\le A)$ with $s=e^{-r}\,\mathrm{sys}(S)$ and $A=e^{2r}\,\mathrm{Area}(S)$. Since $\mathcal{MS}_{g,n}^{\ge s}(\le A)$ was shown above to be compact and $\bar{B}(S,r)$ is closed, it follows that $\bar{B}(S,r)$ is compact. □

Proof of Corollary 6.4 The proof is identical to the proof of Theorem 6.3, where instead of using Theorem 6.23(i) one applies Theorem 6.23(ii). □

Proof of Theorem 6.23 (i) We will prove that, for any $S\in\mathcal{MS}_{g,n}^{\ge s}$, there exists a collection of regular points $x_{n+1},\dots,x_{n+m}\in S$, such that the surface $(S,x_1,\dots,x_{n+m})$ has the following three properties:

(a) For any $i\ne j$, $d(x_i,x_j)\ge\frac{1}{2}s$ for all $i\ne j\in\{1,\dots,n+m\}$.

(b) For each i the injectivity radius of x_i on S is at least $\frac{1}{4}s$.

(c) For any $x\in S$ there is a point x_i such that $d(x,x_i)\le\frac{1}{4}\pi$.

Before proving this claim, let us explain why the statement of the theorem follows from it. Indeed, suppose that we have such a collection of points. Then let us consider the Delaunay triangulation of S with respect to points $x_1,\dots,x_{n+m}$ that exists thanks to Proposition 6.15. We claim that all the triangles of the triangulation are $\left(\frac{1}{2}s,\frac{1}{4}\pi\right)$–bounded. Indeed, by condition (c) and Remark 6.17, each such triangle is isometric to a triangle that can be inscribed in a circle of radius at most $\frac{1}{4}\pi$. At the same time, by conditions (a) and (b), all sides of the triangle have length at least $\frac{1}{2}s$.

Let us now show how to find such a collection of points $x_{n+1},\dots,x_{n+m}\in S$. We will add points $x_{n+1},\dots,x_{n+m}$ by induction. Note first that $x_1,\dots,x_n$ satisfy conditions (a) and (b). Suppose that there is a point $x\in S$ at distance more than $\frac{1}{4}\pi$ from $x_1,\dots,x_n$. Let us denote this x by x_{n+1}, and let us show that $(S,x_1,\dots,x_{n+1})$ satisfies conditions (a) and (b) for $m=1$. Note that by [23, Lemma 3.10] we have $\mathrm{sys}(S)\le\frac{1}{2}\pi$, which means $\frac{1}{4}\pi\ge\frac{1}{2}s$, and so when we add x_{n+1} we don't violate (a). It remains to show that the injectivity radius of x_{n+1} of S is at least $\frac{1}{4}s$. We apply (6) from Proposition 6.11 to get

$$\mathrm{inj}(x_{n+1})\ge\min\bigl(s,\tfrac{1}{4}\pi,\min_i\vartheta_i\tfrac{1}{4}\pi\bigr).$$

But, by [23, Lemma 3.13], we know that $\mathrm{sys}(S)\le\min_i\vartheta_i\pi$. So we get $\mathrm{inj}(x_{n+1})\ge\frac{1}{4}s$. Hence, condition (b) is satisfied for $x_1,\dots,x_{n+1}$. In this way we can go on adding points x_{n+i} until condition (c) is satisfied. Indeed, the process must terminate since the $\frac{1}{8}s$–neighborhoods of points x_{n+i} are disjoint disks on S and the area of S is finite.

(ii) To prove the second part of the theorem we work with the double $S(P)$ of P. We construct a collection of regular points $x_{n+1},\dots,x_{n+m}\in S(P)$ such that the surface $(S(P),x_1,\dots,x_{n+m})$ has the following four properties:

- (o) The set of points x_i is invariant under the isometric involution τ of $S(P)$.
- (a) For any $i\neq j$, $d(x_i,x_j)\ge\frac{1}{4}s$ for all $i\neq j\in\{1,\dots,n+m\}$.
- (b) For each i the injectivity radius of x_i on S is at least $\frac{1}{8}s$.
- (c) For any $x\in S$ there is a point x_i such that $d(x,x_i)\le\frac{1}{4}\pi$.

Let us explain how to make the first step. Consider P and ∂P as subsets of $S(P)$. Suppose there is a point $y\in S(P)$ at distance greater than $\frac{1}{4}\pi$ from $x_1,\dots,x_n$. If its distance from ∂P is more than $\frac{1}{8}\pi$, we set $x_{n+1}=y$ and $x_{n+2}=\tau(y)$. In this case conditions (o)–(b) are still satisfied for points $x_1,\dots,x_{n+2}$, since $d(x_{n+1},x_{n+2})\ge\frac{1}{4}\pi$. Suppose now that $d(y,\partial P)<\frac{1}{8}\pi$. Let y' be a point on ∂P closest to y and set $x_{n+1}=y'$. Clearly the distance from x_{n+1} to $x_1,\dots,x_n$ is at least $\frac{1}{8}\pi$. For this reason, as in (i), conditions (b) and (c) are still satisfied. This finishes the first step.

Now, we repeat the above step until we get a collection of points $x_1,\dots,x_{n+m}$ in $S(P)$ that satisfy conditions (o)–(c). As in the proof of Proposition 6.15, we get a canonical decomposition of $S(P)$ into convex spherical polygons, invariant under the action of τ, and such that each polygon has side lengths at least $\frac{1}{4}s$ and can be inscribed in a circle of radius at most $\frac{1}{4}\pi$. Those polygons whose interior doesn't intersect ∂P should be further cut into triangles by diagonals. Suppose that the interior of a polygon Q intersects ∂P. Then $\tau(Q) = Q$, and using a τ–invariant subset of diagonals of Q, one can cut it into a union of triangles exchanged by τ and either a triangle or a trapezoid Q' satisfying $\tau(Q') = Q'$. If Q' is a triangle, we take $Q' \cap P$ as one of the triangles of the triangulation of P. If Q' is a trapezoid, we subdivide further $Q' \cap P$ into two triangles along a diagonal. It is not hard to see that the resulting triangles are $\bigl(f(s), \frac{1}{4}\pi\bigr)$–bounded for some positive function $f(s)$. That concludes the decomposition of P into triangles. □

6.6 Systole of balanced triangles

In this section we calculate the systole of a balanced triangle and show that, for a balanced triangle Δ, we have $\operatorname{sys}(\Delta) = \operatorname{sys}(T(\Delta))$.

Lemma 6.24 *Let Δ be a balanced spherical triangle with vertices x_1, x_2 and x_3. Then*

$$2\operatorname{sys}(\Delta) = \min_{i,j}\bigl(\min(|x_i x_j|, 2\pi - |x_i x_j|)\bigr). \tag{7}$$

Moreover:

(i) *For any vertex x_i of Δ, the distance to the opposite side $x_i x_j$ is larger than $\operatorname{sys}(\Delta)$.*

(ii) *Let $p \in \partial\Delta$ be a point that is not a vertex of Δ. Suppose that η is a geodesic segment in Δ that joins p with x_i and doesn't belong to $\partial\Delta$. Then $l(\eta) > \operatorname{sys}(\Delta)$.*

(iii) *There exists a geodesic segment $\gamma_\Delta \subset \Delta$ of length $2\operatorname{sys}(\Delta)$ that joins two vertices of Δ.*

Proof We will first prove statements (i)–(iii) and then will deduce (7).

(i) Let us show that, for any p in x_2x_3, we have $d(p, x_1) > \operatorname{sys}(\Delta)$. From Remark 2.12 it follows that p lies either in the Voronoi domain of x_2 or of x_3. Assume the former. Then, by definition of Voronoi domains, $d(p, x_1) \geq d(p, x_2)$.

Suppose first that the strict inequality $d(p, x_1) > d(p, x_2)$ holds. Applying the triangle inequality to the points x_1, x_2 and p and using $d(x_1, x_2) \geq 2\operatorname{sys}(\Delta)$, we get

$$d(p, x_1) \geq d(x_1, x_2) - d(p, x_2) > d(x_1, x_2) - d(p, x_1) \geq 2\operatorname{sys}(\Delta) - d(p, x_1).$$

It follows that $d(p, x_1) > \operatorname{sys}(\Delta)$.

Suppose now that $d(p, x_1) = d(p, x_2)$. Then, by Remark 2.12, Δ is semibalanced, p is the midpoint of the segment x_1x_2, and there is a geodesic segment $x_1 p$ that joins x_1 with p. It is clear then that $2|x_1 p| = |x_1 p| + |x_2 p| > 2d(x_1, x_2) \geq 2\operatorname{sys}(\Delta)$.

(ii) Consider two cases. If p lies on the side of Δ opposite to x_i then by (i) we have $\ell(\eta) \geq d(x_i, p) > \operatorname{sys}(\Delta)$. Suppose now p lies on a side adjacent to x_j. In this case η cuts out of Δ a digon with angles less than π (since p is an interior point of an edge). So $e(\eta) = \pi$ and the statement follows from Corollary 2.15.

(iii) Using (i) and Definition 6.2, we see that $2\operatorname{sys}(S) = \min_{i,j} d(x_i, x_j)$. Hence there is a geodesic segment γ_Δ of length $2\operatorname{sys}(S)$ that joins two vertices of Δ.

To prove (7), take the geodesic γ_Δ given by (iii). It cuts out of Δ a digon, one of whose sides is a side $x_i x_j$ of the triangle Δ. If follows that either $2\operatorname{sys}(\Delta) = |x_i x_j|$ or $2\pi - |x_i x_j|$. This shows that $2\operatorname{sys}(\Delta)$ is no smaller than the right-hand expression in (7). The opposite inequality follows immediately from Corollary 2.15. □

Lemma 6.25 *For any balanced triangle Δ and the corresponding spherical torus $(T(\Delta), x)$, we have $\operatorname{sys}(\Delta) = \operatorname{sys}(T(\Delta))$. Conversely, for any spherical torus T and the corresponding balanced spherical triangle $\Delta(T)$, we have $\operatorname{sys}(T) = \operatorname{sys}(\Delta(T))$.*

Proof The first and the second statements are equivalent, so we prove just the first. By Lemma 6.24(iii), there is a geodesic segment γ_Δ in Δ of length $2\operatorname{sys}(\Delta)$ that joins two vertices of Δ. Such a γ_Δ is embedded as a geodesic loop in $T(\Delta)$, which clearly implies $\operatorname{sys}(\Delta) \geq \operatorname{sys}(T(\Delta))$. To get $\operatorname{sys}(\Delta) \leq \operatorname{sys}(T(\Delta))$, let $\gamma_{T(\Delta)}$ be the systole geodesic loop in $T(\Delta)$, and let Δ_1 and Δ_2 be two balanced triangles isometric to Δ from which $T(\Delta)$ is glued. It will be enough to prove that $\gamma_{T(\Delta)}$ lies entirely in Δ_1 or Δ_2. Assume, for contradiction, that this is not so. Then $\gamma_{T(\Delta)}$ contains two subsegments η and η' whose interiors lie in the interior of Δ_1 or Δ_2 and which satisfy the conditions of Lemma 6.24(ii). Applying this lemma, we get $l(\gamma_{T(\Delta)}) \geq l(\eta) + l(\eta') > 2\operatorname{sys}(\Delta)$, which contradicts the established inequality $\operatorname{sys}(\Delta) \geq \operatorname{sys}(T(\Delta))$. □

Corollary 6.26 *The function $\operatorname{sys}(\Delta)^{-1} = 2\min_{i,j}\bigl(\min(|x_i, x_j|, 2\pi - |x_i, x_j|)\bigr)^{-1}$ is proper on $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$ in the analytic topology.*

Proof The function $\operatorname{sys}(\Delta)^{-1}$ is proper on $\mathcal{MT}(\vartheta)$ in the $\mathcal{L}$–topology by Corollary 6.4. At the same time, by Proposition 6.14, the $\mathcal{L}$–topology and the analytic topology coincide on $\mathcal{MT}_{\mathrm{bal}}(\vartheta)$. □

6.7 Proof of Theorem 6.5

Here we finally prove Theorem 6.5, concerning 2–marked tori. We note first that Theorem 6.3 holds for 2–marked tori as well; namely, the function sys^{-1} is proper in the Lipschitz topology on the space $\mathcal{MS}_{1,1}^{(2)}({\leq}A)$ of such tori of area at most A.

We will use the following standard lemma, whose proof we omit:

Lemma 6.27 *Let X and Y be locally compact Hausdorff topological spaces and let $\varphi\colon X \to Y$ be a continuous bijective map.*

(i) *If φ is proper then it is a homeomorphism.*

(ii) *Suppose there exist proper functions $s_X\colon X \to \mathbb{R}$ and $s_Y\colon Y \to \mathbb{R}$ such that $s_X = s_Y \circ \varphi$. Then φ is a homeomorphism.*

Proof of Theorem 6.5 (i) By Proposition 3.22(i) $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ is a surface, so we need to show that $T^{(2)}$ is a homeomorphism. To show we can apply Lemma 6.27, we note:

(a) Every bi-Lipschitz map $\Delta \to \Delta'$ that restricts to a homothety on the edges gives rise to a σ–equivariant bi-Lipschitz map $T^{(2)}(\Delta) \to T^{(2)}(\Delta')$ with the same Lipschitz constant. Hence, the map $T^{(2)}$ is contracting with respect to the $\overline{\mathcal{L}}$–metric on $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$, namely $d_{\mathcal{L}}(T^{(2)}(\Delta), T^{(2)}(\Delta')) \le d_{\overline{\mathcal{L}}}(\Delta, \Delta')$. It follows that $T^{(2)}$ is continuous. Moreover, $T^{(2)}$ is bijective by Lemma 4.6.

(b) Since $\mathcal{MT}^{\pm}_{\mathrm{bal}}(\vartheta)$ is a surface it is locally compact, and the function sys^{-1} is proper on it by Corollary 6.26.

(c) The space $(\mathcal{MS}^{(2)}_{1,1}(\vartheta), \mathcal{L})$ is locally compact, and the function sys^{-1} is proper on it by Theorem 6.3.

(d) The map $T^{(2)}$ preserves the function sys^{-1} by Lemma 6.25.

To sum up, the map $T^{(2)}$ satisfies all the properties of Lemma 6.27, which proves the claim.

(ii) The proof of this claim is the same and so we omit it. □

Remark 6.28 (orbifold structures on $\mathcal{MS}_{1,1}(\vartheta)$ and $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$) Let $\mathcal{MS}^{(4)}_{1,1}(\vartheta)$ be the set of spherical tori T endowed with a 4–marking, namely an isomorphism $H_1(T;\mathbb{Z}_4) \cong (\mathbb{Z}_4)^2$. We endow $\mathcal{MS}^{(4)}_{1,1}(\vartheta)$ with the Lipschitz distance measured among maps between tori that respect the 4–marking. Since 4–marked tori have no nontrivial conformal automorphisms, $\mathcal{MS}^{(4)}_{1,1}(\vartheta)$ is a moduli space for such 4–marked tori. It is easy to see that the forgetful map $\mathcal{MS}^{(4)}_{1,1}(\vartheta) \to \mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is a local isometry, and in fact an unramified Galois cover with group $K/\{\pm 1\}$, where $K = \ker(\mathrm{SL}(2,\mathbb{Z}_4) \to \mathrm{SL}(2,\mathbb{Z}_2))$.

Assume first that ϑ is not odd. The space $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ is an orientable surfaces of finite type by Theorem 6.5 and Proposition 3.22(i), and so the same holds for $\mathcal{MS}^{(4)}_{1,1}(\vartheta)$. We endow $\mathcal{MS}^{(2)}_{1,1}(\vartheta)$ with the orbifold structure given by $\mathcal{MS}^{(4)}_{1,1}(\vartheta)/K$, and $\mathcal{MS}_{1,1}(\vartheta)$

with the orbifold structure $\mathcal{MS}_{1,1}^{(4)}(\vartheta)/\mathrm{SL}(2,\mathbb{Z}_4)$. As a consequence, $\mathcal{MS}_{1,1}^{(2)}(\vartheta) \to \mathcal{MS}_{1,1}(\vartheta)$ is an unramified Galois cover with group $\mathrm{SL}(2,\mathbb{Z}_2) \cong \mathrm{S}_3$.

Assume that $\vartheta = 2m+1$ is odd. Again, $\mathcal{MS}_{1,1}^{(2)}(2m+1)^\sigma$ is a disjoint union of finitely many 2–dimensional disks by Theorem 6.5 and Proposition 3.25, and so the same holds for the moduli space $\mathcal{MS}_{1,1}^{(4)}(2m+1)^\sigma$. The same argument as in Construction 4.16 shows that $\mathcal{MS}_{1,1}^{(4)}(2m+1)$ fibers over $\mathcal{MS}_{1,1}^{(4)}(2m+1)^\sigma$ with fiber $\mathbb{R}$, and so is a 3–dimensional manifold. We then put on $\mathcal{MS}_{1,1}^{(2)}(2m+1)$ and $\mathcal{MS}_{1,1}(2m+1)$ the orbifold structures induced by $\mathcal{MS}_{1,1}^{(2)}(2m+1) = \mathcal{MS}_{1,1}^{(4)}(2m+1)/K$ and $\mathcal{MS}_{1,1}(2m+1) = \mathcal{MS}_{1,1}^{(4)}(2m+1)/\mathrm{SL}(2,\mathbb{Z}_4)$. We put a similar structure on the moduli spaces of σ–invariant metrics.

In all cases, the orbifold order of a point in such moduli spaces is the number of automorphisms of the corresponding (possibly marked) spherical torus.

Appendix Monodromy and coaxiality

In this section we prove that a spherical torus with one conical point of angle $2\pi\vartheta$ is coaxial if and only if ϑ is an odd integer. This was already shown in [2].

In order to prove this, we recall that monodromy representation of spherical surfaces can be lifted to SU(2):

Proposition A.1 (lift of the monodromy to SU(2)) *Let $(S, \boldsymbol{x})$ be a spherical surface with conical points of angles $2\pi\boldsymbol{\vartheta}$ and let $p \in \dot{S}$ be a basepoint. Let $(\tilde{\dot{S}}, \tilde{p})$ be a universal cover of $(\dot{S}, p)$, endowed with the pullback spherical metric, and let $\iota\colon \tilde{\dot{S}} \to \mathbb{S}^2 \cong \mathbb{CP}^1$ be an associated developing map with monodromy representation $\rho\colon \pi_1(\dot{S}, p) \to \mathrm{SO}_3(\mathbb{R})$. Then there exists a lift $\hat{\rho}\colon \pi_1(\dot{S}, p) \to \mathrm{SU}(2)$ of ρ such that*

- (a) *the developing map ι extends to the completion $\hat{S}$ of $\tilde{\dot{S}}$ and each point of $\hat{S} \setminus \tilde{\dot{S}}$ corresponds to a loop based at p that simply winds about some x_j,*
- (b) *if $\gamma_j \in \pi_1(\dot{S}, p)$ is a loop that simply winds about x_j corresponding to a point $\hat{x}_j$ in $\hat{S} \setminus \tilde{\dot{S}}$, then $\hat{\rho}(\gamma_j) \in \mathrm{SU}(2)$ acts on the complex line $\iota(\hat{x}_j) \subset \mathbb{C}^2$ as multiplication by $e^{i\pi(\vartheta_j - 1)}$.*

Moreover, two such lifts multiplicatively differ by a homomorphism $\pi_1(S, p) \to \{\pm I\}$.

Proof In [22, Proposition 2.12] the statement was proven for a surface S of genus 0. For a surface of arbitrary genus, the proof of existence for a lift is analogous, with minor modifications. In particular, D will be the complement $S \setminus \{q\}$ of an unmarked point q in S, and the vector field V is chosen to be nowhere vanishing on D and have vanishing order $2-2g$ at q, so that the unit normalized vector field $\widehat{V}$ on D has even winding number about q.

Finally, two lifts certainly differ by multiplication by a homomorphism $\pi(\dot{S}, p) \to \{\pm I\}$. Since the eigenvalues of the monodromy about the punctures are fixed by (b), such a homomorphism factors through $\pi(S, p) \to \{\pm I\}$. □

We use the above SU(2)–lifting property to characterize 1–punctured tori (S, x) with integral angles. In order to do that, choose standard generators $\{\alpha, \beta, \gamma\}$ of $\pi_1(\dot{S})$ such that $\gamma = [\alpha, \beta]$. Given a spherical metric on (S, x), its monodromy representation ρ can be lifted to an SU_2–valued representation $\hat{\rho}$ by Proposition A.1. Write $A = \hat{\rho}(\alpha)$, $B = \hat{\rho}(\beta)$ and $C = \hat{\rho}(\gamma)$, and note that C has eigenvalues $e^{\pm i\pi(\vartheta-1)}$.

Corollary A.2 (monodromy of tori with odd ϑ) *Let (S, x) be a spherical torus with one conical point of angle $2\pi\vartheta$. Then its monodromy is nontrivial. Moreover (S, x) has coaxial monodromy if and only if ϑ is an odd integer.*

Proof As for the first claim, if the monodromy of (S, x) were trivial, then the developing map of (S, x) would descend to a cover $S \to \mathbb{S}^2$ ramified at x only. This is clearly absurd.

As for the second claim, the monodromy ρ is coaxial if and only if $\hat{\rho}$ is. On the other hand, since elements in SU(2) are diagonalizable, $\hat{\rho}$ is coaxial if and only if it is abelian. Finally, $\hat{\rho}$ is abelian if and only if $\hat{\rho}(\gamma) = I$, which implies that ϑ is an odd integer. □

Corollary A.3 (monodromy of tori with even ϑ) *Let (S, x) be a spherical torus with one conical point of angle $2\pi\vartheta$. Then the monodromy of (S, x) is isomorphic to the Klein group $K_4 \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$ if and only if ϑ is an even integer. In this case, the three nontrivial elements in the monodromy group are rotations of angle π along mutually orthogonal axes of $\mathbb{S}^2$.*

Proof The monodromy is isomorphic to K_4 if and only if

$$\rho(\alpha)^2 = \rho(\beta)^2 = [\rho(\alpha), \rho(\beta)] = I.$$

If ϑ is even an even integer, then $C = -I$. Up to conjugacy, we can assume that A is diagonal. The relation $AB = -BA$ gives that A has eigenvalues $\pm i$ and B has zero entries on the diagonal. It follows that $A^2 = B^2 = -I$, and so $\rho(\alpha)^2 = \rho(\beta)^2 = [\rho(\alpha), \rho(\beta)] = I$. It can be observed that A, B and AB act on $\mathbb{S}^2$ as rotations of angle π along mutually orthogonal axes.

Conversely, suppose the monodromy is isomorphic to the Klein group. Then $C = \pm I$ and so ϑ must be integral, but ϑ cannot be odd by Corollary A.2. Hence, ϑ is even. □

References

[1] **D Burago**, **Y Burago**, **S Ivanov**, *A course in metric geometry*, Graduate Studies in Math. 33, Amer. Math. Soc., Providence, RI (2001) MR Zbl

[2] **C-L Chai**, **C-S Lin**, **C-L Wang**, *Mean field equations, hyperelliptic curves and modular forms, I*, Camb. J. Math. 3 (2015) 127–274 MR Zbl

[3] **C-C Chen**, **C-S Lin**, *Mean field equation of Liouville type with singular data: topological degree*, Comm. Pure Appl. Math. 68 (2015) 887–947 MR Zbl

[4] **Q Chen**, **W Wang**, **Y Wu**, **B Xu**, *Conformal metrics with constant curvature one and finitely many conical singularities on compact Riemann surfaces*, Pacific J. Math. 273 (2015) 75–100 MR Zbl

[5] **Z Chen**, **T-J Kuo**, **C-S Lin**, *Existence and non-existence of solutions of the mean field equations on flat tori*, Proc. Amer. Math. Soc. 145 (2017) 3989–3996 MR Zbl

[6] **Z Chen**, **C-S Lin**, *Critical points of the classical Eisenstein series of weight two*, J. Differential Geom. 113 (2019) 189–226 MR Zbl

[7] **D Cooper**, **C D Hodgson**, **S P Kerckhoff**, *Three-dimensional orbifolds and cone-manifolds*, MSJ Memoirs 5, Math. Soc. Japan, Tokyo (2000) MR Zbl

[8] **B Delaunay**, *Sur la sphère vide*, Bull. Acad. Sci. URSS (1934) 793–800 Zbl

[9] **A Eremenko**, *Metrics of positive curvature with conic singularities on the sphere*, Proc. Amer. Math. Soc. 132 (2004) 3349–3355 MR Zbl

[10] **A Eremenko**, *Metrics of constant positive curvature with four conic singularities on the sphere*, Proc. Amer. Math. Soc. 148 (2020) 3957–3965 MR Zbl

[11] **A Eremenko**, **A Gabrielov**, *On metrics of curvature* 1 *with four conic singularities on tori and on the sphere*, Illinois J. Math. 59 (2015) 925–947 MR Zbl

[12] **A Eremenko**, **A Gabrielov**, *The space of Schwarz–Klein spherical triangles*, J. Math. Phys. Anal. Geom. 16 (2020) 263–282 MR Zbl

[13] **A Eremenko**, **A Gabrielov**, **G Mondello**, **D Panov**, *Moduli spaces for Lamé functions and abelian differentials of the second kind*, Commun. Contemp. Math. 24 (2022) art. id. 2150028 MR Zbl

[14] **A Eremenko**, **V Tarasov**, *Fuchsian equations with three non-apparent singularities*, Symmetry Integrability Geom. Methods Appl. 14 (2018) art. id. 058 MR Zbl

[15] **M Gromov**, *Metric structures for Riemannian and non-Riemannian spaces*, Birkhäuser, Boston, MA (2007) MR Zbl

[16] **J Harer**, **D Zagier**, *The Euler characteristic of the moduli space of curves*, Invent. Math. 85 (1986) 457–485 MR Zbl

[17] **F Klein**, *Vorlesungen über die hypergeometrische Funktion*, Grundl. Math. Wissen. 39, Springer, Berlin (1933) Zbl

[18] **L Li**, **J Song**, **B Xu**, *Irreducible cone spherical metrics and stable extensions of two line bundles*, Adv. Math. 388 (2021) art. id. 107854 MR Zbl

[19] **C-S Lin**, **C-L Wang**, *Elliptic functions, Green functions and the mean field equations on tori*, Ann. of Math. 172 (2010) 911–954 MR Zbl

[20] **C-S Lin**, **C-L Wang**, *Mean field equations, hyperelliptic curves and modular forms, II*, J. Éc. polytech. Math. 4 (2017) 557–593 MR Zbl

[21] **R S Maier**, *Lamé polynomials, hyperelliptic reductions and Lamé band structure*, Philos. Trans. Roy. Soc. Lond. Ser. A 366 (2008) 1115–1153 MR Zbl

[22] **G Mondello**, **D Panov**, *Spherical metrics with conical singularities on a 2–sphere: angle constraints*, Int. Math. Res. Not. 2016 (2016) 4937–4995 MR Zbl

[23] **G Mondello**, **D Panov**, *Spherical surfaces with conical points: systole inequality and moduli spaces with many connected components*, Geom. Funct. Anal. 29 (2019) 1110–1193 MR Zbl

[24] **I Pak**, *Lectures on discrete and polyhedral geometry* (2010) Available at `https://www.math.ucla.edu/~pak/geompol8.pdf`

[25] **H P de Saint-Gervais**, *Uniformization of Riemann surfaces*, Eur. Math. Soc., Zürich (2016) MR Zbl

[26] **J Song**, **Y Cheng**, **B Li**, **B Xu**, *Drawing cone spherical metrics via Strebel differentials*, Int. Math. Res. Not. 2020 (2020) 3341–3363 MR Zbl

[27] **G Tarantello**, *Analytical, geometrical and topological aspects of a class of mean field equations on surfaces*, Discrete Contin. Dyn. Syst. 28 (2010) 931–973 MR Zbl

[28] **W P Thurston**, *Shapes of polyhedra and triangulations of the sphere*, from "The Epstein birthday schrift" (I Rivin, C Rourke, C Series, editors), Geom. Topol. Monogr. 1, Geom. Topol. Publ., Coventry (1998) 511–549 MR Zbl

[29] **E T Whittaker**, **G N Watson**, *A course of modern analysis*, Cambridge Univ. Press (1996) MR

Department of Mathematics, Purdue University
West Lafayette, IN, United States

Dipartimento di Matematica Guido Castelnuovo, Sapienza Università of Roma
Roma, Italy

Department of Mathematics, King's College London
London, United Kingdom

eremenko@purdue.edu, mondello@mat.uniroma1.it, dmitri.panov@kcl.ac.uk

Proposed: Benson Farb
Seconded: David M Fisher, Leonid Polterovich

Received: 4 May 2021
Revised: 11 February 2022

Geometry & Topology **27**:9 (2023) 3699–3713
DOI: 10.2140/gt.2023.27.3699
Published: 5 December 2023

The derivative map for diffeomorphism of disks: an example

DIARMUID CROWLEY
THOMAS SCHICK
WOLFGANG STEIMLE

We prove that the derivative map $d\colon \mathrm{Diff}_\partial(D^k) \to \Omega^k \mathrm{SO}_k$, defined by taking the derivative of a diffeomorphism, can induce a nontrivial map on homotopy groups. Specifically, for $k = 11$ we prove that the following homomorphism is nonzero:

$$d_*\colon \pi_5 \mathrm{Diff}_\partial(D^{11}) \to \pi_5 \Omega^{11} \mathrm{SO}_{11} \cong \pi_{16} \mathrm{SO}_{11}.$$

As a consequence we give a counterexample to a conjecture of Burghelea and Lashof by giving an example of a nontrivial vector bundle E over a sphere which is trivial as a topological $\mathbb{R}^k$–bundle (the rank of E is $k = 11$ and the base sphere is S^{17}).

The proof relies on a recent result of Burklund and Senger which determines the homotopy 17–spheres bounding 8–connected manifolds, the plumbing approach to the Gromoll filtration due to Antonelli, Burghelea and Kahn, and an explicit construction of low-codimension embeddings of certain homotopy spheres.

57R50, 57S05; 57R60

1 Introduction

The derivative map

$$d\colon \mathrm{Diff}_\partial(D^k) \to \mathrm{Map}((D^k, \partial D^k), (\mathrm{SO}_k, \mathrm{Id})) \simeq \Omega^k \mathrm{SO}_k, \quad f \mapsto (x \mapsto D_x f),$$

is a basic invariant of the diffeomorphism group of the k–disk; in fact the first-order approximation in the embedding calculus approach to the diffeomorphism group. While $d_\mathbb{Q}\colon \mathrm{Diff}_\partial(D^k)_\mathbb{Q} \to (\Omega^k \mathrm{SO}_k)_\mathbb{Q}$, the rationalisation of d, is nullhomotopic, as we explain in Section 3, much less is known about the derivative map d integrally. For example, to the best of our knowledge, it was not yet known whether the map induced by d on homotopy groups,

$$d_*\colon \pi_i \mathrm{Diff}_\partial(D^k) \to \pi_{i+k} \mathrm{SO}_k,$$

was ever nontrivial. Burghelea and Lashof showed that d_* vanishes for $i = 0, 1$. At odd primes p, they also showed that $d_* = 0$ provided $i < k-3$ and they made a conjecture equivalent to the claim that this holds for $p = 2$ as well [6, Conjecture, page 40]. Burghelea and Lashof also report A'Campo informing them about a proof that $d_* = 0$ for $i = 2$ (however, a written proof has not appeared).

Using smoothing theory, or an explicit geometric construction we introduce here, the map d_* admits an interpretation as describing the normal bundle of certain homotopy spheres embedded in euclidean space. Combining this interpretation with recent results of Burklund and Senger and the refined plumbing construction of Antonelli, Burghelea and Kahn, we obtain a counterexample to the conjecture of Burghelea and Lashof.

In more detail, in [3; 4] Antonelli, Burghelea and Kahn constructed families of diffeomorphisms of the disk using a pairing

$$\sigma\colon \pi_p\mathrm{SO}_{q-a} \otimes \pi_q\mathrm{SO}_{p-b} \to \pi_{a+b+1}\mathrm{Diff}_\partial(D^{p+q-a-b-1})$$

for $0 \le a \le q$ and $0 \le b \le p$, refining Milnor's plumbing pairing; see below. Now $\pi_8\mathrm{SO}_6 \cong \mathbb{Z}/24$ (see [12, page 162]) and we have:

Theorem 1.1 *Let $\xi \in \pi_8\mathrm{SO}_6 \cong \mathbb{Z}/24$ be a generator. The image of $\sigma(\xi,\xi)$ under the derivative map*

$$d_*\colon \pi_5\mathrm{Diff}_\partial(D^{11}) \to \pi_{16}\mathrm{SO}_{11}$$

is nonzero.

Using Morlet's smoothing theory isomorphism, the derivative map d_* on π_k is identified with the boundary map

$$\partial\colon \pi_{n+k+1}\mathrm{PL}_k/O_k \to \pi_{n+k}\mathrm{SO}_k$$

of the fibration sequence $\mathrm{SO}_k \to \mathrm{SPL}_k \to \mathrm{PL}_k/O_k$ (and this allows for our interpretation of [6, Conjecture, page 40] in terms of the derivative map). We conclude that the map $\mathrm{SO}_{11} \to \mathrm{SPL}_{11}$ is not injective on π_{16}. More specifically, if $\tau_{11}\colon S^{11} \to B\mathrm{SO}_{11}$ represents the tangent bundle of the 11–sphere and $f\colon S^{17} \to S^{11}$ represents the unique nontrivial homotopy class (see [21, Proposition 5.11]), we have:

Corollary 1.2 *The pullback $f^*\tau_{11}$ is a nontrivial vector bundle which becomes trivial as an $\mathbb{R}^{11}$–bundle, even when considered as a bundle with structure group SPL_{11}.*

To the best of our knowledge, this is the first example of a nontrivial vector bundle over a sphere which is known to be trivial as a topological $\mathbb{R}^n$–bundle. Milnor famously gave examples of nontrivial vector bundles over Moore spaces, for example the Moore space $M(\mathbb{Z}/7,7) = S^7 \cup_7 D^8$, which are trivial as $\mathbb{R}^n$–bundles [17, Lemma 9.1]. These examples are stable bundles over 4–connected spaces and so the vector bundles are trivial as piecewise linear bundles too.

Acknowledgements

We would like to thank Alexander Kupers for helpful comments on an earlier draft, as well as two referees for helpful comments.

2 Proofs

In this section we give the proofs of Theorem 1.1 and Corollary 1.2. We first recall Gromoll's map $A = C \circ \lambda$ from [9],

$$A\colon \pi_{n-k}\mathrm{Diff}_\partial(D^k) \xrightarrow{\lambda} \pi_0\mathrm{Diff}_\partial(D^n) \xrightarrow{C} \Theta_{n+1},$$

where Θ_{n+1} is the group of homotopy $(n+1)$–spheres. The first map λ includes fibrewise diffeomorphisms of $D^{n-k} \times D^k$ into all diffeomorphisms, and the second C uses a diffeomorphism of $D^n \subset S^n$ as a datum to clutch two $(n+1)$–disks and make a homotopy sphere.[1]

Lemma 2.1 *For any* $[\psi] \in \pi_{n-k}\mathrm{Diff}_\partial(D^k)$*, the homotopy sphere* $A([\psi]) \in \Theta_{n+1}$ *admits an embedding into* $\mathbb{R}^{n+k+1}$ *whose normal bundle is classified* (*up to possible sign*) *by* $d_*([\psi]) \in \pi_n\mathrm{SO}_k \cong \pi_{n+1}B\mathrm{SO}_k$.

We will offer two proofs of this result; one by an explicit geometric construction and a more abstract one by the classification of smoothings through Rourke and Sanderson's theory of block bundles.

Next we recall that in [16, Section 1], Milnor constructed exotic spheres by plumbing linear disk bundles and taking the boundary sphere; this construction gives rise to a pairing

$$\sigma_M\colon \pi_p\mathrm{SO}_q \otimes \pi_q\mathrm{SO}_p \to \Theta_{p+q+1}.$$

[1] The map C is denoted by Σ in [8].

By [3, Proposition 3.1], the pairing of Antonelli, Burghelea and Kahn refines this pairing in the sense that we have a commutative diagram

(1)
$$\begin{tikzcd}
\pi_p\mathrm{SO}_{q-a} \otimes \pi_q\mathrm{SO}_{p-b} \arrow[r, "\sigma"] \arrow[d] & \pi_{a+b+1}\mathrm{Diff}_\partial(D^{p+q-a-b-1}) \arrow[d, "A"] \\
\pi_p\mathrm{SO}_q \otimes \pi_q\mathrm{SO}_p \arrow[r, "\sigma_M"] & \Theta_{p+q+1}
\end{tikzcd}$$

where the map on the left is the tensor product of the canonical stabilisations. We now consider the homotopy 17–sphere

$$\Sigma_{\xi,\xi} := A(\sigma(\xi,\xi)),$$

recalling that $\xi \in \pi_8\mathrm{SO}_6 \cong \mathbb{Z}/24$ denotes a generator. By the commutativity of (1) and the definition of σ_M, $\Sigma_{\xi,\xi}$ is the boundary of an 8–connected compact 18–manifold and so by a recent result of Burklund and Senger [7, Theorem 1.4] its image under the normal invariant map $\Theta_{17} \to \mathrm{coker}(J_{17})$ must be either 0 or $[\eta\eta_4]$. We will show:

Lemma 2.2 *The homotopy sphere $\Sigma_{\xi,\xi}$ represents $[\eta\eta_4] \in \mathrm{coker}(J_{17})$.*

We deduce from this that every embedding $\Sigma_{\xi,\xi} \hookrightarrow S^{28}$ has a nontrivial normal bundle. Indeed, recall that the map $\Theta_{17} \to \mathrm{coker}(J_{17})$ is obtained by embedding a homotopy 17–sphere into some euclidean space with trivial normal bundle, and performing the Pontryagin–Thom collapse as to obtain an element in π^s_{17} which is well-defined modulo the image of J. Now, assuming by contradiction that $\Sigma_{\xi,\xi}$ embeds into S^{28} with trivial normal bundle, then $[\eta\eta_4]$ would have a representative in $\pi_{28}S^{11}$. However, this contradicts the computations of Toda [21, Theorem 12.17 and Proposition 12.20] on the stabilisation map $\pi_{28}S^{11} \to \pi^s_{17}$, which we display below:

$$\begin{tikzcd}
\pi_{28}S^{11} \arrow[d] & = & \mathbb{Z}/2((\eta^2\rho)_{11}) \arrow[d] & \oplus & \mathbb{Z}/2((\mu_{17})_{11}) \arrow[d] & \oplus & \mathbb{Z}/2((\nu\kappa)_{11}) \arrow[d] & & \\
\pi^s_{17} & = & \mathbb{Z}/2(\eta^2\rho) & \oplus & \mathbb{Z}/2(\mu_{17}) & \oplus & \mathbb{Z}/2(\nu\kappa) & \oplus & \mathbb{Z}/2(\eta\eta_4)
\end{tikzcd}$$

Here the notation is such that an element $(\delta)_{11}$ stabilises to δ and the stable class $\eta^2\rho$ generates $\mathrm{im}(J_{17}\colon \pi_{17}(\mathrm{SO}) \to \pi^s_{17})$. Lemma 2.1 then implies that $d_*(\sigma(\xi,\xi)) \neq 0$, which concludes the proof of Theorem 1.1, modulo Lemmas 2.1 and 2.2.

To prove Corollary 1.2 we note that by Lemma 2.1 the normal bundle $\nu(\Sigma_{\xi,\xi} \subset S^{28})$ has clutching function $\pm d_*(\sigma(\xi,\xi)) \in \pi_{16}\mathrm{SO}_{11}$ and $d_*(\sigma(\xi,\xi)) \neq 0$ by Theorem 1.1.

Moreover, $d_*(\sigma(\xi,\xi))$ maps to $0\in\pi_{16}\mathrm{SPL}_{11}$; this is explained following Theorem 1.1 using the exact sequence $\pi_{*+1}(\mathrm{PL}_{11}/O_{11})\xrightarrow{d_*}\pi_*(\mathrm{SO}_{11})\longrightarrow\pi_*(\mathrm{SPL}_{11})$. Now by Antonelli [2], the normal bundle of every homotopy 17–sphere embedded in euclidean space in codimension 12 is zero. Hence

$$\nu(\Sigma_{\xi,\xi}\subset S^{28})\in\ker(\pi_{17}B\mathrm{SO}_{11}\to\pi_{17}B\mathrm{SO}_{12})=\mathrm{im}(\pi_{17}S^{11}\to\pi_{17}B\mathrm{SO}_{11}),$$

where the last map is induced by the classifying map of the tangent bundle of the 11–sphere.

It remains to prove Lemmas 2.1 and 2.2.

Proof of Lemma 2.1 Choose a smooth map $\psi\colon D^{n-k}\times D^k\to D^k$ representing the class $[\psi]\in\pi_{n-k}(\mathrm{Diff}_\partial(D^k))$, ie for $x\in D^{n-k}$ we have that $\psi_x:=\psi(x,-)\in\mathrm{Diff}_\partial(D^k)$, and $\psi_x=\mathrm{Id}_{D^k}$ for $x\in\partial D^{n-k}$. Then $\lambda([\psi])$ is represented by

$$\Psi\colon D^k\times D^{n-k}\to D^k\times D^{n-k},\quad (x,y)\mapsto(x,\psi(x,y)),$$

and $A([\psi])$ is represented by the homotopy sphere Σ^{n+1}_Ψ obtained by gluing two copies of D^{n+1} along the boundary using the diffeomorphism Ψ. Note also that the image of $[\psi]$ under the derivative map is represented by $d\psi\colon D^{n-k}\times D^k\to\mathrm{Gl}_k(\mathbb{R})$ with $d\psi(x,y)=D_y\psi_x$.

For technical reasons, we actually assume without loss of generality that the maps are the identity maps in a neighbourhood of the boundaries.

We construct an explicit embedding $\iota_\psi\colon\Sigma^{n+1}_\Psi\hookrightarrow S^{n+k+1}$ of Σ^{n+1}_Ψ, compute the normal bundle of this embedding and show explicitly that it is obtained by clutching with $d\psi$.

As might be expected, given that all our data is on disks (and trivial near the boundary of the disks), we actually produce an interesting embedding ι_ψ of $D^{n+1}=D^{n-k}\times D^k\times[0,1]$ into $D^{n+k+1}=D^{n-k}\times D^k\times D^k\times[0,1]$, which has standard form near the boundary, and then obtain an embedding of Σ^{n+1}_Ψ by gluing with a standard embedding of D^{n+1} into D^{n+k+1} in the appropriate way.

The desired embedding ι_ψ is explicitly given by

$$\begin{aligned}\iota_\psi\colon D^{n-k}\times D^k\times[0,1]&\to D^{n-k}\times D^k\times D^k\times[0,1],\\ (x,y,t)&\mapsto(x,\alpha(t)y,\beta(t)\psi_x(y),t).\end{aligned}$$

Here, $\alpha,\beta\colon[0,1]\to[0,1]$ are smooth maps such that $\alpha(t)=1$ for $t<0.6$ and $\alpha(t)=0$ for $t>0.9$, and $\beta(t)=\alpha(1-t)$.

This is evidently a smooth embedding whose image we denote by S_ψ, and we let $\partial S_\psi = \iota_\psi(\partial D^{n+1})$. Then $\partial S_\psi \subset \partial D^{n+k+1}$: to see this, observe that if either the x– or the t–coordinate is in the boundary, then the first or fourth coordinate of the image point is so, too. For each $t \in [0,1]$, then $\alpha(t) = 1$ or $\beta(t) = 1$. If $y \in \partial D^k$, then therefore either the second or the third component of the image point is in the boundary (or both). As $\partial(D^{n-k} \times D^k \times [0,1])$ is the union of those points with at least one component in the boundary, this proves the claim.

We also note that the subset $\partial S_\psi \subset \partial D^{n+k+1}$ is in fact independent of ψ (as ψ is fixed to be the identity map near the boundary). Let us identify this image set ∂S_ψ with $S^n = \partial(D^{n-k} \times D^k \times [0,1])$ via the restriction of ι_{Id} to $\partial(D^{n-k} \times D^k \times [0,1])$.

Then $\iota_\psi|\colon \partial(D^{n-k} \times D^k \times [0,1]) \to \partial(D^{n-k} \times D^k \times D^k \times [0,1])$ is supported on the disk $D^{n-k} \times D^k \times \{1\}$, where it is given by Ψ. Therefore, we can glue two copies of $D^{n-k} \times D^k \times D^k \times [0,1]$ along the boundary by the identity map to obtain S^{n+k+1}, and the embeddings ι_ψ in one copy and ι_{Id} in the other glue together to form the desired embedding of Σ^{n+1}_Ψ into S^{n+k+1}.

Strictly speaking, one has to round the corners off to get an actual smooth embedding. This can easily be achieved, as ψ is the identity in a neighbourhood of the boundaries. We omit spelling out the somewhat cumbersome details.

It remains to compute the normal bundle of the embedding. To do this, we first compute the differential

$$D\iota_\psi\colon (D^{n-k} \times D^k \times [0,1]) \times (\mathbb{R}^{n-k} \oplus \mathbb{R}^k \oplus \mathbb{R}) \to S_\psi \times (\mathbb{R}^{n-k} \oplus \mathbb{R}^k \oplus \mathbb{R}^k \oplus \mathbb{R})$$

to be given in each fibre by

$$D_{(x,y,t)}\iota_\psi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha(t) & \alpha'(t)y \\ \beta(t)\,\partial_x\psi & \beta(t)\,d\psi & \beta'(t)\psi_x(y) \\ 0 & 0 & 1 \end{pmatrix}.$$

We obtain an explicit trivialisation of the normal bundle of this embedding via the fibrewise linear map covering ι_ψ,

$$\nu\colon D^{n-k} \times D^k \times [0,1] \times \mathbb{R}^k \to S_\psi \times (\mathbb{R}^{n-k} \oplus \mathbb{R}^k \oplus \mathbb{R}^k \oplus \mathbb{R}),$$

$$\nu_{(x,y,t)} = \begin{pmatrix} 0 \\ -\beta(t)(d\psi)^{-1} \\ \alpha(t) \\ 0 \end{pmatrix}.$$

To observe that this really describes the normal bundle, for dimension reasons we just have to check that the image of ν intersects the tangent bundle of S_ψ, ie the image of $D\iota_\psi$, trivially. It is clear that $\nu(v)$ can only be equal to a tangent vector of the form $(0, \alpha(t)w, \beta(t)\, d\psi(w), 0)$ for $w \in \mathbb{R}^k$. This implies $\alpha(t)v = \beta(t)\, d\psi(w)$ and $-\beta(t)v = \alpha(t)\, d\psi(w)$; the two equations imply $\alpha(t)^2 v = -\beta(t)^2 v$ and finally (as $\alpha(t)^2 + \beta(t)^2 > 0$) then $v = 0$ and then also $w = 0$. It follows that the image of ν represents the normal bundle of S_ψ in D^{n+k+1}.

For the other half-disk which produces the embedding of Σ_Ψ^{n+1} into S^{n+k+1}, we obtain a trivialisation of the normal bundle by the same recipe, replacing ψ by Id. We observe then that we obtain the global normal bundle by gluing these two explicitly chosen normal subbundles of TD^{n+k+1} along the boundary, where they coincide. The trivialisations differ precisely on the half-disk $\iota(D^{n-k} \times D^k \times \{1\})$, and there they differ by the derivative map $d\psi$. On the other half-disk, the two trivialisations coincide.

Consequently, the normal bundle of the embedding ι_ψ is obtained by clutching with $d\psi$, precisely as claimed, and the lemma is proved. □

Remark 2.3 It is tempting to hope that the explicit geometric construction of d_* as the normal bundle of the embedding ι_ψ can be used to get some new information about d_*. On the other hand, the information obtained by the formulas in the proof given above seems rather limited. At least in the case where ψ lies in the image of σ, we present Conjecture 3.1 below on $d_* \circ \sigma$.

Proof of Lemma 2.2 Let A_8^{18} denote the group of bordism classes relative boundary of 8–connected 18–manifolds with boundary a homotopy sphere, which are defined in [23, Section 17]. Specifically, elements of A_8^{18} are represented by compact oriented 8–connected 18–manifolds W with boundary a homotopy sphere, and W_1 is bordant to W_2 if there is an h–cobordism Y between their boundaries such that the closed manifold $W_1 \cup Y \cup -W_2$ bounds an 8–connected 19–manifold. According to [23, Section 17] and [22, Theorem 2(5)], we have an isomorphism

$$A_8^{18} \to \mathbb{Z}/2 \oplus \mathbb{Z}/2, \quad [W] \mapsto (\Phi(\varphi_W), \varphi_W(\hat{\chi}_W)). \tag{2}$$

Here $\varphi_W \colon H_9(W;\mathbb{Z}) \to \mathbb{Z}/2$ is a quadratic refinement of the mod 2 intersection form defined as follows. By [22, Lemma 2], representing an integral homology class by an embedded sphere and taking its normal bundle gives rise to a quadratic map

$$\alpha_W \colon H_9(W;\mathbb{Z}) \to \pi_8 \mathrm{SO}_9.$$

Since the stabilisation map $S\colon \pi_8\mathrm{SO}_9 \to \pi_8\mathrm{SO} = \mathbb{Z}/2$ is split surjective with kernel $\mathbb{Z}/2$, from α_W we obtain a quadratic map $\varphi_W\colon H_9(W;\mathbb{Z}) \to \mathbb{Z}/2$ with values in $\mathbb{Z}/2 = \ker(S)$ by fixing a splitting of $\pi_8\mathrm{SO}_9$. The first component of (2) is the Arf invariant of φ_W and we next define the second component. Let $S\alpha_W\colon H_9(W;\mathbb{Z}) \to \mathbb{Z}/2 = \pi_8(\mathrm{SO})$ be the composition of α_W with the stabilisation map S above. Using [22, Lemma 2] again, we see that $S\alpha_W$ is a homomorphism. Define $\chi_W \in H_9(W;\mathbb{Z}/2)$ to be the Poincaré dual of

$$S\alpha_W \in \mathrm{Hom}(H_9(W;\mathbb{Z}),\mathbb{Z}/2) = H^9(W;\mathbb{Z}/2) \cong H^9(W,\partial W;\mathbb{Z}/2).$$

The second component of (2) is given by evaluating φ_W on any integral lift $\hat{\chi}_W$ of χ_W.

Let $S^3(\xi) \in \pi_8\mathrm{SO}_9$ be the image of $\xi \in \pi_8\mathrm{SO}_6$ under the inclusion $\mathrm{SO}_6 \to \mathrm{SO}_9$. By the commutativity of (1), $\Sigma_{\xi,\xi}$ is the boundary of the Milnor plumbing W of $S^3(\xi) \in \pi_8\mathrm{SO}_9$ with itself, and we compute $\varphi_W(\hat{\chi}_W)$ as follows: with $H_9(W;\mathbb{Z}) = \mathbb{Z}(x) \oplus \mathbb{Z}(y)$ the normal bundles obtained from representing x and y by embeddings are both given by $S^3(\xi)$. We conclude that $\varphi_W(x) = \varphi_W(y)$. Moreover, we may use that in this basis the intersection form λ_W of W has matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Furthermore, $S^3(\xi)$ stabilises to a generator of $\pi_8\mathrm{SO}$, by Lemma 2.4 below. Thus, $S\alpha_W$ maps both x and y to a generator and so we may take $\hat{\chi}_W = x + y$. We now compute that

$$\varphi_W(\hat{\chi}_W) = \varphi_W(x+y) = \underbrace{\varphi_W(x) + \varphi_W(y)}_{=0} + \rho_2(\lambda_W(x,y)) = 1,$$

where ρ_2 denotes reduction mod 2.

Now, taking the homotopy sphere on the boundary defines a homomorphism

$$\partial\colon A_8^{18} \to \Theta_{17}. \tag{3}$$

From the short exact sequence

$$0 \to bP_{18} (= \mathbb{Z}/2) \to \Theta_{17} \to \mathrm{coker}(J_{17}) \to 0$$

and [7, Theorem 1.4], we see that the image of the map ∂ from (3) consists of precisely 4 different elements, so the map ∂ is injective. Each of the bP–spheres is the boundary of a manifold P which satisfies $S\alpha_P = 0$ and therefore $\varphi_P(\hat{\chi}_P) = 0$ and it follows that the element W from above must map under ∂ to a non-bP–sphere, which then represents $[\eta\eta_4]$ in view of [7, Theorem 1.4]. □

Lemma 2.4 *The map $\mathbb{Z}/24 \cong \pi_8\mathrm{SO}_6 \to \pi_8\mathrm{SO} \cong \mathbb{Z}/2$ is surjective.*

Proof By [14, Theorem 1.4], $(\mathbb{Z}/2)^3 \cong \pi_8\mathrm{SO}_8 \to \pi_8\mathrm{SO} \cong \mathbb{Z}/2$ is onto and therefore has a kernel of 4 elements. (We refer to [12] for the computation of the relevant homotopy groups.) On the other hand, $(\mathbb{Z}/2)^2 \cong \pi_8\mathrm{SO}_7 \to \pi_8\mathrm{SO}_8$ is injective (its cokernel injects into $\pi_8(S^7) \cong \mathbb{Z}/2$) and so has an image of 4 elements. These two subgroups do not coincide: Since the maximal number of pointwise linearly independent vector fields on S^9 is 1 [1, Theorem 1.1], the tangent bundle of S^9 defines an element in $\pi_8\mathrm{SO}_8$ that is not in the image of $\pi_8\mathrm{SO}_7$ but maps to $0 \in \pi_8\mathrm{SO}$.

Therefore, $(\mathbb{Z}/2)^2 \cong \pi_8\mathrm{SO}_7 \to \pi_8\mathrm{SO}$ is surjective and has a kernel of precisely two elements; similarly the image of $\mathbb{Z}/24 \cong \pi_8\mathrm{SO}_6 \to \pi_8\mathrm{SO}_7 \cong (\mathbb{Z}/2)^2$ consists of precisely two elements (its cokernel injects into $\pi_8 S^6 \cong \mathbb{Z}/2$), and we are left to show that these two subgroups do not agree. To see this, we consider the element $a := (2\gamma)_7\eta_7$ where $(2\gamma)_7$ is a generator of $\mathbb{Z} \cong \pi_7\mathrm{SO}_7$ and $\eta_7\colon S^8 \to S^7$ is the nontrivial class: By [14, Theorem 1.4], $(2\gamma)_7$ stabilises to an element divisible by 2 and so a is in the kernel of the stabilisation; and it does not lift to $\pi_8\mathrm{SO}_6$ by the commutativity of the following diagram with exact rows:

$$\begin{tikzcd}
 & \pi_7\mathrm{SO}_7 \arrow[r] \arrow[d, "\eta_7"] & \pi_7 S^6 \arrow[r] \arrow[d, "\eta_7", "\cong"'] & \pi_6\mathrm{SO}_6\ (=0) \\
\pi_8\mathrm{SO}_6 \arrow[r] & \pi_8\mathrm{SO}_7 \arrow[r] & \pi_8 S^6 &
\end{tikzcd}$$

□

We conclude this section by giving the promised second proof of Lemma 2.1. To this end we recall from [18, Section 6] that a smoothing of S^{n+1} in S^{n+k+1} consists of a smooth manifold W and a PL homeomorphism $H\colon W \to S^{n+k+1}$, such that $\Sigma := H^{-1}(S^{n+1}) \subset W$ is a smooth submanifold, and such that H is concordant to the identity smoothing of S^{n+k+1}; and recall the group $\mathfrak{d}^k_{n+1}$ of concordance classes of such smoothings.[2] We note that up to diffeomorphism, W is a standard sphere mapping to S^{n+k+1} by a PL homeomorphism concordant to the identity, so that elements of $\mathfrak{d}^k_{n+1}$ are represented by PL homeomorphisms $H\colon S^{n+k+1} \to S^{n+k+1}$ which are concordant to the identity (ie orientation-preserving). Note also that Σ is a homotopy $(n+1)$–sphere, oriented through the PL homeomorphism $h := H|_\Sigma$, which is smoothly embedded into S^{n+k+1}.

[2] The group $\mathfrak{d}^k_{n+1}$ is denoted by Γ^k_{n+1} in [18]. We have used different notation, to avoid confusion with the notation Γ^{n+1}_k for the subgroups of the Gromoll filtration.

There are two obvious homomorphisms out of $\mathfrak{d}^k_{n+1}$,

$$\Theta_{n+1} \xleftarrow{F} \mathfrak{d}^k_{n+1} \xrightarrow{\nu} \pi_{n+1} B\mathrm{SO}_k,$$

the left one mapping the class of H to the diffeomorphism class of Σ, and the right one to the classifying map of the normal bundle of $\Sigma \subset S^{n+k+1}$ (where, as usual, we identify a homotopy sphere up to homotopy equivalence with a standard sphere using the given orientation). Then, Lemma 2.1 is clearly implied by the following result:

Lemma 2.5 *There exists a group homomorphism $B\colon \pi_{n-k}\mathrm{Diff}_\partial(D^k) \to \mathfrak{d}^k_{n+1}$ such that the following diagram commutes up to possible signs*:

$$\begin{tikzcd}
 & \pi_{n-k}\mathrm{Diff}_\partial(D^k) \arrow[r, "d_*"] \arrow[dl, "A"'] \arrow[d, "B"] & \pi_n\mathrm{SO}_k \arrow[d, "\cong"] \\
\Theta_{n+1} & \mathfrak{d}^k_{n+1} \arrow[l, "F"'] \arrow[r, "\nu"] & \pi_{n+1} B\mathrm{SO}_k
\end{tikzcd}$$

Proof We recall the homotopy equivalence

$$M_k\colon \mathrm{Diff}_\partial(D^k) \to \Omega^{k+1}\mathrm{PL}_k/\mathrm{SO}_k$$

of Morlet (see [5, Theorem 4.4]) and consider the diagram

(4)
$$\begin{tikzcd}
\pi_0\mathrm{Diff}_\partial(D^n) \arrow[dd, "(M_n)_*"] \arrow[ddd, bend right=60, "C"', "\cong"] & \pi_{n-k}\mathrm{Diff}_\partial(D^k) \arrow[l, "\lambda"'] \arrow[r, "d_*"] \arrow[d, "(M_k)_*"] & \pi_n\mathrm{SO}_k \arrow[d, equal] \\
 & \pi_{n+1}\mathrm{PL}_k/O_k \arrow[r, "\partial"] \arrow[d, "\tilde{i}"] \arrow[dl, "S"'] & \pi_n\mathrm{SO}_k \arrow[d, equal] \\
\pi_{n+1}\mathrm{PL}/O & \pi_{n+1}\widetilde{\mathrm{PL}}_k/O_k \arrow[l, "\tilde{S}"'] \arrow[r, "\partial"] & \pi_n\mathrm{SO}_k \\
\Theta_{n+1} \arrow[u, "\cong"', "\Psi"] & \mathfrak{d}^k_{n+1} \arrow[l, "F"'] \arrow[r, "\nu"] \arrow[u, "\cong"'] & \pi_{n+1} B\mathrm{SO}_k \arrow[u, "\cong"']
\end{tikzcd}$$

Here Ψ is the map which sends a homotopy sphere Σ to the element represented by the tangent PL microbundle of the mapping cylinder $\mathrm{cyl}(h\colon \Sigma \to S^{n+1})$ of an orientation-preserving PL homeomorphism h, along with its linear structure induced by the smooth structure of Σ on the Σ end of the cylinder and its canonical trivialisation at the S^{n+1} end. The map Ψ is an isomorphism by surgery theory; see eg [15, Theorem 6.48]. The map $\mathfrak{d}^k_{n+1} \to \pi_{n+1}\widetilde{\mathrm{PL}}_k/O_k$ is an isomorphism by [18, Corollary 6.7]: it is defined by sending the class of $(H,h)\colon (S^{n+k+1},\Sigma^{n+1}) \to (S^{n+k+1},S^{n+1})$ to the normal block bundle ν_{cyl} of $\mathrm{cyl}(h)$ inside $\mathrm{cyl}(H)$ along with its linear reduction at the Σ^{n+1} end of the cylinder and its canonical trivialisation at the other end. Finally, the map $\widetilde{S}$ is obtained from the fact that the inclusion $\mathrm{PL} \to \widetilde{\mathrm{PL}}$ is an equivalence [19, Corollary 5.5(ii)]; that is, there is no essential difference between stable PL (micro)bundles and stable block bundles.

We claim that the lower left square of (4) is commutative up to sign. To see this, we may assume, increasing k if necessary, that the normal block bundle ν_{cyl} is given by a PL microbundle. Then, the sum of the two composites, applied to $[(H,h)]$, is represented by the direct sum microbundle $T\mathrm{cyl}(h) \oplus \nu_{\mathrm{cyl}}$ over $\mathrm{cyl}(h)$ along with its linear reduction at the front end and its canonical trivialisation at the other end. But now, we have an isomorphism $T\mathrm{cyl}(h) \oplus \nu_{\mathrm{cyl}} \cong T\mathrm{cyl}(H)|_{\mathrm{cyl}(h)}$ of microbundles which extends isomorphisms $T\Sigma \oplus \nu_{\Sigma\subset S^{n+k+1}} \cong TS^{n+k+1}|_{\Sigma}$ and $TS^{n+1} \oplus \nu_{S^{n+1}\subset S^{n+k+1}} \cong TS^{n+k+1}|_{S^{n+1}}$ of vector bundles.

Since H is PL isotopic to the identity (being an orientation-preserving PL homeomorphism of the sphere), we conclude that the sum of the two composite maps, applied to $[(H,h)]$, represents the zero element.

All other parts of this diagram commute up to possible signs: the commutativity of the squares on the right and of the triangle in the middle follows from the definitions. That $M_{n*} \circ \lambda = S \circ M_{k*}$ follows from [8, Lemma 2.5], and that $M_{n*} = \Psi \circ C$ is proven in [8, Lemma 2.7]. The lemma now follows by a diagram chase. □

3 Concluding remarks

In this section we discuss some of the background to our results and state a conjecture about the map $d_* \circ \sigma$.

(1) The homotopy fibre of $d\colon \mathrm{Diff}_\partial(D^k) \to \Omega^k \mathrm{SO}_k$ is the H–space $\mathrm{Diff}^{\mathrm{fr}}_\partial(D^k)$ of framing-preserving diffeomorphisms. It is the loop space of the classifying space $B\mathrm{Diff}^{\mathrm{fr}}_\partial(D^k)$, which features in the recent work of Kupers and Randal-Williams [13] on the rational homotopy groups of $\mathrm{Diff}_\partial(D^k)$. We see that d is rationally trivial because the Alexander trick implies that d becomes nullhomotopic after composition with the natural map $\Omega^k\mathrm{SO}_k \to \Omega^k\mathrm{SPL}_k$. It is well known that $(\mathrm{SO}_k)_{\mathbb{Q}}$ is Eilenberg–Mac Lane, detected by the suspensions of the rational Pontryagin classes and rational Euler class. Since these classes are defined on $(\mathrm{SPL}_k)_{\mathbb{Q}}$, it follows that $(\mathrm{SO}_k)_{\mathbb{Q}}$ is a homotopy

retract of $(\mathrm{SPL}_k)_{\mathbb{Q}}$. If $\Omega^k_0 X$ denotes the connected component of the constant map, then it follows that $(\Omega^k_0\mathrm{SO}_k)_{\mathbb{Q}} \simeq \Omega^k_0(\mathrm{SO}_k)_{\mathbb{Q}}$ is a homotopy retract of $(\Omega^k_0\mathrm{SPL}_k)_{\mathbb{Q}}$, showing that the map $d\colon \mathrm{Diff}_\partial(D^k) \to \Omega^k_0\mathrm{SO}_k$ is rationally nullhomotopic.

(2) The proof of Theorem 1.1 relies on the fact that the normal bundle of any embedding $\Sigma_{\xi,\xi} \hookrightarrow S^{28}$ is nontrivial. Despite the elementary argument we give for this in Section 2, computing the normal bundle of an embedding of a homotopy sphere $g\colon \Sigma^{n+1} \hookrightarrow S^{n+k+1}$ is a subtle problem. Provided one is in the metastable range $n < 2k-4$, Haefliger [10] proved that the isotopy class of g depends only on the diffeomorphism type of Σ, so that, in particular, the normal bundle is independent of the choice of embedding. Hsiang, Levine and Sczarba [11] proved that the latter statement holds even for $n < 2k-2$, defined the homomorphism

$$\phi^k_{n+1}\colon \Theta_{n+1} \to \pi_{n+1}B\mathrm{SO}_k, \quad \Sigma \mapsto \nu(\Sigma \subset S^{n+k+1}), \quad \text{where } n < 2k-2,$$

and proved that $\phi^{13}_{16} \neq 0$; ie the exotic 16–sphere embeds into S^{29} with nontrivial normal bundle. Then Antonelli [2] made a systematic study of normal bundles of homotopy spheres in the metastable range, which includes the statement that $\phi^{11}_{17} \neq 0$.

(3) Concerning A'Campo's claim that d_* vanishes for $i = 2$, we note that, since $\phi^{13}_{16} \neq 0$, Lemma 2.1 entails that if A'Campo's claim holds, then the exotic 16–sphere does not lie in the image of the map $A\colon \pi_2\mathrm{Diff}_\partial(D^{13}) \to \Theta_{16} = \mathbb{Z}/2$. This is consistent with computations we have made for the refined plumbing pairing

$$\sigma\colon \pi_8\mathrm{SO}_6 \otimes \pi_7\mathrm{SO}_8 \to \pi_2\mathrm{Diff}_\partial(D^{13}),$$

which show that $A \circ \sigma = 0$, even though $\sigma_M\colon \pi_8\mathrm{SO}_7 \otimes \pi_7\mathrm{SO}_8 \to \Theta_{16}$ is nontrivial, a statement which can be deduced from [20, Satz 12.1].

(4) Finally, we present a conjectural description of the homomorphism

$$d_* \circ \sigma\colon \pi_p\mathrm{SO}_{q-a} \otimes \pi_q\mathrm{SO}_{p-b} \to \pi_{p+q}\mathrm{SO}_{p+q-a-b-1}$$

in purely homotopy-theoretic terms.

Let $h\colon \pi_i\mathrm{SO}_j \to \pi_i(S^{j-1})$ be the map induced by the canonical projection $\mathrm{SO}_j \to S^{j-1}$. For maps $f\colon W \to X$ and $f\colon Y \to Z$ let $f * g\colon W * Y \to X * Z$ be their *join*. Let $\partial\colon \pi_{m+1}(S^k) \to \pi_m\mathrm{SO}_k$ denote the boundary map in the homotopy long exact sequence of the fibration $\mathrm{SO}_k \to \mathrm{SO}_{k+1} \to S^k$. For compactness, we use the notation

$$p' := p-b \quad \text{and} \quad q' := q-a$$

and let $\xi_1 \in \pi_p \mathrm{SO}_{q'}$ and $\xi_2 \in \pi_q \mathrm{SO}_{p'}$. Then we have

$$h(\xi_1) \in \pi_p(S^{q'-1}), \quad h(\xi_2) \in \pi_q S^{p'-1} \quad \text{and} \quad h(\xi_1) * h(\xi_2) \in \pi_{p+q+1}(S^{p'+q'-1}),$$

so that $\partial\big(h(\xi_1) * h(\xi_2)\big) \in \pi_{p+q}\mathrm{SO}_{p'+q'-1}$.

In addition, we have the J–homomorphisms

$$J_{p,q'} \colon \pi_p \mathrm{SO}_{q'} \to \pi_{p+q'} S^{q'} \quad \text{and} \quad J_{q,p'} \colon \pi_q \mathrm{SO}_{p'} \to \pi_{q+p'} S^{p'},$$

and we can suspend in the target of each of these to get the homomorphisms

$$\Sigma^a \circ J_{p,q'} \colon \pi_p \mathrm{SO}_{q'} \to \pi_{p+q} S^q \quad \text{and} \quad \Sigma^b \circ J_{q,p'} \colon \pi_q \mathrm{SO}_{p'} \to \pi_{p+q} S^p.$$

We then take compositions with the maps induced by ξ_i for $i = 1, 2$ and the inclusions $i_{p'} \colon \mathrm{SO}_{p'} \to \mathrm{SO}_{p'+q'-1}$ and $i_{q'} \colon \mathrm{SO}_{q'} \to \mathrm{SO}_{p'+q'-1}$. Hence we have homomorphisms

$$\bar{\xi}_{2*} \colon \pi_p \mathrm{SO}_{q'} \xrightarrow{\Sigma^a \circ J_{p,q'}} \pi_{p+q} S^q \xrightarrow{\xi_{2*}} \pi_{p+q} \mathrm{SO}_{p'} \xrightarrow{i_{p'*}} \pi_{p+q} \mathrm{SO}_{p'+q'-1},$$
$$\bar{\xi}_{1*} \colon \pi_q \mathrm{SO}_{p'} \xrightarrow{\Sigma^b \circ J_{q,p'}} \pi_{p+q} S^p \xrightarrow{\xi_{1*}} \pi_{p+q} \mathrm{SO}_{q'} \xrightarrow{i_{q'*}} \pi_{p+q} \mathrm{SO}_{p'+q'-1}.$$

Conjecture 3.1 *Up to sign, the homomorphism*

$$d_* \circ \sigma \colon \pi_p \mathrm{SO}_{q'} \otimes \pi_q \mathrm{SO}_{p'} \to \pi_{p+q} \mathrm{SO}_{p'+q'-1}$$

is given by

$$d_*(\sigma(\xi_1, \xi_2)) = \partial\big(h(\xi_1) * h(\xi_2)\big) + \bar{\xi}_{1*}(\xi_2) + \bar{\xi}_{2*}(\xi_1).$$

We briefly discuss Conjecture 3.1 in light of Theorem 1.1 and Corollary 1.2. For $\xi \in \pi_8 \mathrm{SO}_6$ a generator, $h(\xi) \in \pi_8 S^5 \cong \pi_3^s$ is again a generator and we choose ξ so that $h(\xi) = \nu_5$. Hence Conjecture 3.1 gives $d_*(\sigma(\xi, \xi)) = \partial(\nu_5 * \nu_5) + 2\bar{\xi}_*(\xi)$. Now $\pi_{16} S^8 \cong (\mathbb{Z}/2)^4$, which entails that $2\bar{\xi}_*(\xi) = 0$ and the proof of Corollary 1.2 shows that $d_*(\sigma(\xi, \xi)) = \partial(\nu_{11}^2)$. Since $\nu_5 * \nu_5 = \nu_{11}^2$, Conjecture 3.1 is consistent with Theorem 1.1 and Corollary 1.2, with both giving the same nonzero expression for $d_* \circ \sigma \colon \pi_8 \mathrm{SO}_6 \otimes \pi_8 \mathrm{SO}_6 \to \pi_{16} \mathrm{SO}_{11}$.

References

[1] **J F Adams**, *Vector fields on spheres*, Ann. of Math. 75 (1962) 603–632 MR Zbl

[2] **P L Antonelli**, *On stable diffeomorphism of exotic spheres in the metastable range*, Canadian J. Math. 23 (1971) 579–587 MR Zbl

[3] **P Antonelli**, **D Burghelea**, **P J Kahn**, *Gromoll groups,* Diff S^n *and bilinear constructions of exotic spheres*, Bull. Amer. Math. Soc. 76 (1970) 772–777 MR Zbl

[4] **P L Antonelli**, **D Burghelea**, **P J Kahn**, *The nonfinite homotopy type of some diffeomorphism groups*, Topology 11 (1972) 1–49 MR Zbl

[5] **D Burghelea**, **R Lashof**, *The homotopy type of the space of diffeomorphisms, I*, Trans. Amer. Math. Soc. 196 (1974) 1–36 MR Zbl

[6] **D Burghelea**, **R Lashof**, *The homotopy type of the space of diffeomorphisms, II*, Trans. Amer. Math. Soc. 196 (1974) 37–50 MR Zbl

[7] **R Burklund**, **A Senger**, *On the high-dimensional geography problem*, preprint (2020) arXiv 2007.05127

[8] **D Crowley**, **T Schick**, *The Gromoll filtration, KO–characteristic classes and metrics of positive scalar curvature*, Geom. Topol. 17 (2013) 1773–1789 MR Zbl

[9] **D Gromoll**, *Differenzierbare Strukturen und Metriken positiver Krümmung auf Sphären*, Math. Ann. 164 (1966) 353–371 MR Zbl

[10] **A Haefliger**, *Plongements différentiables de variétés dans variétés*, Comment. Math. Helv. 36 (1961) 47–82 MR Zbl

[11] **W C Hsiang**, **J Levine**, **R H Szczarba**, *On the normal bundle of a homotopy sphere embedded in Euclidean space*, Topology 3 (1965) 173–181 MR Zbl

[12] **M A Kervaire**, *Some nonstable homotopy groups of Lie groups*, Illinois J. Math. 4 (1960) 161–169 MR Zbl

[13] **A Kupers**, **O Randal-Williams**, *On diffeomorphisms of even-dimensional discs*, J. Amer. Math. Soc. (online ahead of print October 2023)

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

[15] **W Lück**, *A basic introduction to surgery theory*, from "Topology of high-dimensional manifolds, I, II" (F T Farrell, L Göttsche, W Lück, editors), ICTP Lect. Notes 9, Abdus Salam Int. Cent. Theoret. Phys., Trieste (2002) 1–224 MR Zbl

[16] **J Milnor**, *Differentiable structures on spheres*, Amer. J. Math. 81 (1959) 962–972 MR Zbl

[17] **J Milnor**, *Microbundles, I*, Topology 3 (1964) 53–80 MR Zbl

[18] **C P Rourke**, **B J Sanderson**, *Block bundles, I*, Ann. of Math. 87 (1968) 1–28 MR Zbl

[19] **C P Rourke**, **B J Sanderson**, *Block bundles, III: Homotopy theory*, Ann. of Math. 87 (1968) 431–483 MR Zbl

[20] **S Stolz**, *Hochzusammenhängende Mannigfaltigkeiten und ihre Ränder*, Lecture Notes in Math. 1116, Springer (1985) MR Zbl

[21] **H Toda**, *Composition methods in homotopy groups of spheres*, Annals of Mathematics Studies 49, Princeton Univ. Press (1962) MR Zbl

[22] **C T C Wall**, *Classification of* $(n-1)$*–connected* $2n$*–manifolds*, Ann. of Math. 75 (1962) 163–189 MR Zbl

[23] **C T C Wall**, *Classification problems in differential topology, VI: Classification of* $(s-1)$*–connected* $(2s+1)$*–manifolds*, Topology 6 (1967) 273–296 MR Zbl

School of Mathematics and Statistics, University of Melbourne
Parkville, Victoria, Australia

Mathematisches Institut, Universität Göttingen
Göttingen, Germany

Institut für Mathematik, Universität Augsburg
Augsburg, Germany

dcrowley@unimelb.edu.au, thomas.schick@math.uni-goettingen.de,
wolfgang.steimle@math.uni-augsburg.de

https://www.dcrowley.net,
https://www.uni-math.gwdg.de/schick, https://www.uni-augsburg.de/de/fakultaet/mntf/math/prof/diff/team/wolfgang-steimle

Proposed: Stefan Schwede Received: 6 May 2021
Seconded: Ulrike Tillmann, Nathalie Wahl Revised: 12 January 2022

Geometry & Topology **27**:9 (2023) 3715–3731
DOI: 10.2140/gt.2023.27.3715
Published: 5 December 2023

On self-shrinkers of medium entropy in $\mathbb{R}^4$

Alexander Mramor

We study smooth asymptotically conical self-shrinkers in $\mathbb{R}^4$ with Colding–Minicozzi entropy bounded above by Λ_1.

53A10, 53E10

1 Introduction

Self-shrinkers are basic singularity models for the mean curvature flow, and, in the noncompact case, nongeneric ones (generic ones being generalized round cylinders $S^k(\sqrt{2k}) \times \mathbb{R}^{n-k}$) are expected to often be asymptotically conical. Our purpose is to understand the topology of smooth self-shrinkers $M^3 \subset \mathbb{R}^4$ with Colding–Minicozzi entropy $\lambda(M)$, discussed in Section 2, bounded above by Λ_1, the entropy of the round circle.

Our main result is, in part, inspired by arguments of Bernstein and L Wang [3], Hershkovits and White [21], Ilmanen and White [24], Mramor [28], Mramor and S Wang [29] and White [39]. The basic idea is that by considering renormalized mean curvature flows out of (appropriate perturbations of) asymptotically conical self-shrinkers, we may use the entropy assumption to constrain which types of singularities may occur. This has strong implications for how topology may change under the flow. This is useful because topology can, in a sense, be used to "trap" the flow. On the other hand, the flow must clear out; these two principles can then be combined to constrain the topology of the self-shrinker in question.

Theorem 1.1 *Suppose $M^3 \subset \mathbb{R}^4$ is a smooth 2–sided asymptotically conical self-shrinker with entropy less than Λ_1 and k ends. Then it is diffeomorphic to S^3 with k 3–balls removed and replaced with k copies of $S^2 \times \mathbb{R}_+$ attached along their respective boundaries. If $k = 1$ then $M \simeq \mathbb{R}^3$, and in particular this is the case when $\lambda(M) \leq \Lambda_2$.*

This extends to the noncompact case joint work of the author and S Wang [29] on compact self-shrinkers $M^n \subset \mathbb{R}^{n+1}$ when $n = 3$, where they showed that, for each $n \geq 3$, closed self-shrinkers M^n with $\Lambda(M) < \Lambda_{n-2}$ are diffeomorphic to S^n. This in turn extends a result of Colding, Ilmanen, Minicozzi, and White [9], which says closed self-shrinkers with entropy less than $\Lambda_{n-1} \lneq \Lambda_{n-2}$ are diffeomorphic to S^n, hence weakening the assumed entropy bound. In a similar manner, the result above extends (in a weaker sense than the compact case) a result of Bernstein and L Wang [4] for noncompact shrinkers in $\mathbb{R}^4$, where they showed (amongst other results; see Corollary 1.4 therein), for asymptotically conical self-shrinkers $M^3 \subset \mathbb{R}^4$ satisfying $\lambda(M) \leq \Lambda_2$, the stronger conclusion that they are diffeomorphic to $\mathbb{R}^3$. Our argument does at least recover their statement, as discussed at the end of the proof. With the round cylinder in mind, our conclusion seems likely to be sharp in this sense, although it could be possible that a shrinker in $\mathbb{R}^4$ with this entropy bound has more than one end precisely when it is a cylinder.

In this dimension and under this entropy bound, we remark that generic mean curvature flow through neck-pinch singularities has been established by Chodosh, Choi, Mantoulidis and Schulze [7; 8], so for some applications of the flow (see for instance Daniels-Holgate [12]) the study of self-shrinkers in this regime is unnecessary. However, besides its intrinsic interest, this result might still be of use in understanding singularity along nongeneric flows, which could imaginably occur, for instance, in problems involving families of flows (although to the author's knowledge, potential fattening is a more serious concern). It also paints an explicit picture of how a perturbation of a nongeneric flow might only develop neck-pinch singularities, by some copies of the $S^2 \times \mathbb{R}_+$ in the statement above pinching off before, roughly speaking, the S^3 factor collapses to a point (as opposed to handles prematurely pinching off a more complicated model).

An important extra difficulty to consider in the noncompact case versus the closed case is that, a priori, nontrivial topology may be "lost" to spatial infinity under the flow without being properly understood. To illustrate this concern by an admittedly crude thought experiment, a hypothetical translator asymptotically modeled on $T^2 \times \mathbb{R}$ would never develop a singularity, and hence its topology would never be "encountered" as a high curvature region in the flow. In particular it seems, for $n = 3$, asymptotically conical self-shrinkers could a priori have a complicated link. Our first task, and really most of the work of this paper, will be to show that in fact the link is simple.

Theorem 1.2 *Suppose $M^3 \subset \mathbb{R}^4$ is a smooth 2–sided asymptotically conical self-shrinker with entropy less than Λ_1. Then its link L is homeomorphic to a union of S^2.*

As an indication of why one might argue this is reasonable, consider that, in general, the link L of a shrinker $M^n \subset \mathbb{R}^{n+1}$ is of dimension $n-1$, so an entropy bound of Λ_{n-2} on M^n implies morally that its link is low-entropy; for a submanifold $N^k \subset \mathbb{R}^{k+1}$, we say N is low entropy if $\lambda(N) < \Lambda_{k-1}$ (hence the title of paper, since $\Lambda_{k-2} > \Lambda_{k-1}$, as discussed in the next section). These compact surfaces with this entropy bound, at least in low dimensions ($n = 2, 3$), are known to be spheres.

The dimension bound assumption is for topological reasons[1] that perhaps indicate a deficit in knowledge and finesse more than any true difficulty. Potentially providing a sliver of hope that this is the case, Ilmanen and White [24] showed lower bounds for the densities of area-minimizing cones in terms of the topology of their link in every dimension. The area-minimizing property there is employed by using a foliation near the cone by minimal surfaces, which are used as barriers in a mean curvature flow argument (at a high level our argument is similar to theirs). Since cones are noncompact, this is clearly the same sort of result as ours.

For instance, one might hope to directly modify our argument in the next higher dimension ($n=4$) because simply connected 3–manifolds are spherical by the resolution of the 3D Poincaré conjecture by Perelman [30; 31; 32] — we use the corresponding (much easier) fact for surfaces below to classify the link. As an example of why this simple criterion alone doesn't seem to immediately lead to a proof of the corresponding statement for $n = 4$, a potential issue (to the author's understanding) in this dimension is that the link could be a nontrivial homology sphere — below we use that nonspherical oriented surfaces have nontrivial homology in a seemingly essential way. For higher dimensions, of course, there are higher-dimensional versions of the Poincaré conjecture, as verified by Freedman [14] and Smale [33]; this naturally seems even more complicated, for a number of reasons, than the $n = 4$ case just discussed.

Acknowledgements The author is supported by an AMS–Simons travel grant and thanks them for their generosity, as well as the referee for their careful reading and critique.

2 Preliminaries

Let $X : M \to N^{n+1}$ be an embedding of M realizing it as a smooth closed hypersurface of N, which by abuse of notation we also refer to as M. Then the mean curvature flow

[1] In the argument we use the classification of surfaces, Alexander's theorem, and Dehn's lemma, which are dimension-dependent. In a probably less essential way, the 3D Poincaré conjecture is also used.

M_t of M is given by (the image of) $X\colon M \times [0,T) \to N^{n+1}$ satisfying the following, where ν is the outward normal:

$$\frac{dX}{dt} = \vec{H} = -H\nu, \quad X(M,0) = X(M). \tag{2-1}$$

By the comparison principle, singularities occur often, which makes their study important. To study these singularities, one may parabolically rescale about the developing high curvature region to obtain an ancient flow defined for times $(-\infty, T]$; when the basepoint is fixed, this is called a *tangent flow blowup* which will be modeled on self-shrinkers. By Huisken monotonicity [22] these are surfaces equivalently defined as

(1) $M^n \subset \mathbb{R}^{n+1}$ satisfying $H - \frac{1}{2}\langle X, \nu\rangle = 0$, where X is the position vector;

(2) minimal surfaces in the Gaussian metric $G_{ij} = e^{-|x|^2/(2n)}\delta_{ij}$; or

(3) surfaces M which give rise to ancient flows M_t that move by dilations by setting $M_t = \sqrt{-t}M$.

(These notions all make sense at least when the shrinker is smooth, but some definitions apply in the varifold sense as well.) As is well known, the second variation formula for area shows there are no stable minimal surfaces in Ricci positive manifolds; see, for instance, Chapter 1 of [10]. This turns out to also be true for minimal surfaces of polynomial volume growth in $\mathbb{R}^n$ endowed with the Gaussian metric as discussed in [11]. To see why this is so, the Jacobi operator for the Gaussian metric is given by

$$L = \Delta + |A|^2 - \tfrac{1}{2}\langle X, \nabla(\cdot)\rangle + \tfrac{1}{2}. \tag{2-2}$$

The extra $\frac{1}{2}$ term is essentially the reason such self-shrinkers are unstable in the Gaussian metric. For example, owing to the constant term it's clear in the compact case that one could simply plug in the function "1" to get a variation with $Lu > 0$ which doesn't change sign, implying that the first eigenvalue is negative. In fact, every properly embedded shrinker has polynomial volume growth by Q Ding and Y L Xin:

Theorem 2.1 [13, Theorem 1.1] *Any complete noncompact properly immersed self-shrinker M^n in $\mathbb{R}^{n+m}$ has Euclidean volume growth at most.*

We combine these facts below to conclude that the self-shrinker we find in some cases must in fact be unstable.

The mean curvature flow is best understood in the mean convex case because it turns out, under quite weak assumptions, the only possible shrinkers are generalized cylinders $S^k \times \mathbb{R}^{n-k}$. This is especially so for 2–convex surfaces ($\lambda_1 + \lambda_2 > 0$), and a surgery theory with this convexity condition similar to the Ricci flow with surgery has been carried out. For the mean curvature flow with surgery, one finds, for a 2–convex surface M, curvature scales $H_{\mathrm{th}} < H_{\mathrm{neck}} < H_{\mathrm{trig}}$ such that when $H = H_{\mathrm{trig}}$ at some point p and time t the flow is stopped, and suitable points where $H \sim H_{\mathrm{neck}}$ are found to do surgery where "necks" (at these points the surface will be approximately cylindrical) are cut and caps are glued in. The high curvature regions are topologically identified as S^n or $S^{n-1} \times S^1$ and discarded, and the low curvature regions will have curvature bounded on the order of H_{th}. The flow is then restarted and the process repeated.

It was initially established for compact 2–convex hypersurfaces in $\mathbb{R}^{n+1}$ where $n \geq 3$ by Huisken and Sinestrari in [23], and their approach was later extended to the case $n = 2$ by Brendle and Huisken in [5], where 2–convexity is mean convexity. A somewhat different approach covering all dimensions simultaneously was given later by Haslhofer and Kleiner in [17] shortly afterwards. Haslhofer and Ketover then showed several years later in Section 8 of [15], en route to proving their main result, that the mean curvature flow with surgery can be applied to *compact* mean convex hypersurfaces in general ambient manifolds. Important to this article, the author with S Wang established it for (compact) mean convex hypersurfaces with entropy less than Λ_{n-2} in the sense of Colding and Minicozzi.

In [10], Colding and Minicozzi introduced their important notion of entropy, which is defined as the supremum of translated and rescaled Gaussian densities; indeed, consider a hypersurface $\Sigma^k \subset \mathbb{R}^\ell$. Then, given $x_0 \in \mathbb{R}^\ell$ and $r > 0$, define the functional $F_{x_0,r}$ by

$$F_{x_0,r}(\Sigma) = \frac{1}{(4\pi r)^{k/2}} \int_\Sigma e^{-|x-x_0|^2/(4r)}\, d\mu. \tag{2-3}$$

(When $x_0 = \vec{0}$ and $r = 1$, this is just a normalization of area in the Gaussian metric.) Colding and Minicozzi then define the entropy $\lambda(\Sigma)$ of a submanifold to be the supremum over all $F_{x_0,r}$ functionals:

$$\lambda(\Sigma) = \sup_{x_0,r} F_{x_0,r}(\Sigma). \tag{2-4}$$

The aforementioned Huisken monotonicity [22] implies that this quantity is in fact monotone under the flow, and because it is defined as a supremum over rescalings and recenterings, it also controls the nature of singularities encountered along the flow; see [9; 2; 3; 4] for instance. Note that surfaces of polynomial volume growth have finite entropy.

The current state of knowledge of mean curvature flow singularities approached from an entropy perspective seems to be "quantized" by the entropy Λ_k of round spheres as we now discuss. By a calculation of Stone [34] we have

$$\Lambda_1 > \tfrac{3}{2} > \Lambda_2 > \cdots > \Lambda_n \to \sqrt{2}.$$

So far in the literature, many results using an entropy condition assume that the submanifold M under consideration satisfies $\lambda(M) < \Lambda_{n-1}$, which seems to most often be referred to as a low or small entropy condition. The next natural entropy condition to consider then is a bound by Λ_{n-2}, which we refer to as a medium entropy bound; one might expect studying surfaces with this entropy bound to be tractable because morally it implies that mean convex singularities encountered will be 2–convex, which as implied above in the discussion on surgery are the easiest to consider/flow through (after convex ones). Indeed this philosophy was carried out in the compact case in the joint work with S Wang [29] ("low" in its title refers to what we define as medium). An important observation for our argument is that this philosophy can be extended to the noncompact setting, but there are significant new issues to consider. For instance, in the noncompact case the asymptotics of the submanifold in question matter.

Throughout this article we will say an end E of a self-shrinker is *asymptotically conical* if E satisfies $\lim_{\tau\to\infty} \tau^{-1} E = C(E)$ in $C^\infty_{\mathrm{loc}}(\mathbb{R}^{n+1} \setminus 0)$ for $C(E)$ a regular cone in $\mathbb{R}^{n+1}$. A similar definition can be made for asymptotically cylindrical ends, and by results of L Wang [35], for $n = 2$ every end of a self-shrinker of finite topology is either asymptotically conical or cylindrical (with multiplicity one). Naturally, one says a self-shrinker is asymptotically conical if every end is. Considering singular/GMT extensions of shrinkers and asymptotically conical ends in a natural way, under suitable entropy assumptions and assumptions on the underlying measure, the support of the shrinker and asymptotic cone can be shown to be smooth (so asymptotically conical as in the sense above); see [4, Propositions 3.2 and 3.3; 8, Lemma 2.1]. In particular, our theorem applies to asymptotically conical (in the weak sense) shrinkers which arise as blowups under our entropy assumption in $\mathbb{R}^4$. Note, since $\Lambda_1 < 2$, that the convergence will be with multiplicity one.

In the same paper where they introduced entropy, Colding and Minicozzi showed the only singularities which morally shouldn't be able to be perturbed away are the mean convex ones, the generalized round cylinders $S^k(\sqrt{2k}) \times \mathbb{R}^{n-k}$, called round because the spherical factor is a standard round sphere of a radius appropriate to satisfy the shrinker equation. In particular, other singularity models should be able to be perturbed away, so round cylinders are called generic singularity models. Their numbers are few (only n of them), whereas for instance in $\mathbb{R}^3$ there are self-shrinkers are of arbitrarily large genus by [25], so one could say most self-shrinkers are nongeneric.

Concerning nongeneric singularity models, the no-cylinder conjecture of Ilmanen says that the types of ends shouldn't be "mixed" in that if there is a single cylindrical end then M is a cylinder, so one expects that "most" self-shrinkers in $\mathbb{R}^3$ are asymptotically conical (see [36] for a partial result confirming this). Extending this conjecture to the next higher dimension, this provides our justification with the above paragraph in mind for the claim that self-shrinkers are "often" asymptotically conical — it is also quite convenient for analytical reasons.

Returning to flows through singularities, an important advantage of the mean curvature flow with surgery is that the topological change across discontinuous times, when necks are cut and high curvature regions discarded, is easy to understand. A disadvantage is that it isn't quite a Brakke flow (a geometric measure theory formulation of the mean curvature flow) and so does not immediately inherit some of the consequences thereof, but at least it is closely related to the level set flow by results of Lauer [26] and Head [19; 20] which, in the nonfattening case, is (modulo some technicalities). In their work they show that surgery converges to the level set flow in Hausdorff distance (and in fact in the varifold sense, as Head shows) as the surgery parameters degenerate (ie as one lets $H_{\text{th}} \to \infty$). This connection is useful for us because deep results of White [38] show that a mean convex LSF will converge to a (possibly empty) stable minimal surface long-term.

As mentioned above, mean curvature flow with surgery in a curved ambient setting (at least for 3–manifolds and bounded geometry) has been already accomplished by Haslhofer and Ketover, but some extra care is needed for the Gaussian metric, especially in the noncompact case. This is because the metric is poorly behaved at infinity (as one sees from the calculation of its scalar curvature), which introduces some analytic difficulties for using the flow, so instead we consider the renormalized mean curvature flow (which we'll abbreviate RMCF) defined by

$$\frac{dX}{dt} = \vec{H} + \tfrac{1}{2}X. \tag{2-5}$$

Here, as before, X is the position vector on M. It is related to the regular mean curvature flow by the following reparametrization; this allows one to transfer many deep theorems on the MCF to the RMCF. Suppose that M_t is a mean curvature flow on

$[-1, T)$ for $-1 < T \leq 0$ ($T = 0$ is the case for a self-shrinker). Then the renormalized flow $\widehat{M}_\tau$ of M_t, defined on $[0, -\log(-T))$, is given by

$$\widehat{X}_\tau = e^{\tau/2} X_{-e^{-\tau}}, \quad \tau = -\log(-t). \tag{2-6}$$

Up to any finite time the reparametrization is bounded and preserves many properties of the regular MCF, like the avoidance principle and that entropy is monotone under the RMCF. With this in mind, the author showed in his previous article [28] that one can then construct a flow with surgery using the RMCF on suitable perturbations of noncompact self-shrinkers, and that as one lets the surgery parameters degenerate, indeed the surgery converges to the level set flow when $n = 2$. This can be readily combined with the aforementioned joint work with S Wang [29] to show the following:

Theorem 2.2 *Let $M^n \subset \mathbb{R}^{n+1}$ be a smoothly asymptotically conical hypersurface such that $H - \frac{1}{2}\langle X, \nu\rangle \geq c(1 + |X|^2)^{-\alpha}$ for some constants $c, \alpha > 0$ and choice of normal such that $\lambda(M) < \Lambda_{n-2}$. Then denoting by K the region bounded by M whose outward normal corresponds to the choice of normal on M, the level set flow M_t of M with respect to the renormalized mean curvature flow satisfies*:

(1) *The flow is inward, in that $K_{t_1} \subset K_{t_2}$ for any $t_1 > t_2$, considering the corresponding motion of K.*

(2) *M_t is the Hausdorff limit of surgery flows S^k_t with initial data M.*

(3) *M_t is a forced Brakke flow* (*with forcing term given by position vector*).

Here α–noncollapsedness means there are inner and outer osculating balls of radius proportional to the shrinker mean curvature, and this has many consequences; see [1; 16]. The assumption on the asymptotics are conditions for which shrinker mean convexity is preserved and existence of an entropy-decreasing perturbation of a self-shrinker smoothly asymptotic to a cone can always be assumed to satisfy this by work of Bernstein and L Wang in [3]. We use this theorem (often implicitly) below with $n = 3$ when we discuss the flow of M.

H_G, the mean curvature in the Gaussian metric, is related to the renormalized mean curvature by $H_G = e^{|x|^2/4}\big(H - \frac{1}{2}\langle X, \nu\rangle\big)$, and as a result the time limit of the flow defined in the theorem above by White's theory for mean convex MCF (in particular [38]) will be a stable self-shrinker if nonempty. It will also have finite entropy by Huisken monotonicity, and hence have polynomial volume growth. As a result, either by the instability results mentioned above or by the Frenkel theorem for self-shrinkers given in the appendix of [7], we have the following:

Lemma 2.3 *Let* M_t *be the flow defined in Theorem 2.2. Then* $\lim_{t\to\infty} M_t = \varnothing$.

Lastly, note that by switching our choice of normal and using minimality (in the Gaussian metric) of the original surface we may shrinker mean convex perturb either inward or outward (for a 2–sided surface, of course, the distinction is somewhat arbitrary) to study its topology as observed in [3; 21] — this idea is critical to our argument and we will make our choice of perturbation depending on which case we are considering in the argument below.

3 Proof of Theorem 1.2

Note, by the entropy assumption (in particular that $\Lambda_1 < 2$), that M is embedded, and hence its link is too. Without loss of generality for this section, the link L is connected. Supposing L is not diffeomorphic to S^2, there exists some $R_1 \gg 0$ such that $M \cap S(0,R) := L_R$ is not diffeomorphic to S^2 for $R > R_1$ and that, by the asymptotically conical assumption, $L_R \simeq L_{R_1}$ for all $R > R_1$. By the classification of surfaces (note that L_R is orientable, which can be seen by projecting the normal of $M \cap S(0,R)$ onto $TS(0,R)$, giving a section of the normal bundle of L_R, which has no kernel because the sphere intersects M transversely) L_R is topologically a connect sum of tori which bounds domains (not necessarily handlebodies) $K_R, K_R^c \subset S(0,R)$.

Fixing a choice of $R > R_1$, consider a standard generator $\gamma \subset L_R$ of $H_1(L_R)$; that is, writing L_R as a connect sum of tori, γ is homotopic to one of the two generators of a single one of the tori. Note that γ is also homotopically nontrivial in L_R. We consider two cases: either γ is homotopically trivial in M or not. Without loss of generality, γ is embedded and smooth as well.

3.1 Case 1: γ is nullhomotopic in M

Since γ is homotopically trivial in M, it bounds a disc D in M; suppose $D \subset M \cap B(0,R_2)$. Hence, for any embedded curve $\gamma' \subset L_{R'}$ isotopic (in M) to γ for $R' > R_2$, γ' is nullhomotopic in $M \cap B(0,R')$ and hence bounds an embedded disc $D' \subset M \cap B(0,R')$ by Dehn's lemma (see [18]) — Dehn's lemma gives a PL embedded disc, but when γ' is smooth note that D' can be taken to be smooth as well by the Whitney approximation theorem [27]. The idea is, morally, such discs serve as barriers in a sense to keep the flow of (a perturbation of) M "propped" up. The following indicates which domain M bounds to perturb and flow into:

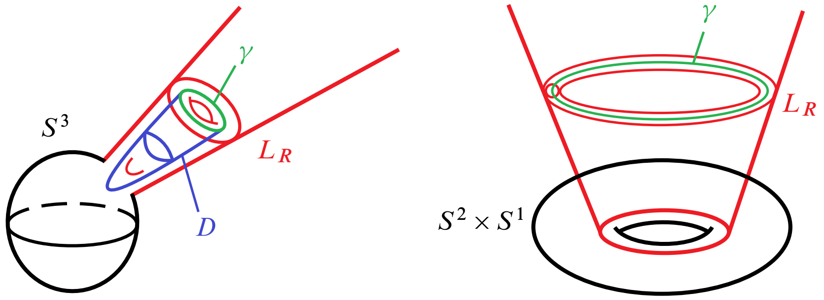


Figure 1: This figure illustrates toy examples one might imagine for Case 1 (left) and Case 2 (right), considered when L_R is a standardly embedded torus. In the first case, $\gamma \in \partial(M \cap B(0,R))$ is nullhomotopic in M, and hence bounds an embedded disc $D \subset M \cap B(0,R)$ by Dehn's lemma.

Lemma 3.1 *In one of K_R or K_R^c the curve γ is not homotopically trivial.*

Proof It seems one could probably use the Mayer–Vietoris sequence and Hurewicz isomorphism here as in the proof of Theorem 1 in [21], but we present a more geometric argument. Suppose, for the sake of contradiction, it were homotopically trivial in both simultaneously. By Dehn's lemma, γ bounds PL embedded (of course, in fact, smooth) discs $D_1 \subset K_R$ and $D_2 \subset K_R^c$ which intersect along γ, giving an embedded $S^2 \subset S^3$. Since γ is smooth their union gives a PL embedded S^2, and so by Alexander's theorem (see [18]) $D_1 \cup D_2$ then bounds a (PL embedded) 3–ball $B \subset S^3$. From its construction, L_R intersects B in one boundary component, namely γ. In particular, γ is homologically trivial in L_R, giving a contradiction. □

Of course, the lemma applies equally for γ' homotopic to γ in $L_{R'}$ for $R' > R > R_1$. After potentially relabeling, γ is homotopically nontrivial in K_R. In this case, consider a shrinker mean convex perturbation of M, as constructed in [3], which descends (ie intersecting with $S(0,R)$) to a perturbation of L_R into K_R, and consider the corresponding renormalized flow M_t (recalling that we can choose which direction to flow into, as discussed in the preliminaries). This flow likewise descends to a flow $(L_R)_t$ of L_R. Note though that, although M_t is an RMCF, $(L_R)_t$ isn't necessarily (to the author's knowledge) an easily described flow in $S(0,R)$, but we will still find it profitable to consider.

By Lemma 2.3, M_t must leave every bounded set in some finite time, and hence $(L_R)_t$ must eventually become empty. Denote this time by T. We will play it off against the next two lemmas, the first essentially that the disc we find by Dehn's lemma persists:

Lemma 3.2 *For any $\tau > 0$, one can pick $\bar{R}$ sufficiently large that, for $t \in [0, \tau]$, there will be a smoothly embedded curve $\bar{\gamma}_t \subset (L_{\bar{R}})_t$ isotopic to γ which bounds a smoothly embedded disc $\bar{D}_t \subset B(0, \bar{R}) \cap M_t$.*

Proof With the construction of the flow by surgery flows given in Theorem 2.2 in mind, we first show, for exposition, that this holds for an approximating surgery flow S_t of M. Since the $L_{R'}$ are all diffeomorphic for R' large, there is clearly an initial choice of curve $\bar{\gamma}$ which is isotopic to γ. Up to the first surgery time and in between surgery times when the surgery flow is smooth, this curve is just given by restricting the motion of $(L_{\bar{R}})_t$ along $\bar{\gamma}$ because, for $\bar{R}$ large enough, $(L_{\bar{R}})_t$ will be a graph over $(L_{\bar{R}})_0$ by pseudolocality and Theorem 2.2(3) on $[0, \tau]$ (in particular, $\bar{\gamma}_t$ can be taken to be embedded and smoothly vary) for times in $[0, \tau]$. Concerning the bounded disc for smooth times, the flow is an isotopy which restricts to an isotopy of the disc (modding out tangential components of the flow). Now consider a surgery time $t_s < \tau$; we must check that after surgery $\bar{\gamma}_{t_s}$ still bounds a disc. Again by pseudolocality for all surgery necks N, $N \cap S(0, \bar{R})$ is empty, and similarly all N must be within $B(0, \bar{R})$. Considering a cap C in the surgery procedure, since it is topologically a ball, the intersection of D_{t_s} with $\partial C \simeq S^2$ is a disjoint union of closed curves which bound discs by the Schoenflies theorem (without loss of generality D_t enters all caps transversely). Surgering along these discs gives a union of S^2 along with a new disc whose boundary is $\bar{\gamma}_{t_s}$ (essentially filling in the part of the disc between the end and the "closest" surgery necks). In particular, $\bar{\gamma}_{t_s}$ continues to bound a disc after surgery. Note that it's conceivable at this stage that the discarded copies of S^2 bound nontrivial topology of M_t, so the D_t do not necessarily form an isotopic family of discs, a priori.

Now we discuss how to show the curve $\bar{\gamma}$ from the previous paragraph always bounds a disc in the limiting flow. What one might first wish for is to take a limit (by compactness) of the discs as the surgery parameters degenerate, but if the limiting disc enters a singular region of the flow it could potentially complicate things, so it's best if the disc is taken to avoid it completely. There is also the matter of boundedness along this sequence of discs needed to apply a compactness theorem, which suggests it's best, in terms of the disc, to work only within the context of the level set flow.

To begin, we consider high curvature points we might encounter as we travel sufficiently deep within a high curvature region (loosely speaking) from a low curvature region, as in our situation of a disc starting from an end (where $\bar{\gamma}$ is) approaching a singularity in the interior of $B(0, \bar{R})$. At points where, say, $H \sim H_{\text{can}}$, referring to parameters in the canonical neighborhood theorem (see [17], noting that here we suppress some notation), one can find nearby "neck-like" points (see Proposition 3.2 in [17]) in any corresponding ancient model that could appear irrespective of surgery parameters. Intuitively a surgery flow near such a point is modeled locally by a neck or a cap bordered by a neck facing towards the low curvature region. With the Hausdorff convergence in mind then, one can use Arzelà–Ascoli to pass to the limit on these bounded curvature regions for the surgery flows to see that the level set flow always has necks where $H \sim H_{\mathrm{can}}$ as one approaches a singular region of the level set flow from a low curvature one (of course, these are smooth points as well). These necks on the level set flow give points to surger the disc as we did for the surgery flows; in this case, we perform the surgery on the disc whenever one of its points is in a region of the level set flow where, say, $H = 10H_{\mathrm{can}}$. Note that between these times the disc varies continuously since it is within a region of the level set flow of bounded curvature, and that the disc can be taken to be smooth at all times since it is surgered on along cross sections of necks of bounded curvature. □

Without using Dehn's lemma (and, in particular, with repeating the argument in higher dimensions in mind), it seems the intersection of the disc with the boundary of the cap could be much more complicated, although naively it seems likely that γ would remain homotopically trivial. We will pit this lemma against the definition of the time T using the following lemma, which says the discs must "leave" the end no matter what:

Lemma 3.3 *With $\overline{\gamma}$ and τ as in Lemma 3.2, after potentially taking R larger, there is an $\epsilon > 0$ such that, in $B(0, R-\epsilon)^c \cap M_t$, the curve $\overline{\gamma}_t$ isn't nullhomotopic and so doesn't bound a disc $B(0, R-\epsilon)^c \cap M_t$. In particular, the disc D_t from the previous lemma satisfies that $D_t \cap S(0, R-\epsilon)$ is nonempty on $[0, \tau]$.*

Proof Denote by K the region M bounds which includes K_R. Note then that, for R large enough, $B(0, R-\epsilon)^c \cap K \simeq K_R \times [R-\epsilon, \infty)$, and in particular $\overline{\gamma}$ is homotopically nontrivial in this domain since it is homotopically nontrivial in K_R. If, for some time $t \in [0, \tau]$, $\overline{\gamma}_t$ is nullhomotopic in $B(0, R-\epsilon)^c \cap M_t$, then in particular $\overline{\gamma}_t$ bounds (the image of) a disc in $B(0, R-\epsilon)^c \cap K_t$. By the set monotonicity of the flow, ie that $K_t \subset K$, we get in fact that $\overline{\gamma}_t$ and hence $\overline{\gamma}$ are nullhomotopic in $B(0, R-\epsilon)^c \cap K$, giving a contradiction. □

Applying the above lemmas with $\tau = T + 1$, we see that we arrive at a contradiction. Considering a time $t \in (T, T+1)$ and the disc D_t given from Lemma 3.2, the disc, by Lemma 3.3, must have nonempty intersection with $S(0, R-\epsilon)$. On the other hand, it cannot pass through $S(0, R)$ because $(L_R)_t = \varnothing$ for $t > T$. This completes the argument in this case.

3.2 Case 2: γ is homotopically nontrivial in M

This case is easier in a sense, because we may directly apply the deep ideas of White [39], in particular Theorems 1.1 and 5.2 therein. Specialized to our setting, a consequence is that if K is a smooth mean convex set (and compact as stated in Theorem 1.1, but this can also apply in the noncompact case as long as singularities occur only in a bounded ball for the time it is applied, by Theorem 5.2) in a Riemannian manifold N of dimension 4 (in particular, less than 7), then if a curve in K^c is initially homotopically nontrivial and later becomes contractible in $(K^c)_t$, a singularity of the form $S^1 \times S^2$ must have occurred, contradicting the entropy bound. Here we will consider N to be a subset of $\mathbb{R}^4$ (possibly all of $\mathbb{R}^4$, depending on which case we are in below) endowed with the Gaussian metric, so the flow constructed is more precisely a mean convex foliation. However, the flow is monotone, satisfies the Brakke regularity theorem and the singular set dimension results of White [37], and all the singularities are modeled on round cylinders so the results of the paper apply — this is essentially the upshot of Hershkovits and White [21] (although they phrase things entirely in terms of the RMCF), where they study the interplay of entropy and topology for compact self-shrinkers.

There are two possible cases for γ: that γ is homotopically nontrivial in one of the components K or K^c of $\mathbb{R}^4$ it bounds, or not. First suppose that γ is homotopically nontrivial in (at least) one of K or K^c, say K^c to align with White's terminology. Consider then a nontrivial curve γ in K^c. Since γ is contractible in $\mathbb{R}^4$, the corresponding homotopy gives that it bounds a (continuous image of, perhaps not embedded) disc D — note that this disc must intersect K. Perturbing and flowing into K by Lemma 2.3, eventually we must have $D \subset K^c_t$, say, by T, implying by this time that γ' is nullhomotopic in K^c_t. By pseudolocality [6] there is an $R \gg 0$ such that, near $S(0,R)$, M_t is a smooth flow which intersects the sphere transversely, so defining $N = B(0,R)$, Section 5 of White [39] implies a singularity modeled on $S^1 \times \mathbb{R}^2$ formed, contradicting the entropy bound.

Now we consider the possibility that γ is homotopically trivial in both K and K^c; this naively seems to be a more exotic case than above, but we are unsure it can be ruled out a priori by purely topological reasoning. Then $\gamma \in M$ bounds a disc in both K and K^c. Picking essentially arbitrarily (only to align with White's notation), we define $\widetilde{N}$ to be the union of K and $K^c \cap M \times [0,\epsilon)\nu$ (ie a collar of M), where ν is the normal pointing away from K and $\epsilon > 0$ is some number small enough that the ϵ level set of the collar is also embedded in $\mathbb{R}^4$. Note that this collar region retracts onto M; the utility of this is that now γ is a homotopically nontrivial curve in $K^c \cap \widetilde{N} \subset \widetilde{N} \subsetneq \mathbb{R}^4$. Consider, as in the previous paragraph, a disc $D \subset K$ bounding γ and flow out of K^c into K (that the disc can be taken to be contained in a single component, and hence in N, is why we split this up into cases). As above, Lemma 2.3 (this still applies since the flow of M_t is the same considered in $\mathbb{R}^4$ or $\widetilde{N}$) gives that eventually $D \subset K_t^c \cap \widetilde{N}$ — call this time T. Let $R \gg 0$ be large enough that M_t intersects $S(0, R)$ only transversely and as a smooth flow; again such an R exists by pseudolocality. Defining $N = \widetilde{N} \cap B(0, R)$ and noting that γ is still homologically nontrivial in $K^c \cap \widetilde{N} \cap B(0, R)$, Section 5 of [39] gives that a singularity modeled on $S^1 \times \mathbb{R}^2$ must have formed (in fact, by time T), giving a contradiction.

4 Proof of Theorem 1.1

By the Frenkel property for self-shrinkers, M must be connected. By Theorem 1.2 there exists R sufficiently large that $M \cap B(0, R)$ is diffeomorphic to a connected 2–sided hypersurface N^3 whose boundary consists of a number of 2–spheres along each of which an end homeomorphic to $S^2 \times \mathbb{R}_+$ is attached, where by ends here we mean, for an appropriate choice of R, disjoint connected components of $M \setminus B(0, R)$ which are diffeomorphic to half cylinders over distinct (for distinct ends) connected components of the link; such an R exists since M is asymptotically conical, and the convergence is multiplicity one. The point is to confirm that N is simply connected. Then, by capping off each component of N (considering N as an intrinsically defined manifold, as a hypersurface in $\mathbb{R}^4$ it seems some ends could be "parallel", which would preclude doing this at least in an embedded way) with a 3–ball, we obtain a closed connected simply connected 3–manifold $\widetilde{N}$ which, by the resolution of the 3D Poincaré conjecture, is diffeomorphic to S^3. If there is a homotopically nontrivial curve on N and hence M, by the Seifert–Van Kampen theorem and Theorem 1.2 we can proceed directly as we do in the second case of the proof above using [39], giving the first part of Theorem 1.1. Note that, with surgery for compact manifolds in mind, one should be able to argue directly with a bit more work that $\widetilde{N}$ is diffeomorphic to either S^3 or a connect sum of $S^2 \times S^1$, the latter of which could subsequently be ruled out, avoiding the use of the Poincaré conjecture — this seems to be naturally a more robust line of reasoning for considering higher-dimensional versions of our statement.

When the number of ends is equal to one, M is diffeomorphic to $\mathbb{R}^3$ as a consequence of Alexander's theorem, as noted in [4]. Now suppose that $\lambda(M) < \Lambda_2$ (we will discuss the case of equality afterwards) and M had (at least) two ends, labeled E_1 and E_2. Fixing an R in our definition of end given in the paragraph above, consider a curve $\gamma : \mathbb{R} \to M$

such that for s sufficiently negative $\gamma(s) \in E_1$ and for s sufficiently positive $\gamma(s) \in E_2$. With this in mind, intersect M with an embedded hypersurface $P \simeq \mathbb{R}^3$ such that

(i) E_1 lies on one side of P and E_2 lies on the other side,

(ii) P intersects M transversely, and

(iii) $P \cap M$ is compact.

This is always possible by the asymptotically conical assumption. Denote by $P \cap M$ the surface S; note that S is closed since its compact and M is boundaryless. Similarly, denote the bounded portion of P that S bounds by K_S. By perturbing and flowing M so that, restricted to N, the flow is into K_S (using (ii)) we see, as above, by Lemma 2.3 that S_t is eventually empty. Because of this, one may argue that a singularity of M_t must occur which disconnects E_1 from E_2 along γ; note that by using large spheres as barriers far along the ends toward spatial infinity (or, alternatively, pseudolocality), for any given finite time there will be points originating from E_1 and E_2 on one side of P and the other, respectively. In other words, one end can't flow from one side of P to the other in finite time, so a singularity which disconnects M must indeed occur. Clearly such a singularity must be modeled on $S^2 \times \mathbb{R}$, which has entropy Λ_2, contradicting $\lambda(M) < \Lambda_2$. In the case $\lambda(M) = \Lambda_2$, we note that the perturbation of Bernstein and Wang we used strictly decreases entropy placing us in the case of strict inequality.

References

[1] **B Andrews**, *Noncollapsing in mean-convex mean curvature flow*, Geom. Topol. 16 (2012) 1413–1418 MR Zbl

[2] **J Bernstein**, **L Wang**, *A sharp lower bound for the entropy of closed hypersurfaces up to dimension six*, Invent. Math. 206 (2016) 601–627 MR Zbl

[3] **J Bernstein**, **L Wang**, *A topological property of asymptotically conical self-shrinkers of small entropy*, Duke Math. J. 166 (2017) 403–435 MR Zbl

[4] **J Bernstein**, **L Wang**, *Topology of closed hypersurfaces of small entropy*, Geom. Topol. 22 (2018) 1109–1141 MR Zbl

[5] **S Brendle**, **G Huisken**, *Mean curvature flow with surgery of mean convex surfaces in* $\mathbb{R}^3$, Invent. Math. 203 (2016) 615–654 MR Zbl

[6] **B-L Chen**, **L Yin**, *Uniqueness and pseudolocality theorems of the mean curvature flow*, Comm. Anal. Geom. 15 (2007) 435–490 MR Zbl

[7] **O Chodosh**, **K Choi**, **C Mantoulidis**, **F Schulze**, *Mean curvature flow with generic initial data*, preprint (2020) arXiv 2003.14344

[8] **O Chodosh**, **K Choi**, **C Mantoulidis**, **F Schulze**, *Mean curvature flow with generic low-entropy initial data*, preprint (2021) arXiv 2102.11978

[9] **T H Colding**, **T Ilmanen**, **W P Minicozzi, II**, **B White**, *The round sphere minimizes entropy among closed self-shrinkers*, J. Differential Geom. 95 (2013) 53–69 MR Zbl

[10] **T H Colding**, **W P Minicozzi, II**, *Generic mean curvature flow, I: Generic singularities*, Ann. of Math. 175 (2012) 755–833 MR Zbl

[11] **T H Colding**, **W P Minicozzi, II**, *Smooth compactness of self-shrinkers*, Comment. Math. Helv. 87 (2012) 463–475 MR Zbl

[12] **J M Daniels-Holgate**, *Approximation of mean curvature flow with generic singularities by smooth flows with surgery*, Adv. Math. 410 (2022) art. id. 108715 MR Zbl

[13] **Q Ding**, **Y L Xin**, *Volume growth, eigenvalue and compactness for self-shrinkers*, Asian J. Math. 17 (2013) 443–456 MR Zbl

[14] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geometry 17 (1982) 357–453 MR Zbl

[15] **R Haslhofer**, **D Ketover**, *Minimal* 2*–spheres in* 3*–spheres*, Duke Math. J. 168 (2019) 1929–1975 MR Zbl

[16] **R Haslhofer**, **B Kleiner**, *Mean curvature flow of mean convex hypersurfaces*, Comm. Pure Appl. Math. 70 (2017) 511–546 MR Zbl

[17] **R Haslhofer**, **B Kleiner**, *Mean curvature flow with surgery*, Duke Math. J. 166 (2017) 1591–1626 MR Zbl

[18] **A Hatcher**, *Notes on basic* 3*–manifold topology*, unpublished manuscript (2007) Available at `https://pi.math.cornell.edu/~hatcher/3M/3Mdownloads.html`

[19] **J Head**, *The surgery and level-set approaches to mean curvature flow*, PhD thesis, Max Planck Institute for Gravitational Physics (2011) Available at `https://d-nb.info/1025939417/34`

[20] **J Head**, *On the mean curvature evolution of two-convex hypersurfaces*, J. Differential Geom. 94 (2013) 241–266 MR Zbl

[21] **O Hershkovits**, **B White**, *Sharp entropy bounds for self-shrinkers in mean curvature flow*, Geom. Topol. 23 (2019) 1611–1619 MR Zbl

[22] **G Huisken**, *Asymptotic behavior for singularities of the mean curvature flow*, J. Differential Geom. 31 (1990) 285–299 MR Zbl

[23] **G Huisken**, **C Sinestrari**, *Mean curvature flow with surgeries of two-convex hypersurfaces*, Invent. Math. 175 (2009) 137–221 MR Zbl

[24] **T Ilmanen**, **B White**, *Sharp lower bounds on density for area-minimizing cones*, Camb. J. Math. 3 (2015) 1–18 MR Zbl

[25] **N Kapouleas**, **S J Kleene**, **N M Møller**, *Mean curvature self-shrinkers of high genus: non-compact examples*, J. Reine Angew. Math. 739 (2018) 1–39 MR Zbl

[26] **J Lauer**, *Convergence of mean curvature flows with surgery*, Comm. Anal. Geom. 21 (2013) 355–363 MR Zbl

[27] **J M Lee**, *Introduction to smooth manifolds*, 2nd edition, Graduate Texts in Math. 218, Springer (2013) MR Zbl

[28] **A Mramor**, *An unknottedness result for noncompact self shrinkers*, preprint (2020) arXiv 2005.01688

[29] **A Mramor**, **S Wang**, *Low entropy and the mean curvature flow with surgery*, Calc. Var. Partial Differential Equations 60 (2021) art. id. 96 MR Zbl

[30] **G Perelman**, *The entropy formula for the Ricci flow and its geometric applications*, preprint (2002) Zbl arXiv math/0211159

[31] **G Perelman**, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, preprint (2003) Zbl arXiv math/0307245

[32] **G Perelman**, *Ricci flow with surgery on three-manifolds*, preprint (2003) Zbl arXiv math/0303109

[33] **S Smale**, *Generalized Poincaré's conjecture in dimensions greater than four*, Ann. of Math. 74 (1961) 391–406 MR Zbl

[34] **A Stone**, *A density function and the structure of singularities of the mean curvature flow*, Calc. Var. Partial Differential Equations 2 (1994) 443–480 MR Zbl

[35] **L Wang**, *Asymptotic structure of self-shrinkers*, preprint (2016) arXiv 1610.04904

[36] **L Wang**, *Uniqueness of self-similar shrinkers with asymptotically cylindrical ends*, J. Reine Angew. Math. 715 (2016) 207–230 MR Zbl

[37] **B White**, *Stratification of minimal surfaces, mean curvature flows, and harmonic maps*, J. Reine Angew. Math. 488 (1997) 1–35 MR Zbl

[38] **B White**, *The size of the singular set in mean curvature flow of mean-convex sets*, J. Amer. Math. Soc. 13 (2000) 665–695 MR Zbl

[39] **B White**, *Topological change in mean convex mean curvature flow*, Invent. Math. 191 (2013) 501–525 MR Zbl

Department of Mathematics, Johns Hopkins University
Baltimore, MD, United States

amramor1@jhu.edu

Proposed: Tobias H Colding
Seconded: Gang Tian, Bruce Kleiner

Received: 30 June 2021
Revised: 22 March 2022

Geometry & Topology **27**:9 (2023) 3733–3800
DOI: 10.2140/gt.2023.27.3733
Published: 5 December 2023

The Gromov–Hausdorff distance between spheres

Sunhyuk Lim
Facundo Mémoli
Zane Smith

We provide general upper and lower bounds for the Gromov–Hausdorff distance $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n)$ between spheres $\mathbb{S}^m$ and $\mathbb{S}^n$ (endowed with the round metric) for $0 \le m < n \le \infty$. Some of these lower bounds are based on certain topological ideas related to the Borsuk–Ulam theorem. Via explicit constructions of (optimal) correspondences, we prove that our lower bounds are tight in the cases of $d_{\mathrm{GH}}(\mathbb{S}^0, \mathbb{S}^n)$, $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^\infty)$, $d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^2)$, $d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^3)$ and $d_{\mathrm{GH}}(\mathbb{S}^2, \mathbb{S}^3)$. We also formulate a number of open questions.

53C23

1 Introduction

Despite being widely used in Riemannian geometry — see, for example, Burago, Burago and Ivanov [4] and Petersen [30] — very little is known in terms of the *exact* value of the Gromov–Hausdorff distance between two given spaces. In a closely related vein, Gromov [16, page 141] poses the question of computing/estimating the value of the *box distance* $\square_1(\mathbb{S}^m, \mathbb{S}^n)$ (a close relative of d_{GH}) between spheres (viewed as metric measure spaces). In [14], Funano provides asymptotic bounds for this distance via an idea due to Colding (see the discussion preceding Proposition 1.2).

The Gromov–Hausdorff distance is also a natural choice for expressing the stability of invariants in applied algebraic topology — see Carlsson and Mémoli [5; 6; 7] — and has also been invoked in applications related to shape matching — see Bronstein, Bronstein and Kimmel [3] and Mémoli and Sapiro [25; 27] — as a notion of dissimilarity between shapes.

We consider the problem of estimating the Gromov–Hausdorff distance $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n)$ between spheres (endowed with their round/geodesic distance). In particular we show that in some cases, topological ideas related to the Borsuk–Ulam theorem yield lower bounds which turn out to be tight.

1.1 Basic definitions

The Gromov–Hausdorff distance — see Edwards [12] and Gromov [16] — between two bounded metric spaces (X, d_X) and (Y, d_Y) is defined as

$$d_{\mathrm{GH}}(X, Y) := \inf d_{\mathrm{H}}(f(X), g(Y)),$$

where d_{H} denotes the Hausdorff distance between subsets of the ambient space Z and the infimum is taken over all isometric embeddings f and g of X and Y, respectively, into Z, and over all metric spaces Z. We will henceforth denote by $\mathcal{M}_b$ the collection of all bounded metric spaces.

It is known that d_{GH} defines a metric on compact metric spaces up to isometry [16]. A standard reference is [4]. A useful property is that whenever (X, d_X) is a compact metric space and, for some $\delta > 0$, a subset $A \subseteq X$ is a δ–net for X, then $d_{\mathrm{GH}}\big((X, d_X), (A, d_X|_{A\times A})\big) \le \delta$.

Given two sets X and Y, a correspondence between them is any relation $R \subseteq X \times Y$ such that $\pi_X(R) = X$ and $\pi_Y(R) = Y$ where $\pi_X\colon X \times Y \to X$ and $\pi_Y\colon X \times Y \to Y$

are the canonical projections. Given two bounded metric spaces (X, d_X) and (Y, d_Y), and any nonempty relation $R \subseteq X \times Y$, its distortion is defined as

$$\operatorname{dis}(R) := \sup_{(x,y),(x',y') \in R} |d_X(x,x') - d_Y(y,y')|.$$

Remark 1.1 In particular, the graph of any map $\psi\colon X \to Y$ is a relation $\operatorname{graph}(\psi)$ between X and Y and this relation is a correspondence whenever ψ is surjective. The distortion of the relation induced by ψ will be denoted by $\operatorname{dis}(\psi)$.

A theorem of Kalton and Ostrovskii [18] proves that the Gromov–Hausdorff distance between any two bounded metric spaces (X, d_Y) and (Y, d_Y) is equal to

$$d_{\mathrm{GH}}(X,Y) := \tfrac{1}{2} \inf_R \operatorname{dis}(R), \tag{1}$$

where R ranges over all correspondences between X and Y. It was also observed in [18] that

$$d_{\mathrm{GH}}(X,Y) = \tfrac{1}{2} \inf_{\varphi,\psi} \max\{\operatorname{dis}(\varphi), \operatorname{dis}(\psi), \operatorname{codis}(\varphi,\psi)\}, \tag{2}$$

where $\varphi\colon X \to Y$ and $\psi\colon Y \to X$ are any (not necessarily continuous) maps, and

$$\operatorname{codis}(\varphi,\psi) := \sup_{x \in X,\, y \in Y} |d_X(x, \psi(y)) - d_Y(\varphi(x), y)|$$

is the *codistortion* of the pair (φ, ψ).

Known results on $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n)$ We will find it useful to refer to the infinite matrix $\mathfrak{g}$ such that for $m, n \in \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$,

$$\mathfrak{g}_{m,n} := d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n);$$

see Figure 2.

The following lower bound for $\mathfrak{g}_{m,n}$, obtained via simple estimates for covering and packing numbers based on volumes of balls, is in the same spirit as a result by Colding [10, Lemma 5.10].[1] By $v_n(\rho)$ we denote the *normalized volume* of an open ball of radius $\rho \in (0, \pi]$ on $\mathbb{S}^n$ (so that the entire sphere has volume 1). Colding's approach yields:

Proposition 1.2 *For all integers* $0 < m < n$,

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \geq \mu_{m,n} := \tfrac{1}{2} \sup_{\rho \in (0,\pi]} \bigl(v_n^{-1} \circ v_m\bigl(\tfrac{1}{2}\rho\bigr) - \rho\bigr).$$

[1] Funano used a similar idea in [14] to estimate Gromov's box distance between metric measure space representations of spheres.

We relegate the proof of this proposition to Section 3.

Example 1.3 (lower bound for $\mathfrak{g}_{1,2}$ via Colding's idea) In this case, $m=1$ and $n=2$, the lower bound provided by Proposition 1.2 above is $\sup_{\rho\in(0,\pi]}(\arccos(1-\rho/\pi)-\rho)$, which is approximately equal to and bounded below by 0.1605. Thus, $\mathfrak{g}_{1,2}\geq 0.0802$. See Remark 1.9 for a comparison with a new lower bound which also arises from covering/packing arguments via the Lyusternik–Schnirelmann theorem.

In contrast, in this paper, via techniques which include both certain topological ideas leading to lower bounds and the precise construction of correspondences with matching (and hence optimal) distortion, we prove results which imply (see Proposition 1.16 below) that, in particular, $\mathfrak{g}_{1,2}=\frac{\pi}{3}\simeq 1.0472$, which is about 13 times larger than the value obtained by the method above. In [26, Example 5.3] the lower bound $\mathfrak{g}_{1,2}\geq\frac{\pi}{12}$ was obtained via a calculation involving Gromov's *curvature sets* $\boldsymbol{K}_3(\mathbb{S}^1)$ and $\boldsymbol{K}_3(\mathbb{S}^2)$. Finally, via considerations based on Katz's precise calculation [19] of the filling radius of spheres — see Lim, Mémoli and Okutan [21, Corollary 9.3] — yields that $\mathfrak{g}_{1,n}\geq\frac{\pi}{6}$ for all $n\geq 2$ as well as other lower bounds for $\mathfrak{g}_{m,n}$ for general $m<n$ which are not tight. In a related vein, in [17] Ji and Tuzhilin determine the precise value of $d_{\mathrm{GH}}([0,\lambda],\mathbb{S}^1)$ between an interval of length $\lambda>0$ and the circle (with geodesic distance).

1.2 Overview of our results

The diameter of a bounded metric space (X,d_X) is the number

$$\operatorname{diam}(X):=\sup_{x,x'\in X} d_X(x,x').$$

For $m\in\overline{\mathbb{N}}$ we view the m–dimensional sphere,

$$\mathbb{S}^m:=\{(x_1,\dots,x_{m+1})\in\mathbb{R}^{m+1}\mid x_1^2+\cdots+x_{m+1}^2=1\},$$

as a metric space by endowing it with the geodesic distance: for any two points $x,x'\in\mathbb{S}^m$,

$$d_{\mathbb{S}^m}(x,x'):=\arccos(\langle x,x'\rangle)=2\arcsin\bigl(\tfrac{1}{2}d_{\mathrm{E}}(x,x')\bigr),$$

where d_{E} denotes the canonical Euclidean metric inherited from $\mathbb{R}^{m+1}$.

Note that for $m=0$ this definition yields that $\mathbb{S}^0$ consists of two points at distance π, and that $\mathbb{S}^\infty$ is the unit sphere in ℓ^2 with distance given in the expression above.

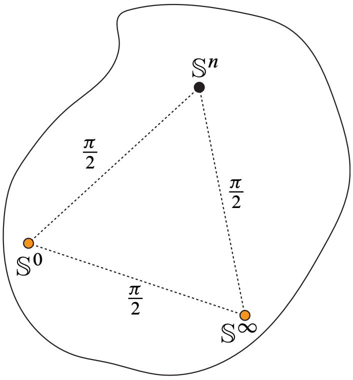


Figure 1: Propositions 1.5 and 1.6 encode the peculiar fact that all triangles in $(\mathcal{M}_b, d_{\mathrm{GH}})$ with vertices $\mathbb{S}^0$, $\mathbb{S}^\infty$, and $\mathbb{S}^n$ (for $0 < n < \infty$) are equilateral.

Remark 1.4 First recall [4, Chapter 7] that, for any two bounded metric spaces X and Y, one always has $d_{\mathrm{GH}}(X,Y) \le \frac{1}{2}\max\{\mathrm{diam}(X), \mathrm{diam}(Y)\}$. This means that

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \le \tfrac{\pi}{2} \quad \text{for all } 0 \le m \le n \le \infty. \tag{3}$$

We first prove the following two propositions, which establish that the above upper bound is tight in certain extremal cases:

Proposition 1.5 (distance to $\mathbb{S}^0$; Chowdhury and Mémoli [9, Proposition 1.2]) *For any integer $n \ge 1$,*

$$d_{\mathrm{GH}}(\mathbb{S}^0, \mathbb{S}^n) = \tfrac{\pi}{2}.$$

Proposition 1.6 (distance to $\mathbb{S}^\infty$) *For any integer $m \ge 0$,*

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^\infty) = \tfrac{\pi}{2}.$$

Proposition 1.5 can be proved as follows: any correspondence between $\mathbb{S}^0$ and $\mathbb{S}^n$ induces a closed cover of $\mathbb{S}^n$ by two sets; then, by the Lyusternik–Schnirelmann theorem, one of these blocks must contain two antipodal points. Proposition 1.6 can be proved in a similar manner; see Figure 1.

Remark 1.7 When taken together, Remark 1.4 and Propositions 1.5 and 1.6 above might suggest that the Gromov–Hausdorff distance between *any* two spheres of different dimension is $\frac{\pi}{2}$. In fact, this is true for the following *continuous* version of d_{GH}:

$$d_{\mathrm{GH}}^{\mathrm{cont}}(X,Y) := \tfrac{1}{2} \inf_{\varphi',\psi'} \max\{\mathrm{dis}(\varphi'), \mathrm{dis}(\psi'), \mathrm{codis}(\varphi',\psi')\},$$

where $\varphi'\colon X \to Y$ and $\psi'\colon Y \to X'$ are *continuous* maps.

Indeed, suppose that $n > m \geq 1$. Then, by the Borsuk–Ulam theorem — see Munkholm [28, Theorem 1] or Matoušek [24, page 29] — for any continuous $\varphi' \colon \mathbb{S}^n \to \mathbb{S}^m$, there must be two antipodal points with the same image under φ'; that is, there is an $x \in \mathbb{S}^n$ such that $\varphi'(x) = \varphi'(-x)$. This implies that $\operatorname{dis}(\varphi') = \pi$, and consequently $d_{\mathrm{GH}}^{\mathrm{cont}}(\mathbb{S}^n, \mathbb{S}^m) \geq \frac{\pi}{2}$. The reverse inequality can be obtained by choosing constant maps φ' and ψ' in the above definition; thus implying that

$$d_{\mathrm{GH}}^{\mathrm{cont}}(\mathbb{S}^m, \mathbb{S}^n) = \tfrac{\pi}{2}.$$

In contrast, we prove the following result for the standard Gromov–Hausdorff distance:

Theorem A $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) < \frac{\pi}{2}$ *for all* $0 < m < n < \infty$.

The Borsuk–Ulam theorem implies that, for any positive integers $n > m$ and for any given continuous function $\varphi \colon \mathbb{S}^n \to \mathbb{S}^m$, there exist two antipodal points in the higher dimensional sphere which are mapped to the *same* point in the lower dimensional sphere. This forces the distortion of any such continuous map to be π. In contrast, in order to prove Theorem A, we exhibit, for every pair of positive numbers m and n with $m < n$, a *continuous antipode-preserving surjection* from $\mathbb{S}^m$ to $\mathbb{S}^n$ with distortion *strictly* bounded above by π, which implies the claim since the graph of any such surjection is a correspondence between $\mathbb{S}^m$ and $\mathbb{S}^n$; see Remark 1.1. The proof relies on ideas related to space-filling curves and spherical suspensions.

The standard Borsuk–Ulam theorem is however still useful for obtaining additional information about the Gromov–Hausdorff distance between spheres. Indeed, via Lemma 3.2 and the triangle inequality for d_{GH}, one can prove the following general lower bound:

Proposition 1.8 *For any* $1 \leq m < n < \infty$,

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \geq \nu_{m,n} := \tfrac{\pi}{2} - \operatorname{cov}_{\mathbb{S}^m}(n+1).$$

Above, for any integer $k \geq 1$, and any compact metric space X, $\operatorname{cov}_X(k)$ denotes the k^{th} *covering radius* of X,

$$\operatorname{cov}_X(k) := \inf\{d_{\mathrm{H}}(X, P) \mid P \subset X \text{ such that } |P| \leq k\}. \tag{4}$$

Remark 1.9 Both of the lower bounds, $\mu_{m,n}$ and $\nu_{m,n}$, from Propositions 1.2 and 1.8, respectively, implement covering/packing ideas and as such it is interesting to compare them:

(1) Note that, since $\operatorname{cov}_{\mathbb{S}^1}(3) = \frac{\pi}{3}$, we have $\nu_{1,2} = \frac{\pi}{6}$, which is about 6.5 times larger than $\mu_{1,2} \approx 0.0802$ (see Example 1.3).

(2) Computing $\nu_{m,n}$ in general requires knowledge of the covering radius $\operatorname{cov}_{\mathbb{S}^m}(k)$ of spheres which is currently only known for $k \le m+2$; see Cho [8, Theorem 3.2]. In contrast, computing $\mu_{m,n}$ can be done (in principle) for any $m < n$ given that we have the explicit formula $v_m(\rho) = (\mathbf{Vol}(\mathbb{S}^{m-1})/\mathbf{Vol}(\mathbb{S}^m)) \int_0^\rho (\sin\theta)^{m-1}\, d\theta$, which is valid for every positive integer m and $\rho \in [0,\pi]$; see Gray [15].

(3) The lower bound $\mu_{m,n}$ is more widely applicable than $\nu_{m,n}$, which originates from the Lyusternik–Schnirelman theorem (see below) and the underlying ideas are in principle only applicable when one of the two metric spaces is a sphere.[2] Indeed, see Colding [10] and Furano [14] for estimates of the Gromov–Hausdorff distance between Riemannian manifolds satisfying upper and lower bounds on curvature obtained by combining volume comparison theorems with techniques similar to those used in proving Proposition 1.2.

(4) Through [8, Theorem 3.2] it is known that $\operatorname{cov}_{\mathbb{S}^m}(m+2) = \pi - \arccos(-1/(m+1))$ for $m \ge 1$. Therefore, when $n = m+1$, the lower bound $\nu_{m,m+1}$ given by Proposition 1.8 becomes $\arccos(-1/(m+1)) - \frac{\pi}{2}$ for $m \ge 1$, which tends to zero as m goes to infinity. It is not known whether or not $\mu_{m,m+1}$ has the same behavior.

As an immediate corollary, we obtain the following result, which complements both Proposition 1.6 and Theorem A:

Corollary 1.10 *Given any positive integer m and $\varepsilon > 0$, there exists an integer $n = n(m,\varepsilon) > m$ such that*

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \ge \tfrac{\pi}{2} - \varepsilon.$$

Remark 1.11 For small $\varepsilon > 0$ one can estimate the value of n above as

$$n = n(m,\varepsilon) = O(\varepsilon^{-m}).$$

The results above motivate the following two questions:

Question I *Is it true that, for fixed $m \ge 1$, $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n)$ is nondecreasing for all $n \ge m$?*

[2] This can be ascertained by inspecting the proof of Proposition 1.8 in Section 3.2.

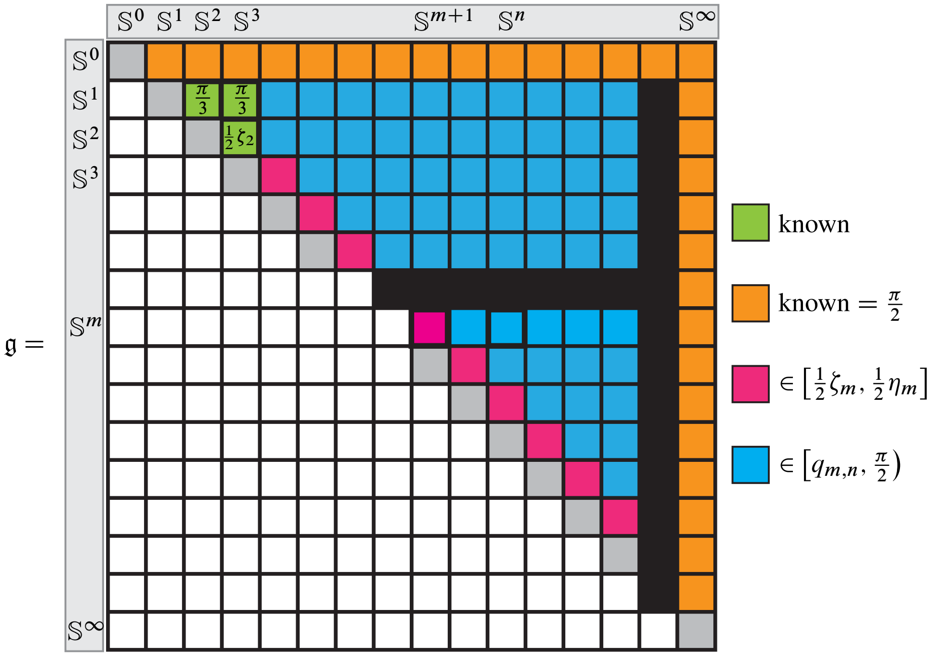

$\mathfrak{g} =$
$\mathbb{S}^0$ $\mathbb{S}^1$ $\mathbb{S}^2$ $\mathbb{S}^3$ $\mathbb{S}^{m+1}$ $\mathbb{S}^n$ $\mathbb{S}^\infty$
$\mathbb{S}^0$ $\mathbb{S}^1$ $\mathbb{S}^2$ $\mathbb{S}^3$ $\mathbb{S}^m$ $\mathbb{S}^\infty$
$\frac{\pi}{3}$
$\frac{\pi}{3}$
$\frac{1}{2}\zeta_2$
known
known $= \frac{\pi}{2}$
$\in \left[\frac{1}{2}\zeta_m, \frac{1}{2}\eta_m\right]$
$\in \left[q_{m,n}, \frac{\pi}{2}\right)$

Figure 2: The matrix $\mathfrak{g}$ such that $\mathfrak{g}_{m,n} := d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n)$. According to Remark 1.4 and Corollary 1.12, all nonzero entries of the matrix $\mathfrak{g}$ are in the range $\left[\frac{\pi}{4}, \frac{\pi}{2}\right]$. In the figure, $\zeta_m = \arccos(-1/(m+1))$ is the edge length of the regular geodesic simplex inscribed in $\mathbb{S}^m$, η_m is the diameter of a face of the regular geodesic simplex in $\mathbb{S}^m$ — see (5) — and $q_{m,n} = \max\left\{\frac{1}{2}\zeta_m, \frac{\pi}{2} - \mathrm{cov}_{\mathbb{S}^m}(n+1)\right\}$.

Question II *Fix $m \geq 1$ and $\varepsilon > 0$. Find (optimal) estimates for*

$$k_m(\varepsilon) := \inf\left\{k \geq 1 \mid d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^{m+k}) \geq \frac{\pi}{2} - \varepsilon\right\}.$$

Via the Lyusternik–Schnirelmann theorem, Proposition 1.8 above depends on the classical Borsuk–Ulam theorem which, in one of its guises [24, Theorem 2.1.1], states that there is no *continuous* antipode-preserving map $g\colon \mathbb{S}^n \to \mathbb{S}^{n-1}$. As a consequence, if $g\colon \mathbb{S}^n \to \mathbb{S}^{n-1}$ is any antipode-preserving map, then g cannot be continuous. A natural question is *how discontinuous* is any such g forced to be. This question was actually tackled in 1981 by Dubins and Schwarz [11], who proved that the *modulus of discontinuity* $\delta(g)$ of any such g needs to be suitably bounded below. These results are instrumental for proving Theorem B below; see Section 5 and Appendix A for details and for a concise proof of the main theorem from [11] (following a strategy outlined by Matoušek in [24]).

For each $m \in \mathbb{N}$ let ζ_m denote the edge length (with respect to the geodesic distance) of a regular $m+1$ simplex inscribed in $\mathbb{S}^m$,

$$\zeta_m := \arccos\left(\frac{-1}{m+1}\right),$$

which is monotonically decreasing in m. For example,

$$\zeta_0 = \pi, \quad \zeta_1 = \tfrac{2\pi}{3}, \quad \zeta_2 = \arccos\left(-\tfrac{1}{3}\right) \approx 0.608\pi, \quad \lim_{m\to\infty} \zeta_m = \tfrac{\pi}{2}.$$

Then we have the following lower bound which will turn out to be optimal in some cases:

Theorem B (lower bound via geodesic simplices) *For all integers* $0 < m < n$,

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \ge \tfrac{1}{2}\zeta_m.$$

Moreover, for any map $\varphi\colon \mathbb{S}^n \to \mathbb{S}^m$, *we have that* $\operatorname{dis}(\varphi) \ge \zeta_m$.

From the above, we have the following general lower bound:

Corollary 1.12 *For all integers* $0 < m < n$, $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \ge \frac{\pi}{4}$.

This corollary of course implies that the sequence of compact metric spaces $(\mathbb{S}^n)_{n\in\overline{\mathbb{N}}}$ is not Cauchy.

Remark 1.13 The lower bound for $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n)$ given by Theorem B coincides with the *filling radius* of $\mathbb{S}^m$; see Katz [19, Theorem 2]. This lower bound is twice the one obtained via the stability of Vietoris–Rips persistent homology [21, Corollary 9.3].

Note that $\operatorname{cov}_{\mathbb{S}^1}(k) \le \pi/k$, which can be seen by considering the vertices of a regular polygon inscribed in $\mathbb{S}^1$ with k sides. Combining this fact with Proposition 1.8, Theorem B, and the fact that $\zeta_1 = \frac{2\pi}{3}$, one obtains the following special claim for the entries in the first row of the matrix $\mathfrak{g}$:

Corollary 1.14 *For all* $n > 1$, $d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^n) \ge \pi \cdot \max\left\{\frac{1}{3}, \frac{1}{2}(n-1)/(n+1)\right\}$.

Remark 1.15 This implies that, whereas $d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^n) \ge \frac{\pi}{3}$ for $n \in \{2,3,4,5\}$, one has the larger lower bound $d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^6) \ge \frac{5\pi}{14} > \frac{\pi}{3}$. Propositions 1.16 and 1.18 below establish that actually $d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^2) = d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^3) = \frac{\pi}{3}$.

Finally, in order to prove that $d_{\mathrm{GH}}(\mathbb{S}^1,\mathbb{S}^2)=\frac{\pi}{3}$, we combine Theorem B with an explicit construction of a correspondence between $\mathbb{S}^1$ and $\mathbb{S}^2$ as follows. Let $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2)$ denote the closed upper hemisphere of $\mathbb{S}^2$. Then the following proposition shows that there exists a correspondence between $\mathbb{S}^1$ and $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2)$ with distortion at most $\frac{2\pi}{3}$. A correspondence between $\mathbb{S}^1$ and $\mathbb{S}^2$ (see Figure 7) with the same distortion is then obtained via a certain *odd* (ie antipode-preserving) extension of the aforementioned correspondence (see Lemma 5.7):

Proposition 1.16 *There exists*

(1) *a correspondence between* $\mathbb{S}^1$ *and* $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2)$, *and*

(2) *a correspondence between* $\mathbb{S}^1$ *and* $\mathbb{S}^2$,

both of which have distortion at most $\frac{2\pi}{3}$. *In particular, together with Theorem B, this implies* $d_{\mathrm{GH}}(\mathbb{S}^1,\mathbb{S}^2)=\frac{\pi}{3}$.

Even though we do not state it explicitly, in a manner similar to Proposition 1.16, all correspondences constructed in Propositions 1.18, 1.19 and 1.20 below also arise from odd extensions of correspondences between the lower dimensional sphere and the upper hemisphere of the larger dimensional sphere (see their respective proofs).

Remark 1.17 Also, by combining the first claim of Proposition 1.16 and Example 1.24(4) below (which is analogous to the claim of Theorem B but tailored to the case of $\mathbb{S}^m$ versus $\boldsymbol{H}_{\geq 0}(\mathbb{S}^m)$), one concludes that $d_{\mathrm{GH}}(\mathbb{S}^1,\boldsymbol{H}_{\geq 0}(\mathbb{S}^2))=\frac{1}{2}\zeta_1=\frac{\pi}{3}$.

Via a construction somewhat reminiscent of the Hopf fibration, we prove that there exists a correspondence between the 3–dimensional sphere and the 1–dimensional sphere with distortion at most $\frac{2\pi}{3}$. By applying suitable rotations in $\mathbb{R}^4$, the proof of the following proposition extends the (a posteriori) optimal correspondence between $\mathbb{S}^1$ and $\mathbb{S}^2$ constructed in the proof of Proposition 1.16 (see Figure 10):

Proposition 1.18 *There exists a correspondence between* $\mathbb{S}^1$ *and* $\mathbb{S}^3$ *with distortion at most* $\frac{2\pi}{3}$. *In particular, together with Theorem B, this implies* $d_{\mathrm{GH}}(\mathbb{S}^1,\mathbb{S}^3)=\frac{\pi}{3}$.

Finally, we were able to compute the exact value of the distance between $\mathbb{S}^2$ and $\mathbb{S}^3$ by producing a correspondence whose distortion matches the one implied by the lower bound in Theorem B. This correspondence is structurally different from the ones constructed in Propositions 1.16 and 1.18 and arises by partitioning $\mathbb{S}^3$ into 32 regions whose diameter is (necessarily) bounded above by ζ_2 and which satisfy suitable pairwise constraints (see Section 2.2):

Proposition 1.19 *There exists a correspondence between $\mathbb{S}^2$ and $\mathbb{S}^3$ with distortion at most ζ_2. In particular, together with Theorem B, this implies $d_{\mathrm{GH}}(\mathbb{S}^2,\mathbb{S}^3)=\frac{1}{2}\zeta_2$.*

Keeping in mind Remark 1.15 and Propositions 1.16 and 1.18, we pose the following:

Question III *Is it true that $d_{\mathrm{GH}}(\mathbb{S}^1,\mathbb{S}^n)=\frac{\pi}{3}$ for $n\in\{4,5\}$?*

Theorem B and Propositions 1.16 and 1.19 lead to formulating the following conjecture:

Conjecture 1 *For all $m\in\mathbb{N}$, $d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^{m+1})=\frac{1}{2}\zeta_m$.*

Note that when $m=1$ and $m=2$, Conjecture 1 reduces to Propositions 1.16 and 1.19, respectively. Moreover, the conjecture would imply that $\lim_{m\to\infty} d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^{m+1})=\frac{\pi}{4}$.

While trying to prove Conjecture 1, we were able to prove the following weaker result via an explicit construction of a certain correspondence generalizing the one constructed in the proof of Proposition 1.16:

Proposition 1.20 *For any positive integer $m>0$, there exists a correspondence between $\mathbb{S}^m$ and $\mathbb{S}^{m+1}$ with distortion at most η_m, where*

$$\eta_m:=\begin{cases}\arccos(-(m+1)/(m+3)) & \text{if } m \text{ is odd},\\ \arccos(-\sqrt{m/(m+4)}) & \text{if } m \text{ is even}.\end{cases}\tag{5}$$

Here, η_m is the diameter of a face of the regular geodesic m–simplex in $\mathbb{S}^m$; see Figure 8 and the discussion in Section 6.2.

This correspondence arises from a partition of $\mathbb{S}^{m+1}$ into $2(m+2)$ regions which are induced by two antipodal regular simplices inscribed in $\mathbb{S}^m$, the equator of $\mathbb{S}^{m+1}$ (see Figure 7 for the case $m=1$, a case in which this correspondence turns out to be optimal).

Corollary 1.21 *For any positive integer $m>0$, $d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^{m+1})\le\frac{1}{2}\eta_m$.*

Remark 1.22 Note that $\eta_m\ge\zeta_m$ for any $m>0$ and $\eta_1=\zeta_1$, so Proposition 1.20 generalizes Proposition 1.16. However, since $1.9106\approx\zeta_2<\eta_2\approx 2.1863$, by Proposition 1.19 we see that Corollary 1.21 is not tight when $m=2$. Also, since $\eta_m<\pi$, Corollary 1.21 gives a quantitative version of the claim in Theorem A when $n=m+1$.

Remark 1.23 Combining Theorem B and Proposition 1.8, we obtain a generalization of the bound given in Corollary 1.14: for all $1\le m<n$,

$$d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^n)\ge\max\left\{\tfrac{1}{2}\zeta_m,\ \tfrac{\pi}{2}-\operatorname{cov}_{\mathbb{S}^m}(n+1)\right\}=:q_{m,n}.\tag{6}$$

Question IV *Formula* (6) *and Remark 1.15 motivate the following question: for $m \geq 1$ large, find the* **rate** *at which the number*[3]

$$n_{\mathrm{diag}}(m) := \max\left\{n > m \;\middle|\; \mathrm{cov}_{\mathbb{S}^m}(n+1) \geq \tfrac{1}{2}\arccos\left(\frac{1}{m+1}\right)\right\}$$

grows with m. The reason for the notation $n_{\mathrm{diag}}(m)$ is that this number provides an estimate for a band around the principal diagonal of the matrix $\mathfrak{g}$ (*see Figure 2*) *inside of which one would hope to prove that*

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) = \tfrac{1}{2}\zeta_m \quad \textit{for all } n \in \{m+1, \dots, n_{\mathrm{diag}}(m)\}.$$

1.3 Additional results and questions

Besides what we have described so far, this paper includes a number of other results about Gromov–Hausdorff distances between spaces closely related to spheres.

1.3.1 Spheres with Euclidean distance Some of the ideas described above (for spheres with geodesic distance) can be easily adapted to provide bounds for the distance between half spheres with geodesic distance, and between spheres with Euclidean distance. However, there is evidence that this phenomenon is subtle and to the best of our knowledge, there is no complete translation between the geodesic and Euclidean cases. This is exemplified by the following.

Let $\mathbb{S}^n_{\mathrm{E}}$ denote the unit sphere with the Euclidean metric d_{E} inherited from $\mathbb{R}^{n+1}$. Then, via Remark 1.17 and Corollary 9.8(2) (which provides a bridge between geodesic distortion and Euclidean distortion via the sine function), we have that

$$d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, \boldsymbol{H}_{\geq 0}(\mathbb{S}^2_{\mathrm{E}})) \leq \sin\bigl(d_{\mathrm{GH}}(\mathbb{S}^1, \boldsymbol{H}_{\geq 0}(\mathbb{S}^2))\bigr) = \tfrac{\sqrt{3}}{2}.$$

Despite this, in Proposition 9.10 we were able to construct a correspondence between these two spaces with distortion *strictly smaller* than $\sqrt{3}$. This suggests that Euclidean analogues of Theorem B may not be direct consequences; see Section 9 for other related results.

This motivates posing the following question:

Question V *Determine $\mathfrak{g}^{\mathrm{E}}_{m,n} := d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}}, \mathbb{S}^n_{\mathrm{E}})$ for all integers $1 \leq m < n$.*

It should however be noted that by Corollary 9.8 we have $\mathfrak{g}^{\mathrm{E}}_{m,n} \leq \sin(\mathfrak{g}_{m,n})$, which renders Proposition 1.20 immediately applicable, yielding $\mathfrak{g}^{\mathrm{E}}_{m,m+1} \leq \sin\bigl(\tfrac{1}{2}\eta_m\bigr)$.

[3] Note that $\zeta_m = \pi - \arccos(1/(m+1))$.

1.3.2 A stronger version of Theorem B By inspecting the proof of Theorem B, we actually have Theorem C which subsumes these results in a much greater degree of generality. Indeed, via this theorem one can obtain the following estimates:

Example 1.24 The following lower bounds hold:

(1) $d_{\mathrm{GH}}([0,\pi],\mathbb{S}^n)\geq\frac{\pi}{3}$ for any $n\geq 2$.

(2) $d_{\mathrm{GH}}(\mathbb{S}^1,\mathbb{S}^2\times\cdots\times\mathbb{S}^2)\geq\frac{\pi}{3}$ for any number of $\mathbb{S}^2$ factors.

(3) $d_{\mathrm{GH}}(\mathbb{S}^m,\boldsymbol{H}_{\geq 0}(\mathbb{S}^n))\geq\frac{1}{2}\zeta_m$ whenever $0<m<n<\infty$.

(4) $d_{\mathrm{GH}}(\boldsymbol{H}_{\geq 0}(\mathbb{S}^m),\boldsymbol{H}_{\geq 0}(\mathbb{S}^n))\geq\frac{1}{2}\zeta_m$ whenever $0<m<n<\infty$.

(5) $d_{\mathrm{GH}}(P,\mathbb{S}^2)\geq\frac{\pi}{3}$ for any finite $P\subset\mathbb{S}^1$. Compare to the $\frac{\pi}{2}$ lower bound given in Lemma 3.2.

(6) $d_{\mathrm{GH}}(P_3,\boldsymbol{H}_{\geq 0}(\mathbb{S}^2))=\frac{\pi}{3}$, where P_3 is the three-point metric space with all interpoint distances equal to $\frac{2\pi}{3}$. Also $d_{\mathrm{GH}}(P_6,\mathbb{S}^2)=\frac{\pi}{3}$, where P_6 is the six-point metric space corresponding to a regular hexagon inscribed in $\mathbb{S}^1$. These are consequences of (5) and small modifications of the correspondences constructed in Proposition 1.16.

Theorem C *Let bounded metric spaces X and Y be such that, for some positive integer m,*

(i) *X can be isometrically embedded into $\mathbb{S}^m$, and*

(ii) *$\boldsymbol{H}_{\geq 0}(\mathbb{S}^{m+1})$ can be isometrically embedded into Y.*

Then:

(1) $d_{\mathrm{GH}}(X,Y)\geq\frac{1}{2}\zeta_m$.

(2) *Moreover,* $\mathrm{dis}(\phi)\geq\zeta_m$ *for any map* $\phi\colon Y\to X$.

Remark 1.25 Example 1.24(1) also holds for $n=1$, albeit this is not implied by Theorem C. The fact that $d_{\mathrm{GH}}([0,\pi],\mathbb{S}^1)\geq\frac{\pi}{3}$ follows from [17, Theorem 4.10] and it also follows from the proof of [20, Lemma 2.3]; see Appendix B.

Organization

In Section 2 we review some preliminaries.

The proof of Proposition 1.2 on a lower bound for $\mathfrak{g}_{m,n}$ involving the normalized volume of open balls is given in Section 3.1, whereas those of Propositions 1.5 (establishing the precise value of $\mathfrak{g}_{0,n}$), 1.6 (establishing the precise value of $\mathfrak{g}_{m,\infty}$), and 1.8 (on a lower bound for $\mathfrak{g}_{m,n}$ involving the covering radius) are given in Section 3.2.

The proof of Theorem A, establishing that $\mathfrak{g}_{m,n} < \frac{\pi}{2}$ (for any $0 < m < n < \infty$), is given in Section 4, whereas those of Theorem B, on a lower bound for $\mathfrak{g}_{m,n}$ deduced from a discontinuous version of the Borsuk–Ulam theorem, and Theorem C (a generalization of Theorem B) are given in Section 5.

The proofs of Propositions 1.16, establishing the precise value of $\mathfrak{g}_{1,2}$, and 1.20, on an upper bound involving the diameter of a face of a geodesic simplex, are given in Section 6.

Proposition 1.18, establishing the precise value of $\mathfrak{g}_{1,3}$, is proved in Section 7, and Proposition 1.19, establishing the precise value of $\mathfrak{g}_{2,3}$, is proved in Section 8.

The case of spheres with Euclidean distance is discussed in Section 9.

Finally, this paper has three appendices. Appendix A provides a succinct and self contained proof of the version of Borsuk–Ulam's theorem due to Dubins and Schwarz [11] which is instrumental for proving Theorem B and related results. Appendix B establishes that the Gromov–Hausdorff distance between the n–dimensional sphere and an interval is always bounded below by $\frac{\pi}{3}$, and Appendix C provides some results about the Gromov–Hausdorff distance between regular polygons.

Additional aspects of this project (such as computational experiments and further constructions of correspondences) are described in [22; 23].

Acknowledgements

We are grateful to Henry Adams for helpful conversations related to this work. We thank Gunnar Carlsson and Tigran Ishkhanov for encouraging Mémoli to tackle the question about the Gromov–Hausdorff distance between spheres via topological methods. This work was supported by NSF grants DMS-1547357, CCF-1526513, IIS-1422400, and CCF-1740761.

2 Preliminaries

Given a metric space (X, d_X) and $\delta > 0$, a δ–net for X is any $A \subset X$ such that for all $x \in X$ there exists $a \in A$ with $d_X(x,a) \le \delta$. The diameter of X is $\operatorname{diam}(X) := \sup_{x,x' \in X} d_X(x,x')$.

Recall [4, Chapter 2] that complete metric space (X, d_X) is a *geodesic space* if and only if it admits midpoints: for all $x, x' \in X$ there exists $z \in X$ such that

$$d_X(x,z) = d_X(x',z) = \tfrac{1}{2} d_X(x,x').$$

We henceforth use the symbol $*$ to denote the one point metric space. It is easy to check that $d_{\mathrm{GH}}(*, X) = \frac{1}{2}\operatorname{diam}(X)$ for any bounded metric space X. From this, and the triangle inequality for the Gromov–Hausdorff distance, it then follows that for all bounded metric spaces X and Y,

$$d_{\mathrm{GH}}(X, Y) \geq \tfrac{1}{2}|\operatorname{diam}(X) - \operatorname{diam}(Y)|. \tag{7}$$

2.1 Notation and conventions about spheres

Finally, let us collect and introduce important notation and conventions which will be used throughout this paper (except for Section 7). For each nonnegative integer $m \in \mathbb{N}$, we define

- $\mathbb{S}^m := \{(x_1, \ldots, x_{m+1}) \in \mathbb{R}^{m+1} \mid x_1^2 + \cdots + x_{m+1}^2 = 1\}$ (m–sphere);
- $\boldsymbol{H}_{\geq 0}(\mathbb{S}^m) := \{(x_1, \ldots, x_{m+1}) \in \mathbb{S}^m \mid x_{m+1} \geq 0\}$ (closed upper hemisphere);
- $\boldsymbol{H}_{>0}(\mathbb{S}^m) := \{(x_1, \ldots, x_{m+1}) \in \mathbb{S}^m \mid x_{m+1} > 0\}$ (open upper hemisphere);
- $\boldsymbol{H}_{\leq 0}(\mathbb{S}^m) := \{(x_1, \ldots, x_{m+1}) \in \mathbb{S}^m \mid x_{m+1} \leq 0\}$ (closed lower hemisphere);
- $\boldsymbol{H}_{<0}(\mathbb{S}^m) := \{(x_1, \ldots, x_{m+1}) \in \mathbb{S}^m \mid x_{m+1} < 0\}$ (open lower hemisphere);
- $\boldsymbol{E}(\mathbb{S}^m) := \{(x_1, \ldots, x_{m+1}) \in \mathbb{S}^m \mid x_{m+1} = 0\}$ (equator of sphere);
- $\mathbb{B}^{m+1} := \{(x_1, \ldots, x_{m+1}) \in \mathbb{R}^{m+1} \mid x_1^2 + \cdots + x_{m+1}^2 \leq 1\}$ (unit closed ball);
- $\widehat{\mathbb{B}}^{m+1} := \{(x_1, \ldots, x_{m+1}) \in \mathbb{R}^{m+1} \mid |x_1| + \cdots + |x_{m+1}| \leq 1\}$ (unit cross-polytope).

Also, $\mathbb{S}^m$, $\boldsymbol{H}_{\geq 0}(\mathbb{S}^m)$, $\boldsymbol{H}_{>0}(\mathbb{S}^m)$, $\boldsymbol{H}_{\leq 0}(\mathbb{S}^m)$, $\boldsymbol{H}_{<0}(\mathbb{S}^m)$ and $\boldsymbol{E}(\mathbb{S}^m)$ are all equipped with the geodesic metric $d_{\mathbb{S}^m}$. Observe that $\mathbb{S}^m$ and $\boldsymbol{E}(\mathbb{S}^{m+1})$ are isometric. We will denote by

$$\iota_m \colon \mathbb{S}^m \to \mathbb{S}^{m+1}, \quad (x_1, \ldots, x_{m+1}) \mapsto (x_1, \ldots, x_{m+1}, 0), \tag{8}$$

the canonical isometric embedding from $\mathbb{S}^m$ into $\mathbb{S}^{m+1}$.

2.2 A general construction of correspondences

Assume X and Y are compact metric spaces such that $X \xhookrightarrow{\phi} Y$ isometrically, eg $\mathbb{S}^m \hookrightarrow \mathbb{S}^n$ for $m \leq n$.

As mentioned in Remark 1.1, any surjection $\psi \colon Y \twoheadrightarrow X$ gives rise to a correspondence between X and Y. The following simple construction of such a ψ will be used throughout this paper. Given $k \in \mathbb{N}$, assume $P_k = \{B_1, \ldots, B_i, \ldots, B_k\}$ is any partition of $Y \setminus \phi(X)$ and $\mathbb{X}_k = \{x_1, \dots, x_i, \dots, x_k\}$ are any k points in X. Then define $\psi\colon Y \twoheadrightarrow X$ by $\psi|_{\phi(X)} := \phi^{-1}$ and $\psi|_{B_i} := x_i$ for each $1 \le i \le k$. It then follows that the distortion of this correspondence is

$$\operatorname{dis}(\psi) = \max\{A, B, C\},$$

where

- $A := \max_i \operatorname{diam}(B_i)$,
- $B := \max_{i \neq j} \max_{y \in B_i, y' \in B_j} |d_X(x_i, x_j) - d_Y(y, y')|$, and
- $C := \max_i \max_{x \in X, y \in B_i} |d_X(x, x_i) - d_Y(\phi(x), y)|$.

This pattern will be used several times in this paper.

3 Some general lower bounds

3.1 The proof of Proposition 1.2

For a metric space X and $\rho > 0$, let $N_X(\rho)$ denote the minimal number of open balls of radius ρ needed to cover X. Also, let $C_X(\rho)$ denote the maximal number of pairwise disjoint open balls of radius $\frac{1}{2}\rho$ that can be placed in X. N_X and C_X are usually referred to as the *covering number* and the *packing number*, respectively.

Note that the covering radius cov_X — see (4) — and the covering number N_X are related by

$$\operatorname{cov}_X(k) = \inf\{\rho > 0 : N_X(\rho) \le k\}.$$

The following *stability* property of $N_X(\cdot)$ and $C_X(\cdot)$ is classical and can be used to obtain estimates for the Gromov–Hausdorff distance between spheres:

Proposition 3.1 [30, page 299] *If X and Y are metric spaces and $d_{\mathrm{GH}}(X, Y) < \eta$ for some $\eta > 0$, then for all $\rho \ge 0$,*

(1) *$N_X(\rho) \ge N_Y(\rho + 2\eta)$, and*

(2) *$C_X(\rho) \ge C_Y(\rho + 2\eta)$.*

Recall that $v_n(\rho)$ is the normalized volume of an open ball or radius ρ on $\mathbb{S}^n$.

Proof of Proposition 1.2 The proof that

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \ge \mu_{m,n} := \frac{1}{2} \sup_{\rho \in (0,\pi]} \left(v_n^{-1} \circ v_m\left(\tfrac{1}{2}\rho\right) - \rho\right)$$

for any $0 < m < n < \infty$ is by contradiction. We first state two claims that we prove at the end.

Claim 1 *For any $\rho > 0$ and $n \geq 1$, the packing number satisfies $C_{\mathbb{S}^n}(\rho) \leq \bigl(v_n\bigl(\frac{1}{2}\rho\bigr)\bigr)^{-1}$.*

Claim 2 *For any $\rho > 0$ and $n \geq 1$, the covering number $N_{\mathbb{S}^n}(\rho)$ satisfies*

$$1 \leq N_{\mathbb{S}^n}(\rho) \cdot v_n(\rho).$$

Assuming the claims above, suppose that $n > m \geq 1$ and $\eta := d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) < \mu_{m,n}$. Pick $\varepsilon > 0$ small enough such that $\eta + \frac{1}{2}\varepsilon < \mu_{m,n}$.

Since $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) < \eta + \frac{1}{2}\varepsilon$, from Proposition 3.1, the fact that for $N_X(\rho) \leq C_X(\rho)$ for any compact metric space X and any $\rho > 0$, and Claim 1, we have that

$$N_{\mathbb{S}^n}(\rho + 2\eta + \varepsilon) \leq N_{\mathbb{S}^m}(\rho) \leq C_{\mathbb{S}^m}(\rho) \leq \bigl(v_m\bigl(\tfrac{1}{2}\rho\bigr)\bigr)^{-1}.$$

Now, from Claim 2 we obtain that, for all $\rho \in [0, \pi]$,

$$1 \leq N_{\mathbb{S}^n}(\rho + 2\eta + \varepsilon) \cdot v_n(\rho + 2\eta + \varepsilon) \leq \frac{v_n(\rho + 2\eta + \varepsilon)}{v_m\bigl(\frac{1}{2}\rho\bigr)}.$$

Then, for all $\rho \in [0, \pi]$, we must have

$$\eta + \tfrac{1}{2}\varepsilon \geq \tfrac{1}{2}\bigl(v_n^{-1} \circ v_m\bigl(\tfrac{1}{2}\rho\bigr) - \rho\bigr).$$

Then, in particular, $\eta + \frac{1}{2}\varepsilon \geq \mu_{m,n}$, a contradiction. □

Proof of Claim 1 Let $k = C_{\mathbb{S}^n}(\rho)$ and let $x_1, \dots, x_k \in \mathbb{S}^n$ be such that

$$B\bigl(x_i, \tfrac{1}{2}\rho\bigr) \cap B\bigl(x_j, \tfrac{1}{2}\rho\bigr) = \varnothing$$

for all $i \neq j$. Thus, $\bigcup_{i=1}^k B(x_i, \frac{1}{2}\rho) \subset \mathbb{S}^n$, and

$$\mathbf{Vol}(\mathbb{S}^n) \geq \mathbf{vol}_{\mathbb{S}^n}\biggl(\bigcup_{i=1}^k B\bigl(x_i, \tfrac{1}{2}\rho\bigr)\biggr) = k \cdot v_n\bigl(\tfrac{1}{2}\rho\bigr) \cdot \mathbf{Vol}(\mathbb{S}^n). \qquad \square$$

Proof of Claim 2 Fix $N = N_{\mathbb{S}^n}(\rho)$ and $x_1, \dots, x_N \in \mathbb{S}^n$ such that $\bigcup_{i=1}^N B(x_i, \rho) = \mathbb{S}^n$. Then

$$\mathbf{Vol}(\mathbb{S}^n) \leq \mathbf{vol}_{\mathbb{S}^n}\biggl(\bigcup_{i=1}^N B(x_i, \rho)\biggr) \leq N \cdot v_n(\rho) \cdot \mathbf{Vol}(\mathbb{S}^n). \qquad \square$$

3.2 Other lower bounds and the proofs of Propositions 1.5 and 1.6

Recall the following corollary to the Borsuk–Ulam theorem [24]:

Theorem D (Lyusternik–Schnirelmann) *Let $n \in \mathbb{N}$ and $\{U_1, \dots, U_{n+1}\}$ be a closed cover of $\mathbb{S}^n$. Then there is an $i_0 \in \{1, \dots, n+1\}$ such that U_{i_0} contains two antipodal points.*

The lemma below will be useful in what follows:

Lemma 3.2 *For any integer $m \geq 1$ and any finite metric space P with cardinality at most $m+1$, we have $d_{\mathrm{GH}}(\mathbb{S}^m, P) \geq \frac{\pi}{2}$.*

Remark 3.3 Lemma 3.2 and Remark 1.4 imply that for each integer $n \geq 1$, we have $d_{\mathrm{GH}}(\mathbb{S}^n, P) = \frac{\pi}{2}$ for any finite metric space P with $|P| \leq n+1$ and $\operatorname{diam}(P) \leq \pi$.

Proof Suppose $m \geq 1$ is given. We prove that $d_{\mathrm{GH}}(\mathbb{S}^m, P) \geq \frac{\pi}{2}$ for any finite set P of size at most $m+1$. Assume that R is an arbitrary correspondence between $\mathbb{S}^m$ and P. We claim that $\operatorname{dis}(R) \geq \pi$, from which the proof will follow. For each $p \in P$, let $R(p) := \{z \in \mathbb{S}^m \mid (z, p) \in R\}$. Then $\{\overline{R(p)} \subseteq \mathbb{S}^m \mid p \in P\}$ is a closed cover of $\mathbb{S}^m$. Since $|P| \leq m+1$, Theorem D yields that $\operatorname{diam}(R(p_0)) = \pi$ for some $p_0 \in P$. Finally, the claim follows since $\operatorname{dis}(R) \geq \max_{p \in P} \operatorname{diam}(R(p))$. □

By a refinement of the proof of Lemma 3.2 above one obtains:

Corollary 3.4 *Let R be any correspondence between a finite metric space P and $\mathbb{S}^\infty$. Then $\operatorname{dis}(R) \geq \pi$. In particular, $d_{\mathrm{GH}}(P, \mathbb{S}^\infty) \geq \frac{\pi}{2}$.*

Proof As in the proof of Lemma 3.2, the correspondence R induces a closed cover of $\mathbb{S}^\infty$. Thus, it induces a closed cover of any finite dimensional sphere $\mathbb{S}^{|P|-1} \subset \mathbb{S}^\infty$. The claim follows from Theorem D. □

Notice that if P has diameter at most π, then $d_{\mathrm{GH}}(P, \mathbb{S}^\infty) = \frac{\pi}{2}$ (see Remarks 1.4 and 3.3). In Appendix C we consider a scenario which is thematically connected with Remark 3.3 and Corollary 3.4, namely the determination of the Gromov–Hausdorff distance between a finite metric space and a sphere. Appendix C fully resolves this question for the case of $\mathbb{S}^1$ and (the vertex set of) inscribed regular polygons.

By a small modification of the proof of Corollary 3.4, we obtain the following stronger claim:

Proposition 3.5 *Let X be any totally bounded metric space. Then $d_{\mathrm{GH}}(X, \mathbb{S}^\infty) \geq \frac{\pi}{2}$.*

Proof Fix any $\varepsilon > 0$ and let $P_\varepsilon \subset X$ be a finite ε–net for X. Then, by the triangle inequality for d_{GH} and Corollary 3.4,

$$d_{\mathrm{GH}}(X, \mathbb{S}^\infty) \geq d_{\mathrm{GH}}(\mathbb{S}^\infty, P_\varepsilon) - d_{\mathrm{GH}}(X, P_\varepsilon) \geq \frac{\pi}{2} - \varepsilon,$$

which implies the claim since $\varepsilon > 0$ was arbitrary. □

Proof of Proposition 1.5 That $d_{\mathrm{GH}}(\mathbb{S}^0,\mathbb{S}^n)=\frac{\pi}{2}$ for any integer $n\geq 1$ follows from Lemma 3.2 and Remark 1.4. □

Proof of Proposition 1.6 That $d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^\infty)=\frac{\pi}{2}$ for any nonnegative integer $m<\infty$ follows from Proposition 3.5 and Remark 1.4. □

Proof of Proposition 1.8 We prove that $d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^n)\geq \nu_{m,n}:=\frac{\pi}{2}-\mathrm{cov}_{\mathbb{S}^m}(n+1)$ for any $1\leq m<n<\infty$.

Let P be any subset $\mathbb{S}^m$ with cardinality not exceeding $n+1$. Since the Hausdorff distance satisfies $d_{\mathrm{H}}(P,\mathbb{S}^m)\geq d_{\mathrm{GH}}(P,\mathbb{S}^m)$, and by the triangle inequality for the Gromov–Hausdorff distance, we have

$$d_{\mathrm{H}}(P,\mathbb{S}^m)+d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^n)\geq d_{\mathrm{GH}}(P,\mathbb{S}^m)+d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^n)\geq d_{\mathrm{GH}}(P,\mathbb{S}^n).$$

Since $\mathrm{diam}(P)\leq\pi$, by Remark 3.3 we have that $d_{\mathrm{GH}}(P,\mathbb{S}^n)=\frac{\pi}{2}$. Hence, from the above,

$$d_{\mathrm{H}}(P,\mathbb{S}^m)+d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^n)\geq\tfrac{\pi}{2}$$

for *any* $P\subset\mathbb{S}^m$ with $|P|\leq n+1$. By the definition of the covering radius (see (4)), we obtain the claim by taking the infimum over all possible such choices of P. □

4 The proof of Theorem A

The Borsuk–Ulam theorem implies that, for any positive integers $n>m$ and for any given continuous map $\varphi\colon\mathbb{S}^n\to\mathbb{S}^m$, there exists two antipodal points in the higher dimensional sphere which are mapped to the same point in the lower dimensional sphere.

We now prove that, in contrast, there always exists a *surjective*, antipode-preserving, and continuous map $\psi_{m,n}$ from the lower dimensional sphere to the higher dimensional sphere.

Theorem E *For all integers $0<m<n<\infty$, there exists an* ***antipode-preserving*** *continuous surjection*

$$\psi_{m,n}\colon\mathbb{S}^m\twoheadrightarrow\mathbb{S}^n,$$

ie $\psi_{m,n}(-x)=-\psi_{m,n}(x)$ for every $x\in\mathbb{S}^m$.

With this theorem, the proof of Theorem A, stating that $d_{\mathrm{GH}}(\mathbb{S}^m,\mathbb{S}^n)<\frac{\pi}{2}$ for all $0<m<n<\infty$, now follows:

Proof of Theorem A Let $\psi_{m,n}\colon \mathbb{S}^m \twoheadrightarrow \mathbb{S}^n$ be the map given in Theorem E. Recall that the graph of a surjective map can be seen as a correspondence and let $R_{m,n} := \operatorname{graph}(\psi_{m,n})$. In order to prove the claim, it is enough to verify that

$$\operatorname{dis}(R_{m,n}) = \operatorname{dis}(\psi_{m,n}) < \pi.$$

Since $\psi_{m,n}$ is continuous and $\mathbb{S}^m$ is compact, the supremum in the definition of distortion is a maximum,

$$\operatorname{dis}(\psi_{m,n}) = \max_{x,x' \in \mathbb{S}^m} |d_{\mathbb{S}^m}(x,x') - d_{\mathbb{S}^n}(\psi_{m,n}(x), \psi_{m,n}(x'))|.$$

Let $x_0, x_0' \in \mathbb{S}^m$ attain the maximum above. Note that we may assume that $x_0 \neq x_0'$, for otherwise we would have $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \leq \frac{1}{2}\operatorname{dis}(R_{m,n}) = \frac{1}{2}\operatorname{dis}(\psi_{m,n}) = 0$, which would imply that $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) = 0$, ie that $\mathbb{S}^m$ and $\mathbb{S}^n$ are isometric, which is a contradiction since $m \neq n$.

Assume first that $x_0' \neq -x_0$. In this case,

$$0 < d_{\mathbb{S}^m}(x_0, x_0') < \pi \quad \text{and} \quad 0 \leq d_{\mathbb{S}^n}(\psi_{m,n}(x_0), \psi_{m,n}(x_0')) \leq \pi.$$

Thus,

$$|d_{\mathbb{S}^m}(x_0, x_0') - d_{\mathbb{S}^n}(\psi_{m,n}(x_0), \psi_{m,n}(x_0'))| < \pi.$$

Assume now that $x_0' = -x_0$. In this case, $d_{\mathbb{S}^m}(x_0, x_0') = d_{\mathbb{S}^n}(\psi_{m,n}(x_0), \psi_{m,n}(x_0')) = \pi$ since $\psi_{m,n}$ is antipode-preserving. Thus, in this case we also have

$$0 = |d_{\mathbb{S}^m}(x_0, x_0') - d_{\mathbb{S}^n}(\psi_{m,n}(x_0), \psi_{m,n}(x_0'))| < \pi. \qquad \square$$

Remark 4.1 The antipode-preserving property of $\psi_{m,n}$ given in Theorem E is stronger than what we need for the purpose of proving Theorem A. Indeed, all one needs is that $\psi_{m,n}(x) \neq \psi_{m,n}(-x)$ for every $x \in \mathbb{S}^m$.

The goal for the rest of this section is to prove Theorem E.

Spherical suspensions and space-filling curves are key technical tools, which we now review.

Space-filling curves

The existence of the space-filling curves is well known [29]:

Theorem F (space-filling curve) *There exists a continuous and surjective map*

$$H\colon [0,1] \twoheadrightarrow [0,1]^2.$$

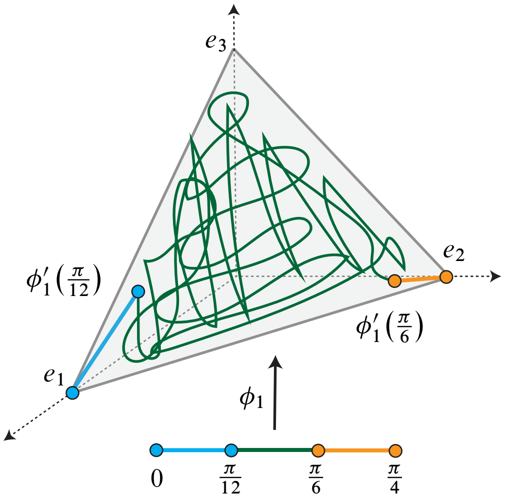


Figure 3: The continuous surjection $\phi_1 \colon \left[0, \frac{\pi}{4}\right] \twoheadrightarrow \mathrm{Conv}(e_1, e_2, e_3)$.

In the sequel, we will use the notation $\mathrm{Conv}(v_1, v_2, \dots, v_d)$ to denote the convex hull of vectors $v_1, v_2, \dots, v_d$.

By resorting to space-filling curves, one can prove the following proposition, which will be crucial in the sequel:

Proposition 4.2 *There exists an antipode-preserving continuous surjection*

$$\psi_{1,2} \colon \mathbb{S}^1 \twoheadrightarrow \mathbb{S}^2.$$

Proof Recall the definition of the 3–dimensional cross-polytope,

$$\widehat{\mathbb{B}}^3 := \mathrm{Conv}(e_1, -e_1, e_2, -e_2, e_3, -e_3) \subset \mathbb{R}^3,$$

where $e_1 = (1,0,0)$, $e_2 = (0,1,0)$, and $e_3 = (0,0,1)$. Then its boundary $\partial\widehat{\mathbb{B}}^3$, which consists of eight triangles

$$\mathrm{Conv}(e_1, e_2, e_3), \quad \mathrm{Conv}(e_1, e_2, -e_3), \quad \dots, \mathrm{Conv}(-e_1, -e_2, -e_3),$$

is homeomorphic to $\mathbb{S}^2$.

Now, divide $\mathbb{S}^1$ into eight closed circular arcs with equal length $\frac{\pi}{4}$. In other words, let

$$\left[0, \tfrac{\pi}{4}\right], \ \left[\tfrac{\pi}{4}, \tfrac{\pi}{2}\right], \ \left[\tfrac{\pi}{2}, \tfrac{3\pi}{4}\right], \ \left[\tfrac{3\pi}{4}, \pi\right], \ \left[\pi, \tfrac{5\pi}{4}\right], \ \left[\tfrac{5\pi}{4}, \tfrac{3\pi}{2}\right], \ \left[\tfrac{3\pi}{2}, \ \tfrac{7\pi}{4}\right], \ \left[\tfrac{7\pi}{4}, 2\pi\right]$$

be those eight regions. Of course, we are identifying 0 and 2π here.

Note that we are able to build a continuous and surjective map

$$\phi_1 : \left[0, \tfrac{\pi}{4}\right] \twoheadrightarrow \mathrm{Conv}(e_1, e_2, e_3) \quad \text{such that } \phi_1(0) = e_1 \text{ and } \phi_1\left(\tfrac{\pi}{4}\right) = e_2$$

as follows: since $\mathrm{Conv}(e_1, e_2, e_3)$ is homeomorphic to $[0,1]^2$, by Theorem F there exists a continuous and surjective map ϕ_1' from $\left[\frac{\pi}{12}, \frac{\pi}{6}\right]$ to $\mathrm{Conv}(e_1, e_2, e_3)$; then, we extend its domain by using linear interpolation between e_1 and $\phi_1'\left(\frac{\pi}{12}\right)$, and e_2 and $\phi_1'\left(\frac{\pi}{6}\right)$ to give rise to ϕ_1 — see Figure 3.

By using an analogous procedure, we construct continuous and surjective maps

$$\begin{aligned}
&\phi_2 : \left[\tfrac{\pi}{4}, \tfrac{\pi}{2}\right] \twoheadrightarrow \mathrm{Conv}(-e_1, e_2, e_3) && \text{such that } \phi_2\left(\tfrac{\pi}{4}\right) = e_2 \text{ and } \phi_2\left(\tfrac{\pi}{2}\right) = e_3,\\
&\phi_3 : \left[\tfrac{\pi}{2}, \tfrac{3\pi}{4}\right] \twoheadrightarrow \mathrm{Conv}(e_1, -e_2, e_3) && \text{such that } \phi_3\left(\tfrac{\pi}{2}\right) = e_3 \text{ and } \phi_3\left(\tfrac{3\pi}{4}\right) = -e_2,\\
&\phi_4 : \left[\tfrac{3\pi}{4}, \pi\right] \twoheadrightarrow \mathrm{Conv}(-e_1, -e_2, e_3) && \text{such that } \phi_4\left(\tfrac{3\pi}{4}\right) = -e_2 \text{ and } \phi_4(\pi) = -e_1.
\end{aligned}$$

Next, we construct the remaining continuous and surjective maps by suitably reflecting the ones already constructed,

$$\begin{aligned}
&\phi_5 : \left[\pi, \tfrac{5\pi}{4}\right] \twoheadrightarrow \mathrm{Conv}(-e_1, -e_2, -e_3) && \text{such that } \phi_5(x) := -\phi_1(-x),\\
&\phi_6 : \left[\tfrac{5\pi}{4}, \tfrac{3\pi}{2}\right] \twoheadrightarrow \mathrm{Conv}(e_1, -e_2, -e_3) && \text{such that } \phi_6(x) := -\phi_2(-x),\\
&\phi_7 : \left[\tfrac{3\pi}{2}, \tfrac{7\pi}{4}\right] \twoheadrightarrow \mathrm{Conv}(e_1, e_2, -e_3) && \text{such that } \phi_7(x) := -\phi_3(-x),\\
&\phi_8 : \left[\tfrac{7\pi}{4}, 2\pi\right] \twoheadrightarrow \mathrm{Conv}(-e_1, e_2, -e_3) && \text{such that } \phi_8(x) := -\phi_4(-x).
\end{aligned}$$

Finally, by gluing all the eight maps ϕ_i, we build an antipode-preserving continuous and surjective map $\overline{\psi}_{1,2} : \mathbb{S}^1 \twoheadrightarrow \partial\widehat{\mathbb{B}}^3$. Using the canonical (closest point projection) homeomorphism between $\partial\widehat{\mathbb{B}}^3$ and $\mathbb{S}^2$, we finally have the announced $\psi_{1,2} : \mathbb{S}^1 \twoheadrightarrow \mathbb{S}^2$. It is clear from its construction that the map $\psi_{1,2}$ is continuous, surjective, and antipode-preserving. Figure 4 depicts the overall structure of the map $\psi_{1,2}$. □

Spherical suspensions

Suppose $m, n \in \mathbb{N}$ and a map $f : \mathbb{S}^m \to \mathbb{S}^n$ are given. Then one can lift this map f to a map from $\mathbb{S}^{m+1}$ to $\mathbb{S}^{n+1}$ in the following way: observe that an arbitrary point in $\mathbb{S}^{m+1}$ can be expressed as $(p \sin\theta, \cos\theta)$ for some $p \in \mathbb{S}^m$ and $\theta \in [0, \pi]$; then the *spherical suspension of f* is the map

$$Sf : \mathbb{S}^{m+1} \to \mathbb{S}^{n+1}, \quad (p \sin\theta, \cos\theta) \mapsto (f(p) \sin\theta, \cos\theta).$$

Lemma 4.3 *If the map $f : \mathbb{S}^m \twoheadrightarrow \mathbb{S}^n$ is continuous, surjective and antipode-preserving, then $Sf : \mathbb{S}^{m+1} \twoheadrightarrow \mathbb{S}^{n+1}$ is also continuous, surjective and antipode-preserving.*

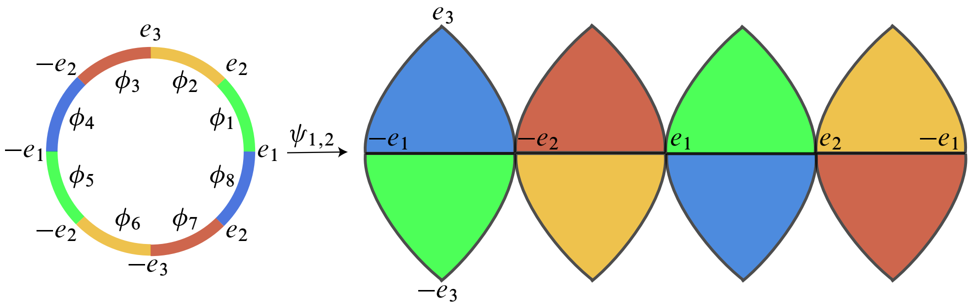


Figure 4: Structure of the map $\psi_{1,2}$ constructed in Proposition 4.2. Inside each arc, the map is defined via a space-filling curve. For simplicity, $\mathbb{S}^2$ is "cartographically" depicted.

Proof Continuity and surjectivity are clear from the construction. Since f is antipode-preserving, we know that $f(-p) = -f(p)$ for every $p \in \mathbb{S}^m$. Hence,

$$\begin{aligned} Sf(-p\sin\theta, -\cos\theta) &= Sf(-p\sin(\pi-\theta), \cos(\pi-\theta)) \\ &= (f(-p)\sin(\pi-\theta), \cos(\pi-\theta)) \\ &= (-f(p)\sin\theta, -\cos\theta) \\ &= -(f(p)\sin\theta, \cos\theta) \\ &= -Sf(p\sin\theta, \cos\theta) \end{aligned}$$

for any $p \in \mathbb{S}^m$ and $\theta \in [0, \pi]$. Thus, Sf is also antipode-preserving. □

We now use induction to obtain:

Corollary 4.4 *For any integer $m > 0$, there exists a continuous, surjective, and antipode-preserving map*

$$\psi_{m,(m+1)} \colon \mathbb{S}^m \twoheadrightarrow \mathbb{S}^{m+1}.$$

Proof Proposition 4.2 guarantees the existence of $\psi_{1,2}$. For general m, it suffices to apply Lemma 4.3 inductively. □

The following lemma is obvious:

Lemma 4.5 *Suppose that $l, m, n \in \mathbb{N}$, and maps $f \colon \mathbb{S}^l \twoheadrightarrow \mathbb{S}^m$ and $g \colon \mathbb{S}^m \twoheadrightarrow \mathbb{S}^n$ are given such that both f and g are continuous, surjective, and antipode-preserving. Then their composition $g \circ f \colon \mathbb{S}^l \twoheadrightarrow \mathbb{S}^n$ is also continuous, surjective, and antipode-preserving.*

The proof of Theorem E

We are now ready to prove Theorem E, which states that there exists an antipode-preserving continuous surjection $\psi_{m,n}\colon \mathbb{S}^m \twoheadrightarrow \mathbb{S}^n$ for any $0 < m < n < \infty$.

Proof of Theorem E By Corollary 4.4, there are continuous, surjective, and antipode-preserving maps $\psi_{m,(m+1)}, \psi_{(m+1),(m+2)}, \dots, \psi_{(n-1),n}$. Then, by Lemma 4.5, the map

$$\psi_{m,n} := \psi_{(n-1),n} \circ \cdots \circ \psi_{(m+1),(m+2)} \circ \psi_{m,(m+1)}$$

is also continuous, surjective, and antipode-preserving. □

5 A Borsuk–Ulam theorem for discontinuous functions and the proof of Theorem B

Definition 5.1 (modulus of discontinuity) Let X be a topological space, Y be a metric space, and $f\colon X \to Y$ be any function. Then we define $\delta(f)$, the *modulus of discontinuity of f*, by

$$\delta(f) := \inf\{\delta \geq 0 \mid \text{each } x \in X \text{ has an open neighborhood } U_x \text{ with } \operatorname{diam}(f(U_x)) \leq \delta\}.$$

Remark 5.2 Of course, $\delta(f) = 0$ if and only if f is continuous.

It turns out that the modulus of discontinuity is a lower bound for distortion:

Proposition 5.3 *Let $\phi\colon (X, d_X) \to (Y, d_Y)$ be a map between two metric spaces. Then*

$$\delta(\phi) \leq \operatorname{dis}(\phi).$$

Proof If $\operatorname{dis}(\phi) = \infty$, then the proof is trivial, so suppose $\operatorname{dis}(\phi) < \infty$. Now, fix arbitrary $x \in X$ and $\varepsilon > 0$. Consider the open ball $U_x := B\big(x, \frac{1}{2}\varepsilon\big)$. Then, for any $x', x'' \in U_x$,

$$d_Y(\phi(x'), \phi(x'')) \leq d_X(x', x'') + |d_X(x', x'') - d_Y(\phi(x'), \phi(x''))| < \operatorname{dis}(\phi) + \varepsilon,$$

so $\operatorname{diam}(\phi(U_x)) \leq \operatorname{dis}(\phi) + \varepsilon$. Since x is arbitrary, this implies $\delta(\phi) \leq \operatorname{dis}(\phi) + \varepsilon$. Since ε is arbitrary, we have the required inequality. □

The following variant of the Borsuk–Ulam theorem, due to Dubins and Schwarz, is the main tool used in the proof of Theorem B.

Theorem G [11, Theorem 1] *For each integer $n > 0$, the modulus of discontinuity of any function $f : \mathbb{B}^n \to \mathbb{S}^{n-1}$ that maps every pair of antipodal points on the boundary of $\mathbb{B}^n$ onto antipodal points on $\mathbb{S}^{n-1}$ is not less than ζ_{n-1}.*

In Appendix A we provide a concise self-contained proof of this result based on ideas by Arnold Waßmer; see Matoušek [24, page 41].

We immediately have:

Corollary 5.4 [11, Corollary 3] *For each integer $n > 0$, the modulus of discontinuity of any function $g : \mathbb{S}^n \to \mathbb{S}^{n-1}$ which maps every pair of antipodal points on $\mathbb{S}^n$ onto antipodal points on $\mathbb{S}^{n-1}$ is not less than ζ_{n-1}.*

We provide a detailed proof of this result for completeness.

Proof Consider the map

$$\Phi : \mathbb{B}^n \to \mathbb{S}^n, \quad (x_1, \dots, x_n) \mapsto \bigl(x_1, \dots, x_n, \sqrt{1-(x_1^2+\cdots+x_n^2)}\bigr).$$

Obviously, Φ is continuous and its image is $\boldsymbol{H}_{\geq 0}(\mathbb{S}^n)$. Now, fix an arbitrary $\delta \geq 0$ such that for every $x \in \mathbb{S}^n$, there exists an open neighborhood U_x of x with $\operatorname{diam}(g(U_x)) \leq \delta$.

Now, fix an arbitrary $x' \in \mathbb{B}^n$. Then $\Phi^{-1}(U_{\Phi(x')})$ is an open neighborhood of x', and

$$\operatorname{diam}\bigl(g \circ \Phi(\Phi^{-1}(U_{\Phi(x')}))\bigr) \leq \operatorname{diam}(g(U_{\Phi(x')})) \leq \delta.$$

Since x' is arbitrary, this means that $\delta \geq \delta(g \circ \Phi)$. Moreover, since $g \circ \Phi$ is antipode-preserving, $\delta(g \circ \Phi) \geq \zeta_{n-1}$ by Theorem G. Hence, we conclude that $\delta \geq \zeta_{n-1}$. Finally, since δ was arbitrary, by taking the infimum we conclude that

$$\delta(g) \geq \zeta_{n-1}. \qquad \square$$

Corollary 5.5 *For each integer $n > 0$, any function $g : \mathbb{S}^n \to \mathbb{S}^{n-1}$ which maps every pair of antipodal points on $\mathbb{S}^n$ onto antipodal points on $\mathbb{S}^{n-1}$ satisfies $\operatorname{dis}(g) \geq \zeta_{n-1}$.*

Proof Apply Corollary 5.4 and Proposition 5.3. □

5.1 The proof of Theorem B

We are almost ready to prove Theorem B, which establishes $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \geq \frac{1}{2}\zeta_m$ for any $0 < m < n < \infty$.

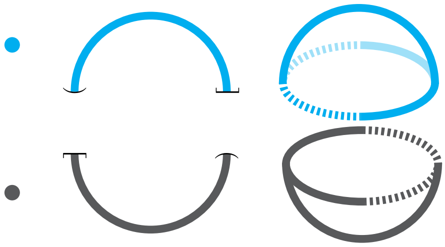


Figure 5: From left to right, the blue sets represent $\boldsymbol{A}(\mathbb{S}^0)$, $\boldsymbol{A}(\mathbb{S}^1)$ and $\boldsymbol{A}(\mathbb{S}^2)$. The figure also shows their antipodes in dark gray. See Definition 5.6.

For each integer $n \geq 1$, recall the natural isometric embedding of $\mathbb{S}^{n-1}$ to the equator $\boldsymbol{E}(\mathbb{S}^n)$ of $\mathbb{S}^n$,

$$\iota_{n-1}\colon \mathbb{S}^{n-1} \hookrightarrow \mathbb{S}^n, \quad (x_1,\ldots,x_n) \mapsto (x_1,\ldots,x_n,0).$$

Also, let us define the sets $\boldsymbol{A}(\mathbb{S}^n) \subset \mathbb{S}^n$ (which we will sometimes refer to as "helmets") for $n \in \mathbb{N}$:

Definition 5.6 (definition of $\boldsymbol{A}(\mathbb{S}^n)$) Let

$$\boldsymbol{A}(\mathbb{S}^0) := \{1\} \quad \text{and} \quad \boldsymbol{A}(\mathbb{S}^1) := \{(\cos\theta, \sin\theta) \in \mathbb{S}^1 \mid \theta \in [0,\pi)\}.$$

Moreover, for general $n \geq 1$, define, inductively,

$$\boldsymbol{A}(\mathbb{S}^n) := \boldsymbol{H}_{>0}(\mathbb{S}^n) \cup \iota_{n-1}(\boldsymbol{A}(\mathbb{S}^{n-1})).$$

See Figure 5 for an illustration. Observe that, for any $n \geq 0$,

$$\boldsymbol{A}(\mathbb{S}^n) \cap (-\boldsymbol{A}(\mathbb{S}^n)) = \varnothing \quad \text{and} \quad \boldsymbol{A}(\mathbb{S}^n) \cup (-\boldsymbol{A}(\mathbb{S}^n)) = \mathbb{S}^n.$$

The following lemma is simple but critical. Given any map $\phi\colon \mathbb{S}^n \to \mathbb{S}^{n-1}$ it will permit constructing an antipode-preserving map ϕ^* with at most the same distortion.

Lemma 5.7 *For any $m,n \geq 0$, let $\varnothing \neq C \subseteq \mathbb{S}^n$ satisfy $C \cap (-C) = \varnothing$ and let $\phi\colon C \to \mathbb{S}^m$ be any map. Then the extension ϕ^* of ϕ to the set $C \cup (-C)$ defined by*

$$\phi^*\colon C \cup (-C) \to \mathbb{S}^m, \qquad x \mapsto \phi(x), \quad -x \mapsto -\phi(x) \quad \textit{for } x \in C,$$

is antipode-preserving and satisfies $\operatorname{dis}(\phi^*) = \operatorname{dis}(\phi)$.

Proof By definition, ϕ^* is antipode-preserving. Now, fix arbitrary $x, x' \in C$. Then

$$\begin{aligned}|d_{\mathbb{S}^n}(x,-x')-d_{\mathbb{S}^m}(\phi^*(x),\phi^*(-x'))| &= |(\pi-d_{\mathbb{S}^n}(x,x'))-(\pi-d_{\mathbb{S}^m}(\phi(x),\phi(x')))| \\ &= |d_{\mathbb{S}^n}(x,x')-d_{\mathbb{S}^m}(\phi(x),\phi(x'))| \\ &\le \operatorname{dis}(\phi)\end{aligned}$$

and

$$|d_{\mathbb{S}^n}(-x,-x')-d_{\mathbb{S}^m}(\phi^*(-x),\phi^*(-x'))| = |d_{\mathbb{S}^n}(x,x')-d_{\mathbb{S}^m}(\phi(x),\phi(x'))| \le \operatorname{dis}(\phi).$$

This implies $\operatorname{dis}(\phi^*) = \operatorname{dis}(\phi)$. □

Corollary 5.8 *For each $n \in \mathbb{Z}_{>0}$ and any map $\phi\colon \mathbb{S}^n \to \mathbb{S}^{n-1}$, there exists an antipode-preserving map $\phi^*\colon \mathbb{S}^n \to \mathbb{S}^{n-1}$ such that* $\operatorname{dis}(\phi^*) \le \operatorname{dis}(\phi)$.

Proof Consider the restriction of ϕ to $\boldsymbol{A}(\mathbb{S}^n)$ and apply Lemma 5.7. □

Finally, we are ready to prove Theorem B.

Proof of Theorem B Let $0 < m < n < \infty$. We first prove the second claim of Theorem B that $\operatorname{dis}(\phi) \ge \zeta_m$ for any map $\phi\colon \mathbb{S}^n \to \mathbb{S}^m$. Suppose to the contrary, so that there is a map $\tilde{g}\colon \mathbb{S}^n \to \mathbb{S}^m$ with $\operatorname{dis}(\tilde{g}) < \zeta_m$. By restriction, this map induces a map $g\colon \mathbb{S}^{m+1} \to \mathbb{S}^m$ such that $\operatorname{dis}(g) < \zeta_m$. By applying Corollary 5.8, one can modify g into an antipode-preserving map $\hat{g}\colon \mathbb{S}^{m+1} \to \mathbb{S}^m$ with $\operatorname{dis}(\hat{g}) < \zeta_m$, which contradicts Corollary 5.5. This yields the proof of the second claim of Theorem B.

Now, in order to prove the first claim of Theorem B that $d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n) \ge \frac{1}{2}\zeta_m$, suppose that Γ is a correspondence between $\mathbb{S}^m$ and $\mathbb{S}^n$ with $\operatorname{dis}(\Gamma) < \zeta_m$. Pick any function $g\colon \mathbb{S}^n \to \mathbb{S}^m$ such that $(g(x), x) \in \Gamma$ for every $x \in \mathbb{S}^n$. This implies that

$$\operatorname{dis}(g) \le \operatorname{dis}(\Gamma) < \zeta_m,$$

which contradicts the second claim. This proves the first claim. □

5.2 The proof of Theorem C

By carefully inspecting the proof of Theorem B, one can extract the much stronger Theorem C.

Proof of Theorem C We will actually prove slightly stronger result. Suppose

(i) X can be isometrically embedded into $\mathbb{S}^m$, and

(ii) $\boldsymbol{A}(\mathbb{S}^{m+1})$ (note that $\boldsymbol{A}(\mathbb{S}^{m+1}) \subset \boldsymbol{H}_{\ge 0}(\mathbb{S}^{m+1})$) can be isometrically embedded into Y.

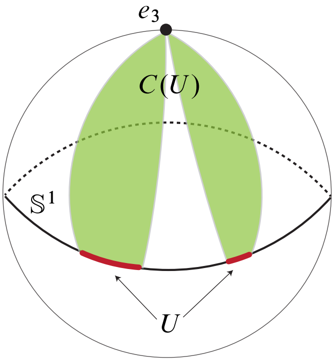


Figure 6: The cone $\mathcal{C}(U)$ for a subset U of $\mathbb{S}^1$.

Now, we prove that $d_{\mathrm{GH}}(X,Y)\geq \frac{1}{2}\zeta_m$. Moreover, $\mathrm{dis}(\phi)\geq\zeta_m$ for any map $\phi\colon Y\to X$.

We first prove the second claim. Suppose to the contrary, so that there is a map $\tilde{g}\colon Y\to X$ with $\mathrm{dis}(\tilde{g})<\zeta_m$. Then, since $\boldsymbol{A}(\mathbb{S}^{m+1})$ is isometrically embedded in Y and X is isometrically embedded in $\mathbb{S}^m$ by the assumption, one can construct a map $g\colon \boldsymbol{A}(\mathbb{S}^{m+1})\to\mathbb{S}^m$ with $\mathrm{dis}(g)<\zeta_m$. Hence, with the aid of Lemma 5.7, one can modify this g into an antipode-preserving map $\hat{g}\colon\mathbb{S}^{m+1}\to\mathbb{S}^m$ with $\mathrm{dis}(\hat{g})<\zeta_m$, which contradicts Corollary 5.5. This yields the proof of the second claim.

Now, in order to prove the first claim, use the same argument used in the proof of Theorem B. □

6 The proofs of Propositions 1.16 and 1.20

To prove Propositions 1.16 and 1.20, we need to define a few notions.

Definition 6.1 For any nonempty $U\subseteq\mathbb{S}^{n-1}$, we define *the cone of U*, as the following subset of $\mathbb{S}^n\subset\mathbb{R}^{n+1}$:

$$\mathcal{C}(U):=\left\{\cos\theta\cdot e_{n+1}+\sin\theta\cdot\iota_{n-1}(u)\in\boldsymbol{H}_{\geq 0}(\mathbb{S}^n)\mid u\in U\text{ and }\theta\in\left[0,\tfrac{\pi}{2}\right]\right\},$$

where $e_{n+1}=(0,0,\cdots,0,1)\in\mathbb{R}^{n+1}$ is the north pole of $\mathbb{S}^n$. See Figure 6.

Lemma 6.2 *For any nonempty $U\subseteq\mathbb{S}^{n-1}$,*

$$\mathrm{diam}(\mathcal{C}(U))=\begin{cases}\frac{\pi}{2} & \textit{if } \mathrm{diam}(U)\leq\frac{\pi}{2},\\ \mathrm{diam}(U) & \textit{if } \mathrm{diam}(U)\geq\frac{\pi}{2}.\end{cases}$$

Proof Recall that

$$\mathcal{C}(U) := \left\{\cos\theta \cdot e_{n+1} + \sin\theta \cdot \iota_{n-1}(u) \in \boldsymbol{H}_{\geq 0}(\mathbb{S}^n) \mid u \in U \text{ and } \theta \in \left[0, \tfrac{\pi}{2}\right]\right\}.$$

Now, for $u, v \in U$ and $\theta, \theta' \in \left[0, \frac{\pi}{2}\right]$, consider the inner product

$$\begin{aligned}\langle \cos\theta \cdot e_{n+1} + \sin\theta \cdot \iota_{n-1}(u), \cos\theta' \cdot e_{n+1} + \sin\theta' \cdot \iota_{n-1}(v)\rangle \\ = \cos\theta\cos\theta' + \langle u, v\rangle \cdot \sin\theta\sin\theta'.\end{aligned}$$

Hence, if $\langle u, v\rangle \geq 0$,

$$\langle \cos\theta \cdot e_{n+1} + \sin\theta \cdot \iota_{n-1}(u), \cos\theta' \cdot e_{n+1} + \sin\theta' \cdot \iota_{n-1}(v)\rangle \geq 0,$$

so $d_{\mathbb{S}^n}(\cos\theta \cdot e_{n+1} + \sin\theta \cdot u, \cos\theta' \cdot e_{n+1} + \sin\theta' \cdot v) \leq \frac{\pi}{2}$.

If $\langle u, v\rangle \leq 0$, $\cos\theta\cos\theta' + \langle u, v\rangle \cdot \sin\theta\sin\theta$ becomes decreasing in θ and θ'. Hence, it is minimized for $\theta = \theta' = \frac{\pi}{2}$. Therefore,

$$\langle \cos\theta \cdot e_{n+1} + \sin\theta \cdot \iota_{n-1}(u), \cos\theta' \cdot e_{n+1} + \sin\theta' \cdot \iota_{n-1}(v)\rangle \geq \langle u, v\rangle,$$

so $d_{\mathbb{S}^n}(\cos\theta \cdot e_{n+1} + \sin\theta \cdot \iota_{n-1}(u), \cos\theta' \cdot e_{n+1} + \sin\theta' \cdot \iota_{n-1}(v)) \leq d_{\mathbb{S}^{n-1}}(u, v)$, which completes the proof. □

Definition 6.3 (geodesic convex hull) Given a nonempty subset $A \subset \mathbb{S}^n$, its *geodesic convex hull* $\mathrm{conv}_{\mathbb{S}^n}(A)$ is defined to be the smallest subset of $\mathbb{S}^n$ containing A such that for any two points in the set, all minimizing geodesics between them are also contained in the set. It is clear that when A is contained in an open hemisphere,

$$\mathrm{conv}_{\mathbb{S}^n}(A) = \{\Pi_{\mathbb{S}^n}(c) \mid c \in \mathrm{conv}(A)\},$$

where $\Pi_{\mathbb{S}^n}(p) := p/\|p\|$ for $p \neq 0$ and $\Pi_{\mathbb{S}^n}(p) := 0$ otherwise.

In what follows we will prove Proposition 1.20 after proving Proposition 1.16. The proof of the former proposition generalizes the construction used in the proof of the latter one, and as a consequence Proposition 1.16 (which exhibits a correspondence between $\mathbb{S}^2$ and $\mathbb{S}^1$) is a special case of Proposition 1.20 (which constructs a correspondence between $\mathbb{S}^{m+1}$ and $\mathbb{S}^m$).

With the goal of making the construction more understandable, we have however decided to first present a detailed proof of Proposition 1.16 since the optimal $R_{2,1}$ correspondence constructed therein is used in the proof of Proposition 1.18 in order to construct an optimal correspondence $R_{3,1}$. After this we provide a streamlined proof of Proposition 1.20.

6.1 The proof of Proposition 1.16

We will find an upper bound for $d_{\mathrm{GH}}(\mathbb{S}^1, \boldsymbol{H}_{\geq 0}(\mathbb{S}^2))$ (resp. $d_{\mathrm{GH}}(\mathbb{S}^1, \mathbb{S}^2)$) by constructing a specific correspondence between $\mathbb{S}^1$ and $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2)$ (resp. $\mathbb{S}^1$ and $\mathbb{S}^2$). This correspondence is inspired by the case $m = 1$ of certain surjective maps from $\mathbb{S}^{m+1}$ to $\mathbb{S}^m$ [11, Scholium 1] developed in the course of the authors' study of the modulus of discontinuity of antipode-preserving maps between spheres. In spite of the fact that these maps will in general fail to yield tight upper bounds for distortion, they still permit giving nontrivial upper bounds for $\mathfrak{g}_{m,m+1}$. This will be explained in Section 6.2.

Proof of Proposition 1.16 We will prove both claims: that there exists

(1) a correspondence between $\mathbb{S}^1$ and $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2)$, and

(2) a correspondence between $\mathbb{S}^1$ and $\mathbb{S}^2$,

both of which have distortion at most $\frac{2\pi}{3}$ in an intertwined way.

In order to prove the first claim, it is enough to find a surjective map $\tilde{\phi}_{2,1} : \boldsymbol{H}_{\geq 0}(\mathbb{S}^2) \twoheadrightarrow \mathbb{S}^1$ (resp. $\phi_{2,1} : \mathbb{S}^2 \twoheadrightarrow \mathbb{S}^1$) such that $\operatorname{dis}(\tilde{\phi}_{2,1}) \leq \zeta_1 = \frac{2\pi}{3}$ (resp. $\operatorname{dis}(\phi_{2,1}) \leq \zeta_1 = \frac{2\pi}{3}$) since this map gives rise to a correspondence $\widetilde{R}_{2,1} := \operatorname{graph}(\tilde{\phi}_{2,1})$ (resp. $R_{2,1} := \operatorname{graph}(\phi_{2,1})$) with $\operatorname{dis}(\widetilde{R}_{2,1}) = \operatorname{dis}(\tilde{\phi}_{2,1}) \leq \zeta_1$ (resp. $\operatorname{dis}(R_{2,1}) = \operatorname{dis}(\phi_{2,1}) \leq \zeta_1$).

Let

$$u_1 := (1,0,0), \quad u_2 := \left(-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}, 0\right), \quad u_3 := \left(-\tfrac{1}{2}, -\tfrac{\sqrt{3}}{2}, 0\right).$$

Note that $\{u_1, u_2, u_3\}$ are the vertices of a regular triangle inscribed in $\boldsymbol{E}(\mathbb{S}^2)$. We divide the open upper hemisphere $\boldsymbol{H}_{>0}(\mathbb{S}^2)$ into three regions by using the Voronoi

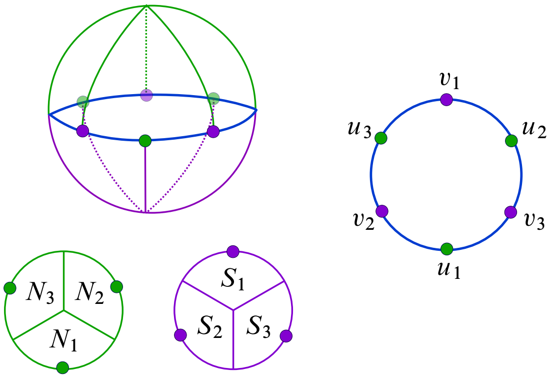


Figure 7: The surjection $\phi_{2,1} : \mathbb{S}^2 \twoheadrightarrow \mathbb{S}^1$ constructed in Proposition 1.16. In the figure, $S_i := -N_i$ and $v_i := -u_1$ for $i = 1,2,3$. The equator of $\mathbb{S}^2$ is mapped to itself under the map (via the identity map).

partitions induced by these three points. Precisely, for each $i = 1, 2, 3$ we define the set

$$N_i := \{x \in \boldsymbol{H}_{>0}(\mathbb{S}^2) \mid d_{\mathbb{S}^2}(x, u_i) \le d_{\mathbb{S}^2}(x, u_j) \text{ if } j \neq i \text{ and } d_{\mathbb{S}^2}(x, u_i) < d_{\mathbb{S}^2}(x, u_j) \text{ if } j < i\}.$$

See Figure 7 for an illustration of the construction.

Observe that $\overline{N_i} = \mathcal{C}\big(\mathrm{Conv}_{\mathbb{S}^1}(\{\iota_1^{-1}(-u_j) \in \mathbb{S}^1 \mid j \neq i\})\big)$ for each $i = 1, 2, 3$. Since $\mathrm{Conv}_{\mathbb{S}^1}(\{\iota_1^{-1}(-u_j) \in \mathbb{S}^1 \mid j \neq i\})$ is just the shortest geodesic between the two points $\{\iota_1(-u_j) \in \mathbb{S}^1 \mid j \neq i\}$ with length $\zeta_1 = \frac{2\pi}{3}$, $\mathrm{diam}(\overline{N_i}) \le \zeta_1$ by Lemma 6.2 for any $i = 1, 2, 3$.

We now construct a map $\tilde{\phi}_{2,1} \colon \boldsymbol{H}_{\ge 0}(\mathbb{S}^2) \to \mathbb{S}^1$,

$$\tilde{\phi}_{2,1}(p) := \begin{cases} \iota_1^{-1}(u_i) & \text{if } p \in N_i, \\ \iota_1^{-1}(p) & \text{if } p \in \boldsymbol{E}(\mathbb{S}^2). \end{cases}$$

Let us prove that the distortion of $\tilde{\phi}_{2,1}$ is less than or equal to ζ_1. We break the study of the value of

$$|d_{\mathbb{S}^2}(p, q) - d_{\mathbb{S}^1}(\tilde{\phi}_{2,1}(p), \tilde{\phi}_{2,1}(q))|$$

for $p, q \in \boldsymbol{H}_{\ge 0}(\mathbb{S}^2)$ into several cases:

(1) Suppose $p \in N_i$ and $q \in N_j$. If $i = j$, then $0 \le d_{\mathbb{S}^2}(p, q) \le \zeta_1$ and $\tilde{\phi}_{2,1}(p) = \tilde{\phi}_{2,1}(q) = \iota_m^{-1}(u_i)$, so $d_{\mathbb{S}^1}(\tilde{\phi}_{2,1}(p), \tilde{\phi}_{2,1}(q)) = 0$. Hence,

$$|d_{\mathbb{S}^2}(p, q) - d_{\mathbb{S}^1}(\tilde{\phi}_{2,1}(p), \tilde{\phi}_{2,1}(q))| \le \zeta_1.$$

If $i \neq j$, then $0 \le d_{\mathbb{S}^2}(p, q) \le \pi$ and $d_{\mathbb{S}^1}(\tilde{\phi}_{2,1}(p), \tilde{\phi}_{2,1}(q)) = \zeta_1$, so

$$|d_{\mathbb{S}^2}(p, q) - d_{\mathbb{S}^1}(\tilde{\phi}_{2,1}(p), \tilde{\phi}_{2,1}(q))| \le \zeta_1.$$

(2) Suppose $p \in N_i$ and $q \in \boldsymbol{E}(\mathbb{S}^2)$. Then

$$\begin{aligned} |d_{\mathbb{S}^2}(p, q) - d_{\mathbb{S}^1}(\tilde{\phi}_{2,1}(p), \tilde{\phi}_{2,1}(q))| &= |d_{\mathbb{S}^2}(p, q) - d_{\mathbb{S}^1}(\iota_1^{-1}(u_i), \iota_1^{-1}(q))| \\ &= |d_{\mathbb{S}^2}(p, q) - d_{\mathbb{S}^2}(u_i, q)| \\ &\le d_{\mathbb{S}^2}(p, u_i) \le \zeta_1. \end{aligned}$$

(3) Suppose $p, q \in \boldsymbol{E}(\mathbb{S}^2)$. Then $\tilde{\phi}_{2,1}(p) = \iota_1^{-1}(p)$ and $\tilde{\phi}_{2,1}(q) = \iota_1^{-1}(q)$. Hence,

$$|d_{\mathbb{S}^2}(p, q) - d_{\mathbb{S}^1}(\tilde{\phi}_{2,1}(p), \tilde{\phi}_{2,1}(q))| = 0 \le \zeta_1.$$

This implies that $\mathrm{dis}(\tilde{\phi}_{2,1}) \le \zeta_1$. Observe that $\tilde{\phi}_{2,1}$ is the identity on $\boldsymbol{E}(\mathbb{S}^2)$, so $\tilde{\phi}_{2,1}$ is surjective.

For the second claim, by applying Lemma 5.7 to $\tilde{\phi}_{2,1}|_{A(\mathbb{S}^2)}$, we construct a map $\phi_{2,1} : \mathbb{S}^2 \twoheadrightarrow \mathbb{S}^1$ such that $\operatorname{dis}(\phi_{2,1}) = \operatorname{dis}(\tilde{\phi}_{2,1}) \leq \zeta_1$. Moreover, by construction, $\phi_{2,1}$ is obviously surjective and antipode-preserving. □

Remark 6.4 The antipode-preserving property of $\phi_{2,1}$ will be useful for the proof of Proposition 1.18.

6.2 The proof of Proposition 1.20

One can prove Proposition 1.20 using a generalization of the approach used in the proof of Proposition 1.16.

Remark 6.5 (diameter of faces of geodesic simplices) Let $\{u_1, \dots, u_{m+2}\}$ be the vertices of a regular $(m{+}1)$–simplex inscribed in $\mathbb{S}^m$. Let

$$F_m := \operatorname{Conv}_{\mathbb{S}^m}(\{u_1, \dots, u_{m+1}\}).$$

In other words, F_m is just a *face* of the geodesic regular simplex inscribed in $\mathbb{S}^m$, where the length of each edge is $\zeta_m = \arccos(-1/(m+1))$.

The diameter of F_m can be determined by applying a result by Santaló [31, Lemma 1]:

$$\operatorname{diam}(F_m) = \eta_m := \begin{cases} \arccos(-(m+1)/(m+3)) & \text{if } m \text{ is odd,} \\ \arccos(-\sqrt{m/(m+4)}) & \text{if } m \text{ is even.} \end{cases}$$

As proved by Santaló, this diameter is realized either by the distance between the circumcenter of the geodesic convex hull of $A_m^{\text{odd}} := \{u_1, \dots, u_{(m+1)/2}\}$ and the circumcenter of the geodesic convex hull of $B_m^{\text{odd}} := \{u_{(m+3)/2}, \dots, u_{m+1}\}$ if m is odd, or by the distance between the circumcenter of the geodesic convex hull of $A_m^{\text{even}} := \{u_1, \dots, u_{m/2}\}$ and the circumcenter of the geodesic convex hull of $B_m^{\text{even}} := \{u_{(m+2)/2}, \dots, u_{m+1}\}$ if m is even. See Figure 8.

Observe that, in general,

$$\zeta_m \leq \eta_m \leq 2(\pi - \zeta_m).$$

Note that as m goes to infinity, ζ_m goes to $\frac{\pi}{2}$, η_m goes to π, and $2(\pi - \zeta_m)$ also goes to π.

Remark 6.6 Let $\{u_1, \dots, u_{m+2}\} \subset \mathbb{S}^m$ be the vertices of a regular $(m{+}1)$–simplex inscribed in $\mathbb{S}^m$. Let $V_1, \dots, V_{m+2}$ be the Voronoi partition of $\mathbb{S}^m$ induced by these vertices. Then $\overline{V_i} = \operatorname{Conv}_{\mathbb{S}^m}(\{-u_j : j \neq i\})$ (so $\overline{V_i}$ is congruent to F_m in Remark 6.5) for each $i = 1, \dots, m+2$. Here is a proof:

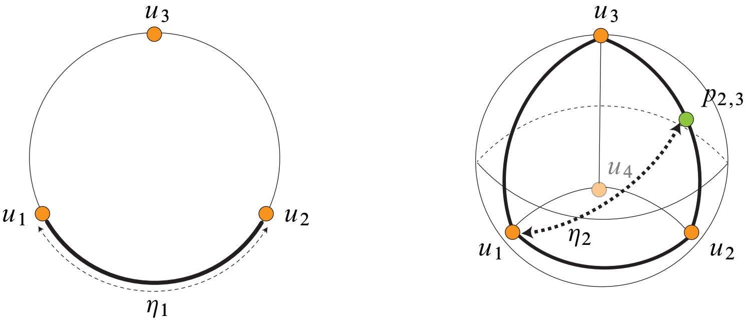


Figure 8: The diameter of a face F_m of a geodesic simplex; the cases $m=1$ and $m=2$. When $m=1$, $A_1^{\text{odd}}=\{u_1\}$ and $B_1^{\text{odd}}=\{u_2\}$. When $m=2$ (on the right), $A_2^{\text{even}}=\{u_1\}$, $B_2^{\text{even}}=\{u_2,u_3\}$ and the circumcenter of the geodesic convex hull of B_2^{even} is the point $p_{2,3}$, ie $\operatorname{diam}(F_2)=\eta_2=d_{\mathbb{S}^2}(u_1,p_{2,3})$.

Without loss of generality, one can assume $i=1$. Observe that

$$\overline{V}_1=\{x\in\mathbb{S}^m \mid d_{\mathbb{S}^m}(x,u_1)\le d_{\mathbb{S}^m}(x,u_j) \text{ for all } j\ne 1\}.$$

Now fix arbitrary $x\in\operatorname{Conv}_{\mathbb{S}^m}(\{-u_j\mid j\ne 1\})$. Then $x=v/\|v\|$ where

$$v=\sum_{j=2}^{m+2}\lambda_j(-u_j)$$

and the λ_j are nonnegative coefficients such that $\sum_{j=2}^{m+2}\lambda_j=1$. Then

$$\langle x,u_1\rangle=\frac{1}{\|v\|}\cdot\frac{1}{m+1}\cdot\sum_{j=2}^{m+2}\lambda_j=\frac{1}{\|v\|}\cdot\frac{1}{m+1}$$

and, for any $k\ne 1$,

$$\langle x,u_k\rangle=\frac{1}{\|v\|}\cdot\biggl(-1+\frac{1}{m+1}\cdot\sum_{\substack{2\le j\le m+2\\ j\ne k}}\lambda_j\biggr).$$

Hence, this implies $\langle x,u_1\rangle\ge\langle x,u_k\rangle$, so $d_{\mathbb{S}^m}(x,u_1)\le d_{\mathbb{S}^m}(x,u_k)$ for any $k\ne 1$. Therefore, $x\in\overline{V}_1$ and $\operatorname{Conv}_{\mathbb{S}^m}(\{-u_j : j\ne 1\})\subseteq\overline{V}_1$.

For the other direction, fix an arbitrary $x\in\overline{V}_1$. Since $\{-u_2,\dots,-u_{m+2}\}$ is a basis of $\mathbb{R}^{m+1}$, there is a unique set of coefficients $\{c_i\}_{i=2}^{m+2}$ such that $x=\sum_{i=2}^{m+2}c_i(-u_i)$. Then one can check $c_i=((m+1)/(m+2))(\langle x,u_1\rangle-\langle x,u_i\rangle)$ for $i=2,\dots,m+2$ by using the fact $\sum_{i=1}^{m+2}\langle x,u_i\rangle=\bigl\langle x,\sum_{i=1}^{m+2}u_i\bigr\rangle=\langle x,0\rangle=0$, and [13, Theorem 5.27] (the fact that $\sum_{i=1}^{m+2} u_i = 0$ can be easily checked by the induction on m). Note that $c_i \geq 0$ since $\langle x, u_1\rangle \geq \langle x, u_i\rangle$. Hence, if we define

$$\lambda_i := \frac{c_i}{\sum_{j=2}^{m+2} c_j} = \frac{1}{m+2}\left(1 - \frac{\langle x, u_i\rangle}{\langle x, u_1\rangle}\right)$$

for each $i = 2, \dots, m+2$ and $v := \sum_{i=2}^{m+2} \lambda_i(-u_i)$, then $x = v/\|v\|$. Therefore, $x \in \mathrm{Conv}_{\mathbb{S}^m}(\{-u_j \mid j \neq 1\})$ and $\overline{V}_1 \subseteq \mathrm{Conv}_{\mathbb{S}^m}(\{-u_j \mid j \neq 1\})$. Hence, $\overline{V}_1 = \mathrm{Conv}_{\mathbb{S}^m}(\{-u_j : j \neq 1\})$, as we claimed.

Proof of Proposition 1.20 We construct a surjective and antipode-preserving map

$$\phi_{(m+1),m}\colon \mathbb{S}^{m+1} \twoheadrightarrow \mathbb{S}^m$$

with

$$\mathrm{dis}(\phi_{(m+1),m}) \leq \eta_m.$$

Let $\{u_1, \dots, u_{m+2}\}$ be the vertices of a regular $(m+1)$–simplex inscribed in $\boldsymbol{E}(\mathbb{S}^{m+1})$. We divide open upper hemisphere $\boldsymbol{H}_{>0}(\mathbb{S}^{m+1})$ into $m+2$ regions by using the Voronoi partitions induced by these $m+2$ vertices. Precisely, for each $i = 1, \dots, m+2$ we define the set

$$N_i := \bigl\{p \in \boldsymbol{H}_{>0}(\mathbb{S}^{m+1}) \mid d_{\mathbb{S}^{m+1}}(p, u_i) \leq d_{\mathbb{S}^{m+1}}(p, u_j) \text{ for all } j \neq i \\ \text{and } d_{\mathbb{S}^{m+1}}(p, u_i) < d_{\mathbb{S}^{m+1}}(p, u_j) \text{ for all } j < i\bigr\}.$$

Observe that $\overline{N_i} = \mathcal{C}(\overline{V_i})$, where $\{V_1, \dots, V_{m+2}\}$ is the Voronoi partition of $\mathbb{S}^m$ induced by

$$\{\iota_m^{-1}(u_1), \dots, \iota_m^{-1}(u_{m+2})\}.$$

Hence, by Lemma 6.2 and Remarks 6.5 and 6.6, one concludes that $\mathrm{diam}(\overline{N_i}) \leq \eta_m$ for any $i = 1, \dots, m+2$.

We now construct a map $\tilde{\phi}_{(m+1),m}\colon \boldsymbol{A}(\mathbb{S}^{m+1}) \to \mathbb{S}^m$ by

$$\tilde{\phi}_{(m+1),m}(p) := \begin{cases} \iota_m^{-1}(u_i) & \text{if } p \in N_i, \\ \iota_m^{-1}(p) & \text{if } p \in \iota_m(\boldsymbol{A}(\mathbb{S}^m)). \end{cases}$$

In order to prove that the distortion of $\tilde{\phi}_{(m+1),m}$ is less than or equal to η_m we break the study of the value of

$$|d_{\mathbb{S}^{m+1}}(p, q) - d_{\mathbb{S}^m}(\tilde{\phi}_{(m+1),m}(p), \tilde{\phi}_{(m+1),m}(q))|$$

for $p, q \in \boldsymbol{A}(\mathbb{S}^{m+1})$ into several cases:

(1) Suppose $p \in N_i$ and $q \in N_j$. If $i = j$, then $d_{\mathbb{S}^{m+1}}(p,q) \le \eta_m$ and $\tilde{\phi}_{(m+1),m}(p) = \tilde{\phi}_{(m+1),m}(q) = \iota_m^{-1}(u_i)$, so $d_{\mathbb{S}^m}(\tilde{\phi}_{(m+1),m}(p), \tilde{\phi}_{(m+1),m}(q)) = 0$. Hence,

$$|d_{\mathbb{S}^{m+1}}(p,q) - d_{\mathbb{S}^m}(\tilde{\phi}_{(m+1),m}(p), \tilde{\phi}_{(m+1),m}(q))| \le \eta_m.$$

If $i \ne j$, then $d_{\mathbb{S}^{m+1}}(p,q) \le \pi$ and $d_{\mathbb{S}^m}(\tilde{\phi}_{(m+1),m}(p), \tilde{\phi}_{(m+1),m}(q)) = \zeta_m$, so that

$$|d_{\mathbb{S}^{m+1}}(p,q) - d_{\mathbb{S}^m}(\tilde{\phi}_{(m+1),m}(p), \tilde{\phi}_{(m+1),m}(q))| \le \zeta_m \le \eta_m.$$

(2) Suppose $p \in N_i$ and $q \in \iota_m(\boldsymbol{A}(\mathbb{S}^m))$. Then

$$\begin{aligned}
&|d_{\mathbb{S}^{m+1}}(p,q) - d_{\mathbb{S}^m}(\tilde{\phi}_{(m+1),m}(p), \tilde{\phi}_{(m+1),m}(q))| \\
&\qquad = |d_{\mathbb{S}^{m+1}}(p,q) - d_{\mathbb{S}^m}(\iota_m^{-1}(u_i), \iota_m^{-1}(q))| \\
&\qquad = |d_{\mathbb{S}^{m+1}}(p,q) - d_{\mathbb{S}^{m+1}}(u_i,q)| \\
&\qquad \le d_{\mathbb{S}^{m+1}}(p,u_i) \le \eta_m.
\end{aligned}$$

(3) Suppose $p, q \in \iota_m(\boldsymbol{A}(\mathbb{S}^m))$. Then $\tilde{\phi}_{(m+1),m}(p) = p$ and $\tilde{\phi}_{(m+1),m}(p) = q$. Hence,

$$|d_{\mathbb{S}^{m+1}}(p,q) - d_{\mathbb{S}^m}(\tilde{\phi}_{(m+1),m}(p), \tilde{\phi}_{(m+1),m}(q))| = 0 \le \eta_m.$$

This implies that $\operatorname{dis}(\tilde{\phi}_{(m+1),m}) \le \eta_m$. Finally, by applying Lemma 5.7 to $\tilde{\phi}_{(m+1),m}$, we construct the map $\phi_{(m+1),m} \colon \mathbb{S}^{m+1} \twoheadrightarrow \mathbb{S}^m$ such that

$$\operatorname{dis}(\phi_{(m+1),m}) = \operatorname{dis}(\tilde{\phi}_{(m+1),m}) \le \eta_m.$$

Moreover, by construction, $\phi_{(m+1),m}$ is obviously surjective and antipode-preserving. Therefore,

$$d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^{m+1}) \le \tfrac{1}{2}\eta_m. \qquad \square$$

Remark 6.7 Even though during the proof of Proposition 1.20 we only established the fact that $\operatorname{dis}(\phi_{(m+1),m}) \le \eta_m$, one can check that $\operatorname{dis}(\phi_{(m+1),m})$ is *exactly equal* to η_m, since one can choose two points $p, q \in N_i$ such that $d_{\mathbb{S}^{m+1}}(p,q)$ is arbitrarily close to η_m.

7 The proof of Proposition 1.18

In this section, we will prove Proposition 1.18 by constructing a specific correspondence between $\mathbb{S}^1$ and $\mathbb{S}^3$ with distortion less than or equal to $\zeta_1 = \frac{2\pi}{3}$. The construction of this correspondence is based on the optimal correspondence $R_{2,1} = \operatorname{graph}(\phi_{2,1})$ between $\mathbb{S}^1$ and $\mathbb{S}^2$ identified in the proof of Proposition 1.16 given in Section 6.1 and some ideas reminiscent of the Hopf fibration. We will define a surjective map $\phi_{3,1} \colon \mathbb{S}^3 \twoheadrightarrow \mathbb{S}^1$ by suitably "rotating" the (optimal) surjection $\phi_{2,1} \colon \mathbb{S}^2 \twoheadrightarrow \mathbb{S}^1$; see Figure 9.

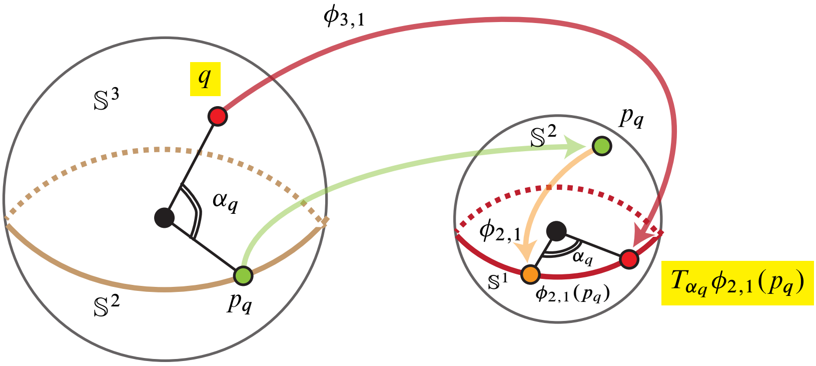


Figure 9: The definition of $\phi_{3,1}$: given $q \in \mathbb{S}^3 \setminus \mathbb{S}^2$ there exists a unique angle $\alpha_q \in (0,\pi)$ and unique point $p_q \in \mathbb{S}^2 \setminus \mathbb{S}^1$ such that $q = T_{\alpha_q} p_q$; then we consider the point $\phi_{2,1}(p_q) \in \mathbb{S}^1$ and define $\phi_{3,1}(q) := T_{\alpha_q}\phi_{2,1}(p_q)$. That $\phi_{3,1}(q) \in \mathbb{S}^1$ follows from Lemma 7.2(2).

Overview of the construction of $\boldsymbol{\phi_{3,1}}$ The diagram below describes the construction of the map $\phi_{3,1}$ at a high level:

$$\begin{tikzcd}
\mathbb{S}^3 \arrow[r, dotted, two heads, "\phi_{3,1}"] \arrow[d, "h"'] & \mathbb{S}^1 \\
\mathbb{S}^2 \times [0,\pi) \arrow[r, two heads, "\phi_{2,1}\times \mathrm{id}"] & \mathbb{S}^1 \times [0,\pi) \arrow[u, two heads, "T_\bullet"']
\end{tikzcd}$$

To an arbitrary $q \in \mathbb{S}^3$, we will be able to assign both a corresponding point $p_q \in \mathbb{S}^2$ and an angle $\alpha_q \in [0,\pi)$ giving rise to a map $h\colon \mathbb{S}^3 \to \mathbb{S}^2 \times [0,\pi)$ such that $h(q) := (p_q, \alpha_q)$. Also, $T_\bullet\colon \mathbb{S}^1 \times [0,\pi) \twoheadrightarrow \mathbb{S}^1$ will be a map such that for each $\alpha \in [0,\pi)$, T_α is a rotation of $\mathbb{S}^1$ by an angle α. Then, as described in the diagram, for $q \in \mathbb{S}^3$, $\phi_{3,1}(q)$ will be defined as $T_{\alpha_q}(\phi_{2,1}(p_q))$. Figures 9 and 10 illustrate the construction.

Note that there is a certain degree of similarity between the map $\pi_1 \circ h\colon \mathbb{S}^3 \twoheadrightarrow \mathbb{S}^2$ (where π_1 is the canonical projection from $\mathbb{S}^2 \times [0,\pi)$ to $\mathbb{S}^2$) and the "Hopf fibration", in the sense that the set $(\pi_1 \circ h)^{-1}(\{p,-p\})$ is isometric to $\mathbb{S}^1$ for $p \in \mathbb{S}^2 \setminus \mathbb{S}^1$ (whereas $(\pi_1 \circ h)^{-1}(\{p\}) = \{p\}$ for $p \in \mathbb{S}^1$).

Details The following coordinate representations will be used throughout this section:[4]

- $\mathbb{S}^1 := \{(x,y,0,0) \in \mathbb{R}^4 : x^2 + y^2 = 1\}$,

[4] In comparison to the coordinate representation specified in Section 2, here we are embedding $\mathbb{S}^1$, $\mathbb{S}^2$, and $\mathbb{S}^3$ into $\mathbb{R}^4$ in such a way that the embeddings $\mathbb{S}^1 \hookrightarrow \mathbb{S}^2 \hookrightarrow \mathbb{S}^3$ are also specific.

- $\mathbb{S}^2 := \{(x,y,z,0) \in \mathbb{R}^4 : x^2+y^2+z^2 = 1\}$,
- $\mathbb{S}^3 := \{(x,y,z,w) \in \mathbb{R}^4 : x^2+y^2+z^2+w^2 = 1\}$.

Also, we will use the map $\phi_{2,1} : \mathbb{S}^2 \twoheadrightarrow \mathbb{S}^1$ and the regions $N_1, N_2, N_3 \subset \mathbb{S}^2$ constructed in the proof of Proposition 1.16; see Section 6.1.

Remark 7.1 The following simple observations will be useful later. See Figure 7.

(1) $\operatorname{diam}(\overline{N}_i) \leq \zeta_1 = \frac{2\pi}{3}$ for any $i = 1,2,3$. (This fact has been already mentioned during the proof of Proposition 1.20.)

(2) If $p = (x,y,z,0) \in N_i$ and $q = (a,b,c,0) \in N_j$ for $(i,j) = (1,2)$, $(2,3)$ or $(3,1)$ (resp. $(i,j) = (2,1)$, $(3,2)$ or $(1,3)$), then $bx - ay \geq 0$ (resp. ≤ 0) and $\phi_{2,1}(p)$ and $\phi_{2,1}(q)$ are in clockwise (resp. counterclockwise) order.

Now, for any $\alpha \in \mathbb{R}$, consider the rotation matrix

$$T_\alpha := \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 & 0 \\ \sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & \cos\alpha & -\sin\alpha \\ 0 & 0 & \sin\alpha & \cos\alpha \end{pmatrix}.$$

For any $p \in \mathbb{S}^3$, $T_\alpha p$ denotes the result of matrix multiplication by viewing p as a 4 by 1 column vector according to the coordinate system described at the beginning of this section.

The following basic properties of these rotation matrices will be useful soon.

Lemma 7.2 *Let $\alpha, \beta \in \mathbb{R}$. Then*:

(1) *For any $q \in \mathbb{S}^3 \setminus \mathbb{S}^1$, there is a unique $p_q \in \mathbb{S}^2 \setminus \mathbb{S}^1$ and a unique $\alpha_q \in [0,\pi)$ such that $q = T_{\alpha_q} p_q$. In particular, $\alpha_q = 0$ if and only if $q \in \mathbb{S}^2 \setminus \mathbb{S}^1$.*

(2) *$\mathbb{S}^1$ and $\mathbb{S}^3$ are invariant with respect to the action of the rotation matrices T_α.*

(3) $T_\alpha T_\beta = T_{\alpha+\beta}$.

(4) *$d_{\mathbb{S}^3}(T_\alpha\, p, T_\alpha\, q) = d_{\mathbb{S}^3}(p,q)$ for any $p,q \in \mathbb{S}^3$.*

(5) *$d_{\mathbb{S}^3}(T_\alpha\, p, p) = \alpha$ for any $p \in \mathbb{S}^3$ and $\alpha \in [0,\pi]$.*

(6) *$d_{\mathbb{S}^3}(T_\alpha(-p), p) = \pi - \alpha$ for any $p \in \mathbb{S}^3$ and $\alpha \in [0,\pi]$.*

Proof (1) Let $q = (x', y', z', w') \in \mathbb{S}^3 \setminus \mathbb{S}^1$. Since q is not in $\mathbb{S}^1$, we know that $(z')^2 + (w')^2 > 0$. Then there exists a unique $\alpha_q \in [0, \pi)$ and $z \in \mathbb{R} \setminus \{0\}$ such that

$$\begin{pmatrix} z' \\ w' \end{pmatrix} = \begin{pmatrix} \cos\alpha_q & -\sin\alpha_q \\ \sin\alpha_q & \cos\alpha_q \end{pmatrix} \begin{pmatrix} z \\ 0 \end{pmatrix};$$

ie $z^2 = (z')^2 + (w')^2$. Then this α_q is the required angle and we choose the unique point $p_q = (x, y, z, 0) \in \mathbb{S}^2 \setminus \mathbb{S}^1$ such that

$$\begin{pmatrix} x' \\ y' \\ z' \\ w' \end{pmatrix} = \begin{pmatrix} \cos\alpha_q & -\sin\alpha_q & 0 & 0 \\ \sin\alpha_q & \cos\alpha_q & 0 & 0 \\ 0 & 0 & \cos\alpha_q & -\sin\alpha_q \\ 0 & 0 & \sin\alpha_q & \cos\alpha_q \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 0 \end{pmatrix}.$$

Since T_{α_q} is the identity matrix when $\alpha_q = 0$, it is clear that $\alpha_q = 0$ if and only if $q \in \mathbb{S}^2 \setminus \mathbb{S}^1$.

(2) Obvious.

(3) Obvious.

(4) This item is equivalent to the condition $\langle T_\alpha p, T_\alpha q\rangle = \langle p, q\rangle$, and it can be easily checked by direct computation.

(5) This item is equivalent to the condition $\langle T_\alpha p, p\rangle = \cos\alpha$, and it can be easily checked by direct computation.

(6) This item is equivalent to the condition $\langle T_\alpha(-p), p\rangle = -\cos\alpha$, and it can be easily checked by direct computation. □

Additional details and the proof of Proposition 1.18 We need a few more definitions and technical lemmas for the proof of Proposition 1.18. We in particular make the following definitions for notational convenience:

- For any $p, q \in \mathbb{S}^2$,

$$E_{p,q} : [0, \pi] \to [-1, 1], \quad \alpha \mapsto \langle T_\alpha p, q\rangle.$$

- For any $p, q \in \mathbb{S}^2$,

$$F_{p,q} : [0, \pi] \to \mathbb{R}, \quad \alpha \mapsto d_{\mathbb{S}^3}(T_\alpha p, q) - \alpha.$$

- For any $p, q \in \mathbb{S}^2$,

$$G_{p,q} : [0, \pi] \to \mathbb{R}, \quad \alpha \mapsto d_{\mathbb{S}^3}(T_\alpha\, p, q) + \alpha.$$

Lemma 7.3 *For any $p=(x,y,z,0)\in\mathbb{S}^2\setminus\mathbb{S}^1$ and $q=(a,b,c,0)\in\mathbb{S}^2$:*

(1) $E_{p,q}(\alpha)\in(-1,1)$ *for any* $\alpha\in(0,\pi)$.

(2) $(E'_{p,q}(\alpha))^2+(E_{p,q}(\alpha))^2\le 1$ *for any* $\alpha\in[0,\pi]$.[5]

(3) $F_{p,q}$ *is a nonincreasing function. Thus,* $-d_{\mathbb{S}^2}(p,q)\le F_{p,q}(\alpha)\le d_{\mathbb{S}^2}(p,q)$ *for any* $\alpha\in[0,\pi]$.

(4) $G_{p,q}$ *is a nondecreasing function. Thus,* $d_{\mathbb{S}^2}(p,q)\le G_{p,q}(\alpha)\le 2\pi-d_{\mathbb{S}^2}(p,q)$ *for any* $\alpha\in[0,\pi]$.

Proof (1) Suppose not, so that $E_{p,q}(\alpha)=\pm1$. This implies that $T_\alpha p=q$ or $-q\in\mathbb{S}^2$. This cannot be true because $\alpha\in(0,\pi)$ and, by Lemma 7.2(1), $T_\alpha p\in\mathbb{S}^3\setminus\mathbb{S}^2$. So this is a contradiction; hence, $E_{p,q}(\alpha)\in(-1,1)$.

(2) As a result of direct computation, we know that

$$E_{p,q}(\alpha)=\langle p,q\rangle\cos\alpha+(bx-ay)\sin\alpha.$$

Here, observe that $bx-ay$ is the 3rd coordinate of the cross product $(x,y,z)\times(a,b,c)$. In particular, this implies $|bx-ay|\le\|(x,y,z)\times(a,b,c)\|=\sin\beta$ where $\langle p,q\rangle=\cos\beta$. Therefore,

$$(E'_{p,q}(\alpha))^2+(E_{p,q}(\alpha))^2=\langle p,q\rangle^2+(bx-ay)^2\le\cos^2\beta+\sin^2\beta=1.$$

(3) Note that $F_{p,q}(\alpha)=\arccos(E_{p,q}(\alpha))-\alpha$. Hence, for any $\alpha\in(0,\pi)$,

$$F'_{p,q}(\alpha)=-\frac{E'_{p,q}(\alpha)}{\sqrt{1-(E_{p,q}(\alpha))^2}}-1.$$

Observe that this expression is well defined by (1). Also, by (2),

$$\begin{aligned}(E'_{p,q}(\alpha))^2+(E_{p,q}(\alpha))^2\le 1 &\iff -E'_{p,q}(\alpha)\le\sqrt{1-(E_{p,q}(\alpha))^2}\\ &\iff F'_{p,q}(\alpha)=-\frac{E'_{p,q}(\alpha)}{\sqrt{1-(E_{p,q}(\alpha))^2}}-1\le 0.\end{aligned}$$

Hence, $F_{p,q}$ is a nonincreasing function. Also, since $F_{p,q}(0)=d_{\mathbb{S}^2}(p,q)$ and

$$F_{p,q}(\pi)=d_{\mathbb{S}^3}(T_\pi p,q)-\pi=d_{\mathbb{S}^2}(-p,q)-\pi=(\pi-d_{\mathbb{S}^2}(p,q))-\pi=-d_{\mathbb{S}^2}(p,q),$$

we have that

$$-d_{\mathbb{S}^2}(p,q)\le F_{p,q}(\alpha)\le d_{\mathbb{S}^2}(p,q).$$

[5] Here E'_{pq} denotes the derivative of $E_{p,q}$.

(4) Note that $G_{p,q}(\alpha) = \arccos(E_{p,q}(\alpha)) + \alpha$. Hence, for any $\alpha \in (0,\pi)$,

$$G'_{p,q}(\alpha) = -\frac{E'_{p,q}(\alpha)}{\sqrt{1-(E_{p,q}(\alpha))^2}} + 1.$$

Observe that this expression is well defined by (1). Also, by (2),

$$\begin{aligned}(E'_{p,q}(\alpha))^2 + (E_{p,q}(\alpha))^2 \le 1 &\iff E'_{p,q}(\alpha) \le \sqrt{1-(E_{p,q}(\alpha))^2}\\ &\iff G'_{p,q}(\alpha) = -\frac{E'_{p,q}(\alpha)}{\sqrt{1-(E_{p,q}(\alpha))^2}} + 1 \ge 0.\end{aligned}$$

Hence, $G_{p,q}$ is nondecreasing function. Also, since $G_{p,q}(0) = d_{\mathbb{S}^2}(p,q)$ and

$$G_{p,q}(\pi) = d_{\mathbb{S}^3}(T_\pi p, q) + \pi = d_{\mathbb{S}^2}(-p,q) + \pi = (\pi - d_{\mathbb{S}^2}(p,q)) + \pi = 2\pi - d_{\mathbb{S}^2}(p,q),$$

we have that

$$d_{\mathbb{S}^2}(p,q) \le G_{p,q}(\alpha) \le 2\pi - d_{\mathbb{S}^2}(p,q). \qquad \square$$

Lemma 7.4 *For any* $p = (x,y,z,0), q = (a,b,c,0) \in \mathbb{S}^2 \setminus \mathbb{S}^1$:

(1) *If* $p \in N_i$ *and* $q \in N_j$ *for* $(i,j) = (1,2)$, $(2,3)$ *or* $(3,1)$, *then*

$$d_{\mathbb{S}^3}(T_{2\pi/3}p, q) \le \tfrac{2\pi}{3}.$$

(2) *If* $p \in N_i$ *and* $q \in N_j$ *for* $(i,j) = (2,1)$, $(3,2)$ *or* $(1,3)$, *then*

$$d_{\mathbb{S}^3}(T_{\pi/3}p, q) \ge \tfrac{\pi}{3}.$$

Proof (1) First, observe that $bx - ay \ge 0$ by Remark 7.1(2). Hence,

$$E_{p,q}\big(\tfrac{2\pi}{3}\big) = \langle T_{2\pi/3}p, q\rangle = -\tfrac{1}{2}\langle p,q\rangle + \tfrac{\sqrt{3}}{2}(bx-ay) \ge -\tfrac{1}{2}\langle p,q\rangle \ge -\tfrac{1}{2}.$$

Therefore,

$$d_{\mathbb{S}^3}(T_{2\pi/3}p, q) = \arccos\big(E_{p,q}\big(\tfrac{2\pi}{3}\big)\big) \le \arccos\big(-\tfrac{1}{2}\big) = \tfrac{2\pi}{3}.$$

(2) The proof of this case is similar to the proof of (1), so we omit it. □

Proof of Proposition 1.18 It is enough to find a surjective map $\phi_{3,1} : \mathbb{S}^3 \twoheadrightarrow \mathbb{S}^1$ such that $\operatorname{dis}(\phi_{3,1}) \le \zeta_1 = \frac{2\pi}{3}$, since this map gives rise to a correspondence $R_{3,1} := \operatorname{graph}(\phi_{3,1})$ with $\operatorname{dis}(R_{3,1}) = \operatorname{dis}(\phi_{3,1}) \le \zeta_1$.

We construct the required surjective map $\phi_{3,1} : \mathbb{S}^3 \twoheadrightarrow \mathbb{S}^1$ as

$$q \mapsto \begin{cases} \phi_{2,1}(q) & \text{if } q \in \mathbb{S}^2, \\ T_{\alpha_q}\phi_{2,1}(p_q) & \text{if } q \in \mathbb{S}^3 \setminus \mathbb{S}^2 \text{ and } q = T_{\alpha_q}p_q \text{ for the } \textit{unique} \text{ such} \\ & \qquad \alpha_q \in (0,\pi) \text{ and } p_q \in \mathbb{S}^2 \setminus \mathbb{S}^1. \end{cases}$$

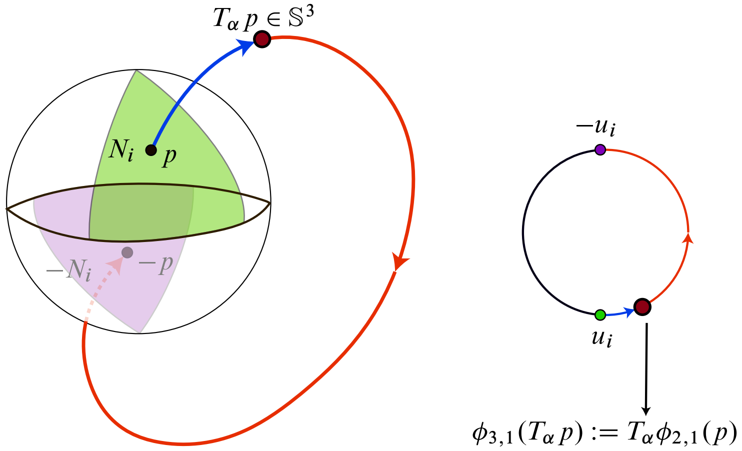

$T_\alpha p \in \mathbb{S}^3$
N_i
p
$-N_i$
$-p$
$-u_i$
u_i
$\phi_{3,1}(T_\alpha p) := T_\alpha \phi_{2,1}(p)$

Figure 10: The definition of the map $\phi_{3,1}$ via the map $\phi_{2,1}$. The point $T_\alpha p$ on $\mathbb{S}^3$ is mapped to the point $T_\alpha\phi_{2,1}(p)$ on $\mathbb{S}^1$. The antipode-preserving map $\phi_{2,1}$ maps the whole region N_i to the point u_i.

Note that $\phi_{3,1}$ is surjective, since $\phi_{3,1}|_{\mathbb{S}^2} = \phi_{2,1}$ and $\phi_{2,1}$ is surjective.

See Figures 9 and 10 for an explanation of the construction of the map $\phi_{3,1}$.

Let us now verify that

$$|d_{\mathbb{S}^3}(q,q') - d_{\mathbb{S}^1}(\phi_{3,1}(q),\phi_{3,1}(q'))| \le \zeta_1$$

for every $q, q' \in \mathbb{S}^3$. Without loss of generality, we can assume that $q = T_\alpha p$ and $q' = T_\beta p'$ for some $p, p' \in \mathbb{S}^2$ and $0 \le \beta \le \alpha < \pi$. Then

$$\begin{aligned}|d_{\mathbb{S}^3}(q,q') - d_{\mathbb{S}^1}(\phi_{3,1}(q),\phi_{3,1}(q'))| \\ &= |d_{\mathbb{S}^3}(T_\alpha p, T_\beta p') - d_{\mathbb{S}^1}(T_\alpha\phi_{2,1}(p), T_\beta\phi_{2,1}(p'))| \\ &= |d_{\mathbb{S}^3}(T_{(\alpha-\beta)}p, p') - d_{\mathbb{S}^1}(T_{(\alpha-\beta)}\phi_{2,1}(p), \phi_{2,1}(p'))|.\end{aligned}$$

Hence, it is enough to prove

$$\bigl|d_{\mathbb{S}^3}(T_\alpha p, q) - d_{\mathbb{S}^1}(T_\alpha\phi_{2,1}(p), \phi_{2,1}(q))\bigr| \le \zeta_1 \tag{9}$$

for any $p, q \in \mathbb{S}^2$ and $\alpha \in [0,\pi)$.

If $p \in \mathbb{S}^1$, then $\phi_{2,1}(p) = p$. Hence,

$$\begin{aligned}|d_{\mathbb{S}^3}(T_\alpha p, q) - d_{\mathbb{S}^1}(T_\alpha\phi_{2,1}(p), \phi_{2,1}(q))| &= |d_{\mathbb{S}^3}(T_\alpha p, q) - d_{\mathbb{S}^1}(T_\alpha p, \phi_{2,1}(q))| \\ &\le d_{\mathbb{S}^3}(q, \phi_{2,1}(q)) \le \zeta_1,\end{aligned}$$

where the last inequality holds by Remark 7.1(1). One can carry out a similar computation if $q \in \mathbb{S}^1$. So let's assume $p = (x, y, z, 0), q = (a, b, c, 0) \in \mathbb{S}^2 \setminus \mathbb{S}^1$. Since $\phi_{2,1}$ is antipode-preserving, it is enough to check inequality (9) only for $p, q \in \boldsymbol{H}_{>0}(\mathbb{S}^2)$. We do this by following the same idea as in the proof of Lemma 5.7.

We do a case-by-case analysis:

(1) Suppose $p \in N_i$ and $q \in N_j$ for $(i, j) = (1, 2)$, $(2, 3)$ or $(3, 1)$. By Remark 7.1(2), the two points $\phi_{2,1}(p)$ and $\phi_{2,1}(q)$ are in clockwise order. Hence,

$$d_{\mathbb{S}^1}(T_\alpha \phi_{2,1}(p), \phi_{2,1}(q)) = \begin{cases} \frac{2\pi}{3} - \alpha & \text{if } \alpha \in \left[0, \frac{2\pi}{3}\right], \\ \alpha - \frac{2\pi}{3} & \text{if } \alpha \in \left[\frac{2\pi}{3}, \pi\right). \end{cases}$$

Consider first the case when $\alpha \in \left[0, \frac{2\pi}{3}\right]$. We have to prove that

$$-\tfrac{2\pi}{3} \le d_{\mathbb{S}^3}(T_\alpha p, q) - \left(\tfrac{2\pi}{3} - \alpha\right) \le \tfrac{2\pi}{3}.$$

Equivalently, we have to prove

$$0 \le G_{p,q}(\alpha) \le \tfrac{4\pi}{3}.$$

The left-hand side inequality is obvious since $G_{p,q}(\alpha) \ge d_{\mathbb{S}^2}(p, q) \ge 0$ by Lemma 7.3(4). The right-hand side inequality is true by Lemmas 7.3(4) and 7.4(1).

Next, consider the case when $\alpha \in \left[\frac{2\pi}{3}, \pi\right)$. We have to prove

$$-\tfrac{2\pi}{3} \le d_{\mathbb{S}^3}(T_\alpha p, q) - \left(\alpha - \tfrac{2\pi}{3}\right) \le \tfrac{2\pi}{3}.$$

Equivalently, we have to prove

$$-\tfrac{4\pi}{3} \le F_{p,q}(\alpha) \le 0.$$

The left inequality is obvious since $F_{p,q}(\alpha) \ge -d_{\mathbb{S}^2}(p, q) \ge -\frac{4\pi}{3}$ by Lemma 7.3(3). The right-hand side inequality is true by Lemmas 7.3(3) and 7.4(1).

(2) Suppose $p \in N_i$ and $q \in N_j$ for $(i, j) = (2, 1)$, $(3, 2)$ or $(1, 3)$. This is almost the same as case (1) except we use Lemma 7.4(2).

(3) Suppose $p, q \in N_i$ for $i = 1, 2, 3$. In this case, $d_{\mathbb{S}^1}(T_\alpha \phi_{2,1}(p), \phi_{2,1}(q)) = \alpha$, which follows from $\phi_{2,1}(p) = \phi_{2,1}(q)$ and Lemma 7.2(5). Hence, we have to show

$$-\tfrac{2\pi}{3} \le d_{\mathbb{S}^3}(T_\alpha p, q) - \alpha = F_{p,q}(\alpha) \le \tfrac{2\pi}{3}.$$

But this is obvious by Remark 7.1(1) and Lemma 7.3(3).

Thus, indeed $\operatorname{dis}(\phi_{3,1}) \le \zeta_1$. □

8 The proof of Proposition 1.19

In this section we provide a construction of an optimal correspondence, $R_{3,2}$, between $\mathbb{S}^3$ and $\mathbb{S}^2$. The structure of this correspondence is different from those described in the proofs of Propositions 1.16 and 1.20. As a matter of fact, as Remark 6.7 mentions, the distortion of the surjection $\phi_{(m+1),m} \colon \mathbb{S}^{m+1} \twoheadrightarrow \mathbb{S}^m$ constructed in Proposition 1.20 is *exactly equal* to η_m. Since $\zeta_2 < \eta_2$, this means that a different construction is required for the case $m = 2$.

Let u_1, u_2, u_3 and u_4 be the vertices of a regular tetrahedron inscribed in $\mathbb{S}^2$ (ie $\langle u_i, u_j \rangle = -\frac{1}{3} = \cos\zeta_2$ for any $i \neq j$). We consider

$$u_1 = (1,0,0), \qquad u_2 = \left(-\tfrac{1}{3}, \tfrac{2\sqrt{2}}{3}, 0\right),$$
$$u_3 = \left(-\tfrac{1}{3}, -\tfrac{\sqrt{2}}{3}, \tfrac{\sqrt{2}}{\sqrt{3}}\right), \quad u_4 = \left(-\tfrac{1}{3}, -\tfrac{\sqrt{2}}{3}, -\tfrac{\sqrt{2}}{\sqrt{3}}\right).$$

Now, let $V_1, V_2, V_3, V_4 \subset \mathbb{S}^2$ be the Voronoi partition of $\mathbb{S}^2$ induced by u_1, u_2, u_3, and u_4. Then, for each i, $\overline{V_i}$ is the spherical convex hull of the set

$$\{-u_j \in \mathbb{S}^2 \mid j \in \{1,2,3,4\} \setminus \{i\}\}.$$

Let

$$r := \arccos\left(\tfrac{2\sqrt{2}}{3}\right).$$

For $i \neq j \in \{1,2,3,4\}$, let $u_{i,j}$ be the point on the shortest geodesic between u_i and $-u_j$ such that $d_{\mathbb{S}^2}(u_i, u_{i,j}) = r$. See Figure 11 for an illustration of V_1.

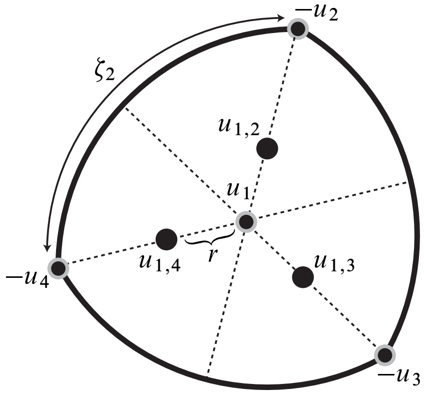


Figure 11: V_1. All the sides of the spherical triangle V_1 (determined by the three points $-u_2$, $-u_3$, and $-u_4$) have the same length ζ_2.

Remark 8.1 One can directly compute the coordinates

$$u_{1,2}=\big(\tfrac{2\sqrt{2}}{3},-\tfrac{1}{3},0\big),\quad u_{1,3}=\big(\tfrac{2\sqrt{2}}{3},\tfrac{1}{6},-\tfrac{1}{2\sqrt{3}}\big),\qquad u_{1,4}=\big(\tfrac{2\sqrt{2}}{3},\tfrac{1}{6},\tfrac{1}{2\sqrt{3}}\big),$$
$$u_{2,1}=\big(-\tfrac{4\sqrt{2}}{9},\tfrac{7}{9},0\big),\quad u_{2,3}=\big(-\tfrac{\sqrt{2}}{9},\tfrac{17}{18},-\tfrac{1}{2\sqrt{3}}\big),\quad u_{2,4}=\big(-\tfrac{\sqrt{2}}{9},\tfrac{17}{18},\tfrac{1}{2\sqrt{3}}\big).$$

Lemma 8.2 *For any $i\neq j\in\{1,2,3,4\}$, the following hold*:

(1) $\langle u_{i,k},u_{i,l}\rangle=\frac{5}{6}$ *for any $k\neq l\in 1,2,3,4\setminus\{i\}$.*

(2) $\langle u_{i,k},u_{j,k}\rangle=\frac{5}{54}$ *for any $k\in 1,2,3,4\setminus\{i,j\}$.*

(3) $\langle u_{i,k},u_{j,l}\rangle=-\frac{2}{27}$ *for any $k\neq l\in 1,2,3,4\setminus\{i,j\}$.*

(4) $\langle u_{i,k},u_{j,i}\rangle=-\frac{25}{54}$ *for any $k\in 1,2,3,4\setminus\{i,j\}$.*

(5) $\langle u_{i,j},u_{j,i}\rangle=-\frac{23}{27}$.

(6) $\langle u_{i},u_{j,k}\rangle=-\frac{\sqrt{2}}{9}$ *for any $k\in 1,2,3,4\setminus\{i,j\}$.*

(7) $\langle u_{i},u_{j,i}\rangle=-\frac{4\sqrt{2}}{9}$.

Proof By symmetry, without loss of generality one can assume $i=1$ and $j=2$. Then use the coordinate values given in Remark 8.1. □

Next, for each i, let $\{V_{i,j}\subset V_i\mid j\in\{1,2,3,4\}\setminus\{i\}\}$ be the Voronoi partition of V_i induced by $\{u_{i,j}\in V_i\mid j\in\{1,2,3,4\}\setminus\{i\}\}$.

From now on, in this section, we will identify $\mathbb{S}^2$ with $\boldsymbol{E}(\mathbb{S}^3)\subset\mathbb{S}^3$. Then obviously

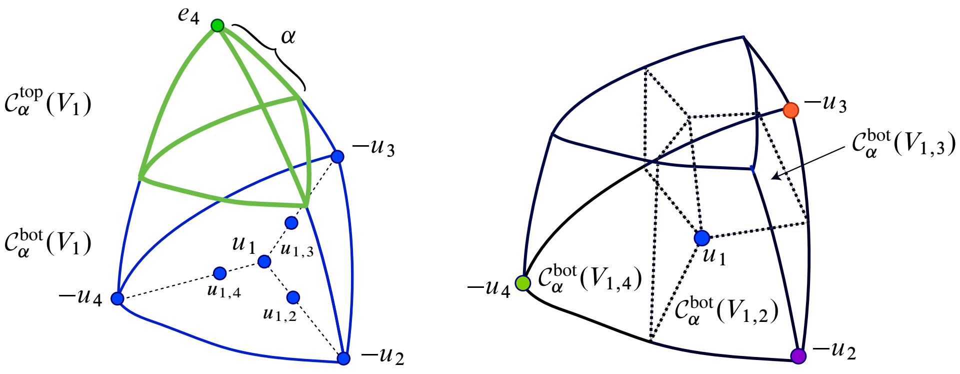
$$\boldsymbol{H}_{\geq 0}(\mathbb{S}^3)=\mathcal{C}(V_1)\cup\mathcal{C}(V_2)\cup\mathcal{C}(V_3)\cup\mathcal{C}(V_4).$$

Moreover, for any $i\in\{1,2,3,4\}$ and $\alpha\in\big[0,\frac{\pi}{2}\big]$, we divide $\mathcal{C}(V_i)$ into

- $\mathcal{C}^{\mathrm{top}}_{\alpha}(V_i):=\{p\in\mathcal{C}(V_i)\mid d_{\mathbb{S}^{n+1}}(e_4,p)\leq\alpha\}$,
- $\mathcal{C}^{\mathrm{bot}}_{\alpha}(V_i):=\{p\in\mathcal{C}(V_i)\mid d_{\mathbb{S}^{n+1}}(e_4,p)>\alpha\}$,
- $\mathcal{C}^{\mathrm{bot}}_{\alpha}(V_{i,j}):=\{p\in\mathcal{C}(V_i)\mid d_{\mathbb{S}^{n+1}}(e_4,p)>\alpha$ and $\Omega(p)\in V_{i,j}\}$ for any j in $\{1,2,3,4\}\setminus\{i\}$,

where

$$\Omega\colon\boldsymbol{H}_{\geq 0}(\mathbb{S}^3)\setminus\{e_4\}\to\boldsymbol{E}(\mathbb{S}^3)=\mathbb{S}^2,\quad (x,y,z,w)\mapsto\frac{1}{\sqrt{1-w^2}}(x,y,z,0),$$


e_4
α
$\mathcal{C}_\alpha^{\mathrm{top}}(V_1)$
$-u_3$
$\mathcal{C}_\alpha^{\mathrm{bot}}(V_1)$
u_1
$u_{1,3}$
$-u_4$
$u_{1,4}$
$u_{1,2}$
$-u_2$
$-u_3$
$\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,3})$
$-u_4$
$\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,4})$
u_1
$\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,2})$
$-u_2$

$\mathcal{C}_\alpha^{\mathrm{top}}(V_1)$ and $\mathcal{C}_\alpha^{\mathrm{bot}}(V_1)$.

$\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,2})$, $\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,3})$, and $\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,4})$.

Figure 12: The regions into which $\mathcal{C}(V_1)$ is split.

is the orthogonal projection onto the equator. Then obviously

$$\mathcal{C}(V_i) = \mathcal{C}_\alpha^{\mathrm{top}}(V_i) \cup \bigcup_{j \in \{1,2,3,4\} \setminus \{i\}} \mathcal{C}_\alpha^{\mathrm{bot}}(V_{i,j})$$

for each $i \in \{1,2,3,4\}$. See Figure 12 for illustrations of $\mathcal{C}_\alpha^{\mathrm{top}}(V_1)$, $\mathcal{C}_\alpha^{\mathrm{bot}}(V_1)$, $\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,2})$, $\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,3})$, and $\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,4})$.

Lemma 8.3 *For $p, q \in \boldsymbol{H}_{\geq 0}(\mathbb{S}^3)$, the following inequalities hold*:

(1) *If $p, q \in \mathcal{C}_\alpha^{\mathrm{top}}(V_i)$ for some $i \in \{1,2,3,4\}$, then*

$$\langle p, q \rangle \geq \cos^2\alpha - \tfrac{1}{\sqrt{3}} \sin^2\alpha = \left(1 + \tfrac{1}{\sqrt{3}}\right)\cos^2\alpha - \tfrac{1}{\sqrt{3}}.$$

In particular, this is equivalent to

$$d_{\mathbb{S}^3}(p,q) \leq \arccos\left(\left(1 + \tfrac{1}{\sqrt{3}}\right)\cos^2\alpha - \tfrac{1}{\sqrt{3}}\right).$$

(2) *If $p \in \mathcal{C}_\alpha^{\mathrm{top}}(V_i)$ and $q \in \mathcal{C}_\alpha^{\mathrm{bot}}(V_{j,i})$ for some $i \neq j \in \{1,2,3,4\}$, then*

$$\langle p, q \rangle \leq \sqrt{\tfrac{2}{3}\cos^2\alpha + \tfrac{1}{3}}.$$

In particular, this is equivalent to

$$d_{\mathbb{S}^3}(p,q) \geq \arccos\left(\sqrt{\tfrac{2}{3}\cos^2\alpha + \tfrac{1}{3}}\right).$$

(3) *If $p \in \mathcal{C}_\alpha^{\mathrm{bot}}(V_{i,k})$ and $q \in \mathcal{C}_\alpha^{\mathrm{bot}}(V_{j,i})$ for distinct $i, j, k \in \{1,2,3,4\}$, then*

$$\langle p, q \rangle \leq \left(1 - \tfrac{1}{\sqrt{3}}\right)\cos^2\alpha + \tfrac{1}{\sqrt{3}}.$$

In particular, this is equivalent to the condition

$$d_{\mathbb{S}^3}(p,q) \geq \arccos\bigl(\bigl(1-\tfrac{1}{\sqrt{3}}\bigr)\cos^2\alpha + \tfrac{1}{\sqrt{3}}\bigr).$$

(4) *If $p \in \mathcal{C}_\alpha^{\mathrm{bot}}(V_{i,j})$ and $q \in \mathcal{C}_\alpha^{\mathrm{bot}}(V_{j,i})$ for some $i \neq j \in \{1,2,3,4\}$, then*

$$\langle p,q\rangle \leq \cos^2\alpha.$$

In particular, this is equivalent to

$$d_{\mathbb{S}^3}(p,q) \geq \arccos(\cos^2\alpha).$$

Proof We express p and q as

$$p = \cos\theta\cdot e_4 + \sin\theta\cdot\iota_2(x), \quad q = \cos\theta'\cdot e_4 + \sin\theta'\cdot\iota_2(y),$$

where $e_4 = (0,0,0,1)$ for some $\theta,\theta' \in \bigl[0,\frac{\pi}{2}\bigr]$ and $x,y \in \mathbb{S}^2$. Then

$$\langle p,q\rangle = \cos\theta\cos\theta' + \langle x,y\rangle\sin\theta\sin\theta'.$$

(1) If $p,q \in \mathcal{C}_\alpha^{\mathrm{top}}(V_i)$ for some $i \in \{1,2,3,4\}$, then we can assume $x,y \in V_i$ and $\theta,\theta' \in [0,\alpha]$. Hence,

$$\langle p,q\rangle \geq \cos\theta\cos\theta' - \tfrac{1}{\sqrt{3}}\sin\theta\sin\theta' \geq \cos^2\alpha - \tfrac{1}{\sqrt{3}}\sin^2\alpha = \bigl(1+\tfrac{1}{\sqrt{3}}\bigr)\cos^2\alpha - \tfrac{1}{\sqrt{3}},$$

where the first inequality holds because $\langle x,y\rangle \geq -\frac{1}{\sqrt{3}}$, by Remark 6.5, and the second holds since $\cos\theta\cos\theta' + \langle x,y\rangle\sin\theta\sin\theta'$ is decreasing in both θ and θ'.

(2) If $p \in \mathcal{C}_\alpha^{\mathrm{top}}(V_i)$ and $q \in \mathcal{C}_\alpha^{\mathrm{bot}}(V_{j,i})$ for some $i \neq j \in \{1,2,3,4\}$, then we can assume $x \in V_i$, $y \in V_{j,i}$, $\theta \in [0,\alpha]$, and $\theta' \in \bigl[\alpha,\frac{\pi}{2}\bigr]$. Now, consider two cases separately.

If $\langle x,y\rangle \leq 0$, then $\cos\theta\cos\theta' + \langle x,y\rangle\sin\theta\sin\theta'$ is decreasing with respect to both θ and θ'. Hence,

$$\langle p,q\rangle \leq \cos 0\cos\alpha + \langle x,y\rangle\sin 0\sin\alpha = \cos\alpha.$$

If $\langle x,y\rangle \geq 0$, observe that

$$\langle p,q\rangle = (1-\langle x,y\rangle)\cos\theta\cos\theta' + \langle x,y\rangle\cos(\theta'-\theta).$$

If we view θ' as a variable on $\bigl[\alpha,\frac{\pi}{2}\bigr]$,

$$\frac{\partial}{\partial\theta'}\bigl((1-\langle x,y\rangle)\cos\theta\cos\theta' + \langle x,y\rangle\cos(\theta'-\theta)\bigr)$$
$$= -(1-\langle x,y\rangle)\cos\theta\sin\theta' - \langle x,y\rangle\sin(\theta'-\theta) \leq 0.$$

Hence, $\langle p,q\rangle$ is maximized when $\theta'=\alpha$. So $\langle p,q\rangle\le\cos\theta\cos\alpha+\langle x,y\rangle\sin\theta\sin\alpha$. Now, if we view θ as a variable and take a derivative,

$$\frac{\partial}{\partial\theta}(\cos\theta\cos\alpha+\langle x,y\rangle\sin\theta\sin\alpha)=-\sin\theta\cos\alpha+\langle x,y\rangle\cos\theta\sin\alpha.$$

One can easily check that

$$-\sin\theta\cos\alpha+\langle x,y\rangle\cos\theta\sin\alpha\begin{cases}\ge 0 & \text{if } \theta\in[0,\theta_0],\\ \le 0 & \text{if } \theta\in[\theta_0,\alpha],\end{cases}$$

where θ_0 is the unique critical point satisfying $\tan\theta_0=\langle x,y\rangle\tan\alpha$. Hence,

$$\cos\theta\cos\alpha+\langle x,y\rangle\sin\theta\sin\alpha$$

is maximized when $\theta=\theta_0$. Hence,

$$\langle p,q\rangle\le\cos\theta\cos\alpha+\langle x,y\rangle\sin\theta\sin\alpha\le\sqrt{\cos^2\alpha+\langle x,y\rangle^2\sin^2\alpha}.$$

Note that $\langle x,y\rangle\le\frac{1}{\sqrt{3}}$ since $x\in V_i$ and $y\in V_{ji}$ (this value $\frac{1}{\sqrt{3}}$ can be achieved when x is the midpoint of $-u_k$ and $-u_l$ for $k\ne l\in\{1,2,3,4\}\setminus\{i,j\}$ and $y=u_j$). Hence, one can conclude

$$\langle p,q\rangle\le\sqrt{\cos^2\alpha+\tfrac{1}{3}\sin^2\alpha}=\sqrt{\tfrac{2}{3}\cos^2\alpha+\tfrac{1}{3}}.$$

Since obviously $\cos\alpha\le\sqrt{\cos^2\alpha+\frac{1}{3}\sin^2\alpha}=\sqrt{\frac{2}{3}\cos^2\alpha+\frac{1}{3}}$, this completes the proof of this case.

(3) If $p\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{i,k})$ and $q\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{j,i})$ for distinct $i,j,k\in\{1,2,3,4\}$, then one can assume $x\in V_{i,k}$, $y\in V_{j,i}$, and $\theta,\theta'\in\left[\alpha,\frac{\pi}{2}\right]$. Now, consider two cases separately.

If $\langle x,y\rangle\le 0$, then $\cos\theta\cos\theta'+\langle x,y\rangle\sin\theta\sin\theta'$ is decreasing with respect to both θ and θ'. Hence,

$$\langle p,q\rangle\le\cos^2\alpha+\langle x,y\rangle\sin^2\alpha\le\cos^2\alpha.$$

If $\langle x,y\rangle\ge 0$, without loss of generality, one can assume $\theta\ge\theta'$. Also, observe that

$$\langle p,q\rangle=(1-\langle x,y\rangle)\cos\theta\cos\theta'+\langle x,y\rangle\cos(\theta-\theta').$$

If we view θ as a variable on $\left[\theta',\frac{\pi}{2}\right]$,

$$\begin{aligned}\frac{\partial}{\partial\theta}\big((1-\langle x,y\rangle)\cos\theta\cos\theta'+\langle x,y\rangle\cos(\theta-\theta')\big)\\ &=-(1-\langle x,y\rangle)\sin\theta\cos\theta'-\langle x,y\rangle\sin(\theta-\theta')\\ &\le 0.\end{aligned}$$

Hence, $\langle p,q\rangle$ is maximized when $\theta=\theta'$. So $\langle p,q\rangle\le\cos^2\theta'+\langle x,y\rangle\sin^2\theta'$. Now, if we view θ' as a variable and take a derivative,

$$\frac{\partial}{\partial\theta'}(\cos^2\theta'+\langle x,y\rangle\sin^2\theta')=-2(1-\langle x,y\rangle)\cos\theta'\sin\theta'\le 0.$$

Therefore, $\cos^2\theta'+\langle x,y\rangle\sin^2\theta'$ is maximized when $\theta'=\alpha$. Hence,

$$\langle p,q\rangle\le\cos^2\alpha+\langle x,y\rangle\sin^2\alpha.$$

Note that $\langle x,y\rangle\le\frac{1}{\sqrt{3}}$ as in the proof of the previous case. Hence, finally we get $\langle p,q\rangle\le\cos^2\alpha+\frac{1}{\sqrt{3}}\sin^2\alpha=\bigl(1-\frac{1}{\sqrt{3}}\bigr)\cos^2\alpha+\frac{1}{\sqrt{3}}$. Since $\cos^2\alpha$ is obviously smaller than $\cos^2\alpha+\frac{1}{\sqrt{3}}\sin^2\alpha=\bigl(1-\frac{1}{\sqrt{3}}\bigr)\cos^2\alpha+\frac{1}{\sqrt{3}}$, this completes the proof of this case.

(4) If $p\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{i,j})$ and $q\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{j,i})$ for some $i\ne j\in\{1,2,3,4\}$, then one can assume $x\in V_{i,j}$, $y\in V_{j,i}$, and $\theta,\theta'\in\bigl[\alpha,\frac{\pi}{2}\bigr]$. Since $\langle x,y\rangle\le 0$ always in this case, $\cos\theta\cos\theta'+\langle x,y\rangle\sin\theta\sin\theta'$ is decreasing with respect to both θ and θ'. Hence, $\langle p,q\rangle$ is maximized when $\theta=\theta'=\alpha$. Therefore,

$$\langle p,q\rangle\le\cos^2\alpha+\langle x,y\rangle\sin^2\alpha\le\cos^2\alpha. \qquad\square$$

Finally, we are ready to construct the map

$$\tilde{\phi}^\alpha_{3,2}\colon \boldsymbol{H}_{>0}(\mathbb{S}^3)\to\mathbb{S}^2,\qquad p\mapsto\begin{cases}u_i & \text{if } p\in\mathcal{C}_\alpha^{\mathrm{top}}(V_i)\text{ for some } i\in\{1,2,3,4\},\\ u_{i,j} & \text{if } p\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{i,j})\text{ for some } i\ne j\in\{1,2,3,4\}.\end{cases}$$

Proposition 8.4 *For $\alpha\in\bigl[0,\frac{\pi}{2}\bigr]$ such that $\cos^2\alpha\in\bigl[\frac{\sqrt{3}-1}{3+\sqrt{3}},\frac{7}{9}\bigr]$,*

$$\operatorname{dis}\bigl(\tilde{\phi}^\alpha_{3,2}\bigr)\le\zeta_2.$$

Proof We need to check

$$\bigl|d_{\mathbb{S}^3}(p,q)-d_{\mathbb{S}^2}\bigl(\tilde{\phi}^\alpha_{3,2}(p),\tilde{\phi}^\alpha_{3,2}(q)\bigr)\bigr|\le\zeta_2$$

for any $p,q\in\boldsymbol{H}_{>0}(\mathbb{S}^3)$. We carry out a case-by-case analysis.

(1) Suppose $p,q\in\mathcal{C}(V_i)$ for some $i\in\{1,2,3,4\}$. Without loss of generality, one can assume $i=1$. Then

$$d_{\mathbb{S}^2}\bigl(\tilde{\phi}^\alpha_{3,2}(p),\tilde{\phi}^\alpha_{3,2}(q)\bigr)\le\operatorname{diam}(\{u_1,u_{1,2},u_{1,3},u_{1,4}\})=\arccos\tfrac{5}{6}<\zeta_2$$

by Lemma 8.2(1). Therefore,

$$d_{\mathbb{S}^2}\bigl(\tilde{\phi}^\alpha_{3,2}(p),\tilde{\phi}^\alpha_{3,2}(q)\bigr)-d_{\mathbb{S}^3}(p,q)\le\arccos\tfrac{5}{6}<\zeta_2.$$

So it is enough to prove $d_{\mathbb{S}^3}(p,q)-d_{\mathbb{S}^2}\bigl(\tilde{\phi}^\alpha_{3,2}(p),\tilde{\phi}^\alpha_{3,2}(q)\bigr)\le\zeta_2$. But, for this direction, we need more subtle case-by-case analysis.

(a) Suppose $p,q\in\mathcal{C}_\alpha^{\mathrm{top}}(V_1)$. Then $\tilde{\phi}_{3,2}^\alpha(p)=\tilde{\phi}_{3,2}^\alpha(q)=u_1$. Also, by Lemma 8.3(1) and the choice of α,

$$d_{\mathbb{S}^3}(p,q)\le\arccos\bigl(\bigl(1+\tfrac{1}{\sqrt{3}}\bigr)\cos^2\alpha-\tfrac{1}{\sqrt{3}}\bigr)\le\zeta_2.$$

Hence,

$$d_{\mathbb{S}^3}(p,q)-d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)=d_{\mathbb{S}^3}(p,q)\le\zeta_2.$$

(b) If $p\in\mathcal{C}_\alpha^{\mathrm{top}}(V_1)$ and $q\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_1)$, then $\tilde{\phi}_{3,2}^\alpha(p)=u_1$ and $\tilde{\phi}_{3,2}^\alpha(q)=u_{1,j}$ for some $j\in\{2,3,4\}$. Therefore,

$$d_{\mathbb{S}^3}(p,q)-d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)\le\arccos\bigl(-\tfrac{1}{\sqrt{3}}\bigr)-\arccos\bigl(\tfrac{2\sqrt{2}}{3}\bigr)<\zeta_2.$$

(c) Suppose $p,q\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_1)$.

(i) If $p,q\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,j})$ for some $j\in\{2,3,4\}$, then $\tilde{\phi}_{3,2}^\alpha(p)=\tilde{\phi}_{3,2}^\alpha(q)=u_{1,j}$. Also, it is easy to check the diameter of $\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,j})$ is $\frac{\pi}{2}$. Hence,

$$d_{\mathbb{S}^3}(p,q)-d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)=d_{\mathbb{S}^3}(p,q)\le\tfrac{\pi}{2}<\zeta_2.$$

(ii) If $p\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,k})$ and $p\in\mathcal{C}_\alpha^{\mathrm{bot}}(V_{1,l})$ for some $k\ne l\in\{2,3,4\}$, then

$$d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)=d_{\mathbb{S}^2}(u_{1,k},u_{1,l})=\arccos\bigl(\tfrac{5}{6}\bigr)$$

by Lemma 8.2(1). Therefore,

$$d_{\mathbb{S}^3}(p,q)-d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)\le\arccos\bigl(-\tfrac{1}{\sqrt{3}}\bigr)-\arccos\bigl(\tfrac{5}{6}\bigr)<\zeta_2.$$

(2) Suppose $p\in\mathcal{C}(V_i)$ and $q\in\mathcal{C}(V_j)$ for some $i\ne j\in\{1,2,3,4\}$. Without loss of generality, one can assume $i=1$ and $j=2$. Then, by Lemma 8.2,

$$d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)\ge\arccos\bigl(\tfrac{5}{54}\bigr)>\arccos\bigl(\tfrac{1}{3}\bigr).$$

Therefore,

$$d_{\mathbb{S}^3}(p,q)-d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)<\pi-\arccos\bigl(\tfrac{1}{3}\bigr)=\zeta_2.$$

So, it is enough to prove $d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)-d_{\mathbb{S}^3}(p,q)\le\zeta_2$. Again we need more subtle case-by-case analysis.

(a) If $p\in\mathcal{C}_\alpha^{\mathrm{top}}(V_1)$ and $q\in\mathcal{C}_\alpha^{\mathrm{top}}(V_2)$, then $\tilde{\phi}_{3,2}^\alpha(p)=u_1$ and $\tilde{\phi}_{3,2}^\alpha(q)=u_2$, so $d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)=d_{\mathbb{S}^2}(u_1,u_2)=\zeta_2$. Thus,

$$d_{\mathbb{S}^2}\bigl(\tilde{\phi}_{3,2}^\alpha(p),\tilde{\phi}_{3,2}^\alpha(q)\bigr)-d_{\mathbb{S}^3}(p,q)\le\zeta_2.$$

(b) Suppose $p \in \mathcal{C}_\alpha^{\text{top}}(V_1)$ and $q \in \mathcal{C}_\alpha^{\text{bot}}(V_2)$.

(i) If $q \in \mathcal{C}_\alpha^{\text{bot}}(V_{2,j})$ for some $j \in \{3,4\}$, then, by Lemma 8.2(6),

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) = d_{\mathbb{S}^3}(u_1, u_{2,j}) = \arccos\big(-\tfrac{\sqrt{2}}{9}\big).$$

Hence,

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) - d_{\mathbb{S}^3}(p,q) \le \arccos\big(-\tfrac{\sqrt{2}}{9}\big) < \zeta_2.$$

(ii) If $q \in \mathcal{C}_\alpha^{\text{bot}}(V_{2,1})$ then $d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) = d_{\mathbb{S}^2}(u_1, u_{2,1}) = \arccos\big(-\frac{4\sqrt{2}}{9}\big)$ by Lemma 8.2(7). Moreover, by Lemma 8.3(2) and the choice of α,

$$d_{\mathbb{S}^3}(p,q) \ge \arccos\Big(\sqrt{\tfrac{2}{3}\cos^2\alpha + \tfrac{1}{3}}\Big) > \arccos\big(\tfrac{2\sqrt{2}}{3}\big),$$

which implies

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) - d_{\mathbb{S}^3}(p,q) < \arccos\big(-\tfrac{4\sqrt{2}}{9}\big) - \arccos\big(\tfrac{2\sqrt{2}}{3}\big) = \zeta_2.$$

(c) Suppose $p \in \mathcal{C}_\alpha^{\text{bot}}(V_1)$ and $q \in \mathcal{C}_\alpha^{\text{bot}}(V_2)$. Considering symmetry, there are basically four subcases:

(i) If $p \in \mathcal{C}_\alpha^{\text{bot}}(V_{1,3})$ and $q \in \mathcal{C}_\alpha^{\text{bot}}(V_{2,3})$, then, by Lemma 8.2(2),

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) = d_{\mathbb{S}^2}(u_{1,3}, u_{2,3}) = \arccos\big(\tfrac{5}{54}\big).$$

Hence,

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) - d_{\mathbb{S}^3}(p,q) \le \arccos\big(\tfrac{5}{54}\big) < \zeta_2.$$

(ii) If $p \in \mathcal{C}_\alpha^{\text{bot}}(V_{1,3})$ and $q \in \mathcal{C}_\alpha^{\text{bot}}(V_{2,4})$, then, by Lemma 8.2(3),

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) = d_{\mathbb{S}^2}(u_{1,3}, u_{2,4}) = \arccos\big(-\tfrac{2}{27}\big).$$

Hence,

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) - d_{\mathbb{S}^3}(p,q) \le \arccos\big(-\tfrac{2}{27}\big) < \zeta_2.$$

(iii) If $p \in \mathcal{C}_\alpha^{\text{bot}}(V_{1,3})$ and $q \in \mathcal{C}_\alpha^{\text{bot}}(V_{2,1})$, then, by Lemma 8.2(4),

$$d_{\mathbb{S}^2}\big(\tilde{\phi}_{3,2}^{\alpha}(p), \tilde{\phi}_{3,2}^{\alpha}(q)\big) = d_{\mathbb{S}^2}(u_{1,3}, u_{2,1}) = \arccos\big(-\tfrac{25}{54}\big).$$

Moreover, by Lemma 8.3(3) and the choice of α,

$$d_{\mathbb{S}^3}(p,q) \ge \arccos\Big(\big(1 - \tfrac{1}{\sqrt{3}}\big)\cos^2\alpha + \tfrac{1}{\sqrt{3}}\Big) > \arccos\big(-\tfrac{25}{54}\big) - \zeta_2.$$

Hence,

$$d_{\mathbb{S}^2}(\tilde{\phi}^\alpha_{3,2}(p),\tilde{\phi}^\alpha_{3,2}(q)) - d_{\mathbb{S}^3}(p,q) < \zeta_2.$$

(iv) If $p \in \mathcal{C}^{\mathrm{bot}}_\alpha(V_{1,2})$ and $q \in \mathcal{C}^{\mathrm{bot}}_\alpha(V_{2,1})$, then, by Lemma 8.2(5),

$$d_{\mathbb{S}^2}\big(\tilde{\phi}^\alpha_{3,2}(p),\tilde{\phi}^\alpha_{3,2}(q)\big) = d_{\mathbb{S}^2}(u_{1,2},u_{2,1}) = \arccos\big(-\tfrac{23}{27}\big).$$

Moreover, by Lemma 8.3(4) and the choice of α,

$$d_{\mathbb{S}^3}(p,q) \geq \arccos(\cos^2\alpha) \geq \arccos\big(\tfrac{7}{9}\big).$$

Hence,

$$d_{\mathbb{S}^2}\big(\tilde{\phi}^\alpha_{3,2}(p),\tilde{\phi}^\alpha_{3,2}(q)\big) - d_{\mathbb{S}^3}(p,q) \leq \arccos\big(-\tfrac{23}{27}\big) - \arccos\big(\tfrac{7}{9}\big) = \zeta_2. \qquad \square$$

Lemma 8.5 *For any $p \in \boldsymbol{H}_{>0}(\mathbb{S}^3)$, $d_{\mathbb{S}^3}\big(p,\tilde{\phi}^\alpha_{3,2}(p)\big) \leq \frac{\pi}{2}$.*

Proof Without loss of generality, one can assume $p \in \mathcal{C}(V_1)$. Then one can express p as $p = \cos\theta\cdot e_4 + \sin\theta\cdot\iota_2(x)$, where $e_4 = (0,0,0,1)$ for some $\theta \in \big[0,\frac{\pi}{2}\big]$ and $x \in V_1$. Moreover, since $\tilde{\phi}^\alpha_{3,2}(p) \in \{u_1,u_{1,2},u_{1,3},u_{1,4}\}$,

$$\big\langle p,\tilde{\phi}^\alpha_{3,2}(p)\big\rangle = \big\langle x,\tilde{\phi}^\alpha_{3,2}(p)\big\rangle\cdot\sin\theta.$$

Also, it is easy to check that $\big\langle x,\tilde{\phi}^\alpha_{3,2}(p)\big\rangle \geq 0$ (more precisely, $\langle u_1,x\rangle \geq \frac{1}{3}$ and $\langle u_{1,j},x\rangle \geq \frac{\sqrt{2}}{9}$ for any $x \in N_1$ with $j \neq 1$). This implies $\big\langle p,\tilde{\phi}^\alpha_{3,2}(p)\big\rangle \geq 0$; hence we have the required inequality. □

We are now ready to prove Proposition 1.19.

Proof of Proposition 1.19 It is enough to find a surjective map $\phi_{3,2}\colon \mathbb{S}^3 \twoheadrightarrow \mathbb{S}^2$ such that $\mathrm{dis}(\phi_{3,2}) \leq \zeta_2$ since this map gives rise to the correspondence $R_{3,2} := \mathrm{graph}(\phi_{3,2})$ with $\mathrm{dis}(R_{3,2}) = \mathrm{dis}(\phi_{3,2}) \leq \zeta_2$.

Let

$$\hat{\phi}^\alpha_{3,2}\colon \boldsymbol{A}(\mathbb{S}^3) \to \mathbb{S}^2, \qquad p \mapsto \begin{cases} \tilde{\phi}^\alpha_{3,2}(p) & \text{if } p \in \boldsymbol{H}_{>0}(\mathbb{S}^3), \\ p & \text{if } p \in \iota_2(\boldsymbol{A}(\mathbb{S}^2)). \end{cases}$$

We claim that $\mathrm{dis}(\hat{\phi}^\alpha_{3,2}) = \mathrm{dis}(\tilde{\phi}^\alpha_{3,2})$. To check this, it is enough to show that

$$\big|d_{\mathbb{S}^3}(p,q) - d_{\mathbb{S}^2}(\hat{\phi}^\alpha_{3,2}(p),\hat{\phi}^\alpha_{3,2}(q))\big| \leq \zeta_2$$

for any $p \in \boldsymbol{H}_{>0}(\mathbb{S}^3)$ and $q \in \iota_2(\boldsymbol{A}(\mathbb{S}^2))$. But, this is true since

$$\begin{aligned}\big|d_{\mathbb{S}^3}(p,q) - d_{\mathbb{S}^2}\big(\hat{\phi}^\alpha_{3,2}(p),\hat{\phi}^\alpha_{3,2}(q)\big)\big| &= \big|d_{\mathbb{S}^3}(p,q) - d_{\mathbb{S}^2}\big(\hat{\phi}^\alpha_{3,2}(p),q\big)\big| \\ &\leq d_{\mathbb{S}^3}\big(p,\hat{\phi}^\alpha_{3,2}(p)\big),\end{aligned}$$

and $d_{\mathbb{S}^3}\big(p,\hat{\phi}^{\alpha}_{3,2}(p)\big)=d_{\mathbb{S}^3}\big(p,\tilde{\phi}^{\alpha}_{3,2}(p)\big)\le\frac{\pi}{2}<\zeta_2$ for any $p\in\boldsymbol{H}_{>0}(\mathbb{S}^3)$ by Lemma 8.5. Hence, $\operatorname{dis}\big(\hat{\phi}^{\alpha}_{3,2}\big)=\operatorname{dis}\big(\tilde{\phi}^{\alpha}_{3,2}\big)$. Finally, apply Lemma 5.7 to construct a surjective map $\phi_{3,2}\colon\mathbb{S}^3\twoheadrightarrow\mathbb{S}^2$. Then

$$\operatorname{dis}(\phi_{3,2})=\operatorname{dis}\big(\hat{\phi}^{\alpha}_{3,2}\big)=\operatorname{dis}\big(\tilde{\phi}^{\alpha}_{3,2}\big)\le\zeta_2$$

by Proposition 8.4. □

9 The Gromov–Hausdorff distance between spheres with Euclidean metric

For any nonempty subset $X\subseteq\mathbb{S}^n$, let X_{E} denote the metric space with the inherited Euclidean metric. In particular, $\mathbb{S}^n_{\mathrm{E}}$ will denote the unit sphere with the Euclidean metric d_{E} inherited from $\mathbb{R}^{n+1}$. A natural question is: what is the value of

$$\mathfrak{g}^{\mathrm{E}}_{m,n}:=d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}},\mathbb{S}^n_{\mathrm{E}})$$

for $0\le m<n\le\infty$? We found that, interestingly, these values do not always directly follow from those of $\mathfrak{g}_{m,n}$.

Any correspondence R between $\mathbb{S}^m$ and $\mathbb{S}^n$ can of course be regarded as a correspondence between $\mathbb{S}^m_{\mathrm{E}}$ and $\mathbb{S}^n_{\mathrm{E}}$. Throughout this section, let $\operatorname{dis}(R)$ denote the distortion with respect to the geodesic metric (as usual), and let $\operatorname{dis}_{\mathrm{E}}(R)$ denote the distortion with respect to the Euclidean metric.

The following are direct extensions of parallel results for spheres with geodesic distance:

Remark 9.1 As in Remark 1.4, for all $0\le m\le n\le\infty$,

$$d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}},\mathbb{S}^n_{\mathrm{E}})\le 1.$$

Lemma 9.2 *For any integer $m\ge 1$ and any finite metric space P with cardinality at most $m+1$, we have $d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}},P)\ge 1$.*

Proof Fix an arbitrary correspondence R between $\mathbb{S}^m_{\mathrm{E}}$ and P. Then one can prove that $\operatorname{dis}_{\mathrm{E}}(R)\ge 2$ as in the proof of Lemma 3.2 (via the aid of Lyusternik–Schnirelmann theorem). Since R is arbitrary, one can conclude $d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}},P)\ge 1$. □

Corollary 9.3 *Let R be any correspondence between a finite metric space P and $\mathbb{S}^{\infty}_{\mathrm{E}}$. Then $\operatorname{dis}_{\mathrm{E}}(R)\ge 2$. In particular, $d_{\mathrm{GH}}(P,\mathbb{S}^{\infty}_{\mathrm{E}})\ge 1$.*

Proof See the proof of Corollary 3.4. □

Proposition 9.4 *Let X be any totally bounded metric space. Then $d_{\mathrm{GH}}(X,\mathbb{S}^\infty_{\mathrm{E}})\ge 1$.*

Proof Follow the idea of the proof of Proposition 3.5. □

Proposition 9.5 *For any $n\ge 1$, $d_{\mathrm{GH}}(\mathbb{S}^0_{\mathrm{E}},\mathbb{S}^n_{\mathrm{E}})=1$.*

Proof Apply Remark 9.1 and Lemma 9.2. □

Proposition 9.6 *For any integer $m\ge 0$, $d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}},\mathbb{S}^\infty_{\mathrm{E}})=1$.*

Proof Apply Remark 9.1 and Proposition 9.4. □

The following lemma permits bounding $\mathrm{dis}_{\mathrm{E}}(R)$ via $\mathrm{dis}(R)$:

Lemma 9.7 *Let $0\le m<n\le\infty$, and let R be an arbitrary nonempty relation between $\mathbb{S}^m_{\mathrm{E}}$ and $\mathbb{S}^n_{\mathrm{E}}$. Then*

$$\mathrm{dis}_{\mathrm{E}}(R)\le 2\sin\bigl(\tfrac{1}{2}\,\mathrm{dis}(R)\bigr).$$

Proof First of all, note that $\mathrm{dis}(R):=\sup_{(x,y),(x',y')\in R}|d_{\mathbb{S}^m}(x,x')-d_{\mathbb{S}^n}(y,y')|\le\pi$, since both $\mathrm{diam}(\mathbb{S}^m)$ and $\mathrm{diam}(\mathbb{S}^n)$ are at most π. Fix arbitrary $(x,y),(x',y')\in R$. Then

$$\begin{aligned}
d_{\mathrm{E}}(x,x') &= 2\sin\bigl(\tfrac{1}{2}d_{\mathbb{S}^m}(x,x')\bigr)\\
&= 2\sin\bigl(\tfrac{1}{2}d_{\mathbb{S}^m}(x,x')-\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')+\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')\bigr)\\
&= 2\sin\bigl(\tfrac{1}{2}d_{\mathbb{S}^m}(x,x')-\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')\bigr)\cos\bigl(\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')\bigr)\\
&\qquad\qquad + 2\cos\bigl(\tfrac{1}{2}d_{\mathbb{S}^m}(x,x')-\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')\bigr)\sin\bigl(\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')\bigr)\\
&\le 2\sin\bigl(\tfrac{1}{2}|d_{\mathbb{S}^m}(x,x')-d_{\mathbb{S}^n}(y,y')|\bigr)+2\sin\bigl(\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')\bigr)\\
&= 2\sin\bigl(\tfrac{1}{2}|d_{\mathbb{S}^m}(x,x')-d_{\mathbb{S}^n}(y,y')|\bigr)+d_{\mathrm{E}}(y,y'),
\end{aligned}$$

where the inequality follows since $\cos\bigl(\tfrac{1}{2}d_{\mathbb{S}^m}(x,x')-\tfrac{1}{2}d_{\mathbb{S}^n}(y,y')\bigr)\in[0,1]$.

Hence,

$$d_{\mathrm{E}}(x,x')-d_{\mathrm{E}}(y,y')\le 2\sin\bigl(\tfrac{1}{2}|d_{\mathbb{S}^m}(x,x')-d_{\mathbb{S}^n}(y,y')|\bigr).$$

Similarly, one can also prove

$$d_{\mathrm{E}}(y,y')-d_{\mathrm{E}}(x,x')\le 2\sin\bigl(\tfrac{1}{2}|d_{\mathbb{S}^m}(x,x')-d_{\mathbb{S}^n}(y,y')|\bigr).$$

Therefore,

$$|d_{\mathrm{E}}(x,x')-d_{\mathrm{E}}(y,y')|\le 2\sin\bigl(\tfrac{1}{2}|d_{\mathbb{S}^m}(x,x')-d_{\mathbb{S}^n}(y,y')|\bigr).$$

Since $(x,y),(x',y')\in R$ were arbitrary, this leads to the required conclusion. □

Corollary 9.8 *For any* $0 \le m < n \le \infty$:

(1) $d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}}, \mathbb{S}^n_{\mathrm{E}}) \le \sin(d_{\mathrm{GH}}(\mathbb{S}^m, \mathbb{S}^n))$.

(2) *In more generality, for any* $X \subseteq \mathbb{S}^m$ *and* $Y \subseteq \mathbb{S}^n$, $d_{\mathrm{GH}}(X_{\mathrm{E}}, Y_{\mathrm{E}}) \le \sin(d_{\mathrm{GH}}(X, Y))$.

Corollary 9.9 $d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}}, \mathbb{S}^n_{\mathrm{E}}) < 1$ *for all* $0 < m \ne n < \infty$.

Proof Invoke Corollary 9.8 and Theorem A. □

Given the above, and that we proved $\mathfrak{g}_{1,2} = \frac{\pi}{3}$ and $\mathfrak{g}_{2,3} = \frac{1}{2}\zeta_2$, one might expect that $\mathfrak{g}^{\mathrm{E}}_{1,2} = d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, \mathbb{S}^2_{\mathrm{E}}) = \sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}$ and similarly that $\mathfrak{g}^{\mathrm{E}}_{2,3} = \frac{\sqrt{2}}{\sqrt{3}}$. However, rather surprisingly, we were able to construct a correspondence R_{E} between $\mathbb{S}^1_{\mathrm{E}}$ and $\boldsymbol{H}_{\ge 0}(\mathbb{S}^2_{\mathrm{E}})$ such that $\mathrm{dis}_{\mathrm{E}}(R_{\mathrm{E}}) < \sqrt{3}$ (see Proposition 9.10 and its proof in Section 9.1). This correspondence then naturally induces a function $\phi_{\mathrm{E}}\colon \boldsymbol{A}(\mathbb{S}^2_{\mathrm{E}}) \to \mathbb{S}^1_{\mathrm{E}}$ from the "helmet" on $\mathbb{S}^2_{\mathrm{E}}$ into $\mathbb{S}^1_{\mathrm{E}}$ also with $\mathrm{dis}_{\mathrm{E}}(\phi_{\mathrm{E}}) < \sqrt{3}$.

Proposition 9.10
$$d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, \boldsymbol{H}_{\ge 0}(\mathbb{S}^2_{\mathrm{E}})) < \frac{\sqrt{3}}{2}.$$

This proposition was motivated by Ilya Bogdanov's answer [2] to a MathOverflow question regarding the Gromov–Hausdorff distance between $\mathbb{S}^1_{\mathrm{E}}$ and the unit disk in $\mathbb{R}^2$.

We now discuss the possibility that the correspondence R_{E} described above permits proving that, in fact, $d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, \mathbb{S}^2_{\mathrm{E}}) < \frac{\sqrt{3}}{2}$ via extending R_{E} into a correspondence between $\mathbb{S}^2_{\mathrm{E}}$ and $\mathbb{S}^1_{\mathrm{E}}$ much in the same way that we did so in the case of spheres with their geodesic distance (see Lemma 5.7).

By the same method of proof as that of Corollary 5.5 (giving the lower bound $\mathrm{dis}(g) \ge \zeta_n$ for any antipode-preserving map $g\colon \mathbb{S}^n \to \mathbb{S}^{n-1}$), one obtains the following Euclidean analogue:

Corollary 9.11 *For each integer* $n > 0$, *any function* $g\colon \mathbb{S}^n_{\mathrm{E}} \to \mathbb{S}^{n-1}_{\mathrm{E}}$ *which maps every pair of antipodal points on* $\mathbb{S}^n_{\mathrm{E}}$ *onto antipodal points on* $\mathbb{S}^{n-1}_{\mathrm{E}}$ *satisfies*
$$\mathrm{dis}_{\mathrm{E}}(g) \ge \sqrt{2 + \frac{2}{n}}.$$

Remark 9.12 (extending Lemma 5.7 to the case of spheres with Euclidean metric) Lemma 5.7 was instrumental in our quest for lower bounds for the Gromov–Hausdorff distance between spheres with the geodesic distance. It is natural to attempt to obtain a suitable version of that result to the case of the Euclidean metric. However, there is a caveat. Indeed, one should *not* expect to be able to prove a version in which $\operatorname{dis}_{\mathrm{E}}(\phi^*)$ is *equal* to $\operatorname{dis}(\phi)$ where $\phi\colon \mathcal{A}(\mathbb{S}^n_{\mathrm{E}}) \to \mathbb{S}^m_{\mathrm{E}}$ and ϕ^* is its antipode-preserving extension obtained via the "helmet trick" (as described in the statement of Lemma 5.7). If this was the case, then the antipode-preserving extension ϕ^*_{E} of the function ϕ_{E} mentioned above would satisfy

$$\operatorname{dis}_{\mathrm{E}}(\phi^*_{\mathrm{E}}) < \sqrt{3}. \tag{10}$$

However, note that Corollary 9.11 implies that, in the case of spheres with Euclidean distance, any antipode-preserving map $\psi\colon \mathbb{S}^{m+1}_{\mathrm{E}} \to \mathbb{S}^m_{\mathrm{E}}$ must satisfy

$$\operatorname{dis}_{\mathrm{E}}(\psi) \geq \sqrt{2 + \frac{2}{m+1}}.$$

In particular, it must be that $\operatorname{dis}_{\mathrm{E}}(\psi) \geq \sqrt{3}$ for any antipode-preserving map $\psi\colon \mathbb{S}^2_{\mathrm{E}} \to \mathbb{S}^1_{\mathrm{E}}$, and this would contradict (10).

Still, as we describe next, there is a suitable generalization of Lemma 5.7 which yields nontrivial lower bounds (see Proposition 9.16).

Lemma 9.13 *If* $|a-b| =: \delta \in [0,2]$ *for some* $a, b \in [0,2]$, *then*

$$\left|\sqrt{4-a^2} - \sqrt{4-b^2}\right| \leq \sqrt{\delta(4-\delta)},$$

and the inequality is tight.

Proof The claim is obvious if $\delta = 0$. Henceforth, we will assume that $\delta > 0$. Observe that

$$\begin{aligned}\left|\sqrt{4-a^2} - \sqrt{4-b^2}\right| &= \frac{|a^2-b^2|}{\sqrt{4-a^2}+\sqrt{4-b^2}} \\ &= |a-b| \cdot \frac{a+b}{\sqrt{4-a^2}+\sqrt{4-b^2}} \\ &\leq \delta \cdot \frac{4-\delta}{\sqrt{4\delta-\delta^2}} \\ &= \sqrt{\delta(4-\delta)}.\end{aligned}$$

Finally, the equality holds if $a = 2$ and $b = 2-\delta$, or $a = 2-\delta$ and $b = 2$. □

Lemma 9.14 *For any* $m, n \geq 0$, *let* $\varnothing \neq C \subseteq \mathbb{S}^n_{\mathrm{E}}$ *satisfy* $C \cap (-C) = \varnothing$ *and let* $\phi\colon C \to \mathbb{S}^m_{\mathrm{E}}$ *be any map. Then the extension* ϕ^* *of* ϕ *to the set* $C \cup (-C)$ *defined by*

$$\phi^*\colon C \cup (-C) \to \mathbb{S}^m, \qquad x \mapsto \phi(x), \quad -x \mapsto -\phi(x) \quad \text{for } x \in C,$$

is antipode-preserving and satisfies $\operatorname{dis}_{\mathrm{E}}(\phi^*) \leq \sqrt{\operatorname{dis}_{\mathrm{E}}(\phi)(4-\operatorname{dis}_{\mathrm{E}}(\phi))}$.

Proof By definition, ϕ^* is antipode-preserving. Now, fix arbitrary $x, x' \in C$. Then

$$\begin{aligned}|d_{\mathrm{E}}(x,-x') - d_{\mathrm{E}}(\phi^*(x),\phi^*(-x'))| \\ &= |\sqrt{4-(d_{\mathrm{E}}(x,x'))^2} - \sqrt{4-\bigl(d_{\mathrm{E}}(\phi(x),\phi(x'))\bigr)^2}| \\ &\le \sqrt{|d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(\phi(x),\phi(x'))|\bigl(4-|d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(\phi(x),\phi(x'))|\bigr)} \\ &\le \sqrt{\mathrm{dis}_{\mathrm{E}}(\phi)(4-\mathrm{dis}_{\mathrm{E}}(\phi))}\end{aligned}$$

and

$$|d_{\mathrm{E}}(-x,-x') - d_{\mathrm{E}}(\phi^*(-x),\phi^*(-x'))| = |d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(\phi(x),\phi(x'))| \le \mathrm{dis}_{\mathrm{E}}(\phi).$$

Hence,

$$\mathrm{dis}_{\mathrm{E}}(\phi^*) \le \max\bigl\{\mathrm{dis}_{\mathrm{E}}(\phi), \sqrt{\mathrm{dis}_{\mathrm{E}}(\phi)(4-\mathrm{dis}_{\mathrm{E}}(\phi))}\bigr\} = \sqrt{\mathrm{dis}_{\mathrm{E}}(\phi)(4-\mathrm{dis}_{\mathrm{E}}(\phi))}. \qquad \square$$

Corollary 9.15 *For each $n \in \mathbb{Z}_{>0}$ and any map $\phi\colon \mathbb{S}^n_{\mathrm{E}} \to \mathbb{S}^{n-1}_{\mathrm{E}}$, there exists an antipode-preserving map $\phi^*\colon \mathbb{S}^n_{\mathrm{E}} \to \mathbb{S}^{n-1}_{\mathrm{E}}$ such that $\mathrm{dis}_{\mathrm{E}}(\phi^*) \le \sqrt{\mathrm{dis}_{\mathrm{E}}(\phi)(4-\mathrm{dis}_{\mathrm{E}}(\phi))}$.*

Proof Consider the restriction of ϕ to the "helmet" $\boldsymbol{A}(\mathbb{S}^n)$ (see Section 5.1) and apply Lemma 9.14. □

Proposition 9.16 *For all integers $0 < m < n$,*

$$d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}}, \mathbb{S}^n_{\mathrm{E}}) \ge \frac{1}{2}\Bigl(2 - \sqrt{2 - \frac{2}{m+1}}\Bigr) \ge \frac{1}{2}.$$

Proof Suppose to the contrary that $d_{\mathrm{GH}}(\mathbb{S}^m_{\mathrm{E}}, \mathbb{S}^n_{\mathrm{E}}) < \frac{1}{2}(2 - \sqrt{2-2/(m+1)})$. This implies that there exist a correspondence Γ between $\mathbb{S}^m_{\mathrm{E}}$ and $\mathbb{S}^n_{\mathrm{E}}$ such that $\mathrm{dis}_{\mathrm{E}}(\Gamma) < \frac{1}{2}(2 - \sqrt{2-2/(m+1)})$. Moreover, since $n \ge m+1$, $\mathbb{S}^{m+1}_{\mathrm{E}}$ can be isometrically embedded in $\mathbb{S}^n_{\mathrm{E}}$, so we are able to construct a map $g\colon \mathbb{S}^{m+1}_{\mathrm{E}} \to \mathbb{S}^m_{\mathrm{E}}$ in the following way: for each $x \in \mathbb{S}^{m+1}_{\mathrm{E}} \subseteq \mathbb{S}^n_{\mathrm{E}}$, choose $g(x) \in \mathbb{S}^m_{\mathrm{E}}$ such that $(g(x), x) \in \Gamma$. Then $\mathrm{dis}_{\mathrm{E}}(g) < (2 - \sqrt{2-2/(m+1)})$ as well. By applying Corollary 9.15, one can modify this g into an antipode-preserving map $\hat{g}\colon \mathbb{S}^{m+1}_{\mathrm{E}} \to \mathbb{S}^m_{\mathrm{E}}$ with

$$\mathrm{dis}_{\mathrm{E}}(\hat{g}) \le \sqrt{\mathrm{dis}_{\mathrm{E}}(g)(4-\mathrm{dis}_{\mathrm{E}}(g))} < \sqrt{2 + \frac{2}{m+1}},$$

which contradicts Corollary 9.11. □

Note that in contrast to the case of geodesic distances (where the upper bound given by Proposition 1.16 and the lower bound given by Theorem B agree when $m = 1$ and $n = 2$), Proposition 9.16 yields $\mathfrak{g}^{\mathrm{E}}_{1,2} \ge \frac{1}{2}$, which is strictly smaller than the upper bound $\frac{\sqrt{3}}{2}$ provided by Corollary 9.8 and Proposition 1.16.

9.1 The proof of Proposition 9.10

The proof will be based on a geometric construction which is illustrated in Figures 13 and 14.

Proof To prove the claim, note that it is enough to construct a correspondence R_{E} between $\mathbb{S}^1_{\mathrm{E}}$ and $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2_{\mathrm{E}})$ such that $\operatorname{dis}_{\mathrm{E}}(R_{\mathrm{E}}) < \sqrt{3}$.

First, let $u_1, \dots, u_7$ be the vertices of a regular heptagon inscribed in $\mathbb{S}^1$. Let $v_i := -u_i$ for $i = 1, \dots, 7$. See Figure 13 for a description.

Second, divide $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2_E)$ into seven regions $A_1, \dots, A_7$ as in Figure 14. The precise "disjointification" (on the boundary) of the seven regions is not relevant to the analysis that follows, as it is easy to check.

Now, choose $a_i \in A_i$ for each $i = 1, \dots, 7$ in the following way, where α is some number which is very close to $\frac{\sqrt{3}}{2}$ but still strictly smaller than $\frac{\sqrt{3}}{2}$ (for example, choose $\alpha = 0.866$):

$$a_1 = \left(\sqrt{1-\left(\sqrt{1-\alpha^2}+2-\sqrt{3}\right)^2}, \sqrt{1-\alpha^2}+2-\sqrt{3}, 0\right) \approx (0.640511, 0.767949, 0),$$
$$a_2 = \left(0, \sqrt{1-\alpha^2}+2-\sqrt{3}, \sqrt{1-\left(\sqrt{1-\alpha^2}+2-\sqrt{3}\right)^2}\right) \approx (0, 0.767949, 0.640511),$$
$$a_3 = \left(0, \sqrt{1-\alpha^2}, \alpha\right) \approx (0, 0.5, 0.866),$$

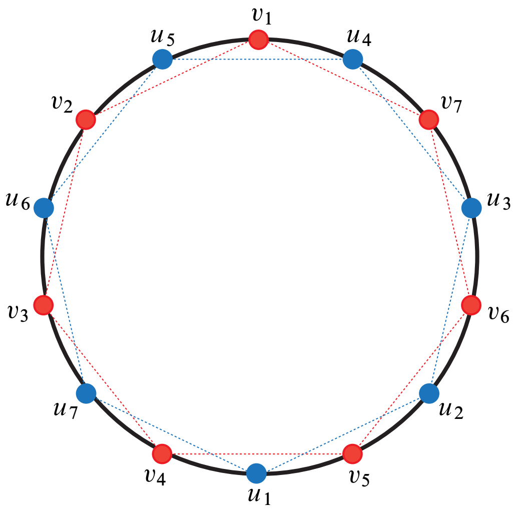


Figure 13: The points $v_1, \dots, v_7$ and $u_1, \dots, u_7$. These arise from two antipodal regular heptagons inscribed in $\mathbb{S}^1$.

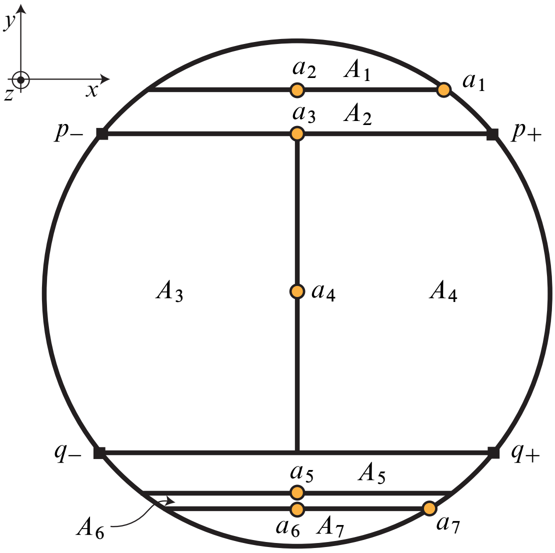


Figure 14: View from above of the seven regions $A_1, \dots, A_7$ of $\boldsymbol{H}_{\geq 0}(\mathbb{S}^2_E)$. All lines shown in the figure (which are projections of circular arcs) are aligned with the either the x or y axis. Also, $p_\pm = (\pm\alpha, \sqrt{1-\alpha^2}, 0)$ and $q_\pm = (\pm\alpha, -\sqrt{1-\alpha^2}, 0)$.

$$\begin{aligned}
a_4 &= (0,0,1),\\
a_5 &= \left(0, -(\sqrt{1-\alpha^2}+\rho_6-\sqrt{3}), \sqrt{1-(\sqrt{1-\alpha^2}+\rho_6-\sqrt{3})^2}\right)\\
&\approx (0, -0.717805, 0.696244),\\
a_6 &= \left(0, -(\sqrt{1-\alpha^2}+\rho_6-\sqrt{3}+\rho_5-\sqrt{3}), \sqrt{1-(\sqrt{1-\alpha^2}+\rho_6-\sqrt{3}+\rho_5-\sqrt{3})^2}\right)\\
&\approx (0, -0.787692, 0.616069),\\
a_7 &= \left(\sqrt{1-(\sqrt{1-\alpha^2}+\rho_6-\sqrt{3}+\rho_5-\sqrt{3})^2}, -(\sqrt{1-\alpha^2}+\rho_6-\sqrt{3}+\rho_5-\sqrt{3}), 0\right)\\
&\approx (0.616069, -0.787692, 0),
\end{aligned}$$

where $\rho_k := \sqrt{2-2\cos(k\pi/7)}$ for $k \in \{1, \dots, 7\}$.

One can directly check that the following seven conditions are satisfied:

(1) $d_{\mathrm{E}}(A_i, A_j) > \rho_6 - \sqrt{3}$ for any $i, j \in \{1, \dots, 7\}$ with $|i-j| = 3$.

(2) $d_{\mathrm{E}}(a_i, a_j) > \rho_6 - \sqrt{3}$ for any $i, j \in \{1, \dots, 7\}$ with $|i-j| = 2$.

(3) $d_{\mathrm{E}}(a_i, a_j) > 2 - \sqrt{3}$ for any $i, j \in \{1, \dots, 7\}$ with $|i-j| = 3$.

(4) $d_{\mathrm{E}}(A_i, a_j) > \rho_5 - \sqrt{3}$ for any $i, j \in \{1, \dots, 7\}$ with $|i - j| = 2$.

(5) $d_{\mathrm{E}}(A_i, a_j) > 2 - \sqrt{3}$ for any $i, j \in \{1, \dots, 7\}$ with $|i - j| = 3$.

(6) $\operatorname{diam}(A_i) < \sqrt{3}$ for any $i \in \{1, \dots, 7\}$.

(7) $d_{\mathrm{E}}(a_i, a_j) < \sqrt{3}$ for any $i, j \in \{1, \dots, 7\}$ with $|i - j| = 1$.

In what follows, for two points $v, w \in \mathbb{S}^1$ with $d_{\mathrm{E}}(v, w) < 2$, $\widehat{vw}$ will denote the (unique) shortest circular arc determined by these two points.

Now we define a correspondence R_{E} by

$$R_{\mathrm{E}} := \bigcup_{i=1}^{7} \{(u_i, y) : y \in A_i\} \cup \bigcup_{i=1}^{7} \{(x, a_i) : x \in \widehat{v_{i+3} v_{i+4}}\}.$$

We now prove that $\operatorname{dis}_{\mathrm{E}}(R_{\mathrm{E}}) < \sqrt{3}$:

First, let us prove that

$$\sup_{(x,y),(x',y') \in R_{\mathrm{E}}} (d_{\mathrm{E}}(x, x') - d_{\mathrm{E}}(y, y')) < \sqrt{3}.$$

For this we verify the inequality $d_{\mathrm{E}}(x, x') - d_{\mathrm{E}}(y, y') < \sqrt{3}$ for all cases induced by the structure of the correspondence R_E:

(1) If $(x, y), (x', y') \in \{u_i\} \times A_i$ for some $i \in \{1, \dots, 7\}$, then $d_{\mathrm{E}}(x, x') = d_{\mathrm{E}}(u_i, u_i) = 0$.

(2) If $(x, y) \in \{u_i\} \times A_i$ and $(x', y') \in \{u_j\} \times A_j$ for some $i, j \in \{1, \dots, 7\}$ with $|i - j| = 1$, then $d_{\mathrm{E}}(x, x') = d_{\mathrm{E}}(u_i, u_j) = \rho_2 < \sqrt{3}$.

(3) If $(x, y) \in \{u_i\} \times A_i$ and $(x', y') \in \{u_j\} \times A_j$ for some $i, j \in \{1, \dots, 7\}$ with $|i - j| = 2$, then $d_{\mathrm{E}}(x, x') = d_{\mathrm{E}}(u_i, u_j) = \rho_4 < \sqrt{3}$.

(4) If $(x, y) \in \{u_i\} \times A_i$ and $(x', y') \in \{u_j\} \times A_j$ for some $i, j \in \{1, \dots, 7\}$ with $|i - j| = 3$, then $d_{\mathrm{E}}(x, x') = d_{\mathrm{E}}(u_i, u_j) = \rho_6 > \sqrt{3}$. However, since $d_{\mathrm{E}}(A_i, A_j) > \rho_6 - \sqrt{3}$ by condition (1) above, we have $d_{\mathrm{E}}(x, x') - d_{\mathrm{E}}(y, y') < \sqrt{3}$.

(5) If $(x, y), (x', y') \in \widehat{v_{i+3} v_{i+4}} \times \{a_i\}$ for some $i \in \{1, \dots, 7\}$, then $d_{\mathrm{E}}(x, x') \leq \operatorname{diam}(\widehat{v_{i+3} v_{i+4}}) = \rho_2 < \sqrt{3}$.

(6) If $(x, y) \in \widehat{v_{i+3} v_{i+4}} \times \{a_i\}$ and $(x', y') \in \widehat{v_{j+3} v_{j+4}} \times \{a_j\}$ for some $i, j \in \{1, \dots, 7\}$ with $|i - j| = 1$, then $d_{\mathrm{E}}(x, x') \leq \operatorname{diam}(\widehat{v_{i+3} v_{i+4}} \cup \widehat{v_{j+3} v_{j+4}}) = \rho_4 < \sqrt{3}$.

(7) If $(x, y) \in \widehat{v_{i+3} v_{i+4}} \times \{a_i\}$ and $(x', y') \in \widehat{v_{j+3} v_{j+4}} \times \{a_j\}$ for some $i, j \in \{1, \dots, 7\}$ with $|i - j| = 2$, then $d_{\mathrm{E}}(x, x') \leq \operatorname{diam}(\widehat{v_{i+3} v_{i+4}} \cup \widehat{v_{j+3} v_{j+4}}) = \rho_6 > \sqrt{3}$.

However, since $d_{\mathrm{E}}(y,y') = d_{\mathrm{E}}(a_i,a_j) > \rho_6 - \sqrt{3}$ by condition (2) above, we have $d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(y,y') < \sqrt{3}$.

(8) If $(x,y) \in \widehat{v_{i+3}v_{i+4}} \times \{a_i\}$ and $(x',y') \in \widehat{v_{j+3}v_{j+4}} \times \{a_j\}$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 3$, then, since $d_{\mathrm{E}}(y,y') = d_{\mathrm{E}}(a_i,a_j) > 2 - \sqrt{3}$ by condition (3) above, we have $d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(y,y') < 2 - (2-\sqrt{3}) = \sqrt{3}$.

(9) If $(x,y) \in \{u_i\} \times A_i$ and $(x',y') \in \widehat{v_{i+3}v_{i+4}} \times \{a_i\}$ for some $i \in \{1,\dots,7\}$, then $u_i \in \widehat{v_{i+3}v_{i+4}}$. Hence, $d_{\mathrm{E}}(x,x') = d_{\mathrm{E}}(u_i,x') \le \operatorname{diam}(\widehat{v_{i+3}v_{i+4}}) < \sqrt{3}$. So $d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(y,y') < \sqrt{3}$.

(10) If $(x,y) \in \{u_i\} \times A_i$ and $(x',y') \in \widehat{v_{j+3}v_{j+4}} \times \{a_j\}$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 1$, then $d_{\mathrm{E}}(x,x') = d_{\mathrm{E}}(u_i,x') \le \operatorname{diam}(\{u_i\} \cup \widehat{v_{j+3}v_{j+4}}) = \rho_3 < \sqrt{3}$.

(11) If $(x,y) \in \{u_i\} \times A_i$ and $(x',y') \in \widehat{v_{j+3}v_{j+4}} \times \{a_j\}$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 2$, then $d_{\mathrm{E}}(x,x') = d_{\mathrm{E}}(u_i,x') \le \operatorname{diam}(\{u_i\} \cup \widehat{v_{j+3}v_{j+4}}) = \rho_5 > \sqrt{3}$. However, since $d_{\mathrm{E}}(A_i,a_j) > \rho_5 - \sqrt{3}$ by condition (4) above, $d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(y,y') < \sqrt{3}$.

(12) If $(x,y) \in \{u_i\} \times A_i$ and $(x',y') \in \widehat{v_{j+3}v_{j+4}} \times \{a_j\}$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 3$, then, since $d_{\mathrm{E}}(A_i,a_j) > 2 - \sqrt{3}$ by condition (5) above, we have $d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(y,y') < \sqrt{3}$.

Next, we prove

$$\sup_{(x,y),(x',y') \in R_{\mathrm{E}}} (d_{\mathrm{E}}(y,y') - d_{\mathrm{E}}(x,x')) < \sqrt{3},$$

for which we verify the inequality $d_{\mathrm{E}}(x,x') - d_{\mathrm{E}}(y,y') < \sqrt{3}$ in a number of cases.

(1) If $(x,y),(x',y') \in \{u_i\} \times A_i$ for some $i \in \{1,\dots,7\}$, then, since $\operatorname{diam}(A_i) < \sqrt{3}$ by condition (6) above, $d_{\mathrm{E}}(y,y') < \sqrt{3}$, so $d_{\mathrm{E}}(y,y') - d_{\mathrm{E}}(x,x') < \sqrt{3}$.

(2) If $(x,y) \in \{u_i\} \times A_i$ and $(x',y') \in \{u_j\} \times A_j$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 1$, then $d_{\mathrm{E}}(x,x') = d_{\mathrm{E}}(u_i,u_j) = \rho_2$ and $d_{\mathrm{E}}(y,y') - d_{\mathrm{E}}(x,x') \le 2 - \rho_2 < \sqrt{3}$.

(3) If $(x,y) \in \{u_i\} \times A_i$ and $(x',y') \in \{u_j\} \times A_j$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 2$, then $d_{\mathrm{E}}(x,x') = d_{\mathrm{E}}(u_i,u_j) = \rho_4$ and $d_{\mathrm{E}}(y,y') - d_{\mathrm{E}}(x,x') \le 2 - \rho_4 < \sqrt{3}$.

(4) If $(x,y) \in \{u_i\} \times A_i$ and $(x',y') \in \{u_j\} \times A_j$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 3$, then $d_{\mathrm{E}}(x,x') = d_{\mathrm{E}}(u_i,u_j) = \rho_6$ and $d_{\mathrm{E}}(y,y') - d_{\mathrm{E}}(x,x') \le 2 - \rho_6 < \sqrt{3}$.

(5) If $(x,y),(x',y') \in \widehat{v_{i+3}v_{i+4}} \times \{a_i\}$ for some $i \in \{1,\dots,7\}$, then $d_{\mathrm{E}}(y,y') = d_{\mathrm{E}}(a_i,a_i) = 0$.

(6) If $(x,y) \in \widehat{v_{i+3}v_{i+4}} \times \{a_i\}$ and $(x',y') \in \widehat{v_{j+3}v_{j+4}} \times \{a_j\}$ for some $i,j \in \{1,\dots,7\}$ with $|i-j| = 1$, then, since $d_{\mathrm{E}}(a_i,a_j) < \sqrt{3}$ by condition (7) above, we have $d_{\mathrm{E}}(y,y') - d_{\mathrm{E}}(x,x') = d_{\mathrm{E}}(a_i,a_j) - d_{\mathrm{E}}(x,x') < \sqrt{3}$.

(7) If $(x,y)\in\widehat{v_{i+3}v_{i+4}}\times\{a_i\}$ and $(x',y')\in\widehat{v_{j+3}v_{j+4}}\times\{a_j\}$ for some $i,j\in\{1,\dots,7\}$ with $|i-j|=2$, then $d_{\mathrm{E}}(x,x')\geq\rho_2$. Hence, $d_{\mathrm{E}}(y,y')-d_{\mathrm{E}}(x,x')\leq 2-\rho_2<\sqrt{3}$.

(8) If $(x,y)\in\widehat{v_{i+3}v_{i+4}}\times\{a_i\}$ and $(x',y')\in\widehat{v_{j+3}v_{j+4}}\times\{a_j\}$ for some $i,j\in\{1,\dots,7\}$ with $|i-j|=3$, then $d_{\mathrm{E}}(x,x')\geq\rho_4$. Hence, $d_{\mathrm{E}}(y,y')-d_{\mathrm{E}}(x,x')\leq 2-\rho_4<\sqrt{3}$.

(9) If $(x,y)\in\{u_i\}\times A_i$ and $(x',y')\in\widehat{v_{i+3}v_{i+4}}\times\{a_i\}$ for some $i\in\{1,\dots,7\}$, then, since $a_i\in A_i$ and $\operatorname{diam}(A_i)<\sqrt{3}$ by condition (6), we have $d_{\mathrm{E}}(y,y')-d_{\mathrm{E}}(x,x')<\sqrt{3}$.

(10) If $(x,y)\in\{u_i\}\times A_i$ and $(x',y')\in\widehat{v_{j+3}v_{j+4}}\times\{a_j\}$ for some $i,j\in\{1,\dots,7\}$ with $|i-j|=1$, then $d_{\mathrm{E}}(x,x')\geq\rho_1$. Hence, $d_{\mathrm{E}}(y,y')-d_{\mathrm{E}}(x,x')\leq 2-\rho_1<\sqrt{3}$.

(11) If $(x,y)\in\{u_i\}\times A_i$ and $(x',y')\in\widehat{v_{j+3}v_{j+4}}\times\{a_j\}$ for some $i,j\in\{1,\dots,7\}$ with $|i-j|=2$, then $d_{\mathrm{E}}(x,x')\geq\rho_3$. Hence, $d_{\mathrm{E}}(y,y')-d_{\mathrm{E}}(x,x')\leq 2-\rho_3<\sqrt{3}$.

(12) If $(x,y)\in\{u_i\}\times A_i$ and $(x',y')\in\widehat{v_{j+3}v_{j+4}}\times\{a_j\}$ for some $i,j\in\{1,\dots,7\}$ with $|i-j|=3$: Observe that $d_{\mathrm{E}}(x,x')\geq\rho_5$. Hence, $d_{\mathrm{E}}(y,y')-d_{\mathrm{E}}(x,x')\leq 2-\rho_5<\sqrt{3}$.

Hence, $\operatorname{dis}_{\mathrm{E}}(R_{\mathrm{E}})<\sqrt{3}$, as required. □

Appendix A A succinct proof of Theorem G

In this appendix we provide a proof of Theorem G following a strategy suggested by Matoušek in [24, page 41] and due to Arnold Waßmer.

Lemma A.1 *If a simplex contains $0\in\mathbb{R}^n$ and all of its vertices lie on $\mathbb{S}^{n-1}$, then there are vertices u and v of the simplex such that $d_{\mathbb{S}^{n-1}}(u,v)\geq\zeta_{n-1}$.*

Proof We give the proof here for completeness — the proof is basically the same as that of [11, Lemma 1]. Let $u_1,\dots,u_{n+1}$ be (not necessarily distinct) vertices of a simplex such that their convex hull contains the origin $0\in\mathbb{R}^n$. Therefore, there are nonnegative numbers $\lambda_1,\dots,\lambda_{n+1}$ such that $\sum_{i=1}^{n+1}\lambda_n=1$ and $0=\sum_{i=1}^{n+1}\lambda_i u_i$. Then

$$0=\left\|\sum_{i=1}^{n+1}\lambda_i u_i\right\|^2=\sum_{i\neq j}\lambda_i\lambda_j\langle u_i,u_j\rangle+\sum_{i=1}^{n+1}\lambda_i^2.$$

Moreover, since $0\leq\sum_{i\neq j}(\lambda_i-\lambda_j)^2=2n\sum_{i=1}^{n+1}\lambda_i^2-2\sum_{i\neq j}\lambda_i\lambda_j$,

$$\sum_{i=1}^{n+1}\lambda_i^2\geq\frac{1}{n}\sum_{i\neq j}\lambda_i\lambda_j.$$

Hence,

$$0 \ge \sum_{i \ne j} \lambda_i \lambda_j \Big(\langle u_i, u_j \rangle + \frac{1}{n} \Big).$$

Thus, there must be some distinct i and j such that $\langle u_i, u_j \rangle \le -1/n$, so that

$$d_{\mathbb{S}^{n-1}}(u_i, u_j) \ge \arccos\Big(-\frac{1}{n}\Big) = \zeta_{n-1}. \qquad \square$$

Below, the notation $V(T)$ for a triangulation T of the cross-polytope $\widehat{\mathbb{B}}^n$ will denote its set of vertices.

Lemma A.2 *Let T be a triangulation of the cross-polytope $\widehat{\mathbb{B}}^n$ which is antipodally symmetric at the boundary (ie if $\Delta \subset \partial\widehat{\mathbb{B}}^n$ is a simplex in T, then $-\Delta \subset \partial\widehat{\mathbb{B}}^n$ is also in T), and let $g \colon V(T) \to \mathbb{S}^{n-1}$ be a mapping that satisfies $g(-v) = -g(v) \in \mathbb{S}^{n-1}$ for all vertices $v \in V(T)$ lying on the boundary of $\widehat{\mathbb{B}}^n$. Then there exist vertices $u, v \in V(T)$ with $d_{\mathbb{S}^{n-1}}(g(u), g(v)) \ge \zeta_{n-1}$.*

Proof By Lemma A.1 it is enough to show that some simplex $\{v_1, \dots, v_m\}$ of T satisfies

$$0 \in \operatorname{Conv}(g(v_1), g(v_2), \dots, g(v_m)).$$

Suppose not; then one can construct the continuous map $\phi \colon \widehat{\mathbb{B}}^n \to \mathbb{R}^n \setminus \{0\}$ such that $\phi(a_1u_1 + \dots + a_mu_m) := a_1g(u_1) + \dots + a_mg(u_m)$, where $\{u_1, \dots, u_m\}$ is a simplex of T, $a_1, \dots, a_m \in [0, 1]$, and $\sum_{i=1}^m a_i = 1$. Next, one can construct the continuous map $\hat{\phi} \colon \widehat{\mathbb{B}}^n \to \mathbb{S}^{n-1}$ such that $\hat{\phi}(x) := \phi(x)/\|\phi(x)\|$ for each $x \in \widehat{\mathbb{B}}^n$. Moreover, this map $\hat{\phi}$ is antipode-preserving on the boundary since if $x \in \partial\widehat{\mathbb{B}}^n$ satisfies $x = a_1v_1 + \dots + a_mv_m$ where $\{v_1, \dots, v_m\}$ is a simplex of $\partial\widehat{\mathbb{B}}^n$, then $\phi(x) = a_1g(v_1) + \dots + a_mg(v_m)$ and $\phi(-x) = a_1g(-v_1) + \dots + a_mg(-v_m)$, so $\phi(-x) = -\phi(x)$. This is a contradiction to the classical Borsuk–Ulam theorem since $\hat{\phi} \circ \alpha^{-1} \colon \mathbb{B}^n \to \mathbb{S}^{n-1}$ is continuous and antipode-preserving on the boundary, where (below, for a vector v we denote by $\|v\|_1$ its 1–norm)

$$\alpha \colon \widehat{\mathbb{B}}^n \to \mathbb{B}^n, \qquad x \mapsto \begin{cases} (0, \dots, 0) & \text{if } x = (0, \dots, 0), \\ x\|x\|_1/\|x\| & \text{otherwise}, \end{cases}$$

is the natural bi-Lipschitz homeomorphism between $\widehat{\mathbb{B}}^n$ and $\mathbb{B}^n$ from the unit cross-polytope to the closed unit ball. □

Now we are ready to prove Theorem G.

Proof of Theorem G Let $f \colon \mathbb{B}^n \to \mathbb{S}^{n-1}$ be a map that is antipode-preserving on the boundary of $\mathbb{B}^n$. Now, fix an arbitrary $\delta \ge 0$ such that for any $x \in \mathbb{B}^n$ there exists an open neighborhood U_x of x with $\operatorname{diam}(f(U_x)) \le \delta$. Fix $\varepsilon > 0$ smaller than the Lebesgue number of the open covering $\{U_x\}_{x \in \mathbb{B}^n}$.

Let $\alpha\colon \widehat{\mathbb{B}}^n \to \mathbb{B}^n$ be the natural (fattening) homeomorphism used in the proof of Lemma A.2. One can construct a triangulation T of $\widehat{\mathbb{B}}^n$ satisfying the following two properties:

(1) T is antipodally symmetric on the boundary of $\widehat{\mathbb{B}}^n$.

(2) T is fine enough that $\|\alpha(u) - \alpha(v)\| \le \varepsilon$ for any two adjacent vertices u and v.

Then, by Lemma A.2, there exist adjacent vertices u and v such that

$$d_{\mathbb{S}^{n-1}}(f \circ \alpha(u), f \circ \alpha(v)) \ge \zeta_{n-1}.$$

Choose $x = \alpha(u)$ and $y = \alpha(v)$. Because of the choice of ε, both x and y are contained in some U_x. Hence, $\delta \ge \operatorname{diam}(f(U_x)) \ge \zeta_{n-1}$, which concludes as in the proof of Corollary 5.4. □

Appendix B The Gromov–Hausdorff distance between a sphere and an interval

To make this paper self-contained, we include a proof of the following proposition.

Proposition B.1 *Let n be any positive integer. Then* $\operatorname{dis}(f) \ge \frac{2\pi}{3}$ *for any function* $f\colon \mathbb{S}^n \to \mathbb{R}$.

As a consequence, $d_{\mathrm{GH}}(\mathbb{S}^n, I) \ge \frac{\pi}{3}$ *for any interval* $I \subseteq \mathbb{R}$.

Proof Note that it is enough to prove the claim for $n = 1$. We adapt an argument from the proof of [20, Lemma 2.3].

Fix an arbitrary $\varepsilon > 0$. Consider an antipodally symmetric triangulation of $\mathbb{S}^1$ with vertex set $V \subset \mathbb{S}^1$ such that $d_{\mathbb{S}^1}(p,q) \le \varepsilon$ for any two adjacent vertices $p, q \in V$. Then let $\tilde{f}\colon \mathbb{S}^1 \to I$ be the linear interpolation of $f|_V\colon V \to I$. Now, by the classical Borsuk–Ulam theorem, there exists $x \in \mathbb{S}^1$ such that $\tilde{f}(x) = \tilde{f}(-x)$. Let $p, q \in V$ be such that $x \in \widehat{pq}$. Then $I \cap J \neq \varnothing$ where I is the closed interval between $f(p)$ and $f(q)$, and J is the closed interval between $f(-p)$ and $f(-q)$ (since I and J both contain $\tilde{f}(x) = \tilde{f}(-x)$). Without loss of generality, we can assume that $f(-p) \in I$. Now, let

$$r := \begin{cases} p & \text{if } |f(-p) - f(p)| \le |f(-p) - f(q)|, \\ q & \text{if } |f(-p) - f(p)| > |f(-p) - f(q)|. \end{cases}$$

Then $|f(-p)-f(r)| \leq \frac{1}{2}\,\mathrm{length}(I) \leq \frac{1}{2}(\mathrm{dis}(f)+\varepsilon)$. Hence,

$$\pi-\varepsilon \leq d_{\mathbb{S}^1}(-p.r) \leq \mathrm{dis}(f)+|f(-p)-f(r)| \leq \tfrac{3}{2}\,\mathrm{dis}(f)+\tfrac{1}{2}\varepsilon,$$

so $\mathrm{dis}(f) \geq \frac{2\pi}{3}-\varepsilon$. □

Appendix C Regular polygons and $\mathbb{S}^1$

In this appendix we compute the distance between regular polygons and also between the circle and a regular polygon.

The following map from metric spaces to metric spaces will be useful. For a metric space (X,d_X), consider the pseudo-ultrametric space (X,u_X) where $u_X\colon X\times X\to\mathbb{R}$ is defined by

$$(x,x') \mapsto u_X(x,x') := \inf\Bigl\{\max_{0\leq i\leq n-1} d_X(x_i,x_{i+1}) \bigm| x=x_0,\dots,x_n=x' \text{ for some } n\geq 1\Bigr\}.$$

Now, define $\boldsymbol{U}(X)$ to be the quotient metric space induced by (X,u_X) under the equivalence $x\sim x'$ if and only if $u_X(x,x')=0$. One then has the following, whose proof we omit:

Proposition C.1 *For any path-connected metric space X it holds that $\boldsymbol{U}(X)=*$.*

We also have the following result, establishing that $\boldsymbol{U}\colon\mathcal{M}_b\to\mathcal{M}_b$ is 1–Lipschitz:

Theorem H [7] *For all bounded metric spaces X and Y,*

$$d_{\mathrm{GH}}(X,Y) \geq d_{\mathrm{GH}}(\boldsymbol{U}(X),\boldsymbol{U}(Y)).$$

For each integer $n\geq 3$, let P_n be the regular polygon with n vertices inscribed in $\mathbb{S}^1$. We also let $P_2=\mathbb{S}^0$. Furthermore, we endow P_n with the restriction of the geodesic distance on $\mathbb{S}^1$. We then have:

Proposition C.2 (d_{GH} between $\mathbb{S}^1$ and inscribed regular polygons) *For all $n\geq 2$,*

$$d_{\mathrm{GH}}(\mathbb{S}^1,P_n)=\frac{\pi}{n}.$$

Proof That $d_{\mathrm{GH}}(\mathbb{S}^1,P_n)\geq \pi/n$ can be obtained as follows: by Theorem H,

$$d_{\mathrm{GH}}(\mathbb{S}^1,P_n) \geq d_{\mathrm{GH}}(\boldsymbol{U}(\mathbb{S}^1),\boldsymbol{U}(P_n)),$$

but, since $\boldsymbol{U}(\mathbb{S}^1) = *$ by Proposition C.1, and $\boldsymbol{U}(P_n)$ is isometric to the metric space over n points with all nonzero pairwise distances equal to $2\pi/n$, from the above inequality and (7) we have $d_{\mathrm{GH}}(\mathbb{S}^1, P_n) \geq \frac{1}{2}\operatorname{diam}(\boldsymbol{U}(P_n)) = \pi/n$. The inequality $d_{\mathrm{GH}}(\mathbb{S}^1, P_n) \leq \pi/n$ follows from the fact that $d_{\mathrm{GH}}(\mathbb{S}^1, P_n) \leq d_{\mathrm{H}}(\mathbb{S}^1, P_n) = \pi/n$. □

Note that, if $\mathbb{S}^1$ and P_n are both endowed with the Euclidean distance (respectively denoted by $\mathbb{S}^1_{\mathrm{E}}$ and $(P_n)_{\mathrm{E}}$), then, in analogy with Proposition C.2, we have the following proposition which solves a question posed in [1]. The proof is slightly different from that of Proposition C.2.

Proposition C.3 *For all* $n \geq 2$, $d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, (P_n)_{\mathrm{E}}) = \sin(\pi/n)$.

Proof One can prove $d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, (P_n)_{\mathrm{E}}) \geq \sin(\pi/n)$ by invoking $\boldsymbol{U}$ as in the proof of Proposition C.2. In order to prove $d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, (P_n)_{\mathrm{E}}) \leq \sin(\pi/n)$, let us construct a specific correspondence R between $\mathbb{S}^1_{\mathrm{E}}$ and $(P_n)_{\mathrm{E}}$. Let $u_1, \dots, u_n$ be the vertices of $(P_n)_{\mathrm{E}}$, and $V_1, \dots, V_n$ be the Voronoi regions of $\mathbb{S}^1$ induced by $u_1, \dots, u_n$. Now let

$$R := \bigcup_{i=1}^{n} V_i \times \{u_i\}.$$

Then we claim $\operatorname{dis}_{\mathrm{E}}(R) \leq 2\sin(\pi/n)$. To prove this, it is enough to check the following two conditions via standard trigonometric identities:

(1) $2\sin(k\pi/n) - 2\sin((k-1)\pi/n) \leq 2\sin(\pi/n)$ for $1 \leq k \leq \lfloor \frac{1}{2}n \rfloor$.

(2) $2 - 2\sin\bigl(\lfloor \frac{1}{2}n \rfloor \pi/n\bigr) \leq 2\sin(\pi/n)$.

Hence, $d_{\mathrm{GH}}(\mathbb{S}^1_{\mathrm{E}}, (P_n)_{\mathrm{E}}) \leq \sin(\pi/n)$. □

We now pose the following question and provide partial information about it in Proposition C.5:

Question VI *Determine, for all* $m, n \in \mathbb{N}$, *the value of* $\mathfrak{p}_{m,n} := d_{\mathrm{GH}}(P_m, P_n)$.

Remark C.4 By simple arguments, which we omit, one can prove that $\mathfrak{p}_{2,3} = \frac{\pi}{3}$, $\mathfrak{p}_{2,4} = \frac{\pi}{4}$, $\mathfrak{p}_{2,5} = \frac{2\pi}{5}$ and $\mathfrak{p}_{2,6} = \frac{\pi}{3}$. Also Proposition C.2 indicates that $\mathfrak{p}_{2,n}$ tends to $\frac{\pi}{2}$ as $n \to \infty$. Then these calculations imply that $n \mapsto \mathfrak{p}_{2,n}$ is *not* monotonically increasing towards $\frac{\pi}{2}$; cf Question I.

Proposition C.5 *For any integer* $0 < m < \infty$, $\mathfrak{p}_{m,m+1} = \pi/(m+1)$.

Proof First, let us prove that $\mathfrak{p}_{m,m+1} \le \pi/(m+1)$. We construct a correspondence R between P_m and P_{m+1} such that $\operatorname{dis}(R) \le 2\pi/(m+1)$. Let $u_1, \dots, u_m$ be the vertices of P_m and $v_1, \dots, v_m, v_{m+1}$ be the vertices of P_{m+1}. Consider the correspondence

$$R := \bigcup_{i=1}^{m} \{(u_m, v_m)\} \cup \{(u_m, v_{m+1})\}.$$

Then, for any $i, j \in \{1, \dots, m\}$,

$$\begin{aligned}
&|d_{\mathbb{S}^1}(u_i, u_j) - d_{\mathbb{S}^1}(v_i, v_j)| \\
&\qquad = \left|\frac{2\pi}{m} \cdot \min\{|i-j|, m-|i-j|\} - \frac{2\pi}{m+1} \cdot \min\{|i-j|, m+1-|i-j|\}\right| \\
&\qquad = \left|\frac{2\pi k}{m} - \frac{2\pi k}{m+1}\right| \text{ or } \left|\frac{2\pi k}{m} - \frac{2\pi(k+1)}{m+1}\right| \qquad \left(\text{for some } 0 \le k \le \left\lfloor \tfrac{1}{2}m \right\rfloor\right) \\
&\qquad = \frac{2\pi k}{m(m+1)} \text{ or } \frac{2\pi}{m+1}\left(1 - \frac{k}{m}\right) \qquad \left(\text{for some } 0 \le k \le \left\lfloor \tfrac{1}{2}m \right\rfloor\right) \\
&\qquad \le \frac{2\pi}{m+1}.
\end{aligned}$$

Also, for any $i \in \{1, \dots, m\}$,

$$\begin{aligned}
&|d_{\mathbb{S}^1}(u_i, u_m) - d_{\mathbb{S}^1}(v_i, v_{m+1})| \\
&\qquad = \left|\frac{2\pi}{m} \cdot \min\{m-i, i\} - \frac{2\pi}{m+1} \cdot \min\{m+1-i, i\}\right| \\
&\qquad = \left|\frac{2\pi k}{m} - \frac{2\pi k}{m+1}\right| \text{ or } \left|\frac{2\pi k}{m} - \frac{2\pi(k+1)}{m+1}\right| \qquad \left(\text{for some } 0 \le k \le \left\lfloor \tfrac{1}{2}m \right\rfloor\right) \\
&\qquad = \frac{2\pi k}{m(m+1)} \text{ or } \frac{2\pi}{m+1}\left(1 - \frac{k}{m}\right) \qquad \left(\text{for some } 0 \le k \le \left\lfloor \tfrac{1}{2}m \right\rfloor\right) \\
&\qquad \le \frac{2\pi}{m+1}.
\end{aligned}$$

Hence, one concludes that $\operatorname{dis}(R) \le 2\pi/(m+1)$.

Next, let us prove that $\mathfrak{p}_{m,m+1} \ge \pi/(m+1)$. Fix an arbitrary correspondence R between P_m and P_{m+1}. Then there must be a vertex u_i of P_m, and two vertices v_j and v_k of P_{m+1} such that $(u_i, v_j), (u_i, v_k) \in R$. Hence,

$$\operatorname{dis}(R) \ge |d_{\mathbb{S}^1}(u_i, u_i) - d_{\mathbb{S}^1}(v_j, v_k)| = \frac{2\pi}{m+1}.$$

Since R is arbitrary, one concludes that $\mathfrak{p}_{m,m+1} \ge \pi/(m+1)$. □

References

[1] **H Adams**, **S Chowdhury**, **A Q Jaffe**, **B Sibanda**, *Vietoris–Rips complexes of regular polygons*, preprint (2018) arXiv 1807.10971

[2] **I Bogdanov**, *Gromov–Hausdorff distance between a disk and a circle*, MathOverflow answer (2018) Available at `https://mathoverflow.net/a/312046/160011`

[3] **A M Bronstein**, **M M Bronstein**, **R Kimmel**, *Numerical geometry of non-rigid shapes*, Springer (2008) MR Zbl

[4] **D Burago**, **Y Burago**, **S Ivanov**, *A course in metric geometry*, Graduate Studies in Mathematics 33, Amer. Math. Soc., Providence, RI (2001) MR Zbl

[5] **G Carlsson**, *Topology and data*, Bull. Amer. Math. Soc. 46 (2009) 255–308 MR Zbl

[6] **G Carlsson**, *Topological pattern recognition for point cloud data*, Acta Numer. 23 (2014) 289–368 MR Zbl

[7] **G Carlsson**, **F Mémoli**, *Characterization, stability and convergence of hierarchical clustering methods*, J. Mach. Learn. Res. 11 (2010) 1425–1470 MR Zbl

[8] **M Cho**, *On the optimal covering of equal metric balls in a sphere*, J. Korea Soc. Math. Educ. Ser. B Pure Appl. Math. 4 (1997) 137–144 MR

[9] **S Chowdhury**, **F Mémoli**, *Explicit geodesics in Gromov–Hausdorff space*, Electron. Res. Announc. Math. Sci. 25 (2018) 48–59 MR Zbl

[10] **T H Colding**, *Large manifolds with positive Ricci curvature*, Invent. Math. 124 (1996) 193–214 MR Zbl

[11] **L E Dubins**, **G Schwarz**, *Equidiscontinuity of Borsuk–Ulam functions*, Pacific J. Math. 95 (1981) 51–59 MR Zbl

[12] **D A Edwards**, *The structure of superspace*, from "Studies in topology" (N M Stavrakas, K R Allen, editors), Academic, New York (1975) 121–133 MR Zbl

[13] **G B Folland**, *Real analysis*: *modern techniques and their applications*, 2nd edition, John Wiley and Sons, New York (1999) MR Zbl

[14] **K Funano**, *Estimates of Gromov's box distance*, Proc. Amer. Math. Soc. 136 (2008) 2911–2920 MR Zbl

[15] **A Gray**, *Tubes*, 2nd edition, Progr. Math. 221, Birkhäuser, Basel (2004) MR Zbl

[16] **M Gromov**, *Metric structures for Riemannian and non-Riemannian spaces*, Progr. Math. 152, Birkhäuser, Boston, MA (1999) MR Zbl

[17] **Y Ji**, **A A Tuzhilin**, *Gromov–Hausdorff distance between interval and circle*, Topology Appl. 307 (2022) art. id. 107938 MR Zbl

[18] **N J Kalton**, **M I Ostrovskii**, *Distances between Banach spaces*, Forum Math. 11 (1999) 17–48 MR Zbl

[19] **M Katz**, *The filling radius of two-point homogeneous spaces*, J. Differential Geom. 18 (1983) 505–511 MR Zbl

[20] **M G Katz**, *Torus cannot collapse to a segment*, J. Geom. 111 (2020) art. id. 13 MR Zbl

[21] **S Lim**, **F Mémoli**, **O B Okutan**, *Vietoris–Rips persistent homology, injective metric spaces, and the filling radius*, preprint (2020) arXiv 2001.07588

[22] **S Lim**, **F Mémoli**, **Z Smith**, GitHub repository Available at `https://github.com/ndag/dgh-spheres`

[23] **S Lim**, **F Mémoli**, **Z Smith**, *The Gromov–Hausdorff distance between spheres*, preprint (2021) arXiv 2105.00611

[24] **J Matoušek**, *Using the Borsuk–Ulam theorem: lectures on topological methods in combinatorics and geometry*, Springer (2003) MR Zbl

[25] **F Mémoli**, *Gromov–Wasserstein distances and the metric approach to object matching*, Found. Comput. Math. 11 (2011) 417–487 MR Zbl

[26] **F Mémoli**, *Some properties of Gromov–Hausdorff distances*, Discrete Comput. Geom. 48 (2012) 416–440 MR Zbl

[27] **F Mémoli**, **G Sapiro**, *A theoretical and computational framework for isometry invariant recognition of point cloud data*, Found. Comput. Math. 5 (2005) 313–347 MR Zbl

[28] **H J Munkholm**, *A Borsuk–Ulam theorem for maps from a sphere to a compact topological manifold*, Illinois J. Math. 13 (1969) 116–124 MR Zbl

[29] **G Peano**, *Sur une courbe, qui remplit toute une aire plane*, Math. Ann. 36 (1890) 157–160 MR Zbl

[30] **P Petersen**, *Riemannian geometry*, Graduate Texts in Math. 171, Springer (1998) MR Zbl

[31] **L A Santaló**, *Convex regions on the n–dimensional spherical surface*, Ann. of Math. 47 (1946) 448–459 MR Zbl

Max Planck Institute for Mathematics in the Sciences
Leipzig, Germany

Department of Mathematics, The Ohio State University
Columbus, OH, United States

Department of Computer Science, University of Minnesota
Minneapolis, MN, United States

`sulim@mis.mpg.de, facundo.memoli@gmail.com, smit9474@umn.edu`

`https://sites.google.com/view/sunhyuklim, http://facundo-memoli.org`

Proposed: Dmitri Burago
Seconded: Dan Abramovich, David Fisher

Received: 7 September 2021
Revised: 7 April 2022

Geometry & Topology **27**:9 (2023) 3801–3831
DOI: 10.2140/gt.2023.27.3801
Published: 5 December 2023

Contact three-manifolds with exactly two simple Reeb orbits

DANIEL CRISTOFARO-GARDINER
UMBERTO HRYNIEWICZ
MICHAEL HUTCHINGS
HUI LIU

It is known that every contact form on a closed three-manifold has at least two simple Reeb orbits, and a generic contact form has infinitely many. We show that if there are exactly two simple Reeb orbits, then the contact form is nondegenerate. Combined with a previous result, this implies that the three-manifold is diffeomorphic to the three-sphere or a lens space, and the two simple Reeb orbits are the core circles of a genus-one Heegaard splitting. We also obtain further information about the Reeb dynamics and the contact structure. For example, the Reeb flow has a disk-like global surface of section and so its dynamics are described by a pseudorotation, the contact structure is universally tight, and in the case of the three-sphere the contact volume and the periods and rotation numbers of the simple Reeb orbits satisfy the same relations as for an irrational ellipsoid.

37J99, 53E50; 53D42

1 Introduction

1.1 Statement of results

Let Y be a closed oriented three-manifold. Recall that a *contact form* on Y is a 1–form λ on Y such that $\lambda \wedge d\lambda > 0$. A contact form λ has an associated *Reeb vector field* R defined by the equations

$$d\lambda(R, \cdot) = 0, \quad \lambda(R) = 1.$$

A *Reeb orbit* is a periodic orbit of R, ie a map

$$\gamma\colon \mathbb{R}/T\mathbb{Z} \to Y, \quad \gamma'(t) = R(\gamma(t)),$$

for some $T > 0$, modulo reparametrization of the domain by translations. The number T is the period, also called the *symplectic action*, of γ. We say that the Reeb orbit γ is *simple* if the map γ is an embedding. Every Reeb orbit is the k–fold cover of a simple Reeb orbit for some positive integer k.

The three-dimensional case of the Weinstein conjecture, which was proved in full generality by Taubes [42], asserts that a contact form on a closed three-manifold has at least one Reeb orbit; see [28] for a survey. It was further shown in [11] that a contact form on a closed three-manifold has at least two simple Reeb orbits. This lower bound is the best possible without further hypotheses:

Example 1.1 Recall that if Y is a compact hypersurface in $\mathbb{R}^4 = \mathbb{C}^2$ which is "star-shaped" (transverse to the radial vector field), then the standard Liouville form

$$\lambda = \frac{1}{2}\sum_{i=1}^{2}(x_i\,dy_i - y_i\,dx_i) \tag{1-1}$$

restricts to a contact form on Y. If Y is the three-dimensional ellipsoid

$$\partial E(a,b) = \left\{ z \in \mathbb{C}^2 \;\middle|\; \frac{\pi|z_1|^2}{a} + \frac{\pi|z_2|^2}{b} = 1 \right\}$$

and if a/b is irrational, then there are exactly two simple Reeb orbits, corresponding to the circles in Y where $z_2 = 0$ and $z_1 = 0$, with periods a and b, respectively.

One can also take quotients of the above irrational ellipsoids by finite cyclic group actions to obtain contact forms on lens spaces with exactly two simple Reeb orbits.

It is conjectured that, in fact, every contact form on a closed connected three-manifold has either two or infinitely many simple Reeb orbits. This was proved by Colin, Dehornoy and Rechtman [8] for contact forms that are nondegenerate (see the definition below), extending a result of [12]. It was also shown by Irie [35] that, for a C^∞–generic contact form on a closed three-manifold, there are infinitely many simple Reeb orbits, and moreover their images are dense in the three-manifold.

The goal of this paper is to give detailed information about the "exceptional" case of contact forms on a closed three-manifold with exactly two simple Reeb orbits.

To state the first result, let $\xi = \operatorname{Ker}(\lambda)$ denote the contact structure determined by λ. This is a rank-2 vector bundle with a linear symplectic form $d\lambda$. If $\gamma\colon \mathbb{R}/T\mathbb{Z} \to Y$ is a Reeb orbit, then the derivative of the time T flow of R restricts to a symplectic linear map

$$P_\gamma\colon (\xi_{\gamma(0)}, d\lambda) \to (\xi_{\gamma(0)}, d\lambda), \tag{1-2}$$

which we call the *linearized return map*. We say that γ is *nondegenerate* if 1 is not an eigenvalue of P_γ; this condition is invariant under reparametrization of γ. We say that the contact form λ is nondegenerate if all Reeb orbits (including nonsimple ones) are nondegenerate. The set of nondegenerate contact forms is residual in the set of all contact forms with the C^∞–topology. The Reeb orbit γ is called *hyperbolic* if P_γ has eigenvalues in $\mathbb{R} \setminus \{\pm 1\}$. The Reeb orbit γ is called *elliptic* if the eigenvalues of P_γ are of the form $e^{\pm 2\pi i\phi}$, and *irrationally elliptic* if moreover ϕ is irrational. If γ is irrationally elliptic, then γ and all of its covers are nondegenerate, because the linearized return map for the k–fold cover of γ has eigenvalues $e^{\pm 2\pi ik\phi}$.

Many results about Reeb dynamics and related questions assume some kind of nondegeneracy hypothesis or allow only certain kinds of degeneracies. One of the main points of the present work is that we can derive our results without making any such assumption.

Theorem 1.2 *Let Y be a closed three-manifold, and let λ be a contact form on Y with exactly two simple Reeb orbits. Then λ is nondegenerate and, moreover, both simple Reeb orbits are irrationally elliptic.*

Theorem 1.2 might seem surprising in view of known results about critical points of real-valued functions on finite-dimensional manifolds. For example, on the two-torus the minimal number of critical points is three, and when there are only three critical points they cannot all be nondegenerate. We refer the reader to Remark 1.12 for related discussion.

As a corollary of Theorem 1.2, we obtain the following topological constraint:

Corollary 1.3 *Let Y be a closed three-manifold, and let λ be a contact form on Y with exactly two simple Reeb orbits. Then Y is diffeomorphic to a lens space.*[1] *Moreover, the two simple Reeb orbits are the core circles of a genus-one Heegaard splitting of Y.*

Proof This was shown in [33, Theorem 1.3 and Section 4.8] under the additional hypothesis that λ is nondegenerate. By Theorem 1.2, this nondegeneracy automatically holds. □

Remark 1.4 A special case of Theorem 1.2, where Y is a compact convex hypersurface in $\mathbb{R}^4$ with the restriction of the standard Liouville form (1-1), was previously shown by Wang, Hu and Long [45, Theorem 1.4].

[1] Here and below our convention is that S^3 is a lens space, but $S^1 \times S^2$ is not.

We also obtain additional dynamical information. To state the result, recall that the *contact volume* of (Y,λ) is defined by

$$\operatorname{vol}(Y,\lambda) := \int_Y \lambda \wedge d\lambda.$$

Theorem 1.5 *Let Y be a lens space and let λ be a contact form on Y with exactly two simple Reeb orbits*, γ_1 *and* γ_2. *Then*:

(a) *Let $p = |\pi_1(Y)| < \infty$, let $T_i \in \mathbb{R}$ denote the period of γ_i, and let $\phi_i \in \mathbb{R}$ denote the "Seifert rotation number" of γ_i; see Definition 4.3. Then*

$$\operatorname{vol}(Y,\lambda) = pT_1T_2 = T_1^2/\phi_1 = T_2^2/\phi_2.$$

(b) *λ is dynamically convex, and the contact structure $\xi = \operatorname{Ker}(\lambda)$ is universally tight.*[2]

Example 1.6 For the ellipsoid in Example 1.1, we have $T_1 = a$, $T_2 = b$, $\phi_1 = a/b$, $\phi_2 = b/a$, $p = 1$, and $\operatorname{vol} = ab$. Thus Theorem 1.5(a) implies that if $Y = S^3$, then the periods T_i, the rotation numbers ϕ_i, and the contact volume satisfy the same relations as for an ellipsoid. For $Y = S^3$, under the additional assumptions that λ is nondegenerate and ξ is the standard contact structure, it was previously shown by Bourgeois, Cieliebak and Ekholm [5] and Gürel [19] that "action–index relations" hold, implying that the periods T_i and rotation numbers ϕ_i satisfy the same relations as for an ellipsoid. The equation $\operatorname{vol} = T_1T_2$ that we prove in this case answers [4, Question 2].

Remark 1.7 There exist contact forms on S^3 with exactly two simple Reeb orbits which are not strictly contactomorphic to ellipsoids. One way to see this is to start from Katok's construction [37] of Finsler metrics on S^2 with exactly two closed geodesics, such that the Liouville measure on the unit tangent bundle is ergodic for the geodesic flow. Such a geodesic flow can then be lifted to a Reeb flow on the standard contact 3–sphere with the same properties. Another way to see this is by Albers, Geiges and Zehmisch [1], who showed that the pseudorotations from Fayad and Katok [15] can be realized as the return map on a disk-like global surface of section for a Reeb flow on the standard contact 3–sphere with precisely two periodic orbits; see Section 1.3 below. On the other hand, Helmut Hofer has suggested to the authors in private

[2] Recall that a contact form on a three-manifold Y with $c_1(\xi)|_{\pi_2(Y)} = 0$ is called *dynamically convex* if $\operatorname{CZ}(\gamma) \geq 3$ for every contractible Reeb orbit γ, where CZ denotes the Conley–Zehnder index (see Section 2.2) computed with respect to a trivialization which extends over a disc bounded by γ. A contact structure on Y is *universally tight* if its pullback to the universal cover of Y is tight.

correspondence (2021) that perhaps imposing the additional condition that the rotation numbers of the two Reeb orbits are Diophantine forces the contact form to be strictly contactomorphic to an ellipsoid; see Fayad and Krikorian [16, Question 6].

Remark 1.8 As shown by Honda [23, Proposition 5.1] (see Cornwell [10, page 17] for more explanation), each lens space has either one or two universally tight contact structures up to isotopy, and when there are two they are contactomorphic (and one is obtained from the other by reversing its orientation). Consequently, in Theorem 1.5(b), the contact structure is contactomorphic to a "standard" contact structure on the lens space obtained as in Example 1.1. In particular, universally tight contact structures on lens spaces are precisely the ones that admit contact forms with exactly two simple Reeb orbits. Some other results obtaining information about contact structures from Reeb dynamics can be found in work by Etnyre and Ghrist [14], Hofer, Wysocki and Zehnder [20; 22], and [24].

Remark 1.9 We also obtain information about the knot types of the simple Reeb orbits γ_1 and γ_2. It follows from the Heegaard splitting in Corollary 1.3 that these are p–unknotted. We further show in Section 5 that their self-linking number is -1 when $p = 1$; similar arguments show that, for general p, their rational self-linking number, as defined by Baker and Etnyre [2], equals $-1/p$.

1.2 Outline of the proofs

We now briefly describe the proofs of Theorems 1.2 and 1.5.

A key ingredient in these proofs, as well as in the related papers [11; 12], is the "volume property" in embedded contact homology, which was proved in [13]. The embedded contact homology (ECH) of (Y, λ) is the homology of a chain complex which is built out of Reeb orbits, and whose differential counts (mostly) embedded pseudoholomorphic curves in $\mathbb{R} \times Y$; see the lecture notes [30] and the review in Section 2. The version of the volume property that we will use here asserts that if Y is a closed connected 3–manifold with a contact form λ, then

$$\lim_{k\to\infty} \frac{c_{\sigma_k}(Y, \lambda)^2}{k} = 2\,\mathrm{vol}(Y, \lambda).$$

Here $\{\sigma_k\}$ is a "U–sequence" in ECH, and c_{σ_k} is a "spectral invariant" associated to σ_k, which is the total symplectic action of a certain finite set of Reeb orbits determined by σ_k; these notions are reviewed in Section 2.

The outline of the proof of Theorem 1.2 is as follows. Let γ_1 and γ_2 denote the two simple Reeb orbits, and let T_1 and T_2 denote their periods. Simple applications of the volume property from [11; 12] (just using the $k^{1/2}$ growth rate of the spectral invariants and not the exact relation with contact volume) show that the homology classes $[\gamma_i] \in H_1(Y)$ are torsion, and the ratio T_1/T_2 is irrational. A more precise use of the volume property then gives the relations

$$\phi_i = \frac{T_i^2}{\mathrm{vol}(Y,\lambda)}, \quad \ell(\gamma_1,\gamma_2) = \frac{T_1T_2}{\mathrm{vol}(Y,\lambda)}, \tag{1-3}$$

where $\phi_i \in \mathbb{R}$ is the Seifert rotation number that appears in Theorem 1.5(a), while $\ell(\gamma_1,\gamma_2) \in \mathbb{Q}$ is the linking number of γ_1 and γ_2; see Definition 4.2. The proof of (1-3) also depends on a new estimate for the behavior of the ECH index (the grading on the ECH chain complex) under perturbations of possibly degenerate contact forms, which is proved in Section 3.

The equations (1-3) imply the relations

$$\phi_1 = \ell(\gamma_1,\gamma_2)\frac{T_1}{T_2}, \quad \phi_2 = \ell(\gamma_1,\gamma_2)\frac{T_2}{T_1}. \tag{1-4}$$

Since $\ell(\gamma_1,\gamma_2)$ is rational and T_1/T_2 is irrational, it follows that ϕ_1 and ϕ_2 are irrational. This implies that γ_1 and γ_2 are irrationally elliptic (see Section 4.4), which completes the proof of Theorem 1.2.

The Heegaard decomposition in Corollary 1.3 implies that $\ell(\gamma_1,\gamma_2) = 1/p$, and combined with (1-4) this proves Theorem 1.5(a). The proof of Theorem 1.5(b) uses additional calculations in Section 5 to deduce dynamical convexity and universal tightness from information about the numbers ϕ_i.

1.3 Pseudorotations

The contact forms studied here are analogous to "pseudorotations", defined in various ways as maps in some class with the minimum number of periodic orbits. For example, according to Ginzburg and Gürel [18] a *Hamiltonian pseudorotation* of $\mathbb{C}P^n$ is defined to be a Hamiltonian symplectomorphism of $\mathbb{C}P^n$ with $n+1$ fixed points and no other periodic points; see eg Le Roux and Seyfaddini [39], Shelukhin [41] and Çineli, Ginzburg and Gürel [7] for generalizations to other symplectic manifolds. More classically, we consider here pseudorotations of the open or closed disk defined as area-preserving homeomorphisms with one fixed point and no other periodic points; see eg Bramham [6] and Fayad and Katok [15].

In fact, there is a direct connection between the contact forms considered here and pseudorotations of the closed disk. Recall that given a closed three-manifold Y with a contact form λ, a *disk-like global surface of section* for the Reeb flow is an immersed disk, with boundary on a Reeb orbit, embedded and transverse to the Reeb flow in the interior, such that the Reeb flow starting at any point in Y hits the disk both forwards and backwards in time.

Corollary 1.10 *Let Y be a closed three-manifold, and let λ be a contact form on Y with exactly two simple Reeb orbits. Then both orbits bound disk-like global surfaces of section whose associated return maps define smooth pseudorotations of the open disk.*

Proof By Theorem 1.2, Corollary 1.3 and Theorem 1.5, Y is a lens space and λ is nondegenerate and dynamically convex. As explained in Remark 1.9, both orbits are p–unknotted, with self-linking number $-1/p$. Hence, the result follows from [24, Theorem 1.12]. □

Remark 1.11 Conversely, as mentioned above, at least some pseudorotations of the closed disk can be "suspended" to contact forms on S^3 with exactly two simple Reeb orbits; see Albers, Geiges and Zehmisch [1].

Remark 1.12 It is shown by Collier, Kerman, Reiniger, Turmunkh and Zimmer [9] — see also Franks [17] — that, for a Hamiltonian pseudorotation of $\mathbb{C}P^1$, each fixed point is strongly nondegenerate, meaning that the linearized return map and its higher powers are nondegenerate, and moreover the fixed points are irrationally elliptic, similarly to Theorem 1.2. It is an open question whether every pseudorotation of $\mathbb{C}P^n$ for $n > 1$ is strongly nondegenerate, and one can ask analogous questions for pseudorotations of more general symplectic manifolds.

Remark 1.13 For pseudorotations of D^2, in many cases, Joly [36] and Pirnapasov [40] proved identities related to Theorem 1.5(a).

Remark 1.14 One can arrange in the statement of Corollary 1.10 that the first return maps on the obtained disk-like global surfaces of section extend smoothly to the boundary and preserve a smooth 2–form that defines an area form in the interior. Moreover, one can conjugate such a return map by a homeomorphism to obtain a pseudorotation of the closed disk which is smooth in the interior.

Acknowledgments Cristofaro-Gardiner was partially supported by NSF grants DMS–1711976 and DMS–2105471. Hryniewicz was partially supported by the DFG SFB/TRR

191 *Symplectic structures in geometry, algebra and dynamics*, 281071066–TRR 191. Hutchings was partially supported by NSF grant DMS–2005437. Liu was partially supported by NSFC (Nos. 12022111 and 11771341) and the Fundamental Research Funds for the Central Universities (No. 2042021kf1059).

Key discussions for this project took place when Cristofaro-Gardiner and Liu were visiting Peking University for the *Workshop on gauge theory and Floer homology* in December 2019. We thank Peking University for their hospitality during this visit. This research was completed while Cristofaro-Gardiner was on a von Neumann fellowship at the Institute for Advanced Study and he thanks the institute for their wonderful support. We thank Viktor Ginzburg and Helmut Hofer for helpful comments.

We thank the referee for careful reading of the paper.

2 Preliminaries

In this section we review the material about embedded contact homology that is needed for the proofs of Theorems 1.2 and 1.5. We include a new, slight extension of the definition of the ECH index to degenerate contact forms.

Throughout this section fix a closed oriented three-manifold Y and a contact form λ on Y, and let $\xi = \operatorname{Ker}(\lambda)$ denote the associated contact structure.

2.1 Topological preliminaries

We now recall some topological notions we will need, following the treatment in [27]. These were originally introduced in a slightly different context in [26].

Definition 2.1 An *orbit set* is a finite set of pairs $\alpha = \{(\alpha_i, m_i)\}$ where the α_i are distinct simple Reeb orbits, and the m_i are positive integers. We define the homology class of the orbit set α by

$$[\alpha] = \sum_i m_i[\alpha_i] \in H_1(Y).$$

Definition 2.2 If $\alpha = \{(\alpha_i, m_i)\}$ and $\beta = \{(\beta_j, n_j)\}$ are orbit sets with $[\alpha] = [\beta]$, define $H_2(Y, \alpha, \beta)$ to be the set of 2–chains Z in Y with $\partial Z = \sum_i m_i\alpha_i - \sum_j n_j\beta_j$, modulo boundaries of 3–chains. The set $H_2(Y, \alpha, \beta)$ is an affine space over $H_2(Y)$.

Given orbit sets α and β as above, let $Z \in H_2(Y, \alpha, \beta)$, and let τ be a homotopy class of symplectic trivialization of the contact structure ξ over the Reeb orbits α_i and β_j.

Definition 2.3 (see [27, Section 2.5]) Define the *relative first Chern class*

$$c_\tau(\alpha,\beta,Z)\in\mathbb{Z}$$

as follows. Let S be a compact oriented surface with boundary and let $f\colon S\to Y$ be a smooth map representing the class Z. Let ψ be a section of $f^*\xi$ which, on each boundary component, is nonvanishing and constant with respect to τ. Define $c_\tau(\alpha,\beta,Z)$ to be the algebraic count of zeroes of ψ.

Definition 2.4 [27, Section 2.7] An *admissible representative* of $Z\in H_2(Y,\alpha,\beta)$ is a smooth map $f\colon S\to[-1,1]\times Y$ where S is a compact oriented surface with boundary, the restriction of f to ∂S consists of positively oriented covers of $\{1\}\times\alpha_i$ with total multiplicity m_i and negatively oriented covers of $\{-1\}\times\beta_j$ with total multiplicity n_j, the composition of f with the projection $[-1,1]\times Y\to Y$ represents the class Z, the restriction of f to the interior of S is an embedding, and f is transverse to $\{-1,1\}\times Y$.

Definition 2.5 [27, Section 2.7] If $Z,Z'\in H_2(Y,\alpha,\beta)$, define the *relative intersection pairing*

$$Q_\tau(Z,Z')\in\mathbb{Z}$$

as follows. Let S and S' be admissible representatives of Z and Z', respectively, whose interiors are transverse and do not intersect near the boundary. Define

$$Q_\tau(Z,Z')=\#(\operatorname{int}(S)\cap\operatorname{int}(S'))-\sum_i\ell_\tau(\zeta_i^+,\zeta_i^{+\prime})+\sum_j\ell_\tau(\zeta_j^-,\zeta_j^{-\prime}).\tag{2-1}$$

Here # denotes the signed count of intersections, while the remaining terms are linking numbers defined as follows. For $\epsilon>0$ small, the intersection of S with $\{1-\epsilon\}\times Y$ consists of the union over i of a braid ζ_i^+ in a neighborhood of α_i (see Section 3.1), while the intersection of S with $\{-1+\epsilon\}\times Y$ consists of the union over j of a braid ζ_j^- in a neighborhood of β_j. Likewise, S' determines braids $\zeta_i^{+\prime}$ and $\zeta_j^{-\prime}$. The notation ℓ_τ indicates the linking number in a neighborhood of α_i or β_j computed using the trivialization τ; see [27, Section 2.6] for details and sign conventions.

When $Z=Z'$, we write[3]

$$Q_\tau(\alpha,\beta,Z)=Q_\tau(Z,Z).$$

[3] An alternative equivalent definition of $Q_\tau(\alpha,\beta,Z)$ is given in [30, Section 3.3], which does not include the linking number terms in (2-1). There the admissible representatives S and S' are required to satisfy additional conditions which force these linking number terms to be zero.

As explained in [27], the relative first Chern class $c_\tau(\alpha,\beta,Z)$ and the relative self-intersection number $Q_\tau(\alpha,\beta,Z)$ depend only on α, β, Z and τ. Moreover, if we change Z by adding $A\in H_2(Y)$, then

$$c_\tau(\alpha,\beta,Z+A)-c_\tau(\alpha,\beta,Z)=\langle c_1(\xi),A\rangle, \tag{2-2}$$

$$Q_\tau(\alpha,\beta,Z+A)-Q_\tau(\alpha,\beta,Z)=2[\alpha]\cdot A. \tag{2-3}$$

Remark 2.6 If γ is a third orbit set, if τ is a trivialization of ξ over the Reeb orbits in α, β and γ, and if $W\in H_2(Y,\beta,\gamma)$, then we have the additivity properties

$$c_\tau(\alpha,\beta,Z)+c_\tau(\beta,\gamma,W)=c_\tau(\alpha,\gamma,Z+W),$$
$$Q_\tau(\alpha,\beta,Z)+Q_\tau(\beta,\gamma,W)=Q_\tau(\alpha,\gamma,Z+W).$$

Note also that the definition of c_τ makes sense more generally if the α_i and β_j are transverse knots. Likewise the definition of Q_τ makes sense if the α_i and β_j are knots and τ is an oriented trivialization of their normal bundles.

2.2 The ECH index

Let $\gamma\colon \mathbb{R}/T\mathbb{Z}\to Y$ be a Reeb orbit and let τ be a symplectic trivialization of $\gamma^*\xi$. The derivative of the time t Reeb flow from $\xi_{\gamma(0)}$ to $\xi_{\gamma(t)}$, with respect to τ, is a 2×2 symplectic matrix $\Phi(t)$. The family of symplectic matrices $\{\Phi(t)\}_{t\in[0,T]}$ induces a family of diffeomorphisms of S^1 in the universal cover of $\mathrm{Diff}(S^1)$, which has a dynamical rotation number, here denoted by $\theta_\tau(\gamma)\in\mathbb{R}$. We call this real number the *rotation number* of γ with respect to τ and denote it by $\theta_\tau(\gamma)\in\mathbb{R}$; it depends only on γ and the homotopy class of τ. When $\theta_\tau(\gamma)\notin\frac{1}{2}\mathbb{Z}$, the eigenvalues of the linearized return map (1-2) are $e^{\pm 2\pi i\theta_\tau(\gamma)}$.

Definition 2.7 Define the *Conley–Zehnder index*

$$\mathrm{CZ}_\tau(\gamma)=\lfloor\theta_\tau(\gamma)\rfloor+\lceil\theta_\tau(\gamma)\rceil\in\mathbb{Z}. \tag{2-4}$$

Remark 2.8 The above definition agrees with the usual Conley–Zehnder index when γ is nondegenerate. When γ is degenerate, it is common to give a different definition of the Conley–Zehnder index, as the minimum of the Conley–Zehnder indices of nondegenerate perturbations of γ, and this will sometimes differ from our definition by 1. For our purposes, especially to obtain an estimate as in Proposition 3.1 below (possibly with a different constant), it does not matter which of these definitions of the Conley–Zehnder index we use for degenerate Reeb orbits.

Notation 2.9 If $\alpha = \{(\alpha_i, m_i)\}$ is an orbit set and if τ is a trivialization of ξ over all of the Reeb orbits α_i, define

$$\mathrm{CZ}^I_\tau(\alpha) = \sum_i \sum_{k=1}^{m_i} \mathrm{CZ}_\tau(\alpha_i^k). \tag{2-5}$$

Here γ^k denotes the k^{th} iterate of γ.

Definition 2.10 Let α and β be orbit sets with $[\alpha] = [\beta] \in H_1(Y)$, and $Z \in H_2(Y, \alpha, \beta)$. Define the *ECH index*

$$I(\alpha, \beta, Z) = c_\tau(\alpha, \beta, Z) + Q_\tau(\alpha, \beta, Z) + \mathrm{CZ}^I_\tau(\alpha) - \mathrm{CZ}^I_\tau(\beta) \in \mathbb{Z}. \tag{2-6}$$

The above agrees with the usual definition of the ECH index — see eg [30, Section 3.4] — when the contact form is nondegenerate. It is explained, for example in [27, Section 2.8], why $I(\alpha, \beta, Z)$ depends only on α, β and Z, and not on τ. Moreover, it follows from (2-2) and (2-3) that, if we change Z by adding $A \in H_2(Y)$, then

$$I(\alpha, \beta, Z + A) - I(\alpha, \beta, Z) = \langle c_1(\xi) + 2\,\mathrm{PD}(\Gamma), A\rangle, \tag{2-7}$$

where $\Gamma = [\alpha] = [\beta] \in H_1(Y)$ and PD denotes the Poincaré dual. By Remark 2.6,

$$I(\alpha, \beta, Z) + I(\beta, \gamma, W) = I(\alpha, \gamma, Z + W). \tag{2-8}$$

2.3 Embedded contact homology

In this subsection assume that the contact form λ is nondegenerate. Let $\Gamma \in H_1(Y)$. We now review how to define the embedded contact homology $\mathrm{ECH}_*(Y, \xi, \Gamma)$. More details may be found in [30].

Definition 2.11 An *ECH generator* is an orbit set $\alpha = \{(\alpha_i, m_i)\}$ such that $m_i = 1$ whenever α_i is hyperbolic.

Definition 2.12 Define $\mathrm{ECC}_*(Y, \lambda, \Gamma)$ to be the vector space[4] over $\mathbb{Z}/2$ generated by ECH generators α with $[\alpha] = \Gamma$. This vector space has a relative $\mathbb{Z}/d$–grading, where d denotes the divisibility of $c_1(\xi) + 2\,\mathrm{PD}(\Gamma) \in H^2(Y; \mathbb{Z})$; if α and β are two generators, then their grading difference is $I(\alpha, \beta, Z) \bmod d$ for any $Z \in H_2(Y, \alpha, \beta)$. This makes sense by (2-7) and (2-8).

[4] It is also possible to use $\mathbb{Z}$ coefficients, as explained in [32, Section 9], but this has not been necessary for the applications of ECH so far.

Remark 2.13 In the special case where $c_1(\xi) \in H^2(Y;\mathbb{Z})$ is torsion and $\Gamma = 0$, the chain complex $\mathrm{ECC}_*(Y,\lambda,0)$ has a canonical absolute $\mathbb{Z}$–grading defined by

$$I(\alpha) = I(\alpha,\varnothing,Z) \in \mathbb{Z}$$

for any $Z \in H_2(Y,\alpha,\varnothing)$. This is well defined by (2-7).

Definition 2.14 An almost complex structure J on $\mathbb{R}\times Y$ is *λ–compatible* if $J\partial_s = R$, where s denotes the $\mathbb{R}$ coordinate, J is invariant under the $\mathbb{R}$ action on $\mathbb{R}\times Y$ by translation of s, and $J(\xi) = \xi$, rotating positively with respect to $d\lambda$.

If J is a generic λ–compatible almost complex structure, one defines a differential

$$\partial_J \colon \mathrm{ECC}_*(Y,\lambda,\Gamma) \to \mathrm{ECC}_{*-1}(Y,\lambda,\Gamma)$$

whose coefficient from α to β is a count of "J–holomorphic currents" that represent classes $Z \in H_2(Y,\alpha,\beta)$ with ECH index $I(\alpha,\beta,Z) = 1$; see [30, Section 3] for details. It is shown in [31] that $\partial_J^2 = 0$. The *embedded contact homology* $\mathrm{ECH}_*(Y,\lambda,\Gamma,J)$ is defined to be the homology of the chain complex $(\mathrm{ECC}_*(Y,\lambda,\Gamma),\partial_J)$. A theorem of Taubes [43], tensored with $\mathbb{Z}/2$, asserts that there is a canonical isomorphism

$$\mathrm{ECH}_*(Y,\lambda,\Gamma,J) = \widehat{HM}^{-*}(Y,\mathfrak{s}_\xi + \mathrm{PD}(\Gamma)) \otimes \mathbb{Z}/2, \tag{2-9}$$

where the right-hand side is a version of Seiberg–Witten Floer cohomology as defined by Kronheimer and Mrowka [38], and $\mathfrak{s}_\xi$ is a spin-c structure on Y determined by ξ. In particular, ECH depends only[5] on the triple (Y,ξ,Γ), and so we can denote it by $\mathrm{ECH}_*(Y,\xi,\Gamma)$.

When Y is connected, there is also a well-defined "U–map"

$$U \colon \mathrm{ECH}_*(Y,\xi,\Gamma) \to \mathrm{ECH}_{*-2}(Y,\xi,\Gamma). \tag{2-10}$$

This is induced by a chain map

$$U_{J,z} \colon (\mathrm{ECC}_*(Y,\lambda,\Gamma),\partial_J) \to (\mathrm{ECC}_{*-2}(Y,\xi,\Gamma),\partial_J)$$

which counts J–holomorphic currents with ECH index 2 passing through a generic basepoint $z \in \mathbb{R}\times Y$. The assumption that Y is connected implies that the induced map on homology does not depend on the choice of basepoint z; see [33, Section 2.5] for details. Taubes showed in [44, Theorem 1.1] that under the isomorphism (2-9), the map on homology induced by $U_{J,z}$ agrees with a corresponding map on Seiberg–Witten Floer cohomology. We thus obtain a well-defined U–map (2-10).

[5] In a sense, ECH does not depend on the contact structure either; see [30, Remark 1.7] for explanation.

Definition 2.15 A *U–sequence* for Γ is a sequence $\{\sigma_k\}_{k\geq 1}$ where each σ_k is a nonzero homogeneous class in $\mathrm{ECH}_*(Y,\xi,\Gamma)$, and $U\sigma_{k+1}=\sigma_k$ for each $k\geq 1$.

We will need the following nontriviality result for the U–map, which is proved by combining Taubes' isomorphism (2-9) with results from Kronheimer and Mrowka [38]:

Proposition 2.16 [12, Proposition 2.3] *If $c_1(\xi)+2\,\mathrm{PD}(\Gamma)\in H^2(Y;\mathbb{Z})$ is torsion, then a U–sequence for Γ exists.*

2.4 Spectral invariants

If $\alpha=\{(\alpha_i,m_i)\}$ is an orbit set, define its *symplectic action* by

$$\mathcal{A}(\alpha)=\sum_i m_i\int_{\alpha_i}\lambda.$$

Note here that $\int_{\alpha_i}\lambda$ agrees with the period of α_i, because $\lambda(R)=1$.

Assume now that λ is nondegenerate. For $L\in\mathbb{R}$, define $\mathrm{ECC}^L_*(Y,\lambda,\Gamma)$ to be the subspace of $\mathrm{ECC}_*(Y,\lambda,\Gamma)$ spanned by ECH generators α with symplectic action $\mathcal{A}(\alpha)<L$. It follows from the definition of "λ–compatible almost complex structure" that ∂_J maps ECC^L to itself; see [30, Section 1.4]. We define the *filtered ECH* to be the homology of this subcomplex, which we denote by $\mathrm{ECH}^L_*(Y,\lambda,\Gamma)$. The inclusion of chain complexes induces a map

$$\imath_L\colon \mathrm{ECH}^L_*(Y,\lambda,\Gamma)\to\mathrm{ECH}_*(Y,\xi,\Gamma).$$

It is shown in [34, Theorem 1.3] that the filtered homology $\mathrm{ECH}^L_*(Y,\lambda,\Gamma)$ and the map $\imath_L$ do not depend on the choice of J. However, unlike the usual ECH, filtered ECH does depend on the contact form λ and not just on the contact structure ξ.

Definition 2.17 [29] If $0\neq\sigma\in\mathrm{ECH}_*(Y,\xi,\Gamma)$, define the *spectral invariant*

$$c_\sigma(Y,\lambda)=\inf\{L\mid\sigma\in\mathrm{Im}(\imath_L)\}\in\mathbb{R}.$$

An equivalent definition is that $c_\sigma(Y,\lambda)$ is the minimum L such that the class σ can be represented by a cycle in the chain complex $(\mathrm{ECC}_*(Y,\lambda,\Gamma),\partial_J)$ which is a sum of ECH generators each having symplectic action $\leq L$. In particular, by definition, $c_\sigma(Y,\lambda)=\mathcal{A}(\alpha)$ for some ECH generator α with $[\alpha]=\Gamma$.

We can change the contact form λ, without changing the contact structure ξ, by multiplying λ by a smooth function $f\colon Y \to \mathbb{R}^{>0}$. As explained in [11, Section 2.5], it turns out that even when λ is degenerate, one can still define $c_\sigma(Y,\lambda)$ as a limit of spectral invariants $c_\sigma(Y, f_n\lambda)$ where $f_n\lambda$ is nondegenerate and $f_n \to 1$ in C^0.

These spectral invariants have the following important properties:

Proposition 2.18 *Let Y be a closed connected three-manifold, and let λ be a (possibly degenerate) contact form on Y. Then:*

(a) *If $0 \neq \sigma \in \mathrm{ECH}_*(Y,\xi,\Gamma)$, then*

$$c_\sigma(Y,\lambda) = \mathcal{A}(\alpha)$$

for some orbit set α with $[\alpha] = \Gamma$.

(b) *If $\sigma \in \mathrm{ECH}_*(Y,\xi,\Gamma)$ and $U\sigma \neq 0$, then*

$$c_{U\sigma}(Y,\lambda) \le c_\sigma(Y,\lambda). \tag{2-11}$$

If there are only finitely many simple Reeb orbits, then the inequality (2-11) *is strict.*

(c) ***Volume property*** *If $c_1(\xi) + 2\,\mathrm{PD}(\Gamma) \in H^2(Y;\mathbb{Z})$ is torsion, and if $\{\sigma_k\}_{k\ge1}$ is a U–sequence for Γ, then*

$$\lim_{k\to\infty} \frac{c_{\sigma_k}(Y,\lambda)^2}{k} = 2\,\mathrm{vol}(Y,\lambda).$$

Proof As noted above, part (a) holds by definition when λ is nondegenerate, and in the degenerate case it follows from a compactness argument for Reeb orbits; see [11, Lemma 3.1(a)].

If λ is nondegenerate, then since the chain map $U_{J,z}$ counts J–holomorphic curves it decreases symplectic action like the differential, so strict inequality in (2-11) holds. The not necessarily strict inequality (2-11) in the degenerate case follows by a limiting argument. The fact that (2-11) is strict for degenerate contact forms with only finitely many simple Reeb orbits[6] is proved by a more subtle compactness argument for holomorphic curves in [11, Lemma 3.1(b)].

Part (c), the most nontrivial part, is a special case of [13, Theorem 1.3]. □

[6] The equality $c_{U\sigma}(Y,\lambda) = c_\sigma(Y,\lambda)$ is possible for degenerate contact forms with infinitely many simple Reeb orbits. This happens, for example, for some classes σ when Y is an ellipsoid $\partial E(a,b)$ with a/b rational.

3 The ECH index and perturbations

The goal of this section is to prove Proposition 3.1 below, which gives an upper bound on how much the ECH index can change when one perturbs the contact form. This is an important ingredient in the proof of Theorems 1.2 and 1.5.

To state the proposition, let λ be a contact form on a closed three-manifold Y, and let $\lambda_n = f_n\lambda$ be a sequence of contact forms with $f_n \to 1$ in C^2. In the case of interest, λ will be degenerate, while each of the contact forms λ_n will be nondegenerate.

Fix an orbit set $\alpha = \{(\alpha_i, m_i)\}$ for λ, and let N be a disjoint union of tubular neighborhoods N_i of the simple Reeb orbits α_i. Consider a sequence of orbit sets $\alpha(n)$ for λ_n that converges to α as currents. In particular this implies that if n is sufficiently large, and if we write $\alpha' = \alpha(n)$, then α' is contained in N, and its intersection with N_i is homologous in N_i to $m_i\alpha_i$. There is then a unique $W_\alpha \in H_2(Y, \alpha', \alpha)$ that is contained in N.

Likewise fix an orbit set $\beta = \{(\beta_j, n_j)\}$ for λ along with disjoint tubular neighborhoods of the simple Reeb orbits β_j, and consider a sequence of orbit sets $\beta(n)$ for λ_n that converges to β as currents. Then, for k sufficiently large, writing $\beta' = \beta(n)$, we obtain a distinguished $W_\beta \in H_2(Y, \beta', \beta)$.

For fixed large n there is now a bijection

$$H_2(Y, \alpha, \beta) \simeq H_2(Y, \alpha', \beta')$$

sending $Z \in H_2(Y, \alpha, \beta)$ to

$$Z' = Z + W_\alpha - W_\beta \in H_2(Y, \alpha', \beta').$$

Proposition 3.1 *With the notation as above, for fixed orbit sets α and β, if n is sufficiently large, then*

$$|I(\alpha, \beta, Z) - I(\alpha', \beta', Z')| \le 2\left(\sum_i m_i + \sum_j n_j\right).$$

Here $I(\alpha, \beta, Z)$ denotes the ECH index for λ, and $I(\alpha', \beta', Z')$ denotes the ECH index for λ_n.

3.1 Reduction to a local statement

We now reduce Proposition 3.1 to a local statement, Proposition 3.3, below.

Let γ be an oriented knot in Y, and let N be a tubular neighborhood of γ with an identification $N \simeq S^1 \times D^2$. By a "braid in N with d strands", we mean an oriented knot in N which is positively transverse to the D^2 fibers and which intersects each fiber d times.

Definition 3.2 • A *weighted braid* in N with m strands is a finite set of pairs $\zeta = \{(\zeta_i, m_i)\}$ where the ζ_i are disjoint braids in N with d_i strands, the m_i are positive integers and $\sum_i m_i d_i = m$.

• If τ is an oriented trivialization of the normal bundle of γ, then for $i \neq j$ there is a well-defined linking number $\ell_\tau(\zeta_i, \zeta_j) \in \mathbb{Z}$, as discussed in Section 2.1. Similarly, for each i there is a well-defined writhe $w_\tau(\gamma_i) \in \mathbb{Z}$; see [27, Section 2.6]. Define the *writhe* of the weighted braid ζ by

$$w_\tau(\zeta) = \sum_i m_i^2 w_\tau(\zeta_i) + \sum_{i \neq j} m_i m_j \ell_\tau(\zeta_i, \zeta_j). \tag{3-1}$$

Suppose now that γ is a simple Reeb orbit for λ, and that the normal bundle identification $N \simeq S^1 \times D^2$ above is chosen so that the Reeb vector field for λ is transverse to the D^2 fibers. If $\lambda' = f\lambda$ with f sufficiently C^2 close to 1, then the Reeb vector field for λ' in N is also transverse to the D^2 fibers. Suppose that this is the case.

Let $\gamma' = \{(\gamma'_k, m_k)\}$ be an orbit set for λ' which is contained in N. We can regard γ' as a weighted braid with m strands for some positive integer m. Also note that a trivialization τ of $\gamma^*\xi$ extends to a trivialization of ξ over the entire tubular neighborhood N, and thus canonically induces a homotopy class of trivialization τ' of ξ over the Reeb orbits γ'_k. We can now state:

Proposition 3.3 *With the notation as above, if λ' is sufficiently C^2 close to λ and if γ' is sufficiently close to $m\gamma$ as a current, then*

$$\left| -w_\tau(\gamma') - \mathrm{CZ}^I_{\tau'}(\gamma') + \sum_{l=1}^{m} \mathrm{CZ}_\tau(\gamma^l) \right| \leq 2m. \tag{3-2}$$

Proof of Proposition 3.1 assuming Proposition 3.3 By shrinking the tubular neighborhoods, we can assume without loss of generality that the chosen tubular neighborhood of each orbit α_i or β_j has an identification with $S^1 \times D^2$, in which the Reeb flow of λ is transverse to the D^2 fibers.

In the orbit set α', each pair (α_i, m_i) gets replaced by an orbit set α'_i which represents a weighted braid with m_i strands in the tubular neighborhood of α_i. Likewise, each pair (β_j, n_j) gets replaced by an orbit set β'_j which represents a weighted braid with n_j strands in the tubular neighborhood of β_j. Let τ be a homotopy class of symplectic trivializations of ξ over the Reeb orbits α_i and β_j. As in Proposition 3.3, this canonically induces a homotopy class of symplectic trivializations τ' over the Reeb orbits in the orbit sets α_i' and β_j'.

Because τ and τ' extend to a trivialization of ξ over the tubular neighborhoods containing W_α and W_β, it follows from the definition of the relative first Chern class that

$$c_{\tau'}(\alpha',\beta',Z') = c_\tau(\alpha,\beta,Z). \tag{3-3}$$

By Proposition 3.3, if n is sufficiently large, then

$$\begin{gathered}\Bigl|-w_\tau(\alpha_i') - \mathrm{CZ}^I_{\tau'}(\alpha_i') + \sum_{k=1}^{m_i}\mathrm{CZ}_\tau(\alpha_i^k)\Bigr| \le 2m_i,\\ \Bigl|-w_\tau(\beta_j') - \mathrm{CZ}^I_{\tau'}(\beta_j') + \sum_{l=1}^{n_j}\mathrm{CZ}_\tau(\beta_j^l)\Bigr| \le 2n_j.\end{gathered}\tag{3-4}$$

By (2-6), (3-3) and (3-4), to complete the proof of Proposition 3.1 it is enough to show

$$Q_{\tau'}(\alpha',\beta',Z') = Q_\tau(\alpha,\beta,Z) + \sum_i w_\tau(\alpha_i') - \sum_j w_\tau(\beta_j'). \tag{3-5}$$

To prove (3-5), by Remark 2.6 it is enough to show

$$Q_\tau(\alpha',\alpha,W_\alpha) = \sum_i w_\tau(\alpha_i'), \quad Q_\tau(\beta',\beta,W_\beta) = \sum_j w_\tau(\beta_j').$$

Since the chosen tubular neighborhoods of the Reeb orbits of α_i are disjoint, and the chosen tubular neighborhoods of the Reeb orbits of β_j are disjoint, the above equations follow from Lemma 3.4. □

Lemma 3.4 *Let $\zeta = \{(\zeta_i, m_i)\}$ be a weighted braid with m strands as in Definition 3.2. Let W be the unique relative homology class in $H_2(N,\zeta,(\gamma,m))$. Then*

$$Q_\tau(\zeta,(\gamma,m),W) = w_\tau(\zeta). \tag{3-6}$$

Here τ defines a trivialization of the vertical tangent bundle of $N \to \gamma$ which then induces a trivialization of the normal bundle of each braid ζ_i.

Proof We can make an admissible representative S for W — see Definition 2.4 — whose intersection with $\{1-\epsilon\}\times N$ consists of m_i parallel (with respect to τ) copies of each ζ_i, and which shrinks radially towards γ as the $[-1,1]$ coordinate on $[-1,1]\times N$ goes down to -1. We can make another such admissible representative S', disjoint from S, whose intersection with $\{1-\epsilon\}\times N$ is parallel to the first and which likewise shrinks radially towards γ. Then in (2-1), the intersection number term vanishes. The first linking number term in (2-1) also vanishes, as it is a sum of linking numbers of braids in neighborhoods of the ζ_i; for each i, the braid from S and the braid from S', with respect to τ, are trivial and parallel, and thus have linking number zero. The second linking number term in (2-1) is a linking number in a neighborhood of γ and equals $w_\tau(\zeta)$. □

3.2 The structure of the braids

To prove Proposition 3.3, let γ be a simple Reeb orbit of λ, let N be a tubular neighborhood of γ as in Definition 3.2, and let τ be a trivialization of $\gamma^*\xi$. Let θ denote the rotation number $\theta_\tau(\gamma) \in \mathbb{R}$.

Suppose first that θ is irrational. Then the Reeb orbit γ and all of its covers are nondegenerate. Consequently, when λ' is sufficiently C^2 close to λ, there is a unique Reeb orbit γ_0' for λ' close (as a current) to γ, and for n large the only possibility for the orbit set γ' is that it is the singleton set $\gamma' = \{(\gamma_0', m)\}$. In this case Proposition 3.3 holds because $w_\tau(\gamma') = 0$ and the left-hand side of (3-2) is zero.

The nontrivial case of Proposition 3.3 is when the rotation number θ is rational. In this case we need to investigate the braids that can arise in γ'. The idea in what follows is to first analyze the case where the rotation number is an integer, and then reduce the general case to this one by taking an appropriate cover of a neighborhood of γ.

We start with the case where the rotation number is an integer. Here the picture is simple: each braid has just one strand, and the linking number of any two braids is given by the rotation number. More precisely:

Lemma 3.5 *With the above notation, suppose that the rotation number θ is an integer a. Let $\lambda_n = f_n\lambda$ with $f_n \to 1$ in C^2. Then:*

(a) *For a fixed positive integer d, if $\{\alpha_n\}$ is a sequence where each α_n is a simple Reeb orbit for λ_n in N which is a braid with d strands, with α_n converging as currents to $d\gamma$ as $n \to \infty$, then $d = 1$, and in particular the writhe $w_\tau(\alpha_n)$ equals 0 for n large enough.*

(b) *Given two sequences of simple Reeb orbits $\{\alpha_n\}$ and $\{\beta_n\}$ as in* (a) *with $\alpha_n \neq \beta_n$ for each n, if n is sufficiently large, then the linking number $\ell_\tau(\alpha_n, \beta_n)$ equals a.*

This lemma is proved in Section 3.3 below. We now consider the case where the rotation number is a rational number a/b that is not an integer. Here there is a similarly nice picture: each new simple Reeb orbit that can appear can be treated, for our purposes, like an (a,b) torus braid; see also Remark 3.7. More precisely:

Lemma 3.6 *With the above notation, suppose that the rotation number is $\theta = a/b$, where a and b are relatively prime integers with $b > 1$. Let $\lambda_n = f_n\lambda$ with $f_n \to 1$ in C^2. Then:*

(a) *For n sufficiently large, there is a unique simple Reeb orbit γ_0' for λ_n that is close to γ as a current.*

(b) *For a fixed integer $d > 1$, if $\{\alpha_n\}$ is a sequence where each α_n is a simple Reeb orbit for λ_n in N which is a braid with d strands, with α_n converging as currents to $d\gamma$ as $n \to \infty$, then $d = b$, the writhe $w_\tau(\alpha_n)$ equals $a(b-1)$, and the linking number $\ell_\tau(\gamma_0', \alpha_n)$ equals a.*

(c) *Given two sequences of Reeb orbits $\{\alpha_n\}$ and $\{\beta_n\}$ as in* (b) *with $\alpha_n \neq \beta_n$ for each n, if n is sufficiently large, then the linking number $\ell_\tau(\alpha_n, \beta_n)$ equals ab.*

Remark 3.7 In Lemma 3.6(b), we expect that one can further show that if n is sufficiently large then α_n is an (a,b) torus braid around γ_0'; however, we do not need this.

Proof of Lemma 3.6 assuming Lemma 3.5 Part (a) holds because the Reeb orbit γ is nondegenerate.

To prove part (b), we first note that, by the same argument as for (a), we must have that $d \geq b$, because for $0 < d < b$ the d^{th} iterate of γ has rotation number $da/b \notin \mathbb{Z}$.

Now let $\widetilde{N}$ denote the b–fold cyclic cover of the tubular neighborhood N, with the pullback of the contact form λ_n. There is a unique simple Reeb orbit $\widetilde{\gamma_0'}$ in $\widetilde{N}$ whose projection to N is a b–fold cover of γ_0'. In addition, by lifting the Reeb orbit α_n to a Reeb trajectory in $\widetilde{N}$ and extending it by the Reeb flow if needed, we obtain a simple Reeb orbit $\tilde{\alpha}_n$ in $\widetilde{N}$ whose projection to N is a cover of α_n. By Lemma 3.5(a), if n is sufficiently large, then $\tilde{\alpha}_n$ is a braid with one strand in $\widetilde{N}$, hence α_n has at most b strands. Thus, $d = b$. By Lemma 3.5(b) we have $\ell_\tau(\widetilde{\gamma_0'}, \tilde{\alpha}_n) = a$ in $\widetilde{N}$, and it follows that $\ell_\tau(\gamma_0', \alpha_n) = a$.

We now compute the writhe $w_\tau(\alpha_n)$. There are b possibilities for the Reeb orbit $\tilde{\alpha}_n$ in the previous paragraph, which we denote by η_l for $l \in \mathbb{Z}/b$, ordered so that the $\mathbb{Z}/b$ action on $\widetilde{N}$ by deck transformations sends η_l to η_{l+1}. The writhe $w_\tau(\alpha_n)$ is a signed count of crossings of two strands of α_n. Each such crossing corresponds to a crossing of some η_l with some $\eta_{l'}$ for $l \neq l'$, as well as crossings of η_{l+p} with $\eta_{l'+p}$

for $p = 1, \dots, b-1$, obtained from the first crossing by deck transformations. On the other hand, the linking number of η_l with $\eta_{l'}$ is one half the signed count of crossings of η_l with $\eta_{l'}$. Thus we obtain

$$w_\tau(\alpha_n) = \frac{1}{b}\sum_{l \neq l'} \ell_\tau(\eta_l, \eta_{l'}) = \frac{1}{b} b(b-1)a = a(b-1).$$

Here we are using Lemma 3.5(b) to get that $\ell_\tau(\eta_l, \eta_{l'}) = a$ when $l \neq l'$.

We now prove (c). Similarly to the previous calculation, each crossing counted by the linking number $\ell_\tau(\alpha_n, \beta_n)$ corresponds to b crossings of some lift of α_n (extended to a simple Reeb obit) with some lift of β_n (extended to a simple Reeb orbit). Thus the linking number we want is $1/b$ times the sum of the linking number of each of the b extended lifts of α_n with each of the b extended lifts of β_n, which is $(1/b)b^2a = ab$. □

Proof of Proposition 3.3 As explained above, we can assume that $\theta = a/b$, where a and b are relatively prime integers with $b > 0$. When $a/b \notin \mathbb{Z}$, the orbit set γ' consists of the orbit γ'_0 from Lemma 3.6(a) with multiplicity m_0 for some $m_0 \geq 0$, together with orbits γ'_k for $k \neq 0$ with multiplicities $m_k > 0$. When $a/b \in \mathbb{Z}$ the same is true except that we do not necessarily have a unique γ'_0 and we can take $m_0 = 0$. Since each γ'_k for $k \neq 0$ is close to a b–fold cover of γ, we have

$$m_0 + b\sum_{k\neq 0} m_k = m. \tag{3-7}$$

By (3-1) and Lemmas 3.5 and 3.6, if λ' is sufficiently C^2 close to λ and if γ' is sufficiently close to $m\gamma$ as a current, then

$$w_\tau(\gamma') = a(b-1)\sum_{k\neq 0} m_k^2 + 2am_0\sum_{k\neq 0} m_k + ab\sum_{0\neq k\neq k'\neq 0} m_k m'_k. \tag{3-8}$$

Now we consider Conley–Zehnder indices. By (2-5),

$$\mathrm{CZ}^I_{\tau'}(\gamma') = \sum_{l=1}^{m_0} \mathrm{CZ}_{\tau'}((\gamma'_0)^l) + \sum_{k\neq 0}\sum_{l=1}^{m_k} \mathrm{CZ}_{\tau'}((\gamma'_k)^l). \tag{3-9}$$

For a positive integer $l \leq m$, if λ' is sufficiently close to λ, then with respect to τ the Reeb orbit $(\gamma'_0)^l$ has rotation number close to $(a/b)l$, and each Reeb orbit $(\gamma'_k)^l$ for $k \neq 0$ has rotation number close to al. Then, by (2-4) and (3-9),

$$\left|\mathrm{CZ}^I_{\tau'}(\gamma') - \sum_{l=1}^{m_0}\frac{2al}{b} - \sum_{k\neq 0}\sum_{l=1}^{m_k} 2al\right| \leq m_0 + \sum_{k\neq 0} m_k.$$

It follows from this and (3-7) that

$$\left|\mathrm{CZ}^{I}_{\tau'}(\gamma') - \frac{a}{b}(m_0^2 + m_0) - a\sum_{k\neq 0}(m_k^2 + m_k)\right| \le m. \tag{3-10}$$

Finally, by (2-4),

$$\left|\sum_{l=1}^{m}\mathrm{CZ}_\tau(\gamma^l) - \frac{a}{b}(m^2 + m)\right| \le m.$$

Then, by (3-7),

$$\left|\sum_{l=1}^{m}\mathrm{CZ}_\tau(\gamma^l) - \frac{a}{b}(m_0^2 + m_0) - a(2m_0+1)\sum_{k\neq 0} m_k - ab\Bigl(\sum_{k\neq 0} m_k\Bigr)^2\right| \le m. \tag{3-11}$$

Combining (3-8), (3-10) and (3-11) gives the desired estimate (3-2). □

3.3 Perturbations of degenerate flows

To conclude the proof of Proposition 3.3 we now prove Lemma 3.5.

As in the statement of the lemma, let γ be a simple Reeb orbit of λ of period T, and let $\lambda_n = f_n\lambda$, where $f_n \to 1$ in C^2. Let ϕ^t and ϕ^t_n denote the time t flows of the Reeb vector fields for λ and λ_n, respectively. Let $p \in \gamma$, and let $P_\gamma\colon \xi_p \to \xi_p$ denote the linearized return map (1-2).

Lemma 3.8 *Let $\{(p_n, T_n)\}_{n=1,\dots}$ be a sequence in $Y \times (0,\infty)$ satisfying*:

(c1) $\phi_n^{T_n}(p_n) = p_n \to p$.

(c2) $\phi_n^{T_n/j}(p_n) \neq p_n$ *for all integers* $j \ge 2$ *and all* n.

(c3) $T_n \to T_\infty \in [0,\infty)$.

Then one of the following alternatives holds:

(a1) $T_\infty = T$.

(a2) $T_\infty = Td$ *for some integer* $d \ge 2$, *and the eigenvalues of* P_γ *that are roots of unity of degree* d *generate multiplicatively all roots of unity of order* d.

Proof This is a special case of a result of Bangert [3, Proposition 1] for C^1 flows. □

In the situation of Lemma 3.5, more can be said:

Corollary 3.9 *Suppose that the eigenvalues of* P_γ *are real and positive. Let* $\{(p_n, T_n)\}$ *be a sequence satisfying conditions* (c1), (c2) *and* (c3) *of Lemma 3.8. Then alternative* (a2) *does not hold.*

Proof The only root of unity that can be an eigenvalue of P_γ is 1, hence the set of eigenvalues of P_γ does not generate multiplicatively the group of roots of unity of order d when $d \geq 2$. □

Proof of Lemma 3.5 Part (a) follows from Corollary 3.9.

To prove part (b), fix a diffeomorphism Φ from the tubular neighborhood N of γ to $(\mathbb{R}/T\mathbb{Z})\times\mathbb{C}$ such that γ corresponds to $(\mathbb{R}/T\mathbb{Z})\times\{0\}$, the Reeb vector field R_n of λ_n is transverse to the $\mathbb{C}$ fibers for n sufficiently large (assume that n is this large below), and the derivative of Φ in the normal direction along γ agrees with the trivialization τ. We omit the diffeomorphism Φ from the notation below and write points in N using the coordinates $(t,z)\in(\mathbb{R}/T\mathbb{Z})\times\mathbb{C}$.

By part (a), by taking n large enough we can assume that α_n and β_n have the same period as γ. After reparametrization, the Reeb orbit α_n is given by a map

$$\mathbb{R}/T\mathbb{Z}\to(\mathbb{R}/T\mathbb{Z})\times\mathbb{C},\quad t\mapsto(t,\hat{\alpha}_n(t)),$$

where $\hat{\alpha}_n:\mathbb{R}/T\mathbb{Z}\to\mathbb{C}$. Likewise the Reeb orbit β_n is given by a map $\hat{\beta}_n:\mathbb{R}/T\mathbb{Z}\to\mathbb{C}$. We have

$$\ell_\tau(\alpha_n,\beta_n)=\operatorname{wind}(\hat{\alpha}_n-\hat{\beta}_n), \tag{3-12}$$

where the right-hand side denotes the winding number of the loop

$$\hat{\alpha}_n-\hat{\beta}_n:\mathbb{R}/T\mathbb{Z}\to\mathbb{C}^*.$$

We now compute the right-hand side of (3-12). There is a convex neighborhood U of 0 in $\mathbb{C}$ such that, if n is sufficiently large (which we assume below), then the following two conditions hold: First, $\hat{\alpha}_n(0),\hat{\beta}_n(0)\in U$. Second, for each $t\in[0,T]$ there is a well-defined map $\psi_n^t:U\to\mathbb{C}$ such that, for $z\in U$, the flow of the Reeb vector field R_n starting at $(0,z)$ first hits $\{t\}\times\mathbb{C}$ at the point $(t,\psi_n^t(z))$. In particular, it follows from the definition that

$$\hat{\alpha}_n(t)=\psi_n^t(\hat{\alpha}_n(0)),\quad \hat{\beta}_n(t)=\psi_n^t(\hat{\beta}_n(0)). \tag{3-13}$$

Now consider the derivative of ψ_n^t, which we denote by

$$D\psi_n^t:U\times\mathbb{C}\to\mathbb{C}.$$

By (3-13), we may apply the fundamental theorem of calculus to the function

$$s\mapsto\psi_n^t(s\hat{\alpha}_n(0)+(1-s)\hat{\beta}_n(0))$$

to obtain

$$\hat{\alpha}_n(t)-\hat{\beta}_n(t)=\int_0^1 D\psi_n^t(s\hat{\alpha}_n(0)+(1-s)\hat{\beta}_n(0),\hat{\alpha}_n(0)-\hat{\beta}_n(0))\,ds. \tag{3-14}$$

By the convergence of λ_n, if n is sufficiently large (which we assume below), then the amount that $D\psi_n^t(s\hat{\alpha}_n(0)+(1-s)\hat{\beta}_n(0),\cdot)$ rotates any vector as compared to $D\psi_n^t(0,\cdot)$ can be made arbitrarily small. It follows that the integrand in (3-14), and hence $\hat{\alpha}_n(t)-\hat{\beta}_n(t)$, has positive inner product with $D\psi_n^t(0,\hat{\alpha}_n(0)-\hat{\beta}_n(0))$. Thus, the right-hand side of (3-12) differs by less than $\frac{1}{4}$ from the rotation number (the change in argument divided by 2π) of the path

$$[0,T]\to\mathbb{C}^*,\quad t\mapsto D\psi_n^t(0,\hat{\alpha}_n(0)-\hat{\beta}_n(0)). \tag{3-15}$$

The rotation number of the linearized Reeb flow along γ differs from the rotation number of any individual vector by less than $\frac{1}{2}$. Hence, by again applying convergence of the λ_n as above, if n is sufficiently large then the rotation number of the path (3-15) differs by less than $\frac{1}{2}$ from a. Since the right-hand side of (3-12) is an integer which differs by less than $\frac{3}{4}$ from a, it must equal a. □

4 Two simple Reeb orbits implies nondegenerate

We now prove Theorem 1.2. Throughout this section assume that Y is a closed connected three-manifold and λ is a contact form on Y with exactly two simple Reeb orbits γ_1 and γ_2 of periods T_1 and T_2, respectively.

4.1 The homology classes of the Reeb orbits

Lemma 4.1 *The classes $[\gamma_i]\in H_1(Y)$ and $c_1(\xi)\in H^2(Y;\mathbb{Z})$ are torsion.*

Proof We use a similar argument to the proof of [12, Theorem 1.7].

Since every oriented three-manifold is spin, we can choose $\Gamma\in H_1(Y)$ such that $c_1(\xi)+2\,\mathrm{PD}(\Gamma)=0\in H^2(Y;\mathbb{Z})$. By Proposition 2.16, there exists a U–sequence $\{\sigma_k\}_{\geq 1}$ for Γ. Write $c_k=c_{\sigma_k}(Y,\lambda)\in\mathbb{R}$.

By Proposition 2.18(a),

$$c_k=m_{1,k}T_1+m_{2,k}T_2$$

for some nonnegative integers $m_{1,k}$ and $m_{2,k}$, and furthermore

$$m_{1,k}[\gamma_1]+m_{2,k}[\gamma_2]=\Gamma\in H_1(Y). \tag{4-1}$$

By Proposition 2.18(b), the sequence $\{c_k\}$ is strictly increasing. It then follows from (4-1) that there are infinitely many integral linear combinations of $[\gamma_1]$ and $[\gamma_2]$ that have the same value in $H_1(Y)$. Thus the kernel of the map

$$\mathbb{Z}^2 \to H_1(Y), \quad (m_1, m_2) \mapsto m_1[\gamma_1] + m_2[\gamma_2], \tag{4-2}$$

has rank at least 1.

In fact, the kernel of the map (4-2) must have rank at least 2; otherwise c_k would grow at least linearly in k, contradicting the sublinear growth in the volume property in Proposition 2.18(c). It follows that $[\gamma_1]$ and $[\gamma_2]$ are torsion. Since $c_1(\xi) + 2\,\mathrm{PD}(\Gamma) = 0$, we deduce that $c_1(\xi)$ is also torsion. □

4.2 Computing the ECH index

If m_1 and m_2 are nonnegative integers, we use the notation $\gamma_1^{m_1}\gamma_2^{m_2}$ to indicate the orbit set $\{(\gamma_1, m_1), (\gamma_2, m_2)\}$, with the element (γ_i, m_i) omitted when $m_i = 0$. Write $\alpha = \gamma_1^{m_1}\gamma_2^{m_2}$. If $[\alpha] = 0$, then it follows from Remark 2.13 and Lemma 4.1 that $I(\alpha) \in \mathbb{Z}$ is defined. We now give an explicit computation of $I(\alpha)$, following [33, Section 4.7].

Definition 4.2 Define the *linking number*

$$\ell(\gamma_1, \gamma_2) := \frac{\ell(\gamma_1^{l_1}, \gamma_2^{l_2})}{l_1 l_2} \in \mathbb{Q}, \tag{4-3}$$

where l_1 and l_2 are positive integers such that $l_i[\gamma_i] = 0 \in H_1(Y)$, and on the right-hand side ℓ denotes the usual integer-valued linking number of disjoint nullhomologous loops.

Definition 4.3 For $i = 1, 2$, define the *Seifert rotation number* $\phi_i \in \mathbb{R}$ as follows. Let τ be a trivialization of ξ over γ_i. Let $\theta_{i,\tau} = \theta_\tau(\gamma_i) \in \mathbb{R}$ denote the rotation number of γ_i with respect to τ. Let l_i be a positive integer such that $l_i[\gamma_i] = 0$. Define

$$Q_{i,\tau} := \frac{Q_\tau(\gamma_i^{l_i})}{l_i^2} \in \mathbb{Q}, \tag{4-4}$$

where $Q_\tau(\gamma_i^{l_i})$ is shorthand for $Q_\tau(\gamma_i^{l_i}, \varnothing, Z)$ for any $Z \in H_2(Y, \gamma_i^{l_i}, \varnothing)$. Note that $Q_{i,\tau}$ does not depend on Z by (2-3), and it does not depend on l_i either because Q_τ is quadratic in the relative homology class. Finally, define

$$\phi_i := Q_{i,\tau} + \theta_{i,\tau} \in \mathbb{R}. \tag{4-5}$$

The number ϕ_i does not depend on the choice of trivialization τ by the change of trivialization formulas in [27, Section 2].

Remark 4.4 When γ_i is nullhomologous, one can alternatively describe ϕ_i as follows. Let Σ be a Seifert surface spanned by γ_i. There is a distinguished homotopy class of trivialization τ' of ξ over γ_i, the "Seifert framing", for which the normal vector to Σ has winding number zero around γ_i. We have $Q_{i,\tau'} = 0$ by [27, Lemma 3.10]. It then follows that $\phi_i = \theta_{\tau'}(\gamma_i)$. In the general case when γ_i is rationally nullhomologous, one can similarly describe ϕ_i as the rotation number with respect to a rational framing of γ_i determined by a rational Seifert surface.

Lemma 4.5 *If* $m_1[\gamma_1] + m_2[\gamma_2] = 0 \in H_1(Y)$, *then*

$$I(\gamma_1^{m_1}\gamma_2^{m_2}) = \phi_1 m_1^2 + \phi_2 m_2^2 + 2\ell(\gamma_1,\gamma_2)m_1m_2 + O(m_1+m_2). \tag{4-6}$$

Proof Let l_i be a positive integer with $l_i[\gamma_i] = 0$. Similarly to (4-4), define

$$c_{i,\tau} := \frac{c_\tau(\gamma_i^{l_i})}{l_i} \in \mathbb{Q},$$

where $c_\tau(\gamma_i^{l_i})$ is shorthand for $c_\tau(\gamma_i^{l_i},\varnothing,Z)$ for any $Z \in H_2(Y,\gamma_i^{l_i},\varnothing)$. Then $c_{i,\tau}$ does not depend on Z by (2-2) since $c_1(\xi)$ is torsion, and it is independent of the choice of l_i because c_τ is linear in the relative homology class Z.

It follows from the definition of the ECH index and the facts that c_τ and Q_τ are linear and quadratic in the relative homology class (see [33, Section 4.2]) that

$$I(\gamma_1^{m_1}\gamma_2^{m_2}) = \sum_{i=1}^{2}(m_i c_{i,\tau} + m_i^2 Q_{i,\tau}) + 2m_1m_2\ell(\gamma_1,\gamma_2) + \sum_{i=1}^{2}\sum_{k=1}^{m_i}(\lfloor k\theta_{i,\tau}\rfloor + \lceil k\theta_{i,\tau}\rceil),$$

where $Q_{i,\tau}$ and $\theta_{i,\tau}$ are as in (4-5). Plugging in the approximation

$$\sum_{i=1}^{2}\sum_{k=1}^{m_i}(\lfloor k\theta_{i,\tau}\rfloor + \lceil k\theta_{i,\tau}\rceil) = \sum_{i=1}^{2} m_i^2\theta_{i,\tau} + O(m_1+m_2)$$

then gives (4-6). □

4.3 Using the volume property

Lemma 4.6 *The Seifert rotation numbers and linking number are given by*

$$\phi_i = \frac{T_i^2}{\mathrm{vol}(Y,\lambda)}, \quad \ell(\gamma_1,\gamma_2) = \frac{T_1T_2}{\mathrm{vol}(Y,\lambda)}.$$

Proof Both sides of the above equations are invariant under scaling the contact form by a positive constant, so we may assume without loss of generality that $\mathrm{vol}(Y,\lambda) = 1$.

By Proposition 2.16 and Lemma 4.1, we can choose a U–sequence $\{\sigma_k\}_{k\geq 1}$ for $\Gamma = 0$. Since the U map has degree -2, there is a constant $C \in \mathbb{Z}$ such that, for each positive integer k, the class σ_k has grading $C + 2k$. By Proposition 2.18(a), for each positive integer k there are nonnegative integers $m_{1,k}$ and $m_{2,k}$ such that

$$c_{\sigma_k}(Y,\lambda) = m_{1,k}T_1 + m_{2,k}T_2. \tag{4-7}$$

By the volume property of Proposition 2.18(c),

$$2k = (m_{1,k}T_1 + m_{2,k}T_2)^2 + o(k). \tag{4-8}$$

Fix k and write $\alpha_k = \gamma_1^{m_{1,k}}\gamma_2^{m_{2,k}}$. If λ' is a sufficiently C^2 close nondegenerate perturbation of λ, then by the same compactness argument that proves Proposition 2.18(a) there is an orbit set α'_k close to α_k as a current such that $I(\alpha'_k) = C + 2k$ (and also $\int_{\alpha'_k}\lambda'$ is close to $c_{\sigma_k}(Y,\lambda)$, although we do not need this). By Proposition 3.1,

$$C + 2k = I(\alpha_k) + O(m_{1,k} + m_{2,k}).$$

Combining this with Lemma 4.5, we get

$$2k = \phi_1 m_{1,k}^2 + \phi_2 m_{2,k}^2 + 2\ell(\gamma_1,\gamma_2)m_{1,k}m_{2,k} + O(m_{1,k} + m_{2,k}). \tag{4-9}$$

Putting together (4-8) and (4-9),

$$\begin{aligned}(\phi_1 - T_1^2)m_{1,k}^2 + (\phi_2 - T_2^2)m_{2,k}^2 + 2(\ell(\gamma_1,\gamma_2) - T_1T_2)m_{1,k}m_{2,k}&\\ = O(m_{1,k} + m_{2,k}) + o(k).&\end{aligned}$$

Consequently, if the sequence $(m_{2,k}/m_{1,k})_{k\geq 1}$ has an accumulation point $S \in [0,\infty]$, then the line in the (x,y)–plane of slope S through the origin is in the null space of the quadratic form

$$f(x,y) = (\phi_1 - T_1^2)x^2 + (\phi_2 - T_2^2)y^2 + 2(\ell(\gamma_1,\gamma_2) - T_1T_2)xy.$$

To complete the proof of the lemma, it now suffices to show that the sequence $(m_{2,k}/m_{1,k})_{k\geq 1}$ has at least three accumulation points, as then the quadratic form f must vanish identically. We claim that in fact this sequence has infinitely many accumulation points.

If the sequence has only finitely many accumulation points $S_1,\dots,S_n$, then for every $\epsilon > 0$ there exists $R > 0$ such that every point $(m_{1,k}, m_{2,k})$ is contained in the union of the disk $x^2 + y^2 \leq R^2$ and the cones around the lines of slope $S_1,\dots,S_n$ with angular width ϵ.

Since $\lim_{k\to\infty} c_{\sigma_k}^2/k = 2$, and since the points $(m_{1,k}, m_{2,k})$ are pairwise distinct by Proposition 2.18(b), by (4-7) it follows that, for large L, the number of points $(m_{1,k}, m_{2,k})$ contained in the triangle $T_1 x + T_2 y \leq L$ for $x \geq 0$ and $y \geq 0$ is approximately $\frac{1}{2}L^2$. As a result, there exists $\delta > 0$ such that, for all L sufficiently large, the fraction of lattice points in the above triangle that are contained in the sequence $(m_{1,k}, m_{2,k})_{k\geq 1}$ is at least δ. This gives a contradiction if ϵ in the previous paragraph is chosen sufficiently small. □

4.4 Completing the proof of nondegeneracy

Proof of Theorem 1.2 The ratio T_1/T_2 is irrational[7] by [11, Theorem 1.3]. Also, $\ell(\gamma_1, \gamma_2)$ is rational by the definition (4-3). It then follows from Lemma 4.6 that ϕ_1 and ϕ_2 are irrational.

By (4-5), since $Q_{i,\tau}$ is rational, it follows that the rotation number $\theta_{i,\tau}$ is irrational. Then P_{γ_i} has eigenvalues $e^{\pm 2\pi i \theta_{i,\tau}}$, so the Reeb orbits γ_i are irrationally elliptic. As explained in Section 1.1, it follows that all covers of γ_i are nondegenerate, so λ is nondegenerate. □

5 Additional dynamical information

To finish up, we now prove Theorem 1.5.

To prepare for the proof, recall that if Y is a closed oriented three-manifold, if ξ is a contact structure on Y with $c_1(\xi) = 0 \in H^2(Y;\mathbb{Z})$, and if γ is a nullhomologous transverse knot, then the *self-linking number* $\operatorname{sl}(\gamma) \in \mathbb{Z}$ is defined to be the difference between the Seifert framing (see Remark 4.4) and the framing given by a global trivialization of ξ. In the notation of Section 4.2,

$$\operatorname{sl}(\gamma) = Q_\tau(\gamma) - c_\tau(\gamma), \tag{5-1}$$

where τ is any trivialization of $\xi|_\gamma$.

Now suppose that γ above is a simple Reeb orbit. Let $\phi(\gamma) \in \mathbb{R}$ denote the rotation number of γ with respect to the Seifert framing as in Section 4.2, and let $\theta(\gamma) \in \mathbb{R}$ denote the rotation number of γ with respect to a global trivialization of ξ. Also, let

$$\mathrm{CZ}(\gamma) = \lfloor \theta(\gamma) \rfloor + \lceil \theta(\gamma) \rceil \in \mathbb{Z}$$

[7] The proof is simple: if T_1/T_2 is rational, T_1 and T_2 are both integer multiples of a single number, so Proposition 2.18(b) implies that the spectral invariants associated to a U–sequence grow at least linearly, contradicting the volume property.

denote the Conley–Zehnder index of γ with respect to a global trivialization. It follows from (5-1) that

$$\phi(\gamma) = \theta(\gamma) + \mathrm{sl}(\gamma). \tag{5-2}$$

Proof of Theorem 1.5 By Corollary 1.3, γ_1 and γ_2 are the core circles of a genus-one Heegaard splitting of Y. It follows from this topological description that $\ell(\gamma_1,\gamma_2) = 1/p$. Part (a) of the theorem then follows from Lemma 4.6.

To prove part (b), suppose first that $Y = S^3$. We know from Theorem 1.2 that λ is nondegenerate and there are no hyperbolic Reeb orbits. Then ξ is tight, because otherwise [21, Theorem 1.4] would give a hyperbolic Reeb orbit. Moreover, it follows from [25, Theorem 1.3], combined with [21, Theorem 1.4] and the fact that there are no Reeb orbits with $\mathrm{CZ} = 2$ (since Reeb orbits with even Conley–Zehnder index have integer rotation number and thus are hyperbolic), that one of the simple Reeb orbits, say γ_1, satisfies $\mathrm{sl}(\gamma_1) = -1$ and $\mathrm{CZ}(\gamma_1) = 3$, and is the binding of an open book decomposition with pages that are disk-like global surfaces of section for the Reeb flow. The return map on a page preserves an area form with finite total area, and hence it has a fixed point by Brouwer's translation theorem. This fixed point corresponds to the simple Reeb orbit γ_2, which is transverse to the pages of the open book. Since, on $S^3 \setminus \gamma_1$, the tangent spaces of the pages define a distribution that is isotopic to ξ keeping transversality with the Reeb direction, $\mathrm{sl}(\gamma_2) = -1$. Since $\mathrm{CZ}(\gamma_1) = 3$, we have $\theta(\gamma_1) \in (1,2)$, so, by (5-2), $\phi_1 \in (0,1)$. By Lemma 4.6 as used in (1-4), we have $\phi_1\phi_2 = 1$, so $\phi_2 > 1$. By (5-2), again, $\theta(\gamma_2) > 2$. It follows that all iterates of γ_1 and γ_2 have $\theta > 1$, so λ is dynamically convex.

To prove part (b) in the general case, let $\tilde{\lambda}$ denote the pullback of the contact form λ to the universal cover S^3 of Y. It follows from the Heegaard decomposition that γ_1 and γ_2 each have order p in $\pi_1(Y)$. Consequently $\tilde{\lambda}$ has exactly two simple Reeb orbits $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, which project to γ_1 and γ_2 as p–fold coverings. By the previous paragraph, $(S^3,\tilde{\lambda})$ is dynamically convex and tight, and it follows that (Y,λ) is dynamically convex and universally tight. □

References

[1] **P Albers**, **H Geiges**, **K Zehmisch**, *Pseudorotations of the* 2*–disc and Reeb flows on the* 3*–sphere*, Ergodic Theory Dynam. Systems 42 (2022) 402–436 MR Zbl

[2] **K Baker**, **J Etnyre**, *Rational linking and contact geometry*, from "Perspectives in analysis, geometry, and topology" (I Itenberg, B Jöricke, M Passare, editors), Progr. Math. 296, Springer (2012) 19–37 MR Zbl

[3] **V Bangert**, *On the lengths of closed geodesics on almost round spheres*, Math. Z. 191 (1986) 549–558 MR Zbl

[4] **D Bechara, Sr**, **U L Hryniewicz**, **P A S Salomão**, *On the relation between action and linking*, J. Mod. Dyn. 17 (2021) 319–336 MR Zbl

[5] **F Bourgeois**, **K Cieliebak**, **T Ekholm**, *A note on Reeb dynamics on the tight* 3*–sphere*, J. Mod. Dyn. 1 (2007) 597–613 MR Zbl

[6] **B Bramham**, *Periodic approximations of irrational pseudo-rotations using pseudoholomorphic curves*, Ann. of Math. 181 (2015) 1033–1086 MR Zbl

[7] **E Çineli**, **V L Ginzburg**, **B Z Gürel**, *Pseudo-rotations and holomorphic curves*, Selecta Math. 26 (2020) art. id. 78 MR Zbl

[8] **V Colin**, **P Dehornoy**, **A Rechtman**, *On the existence of supporting broken book decompositions for contact forms in dimension* 3, Invent. Math. 231 (2023) 1489–1539 MR Zbl

[9] **B Collier**, **E Kerman**, **B M Reiniger**, **B Turmunkh**, **A Zimmer**, *A symplectic proof of a theorem of Franks*, Compos. Math. 148 (2012) 1969–1984 MR Zbl

[10] **C R Cornwell**, *Berge duals and universally tight contact structures*, Topology Appl. 236 (2018) 26–43 MR Zbl

[11] **D Cristofaro-Gardiner**, **M Hutchings**, *From one Reeb orbit to two*, J. Differential Geom. 102 (2016) 25–36 MR Zbl

[12] **D Cristofaro-Gardiner**, **M Hutchings**, **D Pomerleano**, *Torsion contact forms in three dimensions have two or infinitely many Reeb orbits*, Geom. Topol. 23 (2019) 3601–3645 MR Zbl

[13] **D Cristofaro-Gardiner**, **M Hutchings**, **V G B Ramos**, *The asymptotics of ECH capacities*, Invent. Math. 199 (2015) 187–214 MR Zbl

[14] **J Etnyre**, **R Ghrist**, *Tight contact structures via dynamics*, Proc. Amer. Math. Soc. 127 (1999) 3697–3706 MR Zbl

[15] **B Fayad**, **A Katok**, *Constructions in elliptic dynamics*, Ergodic Theory Dynam. Systems 24 (2004) 1477–1520 MR Zbl

[16] **B Fayad**, **R Krikorian**, *Some questions around quasi-periodic dynamics*, from "Proceedings of the International Congress of Mathematicians" (B Sirakov, P N de Souza, M Viana, editors), volume 3, World Sci., Hackensack, NJ (2018) 1927–1950 MR Zbl

[17] **J Franks**, *Geodesics on S^2 and periodic points of annulus homeomorphisms*, Invent. Math. 108 (1992) 403–418 MR Zbl

[18] **V L Ginzburg**, **B Z Gürel**, *Hamiltonian pseudo-rotations of projective spaces*, Invent. Math. 214 (2018) 1081–1130 MR Zbl

[19] **B Z Gürel**, *Perfect Reeb flows and action-index relations*, Geom. Dedicata 174 (2015) 105–120 MR Zbl

[20] **H Hofer**, **K Wysocki**, **E Zehnder**, *A characterisation of the tight three-sphere*, Duke Math. J. 81 (1995) 159–226 MR Zbl

[21] **H Hofer**, **K Wysocki**, **E Zehnder**, *Unknotted periodic orbits for Reeb flows on the three-sphere*, Topol. Methods Nonlinear Anal. 7 (1996) 219–244 MR Zbl

[22] **H Hofer**, **K Wysocki**, **E Zehnder**, *A characterization of the tight* 3*–sphere, II*, Comm. Pure Appl. Math. 52 (1999) 1139–1177 MR Zbl

[23] **K Honda**, *On the classification of tight contact structures, I*, Geom. Topol. 4 (2000) 309–368 MR Zbl

[24] **U L Hryniewicz**, **J E Licata**, **P A S Salomão**, *A dynamical characterization of universally tight lens spaces*, Proc. Lond. Math. Soc. 110 (2015) 213–269 MR Zbl

[25] **U Hryniewicz**, **P A S Salomão**, *On the existence of disk-like global sections for Reeb flows on the tight* 3*–sphere*, Duke Math. J. 160 (2011) 415–465 MR Zbl

[26] **M Hutchings**, *An index inequality for embedded pseudoholomorphic curves in symplectizations*, J. Eur. Math. Soc. 4 (2002) 313–361 MR Zbl

[27] **M Hutchings**, *The embedded contact homology index revisited*, from "New perspectives and challenges in symplectic field theory" (M Abreu, F Lalonde, L Polterovich, editors), CRM Proc. Lecture Notes 49, Amer. Math. Soc., Providence, RI (2009) 263–297 MR Zbl

[28] **M Hutchings**, *Taubes's proof of the Weinstein conjecture in dimension three*, Bull. Amer. Math. Soc. 47 (2010) 73–125 MR Zbl

[29] **M Hutchings**, *Quantitative embedded contact homology*, J. Differential Geom. 88 (2011) 231–266 MR Zbl

[30] **M Hutchings**, *Lecture notes on embedded contact homology*, from "Contact and symplectic topology" (F Bourgeois, V Colin, A Stipsicz, editors), Bolyai Soc. Math. Stud. 26, János Bolyai Math. Soc., Budapest (2014) 389–484 MR Zbl

[31] **M Hutchings**, **C H Taubes**, *Gluing pseudoholomorphic curves along branched covered cylinders, I*, J. Symplectic Geom. 5 (2007) 43–137 MR Zbl

[32] **M Hutchings**, **C H Taubes**, *Gluing pseudoholomorphic curves along branched covered cylinders, II*, J. Symplectic Geom. 7 (2009) 29–133 MR Zbl

[33] **M Hutchings**, **C H Taubes**, *The Weinstein conjecture for stable Hamiltonian structures*, Geom. Topol. 13 (2009) 901–941 MR Zbl

[34] **M Hutchings**, **C H Taubes**, *Proof of the Arnold chord conjecture in three dimensions, II*, Geom. Topol. 17 (2013) 2601–2688 MR Zbl

[35] **K Irie**, *Dense existence of periodic Reeb orbits and ECH spectral invariants*, J. Mod. Dyn. 9 (2015) 357–363 MR Zbl

[36] **B Joly**, *The Calabi invariant for Hamiltonian diffeomorphisms of the unit disk*, preprint (2021) arXiv 2102.09352

[37] **A B Katok**, *Ergodic perturbations of degenerate integrable Hamiltonian systems*, Izv. Akad. Nauk SSSR Ser. Mat. 37 (1973) 539–576 MR Zbl In Russian; translated in Math. USSR Izv. 7 (1973) 535–571

[38] **P Kronheimer**, **T Mrowka**, *Monopoles and three-manifolds*, New Math. Monogr. 10, Cambridge Univ. Press (2007) MR Zbl

[39] **F Le Roux**, **S Seyfaddini**, *The Anosov–Katok method and pseudo-rotations in symplectic dynamics*, J. Fixed Point Theory Appl. 24 (2022) art. id. 36 MR Zbl

[40] **A Pirnapasov**, *Hutchings' inequality for the Calabi invariant revisited with an application to pseudo-rotations*, preprint (2021) arXiv 2102.09533

[41] **E Shelukhin**, *Pseudo-rotations and Steenrod squares*, J. Mod. Dyn. 16 (2020) 289–304 MR Zbl

[42] **C H Taubes**, *The Seiberg–Witten equations and the Weinstein conjecture*, Geom. Topol. 11 (2007) 2117–2202 MR Zbl

[43] **C H Taubes**, *Embedded contact homology and Seiberg–Witten Floer cohomology, I*, Geom. Topol. 14 (2010) 2497–2581 MR Zbl

[44] **C H Taubes**, *Embedded contact homology and Seiberg–Witten Floer cohomology, V*, Geom. Topol. 14 (2010) 2961–3000 MR Zbl

[45] **W Wang**, **X Hu**, **Y Long**, *Resonance identity, stability, and multiplicity of closed characteristics on compact convex hypersurfaces*, Duke Math. J. 139 (2007) 411–462 MR Zbl

DCG: *Department of Mathematics, University of California, Santa Cruz*
Santa Cruz, CA, United States

DCG: *School of Mathematics, Institute for Advanced Study*
Princeton, NJ, United States

Current address for DCG: *Department of Mathematics, University of Maryland*
College Park, MD, United States

UH: *RWTH Aachen*
Aachen, Germany

MH: *Department of Mathematics, University of California, Berkeley*
Berkeley, CA, United States

HL: *School of Mathematics and Statistics, Wuhan University*
Wuhan, China

dcristof@umd.edu, hryniewicz@mathga.rwth-aachen.de,
hutching@berkeley.edu, huiliu00031514@whu.edu.cn

Proposed: András I Stipsicz Received: 29 September 2021
Seconded: Ciprian Manolescu, Leonid Polterovich Revised: 24 January 2022

Guidelines for Authors

Submitting a paper to Geometry & Topology

Papers must be submitted using the upload page at the GT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Geometry & Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in GT is available on the GT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

GEOMETRY & TOPOLOGY

Volume 27 Issue 9 (pages 3387–3831) 2023