Geometry & Topology

Volume 28 (2024)

Issue 1 (pages 1–496)

msp

GEOMETRY & TOPOLOGY
msp.org/gt

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

The subscription price for 2024 is US $805/year for the electronic version, and $1135/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 **28**:1 (2024) 1–125
DOI: 10.2140/gt.2024.28.1
Published: 27 February 2024

Homological invariants of codimension 2 contact submanifolds

LAURENT CÔTÉ
FRANÇOIS-SIMON FAUTEUX-CHAPLEAU

Codimension 2 contact submanifolds are the natural generalization of transverse knots to contact manifolds of arbitrary dimension. We construct new invariants of codimension 2 contact submanifolds. Our main invariant can be viewed as a deformation of the contact homology algebra of the ambient manifold. We describe various applications of these invariants to contact topology. In particular, we exhibit examples of codimension 2 contact embeddings into overtwisted and tight contact manifolds which are formally isotopic but fail to be isotopic through contact embeddings. We also give new obstructions to certain relative symplectic and Lagrangian cobordisms.

53D10, 53D42

1 Introduction

1.1 Overview

The purpose of this paper is to introduce new invariants of codimension 2 contact submanifolds. Given a closed, co-oriented contact manifold (Y, ξ) and a codimension 2 contact submanifold V with trivial normal bundle, our main construction produces a unital, $\mathbb{Z}/2$-graded $\mathbb{Q}[U]$-algebra

$$CH_\bullet(Y, \xi, V; \mathfrak{r}). \tag{1-1}$$

Here $\mathfrak{r} = (\alpha_V, \tau, r)$ is a triple consisting of

(i) a nondegenerate contact form α_V on V inducing $\xi_V := \xi|_V \cap TV$,

(ii) a trivialization τ of $N_{Y/V}$, and

(iii) a real number $r > 0$.

This triple is required to satisfy certain conditions (stated in Definition 3.3), and should be viewed as encoding a choice of Reeb dynamics in an infinitesimally small neighborhood of V. The (nonempty) set of all such triples is denoted by $\mathfrak{R}(Y, \xi, V)$.

The invariant (1-1) can be viewed as a deformation of the contact homology algebra $CH_\bullet(Y, \xi)$, as explained in Remark 1.1 below. In particular, U is a formal variable and there is a natural map

$$\mathrm{ev}_{U=1}\colon CH_\bullet(Y, \xi, V; \mathfrak{r}) \to CH_\bullet(Y, \xi) \tag{1-2}$$

obtained by setting $U = 1$.

The algebra $CH_\bullet(Y, \xi, V; \mathfrak{r})$ is generated by (good) Reeb orbits for an auxiliary nondegenerate contact form λ on (Y, ξ). See Pardon [61, Definition 2.59] for the notion of a good Reeb orbit; all nondegenerate Reeb orbits are good except for certain even-degree covers of simple Reeb orbits with odd Conley–Zehnder index. The form λ is required to be *adapted* to $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$, which means in particular that V is preserved by the Reeb flow of λ; see Definition 3.4.

The differential is defined as in ordinary contact homology by counting pseudoholomorphic curves in the symplectization $\widehat{Y}$, where the additional U variable keeps track of the intersection number of curves with the symplectization $\widehat{V} \subset \widehat{Y}$. More precisely, we fix an almost complex structure $J : \xi \to \xi$ which is compatible with the symplectic form $d\lambda$ and preserves $\xi|_V \cap TV \subset TY|_V$. We then consider $\widehat{J}$–holomorphic curves in $\widehat{Y}$, where $\widehat{J} = -\partial_t \otimes \lambda + R_\lambda \otimes dt + J$. The differential is defined on generators by (roughly)[1] the formula

$$d(\gamma) = \sum_{\beta \in \pi_2(\widehat{Y}, \gamma \sqcup \gamma_1 \sqcup \cdots \sqcup \gamma_l)} \#\overline{\mathcal{M}}(\beta)\, U^{\widehat{V} * \beta + \Gamma^-(\beta, V)} \gamma_1 \dots \gamma_l, \tag{1-3}$$

[1] Strictly speaking, (1-3) should be refined as follows: (i) one should indicate that the virtual moduli counts depend on a choice of "virtual perturbation data"; (ii) one should indicate that the counts depend on the order of a certain group of automorphisms of the triple $(\gamma, \gamma_1 \sqcup \cdots \sqcup \gamma_l, \beta)$ which acts on γ by the identity; (iii) one must specify signs (or, more invariantly and following Pardon [61], orientations lines). See (6-7) for the precise formula.

where γ is a positive orbit (ie associated to the convex end of the symplectization) and $\gamma_1, \dots, \gamma_l$ are negative orbits (ie associated to the concave end). We define $\Gamma^-(\beta, V) = \#\{\gamma_i \subset V\}$ to be the number of negative orbits of β which are contained in V.

We denote by $\overline{\mathcal{M}}(\beta)$ the compactification of the moduli space $\mathcal{M}(\beta)$ of $\widehat{J}$–holomorphic curves in the class β. Such moduli spaces are in general nontransversally cut out, so the moduli counts $\#\overline{\mathcal{M}}(\beta)$ appearing in (1-3) are defined as in Pardon's construction [60; 61] of contact homology, via his theory of virtual fundamental cycles. In particular, $\#\overline{\mathcal{M}}(\beta)$ is nonzero only for moduli spaces of virtual dimension zero. Finally, the pairing $(- * -)$ is a count of intersections between $\widehat{V}$ and β, which was introduced by Siefring [64; 54].

We also construct a closely related invariant

$$\widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r}), \tag{1-4}$$

which we call *reduced*. This is a unital, $\mathbb{Z}/2$–graded $\mathbb{Q}$–algebra which is generated by Reeb orbits in the complement of V. The differential counts pseudoholomorphic curves which do not intersect $\widehat{V}$. For appropriately chosen pairs $\mathfrak{r}, \mathfrak{r}' \in \mathfrak{R}(Y, \xi, V)$, we have a morphism of $\mathbb{Q}$–algebras

$$CH_\bullet(Y, \xi, V; \mathfrak{r}') \to \widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r}). \tag{1-5}$$

The invariant $\widetilde{CH}_\bullet(-; -)$ is called *reduced* because it carries less information (in particular, it does not involve taking the kernel of an augmentation as in, for instance, reduced singular homology). However, it is easier to compute.

Remark 1.1 The $\mathbb{Q}[U]$–algebra $CH_\bullet(Y, \xi, V; \mathfrak{r})$ can be viewed as a deformation of the contact homology $\mathbb{Q}$–algebra $CH_\bullet(Y, \xi)$ in the following way. First, recall that for a ring R and a differential graded R–algebra (A, d), a (*formal*) *deformation* of (A, d) is the data of a differential $d_t := d + t d_1 + t^2 d_2 + \cdots$ on the $R[\![t]\!]$–algebra $A[\![t]\!]$ satisfying the graded Leibniz rule, where each d_i is an endomorphism of A; see Gerstenhaber and Wilkerson [30]. Now, let us set $U = e^t$ in (1-3) and expand in t. We then get

$$d(\gamma) = \sum_\beta \sum_{k=0}^{\infty} t^k \cdot (\#\overline{\mathcal{M}}(\beta)) \cdot \frac{(\widehat{V} * \beta + \Gamma^-(\beta, V))^k}{k!} \gamma_1 \cdots \gamma_l. \tag{1-6}$$

Thus $CH_\bullet(Y, \xi, V; \mathfrak{r})$ is indeed a deformation of ordinary contact homology, which can be recovered by sending $t \to 0$. In the case where $\Gamma^-(\beta, V) = 0$, the coefficient

$$\#\overline{\mathcal{M}}(\beta) \cdot \frac{(\widehat{V} * \beta + \Gamma^-(\beta, V))^k}{k!} = \#\overline{\mathcal{M}}(\beta) \cdot \frac{(\widehat{V} * \beta)^k}{k!}$$

could be interpreted as a count of rigid pseudoholomorphic curves which send k marked points in the source to the pseudoholomorphic divisor $\widehat{V}$.

1.2 Energy and positivity of intersection

In order to ensure that (1-3) defines a differential over $\mathbb{Q}[U]$, we need to ensure that $\widehat{V} * \beta + \Gamma^-(\beta, V) \geq 0$ whenever $\#\overline{\mathcal{M}}(\beta) \neq 0$. If $\mathcal{M}(\beta)$ is nonempty[2] and at least one of the asymptotic orbits of β is disjoint from V, then this is a consequence of the familiar phenomenon of positivity of intersection. Indeed, in this case, β admits a $\widehat{J}$–holomorphic representative u which is not contained in $\widehat{V}$. Positivity of intersection then implies that $\widehat{V} * \beta = \widehat{V} * u \geq 0$.

The situation is more complicated when all of the asymptotic orbits of β are contained in V. Indeed, in this case, the $\widehat{J}$–holomorphic representatives of β may be contained in $\widehat{V}$ and positivity of intersection fails in general. However, one can show that there is a universal lower bound on the intersection number,

$$\widehat{V} * \beta \geq -\Gamma^-(\beta, V). \tag{1-7}$$

This explains the appearance of the correction term $\Gamma^-(\beta, V)$ in (1-3).

In order to construct $CH_\bullet(Y, \xi, V; \mathfrak{r})$, it is not enough to define a differential: one also needs to define continuation maps, composition homotopies, etc. These maps are defined by counting curves in more complicated setups. For example, the continuation map is obtained by counting curves in a suitably marked exact relative symplectic cobordism $(\widehat{X}, \widehat{\lambda}, H)$.[3] More precisely, one obtains an algebra map similar to (1-3) by counting $\widehat{J}$–holomorphic curves in $(\widehat{X}, \widehat{\lambda})$ weighted by their intersection number with H, for a compatible almost complex structure $\widehat{J}$ which agrees with $\widehat{J}^\pm$ near the ends.

Unfortunately, for an arbitrary relative symplectic cobordism, a lower bound of the type (1-7) fails to hold. A key step in constructing the invariants (1-1) is to identify a sufficiently large class of relative symplectic cobordisms for which such a lower bound does hold. This leads us to introduce notions of energy for exact symplectic cobordisms and almost complex structures on exact relative symplectic cobordisms. These energy notions are developed in Section 6 and are of central importance in this paper.

We prove that a lower bound as in (1-7) holds under a certain condition which relates the behavior of $\lambda^\pm$ near $V^\pm$ to the energy of $\widehat{J}$. We also prove analogous statements for other related setups. This allows us to prove that $CH_\bullet(Y, \xi, V; \mathfrak{r})$ is well-defined, ie it does not depend on the auxiliary contact form and almost complex structure. We also prove that an exact relative symplectic cobordism $(\widehat{X}, \widehat{\lambda}, H)$ induces a map

$$CH_\bullet(Y^+, \xi^+, V^+; \mathfrak{r}^+) \to CH_\bullet(Y^-, \xi^-, V^-; \mathfrak{r}^-) \tag{1-8}$$

provided that a certain inequality is satisfied, where the inequality involves $\mathfrak{r}^\pm$ and the energy of the (sub)cobordism $H \subset (\widehat{X}, \widehat{\lambda})$.

[2] Since [61] uses virtual techniques to define contact homology without making any transversality assumptions, it is possible for the compactification $\overline{\mathcal{M}}(\beta)$ to be nonempty even if $\mathcal{M}(\beta)$ is empty. Positivity of intersection still holds when this happens, but the proof requires a bit more work. Details can be found in Sections 5.3 and 7.2.

[3] An exact relative symplectic cobordism is the data of an exact symplectic cobordism $(\widehat{X}, \widehat{\lambda})$ which looks like $(\widehat{Y}^\pm, \widehat{\lambda}^\pm)$ near the ends, together with a codimension 2 symplectic submanifold $H \subset \widehat{X}$ which looks like $\widehat{V}^\pm$ near the ends; see Definition 2.17 for the details.

Energy considerations play a similarly central role in our construction of the reduced invariant $\widetilde{CH}_\bullet(-;-)$. Although the continuation map for the reduced invariant does not count curves contained in H, one needs to ensure that sequences of curves disjoint from H does not degenerate into H. This requires hypotheses on the energy of the relevant cobordism. In general, the arguments involved in constructing $CH_\bullet(-;-)$ and $\widetilde{CH}_\bullet(-;-)$ turn out to be very similar.

Energy is not in general well-behaved under gluing symplectic cobordisms, unless one of them happens to be a symplectization. As a result, cobordism maps cannot be composed arbitrarily. This lack of functoriality of the invariants (1-1) and (1-4) can be remedied by considering variants of these invariants which are obtained by taking certain (co)limits over $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$; see Section 8.3. These variants are fully functorial but also seem harder to compute.

Remark 1.2 The apparent failure of positivity of intersection in the absence of energy bounds is not a deficiency of our method: if one could define maps (1-8) without additional hypotheses involving energy and the $\mathfrak{r}^{\pm}$, then this would imply that (1-1) and (1-4) are independent of $\mathfrak{r}$. To see that this cannot hold in general, consider an "irrational ellipsoid" $E(r_1, r_2) = \{z \in \mathbb{C}^2 \mid \pi|z_1|^2/r_1 + \pi|z_2|^2/r_2 \leq 1\} \subset \mathbb{R}^4$, where r_1/r_2 is irrational. Following Hutchings [41, Section 4.2], there are exactly two families of Reeb orbits $\{\gamma_1^k\}_{k \in \mathbb{N}_+}$ and $\{\gamma_2^k\}_{k \in \mathbb{N}_+}$, where $\gamma_i^1 \subset \{z_i = 0\} \cap \partial E(r_1, r_2)$ and γ_i^k denotes the k–fold cover of γ_i^1. These orbits generate the (ordinary) contact homology of the 3–sphere, but the Conley–Zehnder indices of the γ_i^k are highly sensitive to r_1 and r_2. If we could define the contact homology of the complement of (say) γ_1^1, then it would be generated by the γ_2^k. One can verify that this is not an invariant, since changing the r_i does not change the contactomorphism type of $S^3 - \gamma_1^1$, but drastically changes the indices of the γ_2^k.

1.3 Legendrian invariants and the surgery formula

Contact homology is one of many invariants which can be constructed using the framework of Symplectic Field Theory (SFT). SFT was first introduced by Eliashberg, Givental and Hofer [24] and provides (among other things) a conjectural mechanism for constructing invariants in symplectic and contact topology by counting punctured pseudoholomorphic curves in symplectic manifolds with cylindrical ends.

In some of the later sections of this paper, we discuss how the invariants (1-1) and (1-4) are related to other SFT-type invariants. For computational purposes, it is particularly useful to explore the behavior of the invariants (1-1) and (1-4) under Weinstein handle attachment, following the work of Bourgeois, Ekholm and Eliashberg [7].

To this end, we introduce analogs of (1-1) and (1-4) for Legendrian submanifolds. With (Y, ξ, V) as above, suppose that $\Lambda \subset (Y - V, \xi)$ is a Legendrian submanifold. We then define (under mild topological assumptions) invariants

$$\mathcal{L}(Y, \xi, V, \Lambda; \mathfrak{r}) \quad \text{and} \quad \widetilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r}). \tag{1-9}$$

The first invariant can be thought of as a deformation of the Chekanov–Eliashberg dg algebra of $\Lambda \subset (Y, \xi)$, while the second invariant is a reduced version.

We describe a conjectural surgery exact sequence which relates (linearized versions of) the invariants (1-1), (1-4) and (1-9) under Weinstein handle attachments. This surgery exact sequence is an analog of the conjectural surgery exact sequence for linearized contact homology of Bourgeois, Ekholm and Eliashberg [7, Theorem 5.2].

Remark 1.3 The main invariants of this paper, (1-1) and (1-4), are constructed fully rigorously using Pardon's virtual perturbation framework [61]. However, our discussion of the surgery formula (and of the Legendrian invariants therein) requires certain transversality assumptions which are essentially the same as in [7]. We fully expect that if [7] can be made rigorous using Pardon's techniques [61], then extending this to our context should pose no substantial additional difficulties.

For the reader's convenience, *all statements in this paper which depend on unproved assumptions are labeled by a star.* The proofs of starred statements depend only on a limited set of assumptions, which are clearly identified in the text.

One could also attempt to define the invariants in this paper using other perturbation frameworks such as the polyfolds of Hofer, Wysocki and Zehnder [36; 37; 38] or the techniques of Ekholm [20], but we do not pursue this here.

1.4 Applications to contact and Legendrian embeddings

Transverse knots are important objects of study in three-dimensional contact topology. The notion of a codimension 2 contact embedding generalizes transverse knots to contact manifolds of arbitrary dimension. However, until recently, it was not understood whether the high-dimensional theory of codimension 2 contact embeddings is interesting from the perspective of contact topology, or whether it reduces entirely to differential topology.

Definition 1.4 Given a pair of contact manifolds (V^{2m-1},ζ) and (Y^{2n-1},ξ), a *contact embedding* is a smooth embedding

$$i:(V,\zeta)\to(Y,\xi) \tag{1-10}$$

such that $i^*(\xi_{i(V)}\cap di(TV))=\zeta$. Such a map is also referred to as an isocontact embedding in the literature (see eg Casals and Etnyre [10], Eliashberg and Mishachev [25] and Pancholi and Pandit [59]), but we will not use this terminology. A contact submanifold $V\subset(Y,\xi)$ is a submanifold with the property that $\xi|_V\cap TV$ is a contact structure.

Observe that if $2n-1=3$ and $2m-1=1$ in the above definition, then we recover the familiar notion of a (parametrized) transverse knot. The following basic examples of codimension 2 contact embeddings will play an important role in this paper.

Example 1.5 (cf Definition 3.11 and [31]) Let $\pi:Y-B\to S^1$ be an open book decomposition which supports the contact structure ξ on Y. Then the binding $B\subset(Y,\xi)$ is a codimension 2 contact submanifold.

Example 1.6 (see Definition 10.1 and Definition 3.1 in [10]) Let (Y, ξ) be a contact manifold and let $\Lambda \hookrightarrow Y$ be a Legendrian embedding. Then the Weinstein neighborhood theorem furnishes an embedding

$$\tau(\Lambda)\colon (\partial(D^*\Lambda), \xi_{\mathrm{std}}) \hookrightarrow (Y, \xi), \tag{1-11}$$

which is canonical up to isotopy through codimension 2 contact embeddings. We refer to $\tau(\Lambda)$ as the *contact pushoff* of $\Lambda \hookrightarrow Y$. By abuse of notation, we will routinely identify $\tau(\Lambda)$ with its image.

As is customary in contact and symplectic topology, there is a notion of a *formal* contact embedding. This notion encodes certain necessary bundle-theoretic conditions which must be satisfied by any (genuine) contact embedding. It is then natural to seek to understand to what extent the space of genuine contact embeddings of (V, ζ) into (Y, ξ) differs from the space of formal contact embeddings.

In the case that V is a closed manifold of codimension at least 4 with respect to Y, or open and of codimension at least 2, then an h-principle due to Gromov (see Theorem 12.3.1 and Remark 12.3 of Eliashberg and Mishachev [25]) implies that the space of contact embeddings is essentially equivalent to the space of formal contact embeddings. Thus, in these settings, the theory of contact embeddings reduces to differential topology.

In contrast, a breakthrough result due to Casals and Etnyre [10] shows that this h-principle fails in general for codimension 2 contact embeddings of closed manifolds. More precisely, for $n \geq 3$, Casals and Etnyre [10, Theorem 1] exhibited a pair of contact embeddings of $(D^*S^{n-1}, \xi) = \partial_\infty(T^*S^{n-1}, \lambda_{\mathrm{can}})$ into the standard contact sphere $(S^{2n-1}, \xi_{\mathrm{std}})$ which are formally isotopic but are not isotopic through contact embeddings (here and throughout the paper, $\partial_\infty(-)$ denotes the ideal contact boundary). Building on these methods, Zhou [70] recently proved that there are in fact infinitely many formally isotopic contact embeddings of $\partial_\infty(T^*S^{n-1}, \lambda_{\mathrm{can}})$ into the standard contact sphere which are not isotopic through contact embeddings, provided $n \geq 4$.

There has also been recent work to establish existence results for codimension 2 contact embeddings under certain conditions; see Lazarev [48] and Pancholi and Pandit [59]. This culminates in a full existence h-principle for codimension 2 contact embeddings due to Casals, Pancholi and Presas [13], which states that any formal codimension 2 contact embedding is formally isotopic to a genuine contact embedding.

The invariants constructed in this paper can be used to distinguish pairs of formally isotopic contact embeddings which are not isotopic through contact embeddings. We illustrate two types of applications, applying, respectively, to contact embeddings into overtwisted contact manifolds, and into the standard contact sphere.

Let us begin with the overtwisted case. In Construction 12.6, we describe a procedure for constructing pairs of formally isotopic contact embeddings into certain overtwisted contact manifolds which are not isotopic through contact embeddings. Here is a special case of this construction: let (Y, ξ) be an overtwisted contact manifold and fix an open book decomposition for (Y, B, π) which supports ξ; see Section 3.3.

Let $i\colon B \to Y$ be the tautological inclusion of the binding. Using the relative h–principle for contact structures of Borman, Eliashberg and Murphy [5, Theorem 1.2], it can be shown that there exists a codimension 2 contact embedding $j\colon B \to Y$ which is formally isotopic to i, and such that $(Y - j(B), \xi)$ is overtwisted. (Construction 12.6 is more general, but this is the most important example.)

Theorem 1.7 (see Theorem 12.7) *Let i, j and (Y, ξ) be constructed according to Construction 12.6, where (Y, ξ) is an overtwisted contact manifold and i and j are (formally isotopic) contact embeddings. Then i and j are not isotopic through contact embeddings. In fact, i is not isotopic to any reparametrization of j in the source, meaning that $i(V)$ and $j(V)$ are not isotopic as codimension 2 contact submanifolds of (Y, ξ).*

Theorem 1.7 can be proved using either of the invariants (1-1) or (1-4). To the best of our knowledge, it cannot be proved in general using invariants already in the literature. However, in the special case where $i(B)$ is the binding of an open book decomposition (ie the example sketched above), then the conclusion of Theorem 1.7 follows from the fact that the complement of the binding of an open book decomposition is tight. This later fact is due to Etnyre and Vela-Vick [28, Theorem 1.2] in dimension 3; in higher dimensions, it follows from work of Klukas [45, Corollary 3], who proved (following an outline of Wendl [67, Remark 4.1]) the stronger statement that any local filling obstruction (such as an overtwisted disk) in a closed contact manifold must intersect the binding of any supporting open book.

In some cases (see Corollary 12.10), the embeddings i and j in fact coincide with the contact pushoffs of Legendrian embeddings. It is not hard to show that an isotopy of Legendrian embeddings induces an isotopy of their contact pushoffs. Thus the invariants (1-1) and (1-4) also distinguish certain Legendrian embeddings in overtwisted contact manifolds. To our knowledge, these embeddings cannot in general be distinguished using invariants already in the literature; see Remark 1.11.

Our second application concerns codimension 2 contact embeddings into the standard contact spheres $(S^{4n-1}, \xi_{\mathrm{std}})$. More precisely, we use the reduced invariant (1-4) to distinguish formally isotopic contact embeddings of $(S^*S^{2n-1}, \xi) = \partial_\infty(T^*S^{2n-1}, \lambda_{\mathrm{can}})$ into $(S^{4n-1}, \xi_{\mathrm{std}})$, thus reproving the main result of Casals and Etnyre [10, Theorem 1] in dimensions $4n-1$ for $n > 1$.

Theorem* 1.8 (see Theorem* 12.18) *Let (V, ξ) be the ideal boundary of $(T^*S^{2n-1}, \lambda_{\mathrm{can}})$. Then for $n > 1$, there exists a pair of formally isotopic contact embeddings*

$$i_0, i_1\colon (V, \xi) \to (S^{4n-1}, \xi_{\mathrm{std}}) \tag{1-12}$$

which are not isotopic through contact embeddings.

The embeddings we exhibit turn out to coincide with those exhibited by Casals and Etnyre in their proof of [10, Theorem 1.1], although this fact is not entirely obvious; see Remark 12.21. However, the methods for distinguishing them are completely different. Casals and Etnyre consider double branched covers along the contact submanifolds $i_0(V)$ and $i_1(V)$. Using symplectic homology, they prove that these branched covers do not admit the same fillings. This implies that $i_0(V)$ and $i_1(V)$ cannot be isotopic, since otherwise they would have contactomorphic branched covers.

In contrast, our proof of Theorem* 1.8 uses the invariant $\widetilde{CH}_\bullet(-;-)$ introduced in this paper. Roughly speaking, we prove Theorem* 1.8 by partially computing (linearizations of) $\widetilde{CH}_\bullet(-;-)$ associated to the two embeddings under consideration, and observing that they do not match. Our computations rely crucially on our version of the surgery formula discussed in Section 1.3 as well as the well-definedness of the invariants therein. This explains why this theorem statement is starred, following the convention stated in Remark 1.3. We also remark that although Theorem* 1.8 only applies to spheres of dimension $4n-1$, we expect that the same invariant also distinguishes embeddings into spheres of dimension $4n-3$. However, proving this would likely require more involved computations than those carried out in this paper.

1.5 Applications to symplectic and Lagrangian cobordisms

Consider a pair of contact manifolds $(Y^\pm, \xi^\pm)$ and codimension 2 contact submanifolds $(V^\pm, \xi^\pm|_{V^\pm}) \subset (Y^\pm, \xi^\pm)$. An exact relative symplectic cobordism from (Y^+, ξ^+, V^+) to (Y^-, ξ^-, V^-) is a triple $(\widehat{X}, \widehat{\lambda}, H)$, where $(\widehat{X}, \widehat{\lambda})$ is an exact symplectic cobordism from (Y^+, ξ^+) to (Y^-, ξ^-) and $H \subset \widehat{X}$ is a codimension 2 symplectic submanifold which coincides near the ends with the symplectization of $V^\pm$; see Definition 2.17.[4] In the special case where $\widehat{X}$ is the symplectization of $Y^\pm$ and H is diffeomorphic to $\mathbb{R} \times V^\pm$, we speak of a *symplectic concordance* from V^+ to V^-. These notions were first considered by Bowden in his PhD thesis [9]. Using gauge theory, he exhibited certain restrictions on symplectic cobordisms between transverse links in contact 3–manifolds [9, Section 7].

The following theorem gives a constraint on exact symplectic cobordisms between certain pairs of codimension 2 contact submanifolds of an ambient overtwisted manifold. To the best of our knowledge, this is the first negative result in the literature on relative symplectic cobordisms in dimensions greater than three.

Theorem 1.9 *Let $V = i(B)$ and $V' = j(B)$ be the codimension 2 contact submanifolds of the overtwisted contact manifold (Y, ξ) as described in Construction 12.6. Then there does not exist an exact relative symplectic cobordism $(\widehat{X}, \widehat{\lambda}, H)$ from (Y, ξ, V') to (Y, ξ, V) with $H^1(H, (-\infty, 0] \times V; \mathbb{Z}) = H^2(H, (-\infty, 0] \times V; \mathbb{Z}) = 0$. In particular, there is no symplectic concordance from V' to V.*

One can similarly consider Lagrangian cobordisms and concordances between Legendrian submanifolds. An exact Lagrangian cobordism from (Y^+, ξ^+, Λ^+) to (Y^-, ξ^-, V^-) is a triple $(\widehat{X}, \widehat{\lambda}, L)$, where $(\widehat{X}, \widehat{\lambda})$ is an exact symplectic cobordism from (Y^+, ξ^+) to (Y^-, ξ^-) and $L \subset \widehat{X}$ is a Lagrangian submanifold which

[4] Note that our convention of regarding a cobordism as going *from* the convex end *to* the concave end is consistent with [61], but differs from most of the contact topology literature.

coincides near the ends with the Lagrangian cone of $\Lambda^{\pm}$; see Definition 2.20. If $\hat{X}$ is the symplectization of Y^- and $L = \mathbb{R} \times \Lambda^-$, one speaks of a *Lagrangian concordance* from Λ^+ to Λ^-.

The theory of Lagrangian cobordisms has been extensively developed in the literature from various perspectives; see eg Chantraine, Dimitroglou Rizell, Ghiggini and Golovko [15], Ekholm [19], Ekholm, Honda and Kálmán [23] and Sabloff and Traynor [63]. While much is known in $(\mathbb{R}^{2n+1}, \xi_{\mathrm{std}})$ and certain other tight contact manifolds, we are not aware of any results constraining cobordisms and concordances in overtwisted contact manifolds; see Remark 1.11. The next theorem provides a first result in this direction.

Theorem 1.10 *Let Λ and Λ' be the Legendrian submanifolds of the overtwisted contact manifold (Y, ξ) as constructed in Construction 12.9. Then Λ' is not concordant to Λ.*

In contrast, a result of Eliashberg and Murphy [26, Theorem 2.2] implies that Λ is concordant to Λ'.

Remark 1.11 It is a basic fact that exact Lagrangian cobordisms induce morphisms on Legendrian contact homology which behave well under composition of cobordisms; see Etnyre and Ng [27, Section 5.1]. This leads to a myriad of interesting obstructions to the existence of Lagrangian cobordisms and concordances. One can also obtain many interesting obstructions using finite-dimensional invariants (which are closely related to Legendrian contact homology) coming from generating functions or sheaf theory; see eg Li [49] and Sabloff and Traynor [63].

One drawback of these approaches is that they are necessarily blind on overtwisted contact manifolds. Indeed, even if Legendrian contact homology could be rigorously defined in full generality following the framework of Eliashberg, Givental and Hofer [24, Section 2.8], it would provide no information for Legendrians in overtwisted contact manifolds: being a module over the contact homology algebra, it would vanish. In contrast, the invariants developed in this paper do give information about Legendrians even in the overtwisted case.

Our final application states that certain Lagrangian concordances cannot be displaced from a codimension 2 symplectic submanifold. More precisely, let $(Y, \xi) = \mathrm{obd}(T^*S^{n-1}, \mathrm{id})$ and let $V \subset Y$ be the binding of the open book. Let $\Lambda \subset Y$ be the zero section of a page and let Λ' be obtained by stabilizing Λ in the complement of V. It can be shown (see Casals and Murphy [11, Proposition 2.9]) that $\Lambda \subset (Y, \xi)$ is a loose Legendrian; hence Λ and Λ' are Legendrian isotopic in (Y, ξ) and, in particular, concordant.

Theorem* 1.12 *Any Lagrangian concordance from Λ' to Λ must intersect the symplectization of V.*

In contrast, work of Eliashberg and Murphy [26, Theorem 2.2] implies that there exists a Lagrangian concordance from Λ to Λ' which is disjoint from the symplectization of V. Our proof of Theorem* 1.12 uses the deformed versions of the Chekanov–Eliashberg dg algebra in (1-9). Hence the statement is starred according to the convention stated in Section 1.3.

1.6 Context and related invariants

The invariants constructed in this paper, when specialized to contact 3–manifolds, are related to other invariants in the literature. The most closely related invariant is due to Momin [53]. Given a contact 3–manifold (Y^3, ξ), Momin considers the set of pairs (λ, L) where λ is a contact form and $L \subset Y$ is a link of Reeb orbits of λ. Two such pairs (λ, L) and (λ', L') are said to be equivalent if $L = L'$ and each component orbit (and all its multiple covers) has the same Conley–Zehnder index. Under certain assumptions on (Y, λ, L), Momin defines an invariant which we denote by $CH^{\mathrm{mo}}_{\bullet}(Y, [(\lambda, L)])$. This is a $\mathbb{Z}$–graded $\mathbb{Q}$–vector space which depends only on Y and the equivalence class of (λ, L).

The invariant constructed by Momin is in general distinct from the invariants described in this paper. In particular, he considers cylindrical contact homology, whereas we work with ordinary contact homology. However, in the special case where (Y^3, ξ) is the standard contact sphere — or more generally a subcritical Stein manifold with $c_1(\xi) = 0$ — and $L \subset (Y, \xi)$ is a collection of Reeb orbits which bound a symplectic submanifold $H \subset B^4$, then we expect that

$$CH^{\mathrm{mo}}_{\bullet}(Y, [\lambda, L]) = \widetilde{CH}^{\tilde{\epsilon}}_{\bullet}(Y, \xi, L; \mathfrak{r}) \tag{1-13}$$

for suitable $\mathfrak{r}$ which depends on the equivalence class of (λ, L). Here the right-hand side of (1-13) denotes the linearization of $\widetilde{CH}_{\bullet}(Y, \xi, L; \mathfrak{r})$ with respect to an augmentation $\tilde{\epsilon}$ induced by the relative filling $(B^4, \lambda_{\mathrm{std}}, H)$; recall that an augmentation of a dg algebra is a morphism to the ground ring, viewed as a dg algebra concentrated in degree zero. See Section 9.3 for details.

Momin's work has led to beautiful applications to Reeb dynamics on contact 3–manifolds; see eg Alves and Pirnapasov [3] and Hryniewicz, Momin and Salomão [40]. It would be interesting to explore whether the invariants developed in this paper can be used in studying Reeb dynamics in higher dimensions.

Another related invariant is Hutchings' "knot-filtered" embedded contact homology [41]. The setting for this invariant is a contact 3–manifold (Y, ξ) with $H_1(Y; \mathbb{Z}) = 0$. Given a transverse knot $L \subset (Y, \xi)$ and an irrational parameter $\theta \in \mathbb{R} - \mathbb{Q}$, Hutchings defines a filtration on embedded contact homology with values in $\mathbb{Z} + \mathbb{Z}\theta$ which is an invariant of (L, θ). The basic idea is to choose a contact form $\xi = \ker \lambda$ so that L is a Reeb orbit, and to filter the generators of embedded contact homology by their linking number with L. Positivity of intersection considerations imply that the differential decreases the linking number for orbits which are disjoint from L. However, the situation is more complex when the differential involves L, which explains why the filtration is only valued in $\mathbb{Z} + \mathbb{Z}\theta$.

One could presumably carry over Hutchings' construction to the context of (cylindrical) contact homology in dimension 3. We expect that the resulting invariant would carry related information to the one defined by Momin or to the invariants constructed in this paper. However, we do not have a precise formulation of what this relationship should be.

We remark that the invariants introduced by Momin and Hutchings are built using techniques from 4–dimensional symplectic topology which cannot be generalized to higher dimensions. In contrast, the invariants introduced in this paper are constructed by a different approach, which ultimately relies on Pardon's robust virtual fundamental cycles package [60].

In a slightly different direction, we also wish to highlight work of Ekholm, Etnyre, Ng and Sullivan [21], which is similar in spirit to the present work. Recall that the knot contact homology of a framed link $K \subset \mathbb{R}^3$ is an invariant which can be defined as the Legendrian dg algebra of the conormal lift of K; see Ekholm, Etnyre, Ng and Sullivan [22] and Ng [56]. If K is a transverse knot with respect to the standard contact structure, the authors define in [21] a two-parameter deformation of knot contact homology by weighting holomorphic curve counts by their intersection number with a pair of canonically defined complex submanifolds in the symplectization. The resulting deformed dg algebra is an invariant of the transverse knot type of K. Unfortunately, we do not know a precise relationship between this invariant and the ones introduced in this paper.

Notation and conventions All manifolds in this paper are assumed to be smooth. If M is a manifold, a *ball* $B \subset M$ is an open subset diffeomorphic to the open unit disk and whose closure is embedded and diffeomorphic to the closed unit disk. If (M, ω) is symplectic, a *Darboux ball* $B \subset M$ is a ball which is symplectomorphic to the open unit disk equipped with (some rescaling of) the standard symplectic form.

Unless otherwise specified, all contact manifolds considered in this paper are compact without boundary and co-oriented. Given such a contact manifold $(Y, \xi = \ker \lambda)$, the Reeb vector field associated to the contact form λ will be denoted by R_λ. We will let $\xi_V := \xi|_V \cap TV$ denote the contact structure induced by ξ on a contact submanifold $V \subset (Y, \xi)$.

Acknowledgements We thank Yasha Eliashberg for suggesting this project, and for many helpful discussions. We also benefited from discussions and correspondence with Cédric De Groote, Georgios Dimitroglou Rizell, Sheel Ganatra, Oleg Lazarev, Josh Sabloff and Kyler Siegel. We are grateful to John Etnyre for pointing us to multiple useful references. Finally, we wish to thank the referee for many helpful comments and suggestions.

Côté was partially supported by a Stanford University Benchmark Graduate Fellowship and by the National Science Foundation under grant DMS-1926686.

2 Geometric preliminaries

2.1 Symplectic cobordisms

Let (Y, ξ) be a closed co-oriented contact manifold. The *symplectization* of (Y, ξ) is the exact symplectic manifold (SY, λ_Y) where $SY \subset T^*Y$ is the total space of the bundle of positive contact forms on Y (ie a point $(p, \alpha) \in T^*Y$ is in SY if and only if $\alpha : T_pY \to \mathbb{R}$ vanishes on ξ_p and the induced map $T_pY/\xi_p \to \mathbb{R}$

is an orientation-preserving isomorphism) and λ_Y is the restriction of the tautological Liouville form on T^*Y. Given a choice of positive contact form α for (Y,ξ), there is a canonical identification

$$\sigma_\alpha\colon (\mathbb{R}\times Y, e^s\alpha)\to (SY,\lambda_Y) \tag{2-1}$$

given by $\sigma_\alpha(s,p)=(p,e^s\alpha_p)$. We will refer to $(\widehat{Y},\widehat{\alpha}):=(\mathbb{R}\times Y,e^s\alpha)$ as the symplectization of (Y,α).

A subset $U\subset SY$ will be called a *neighborhood of* $+\infty$ (resp. of $-\infty$) if it contains $\sigma_\alpha([N,\infty)\times Y)$ (resp. $\sigma_\alpha((-\infty,-N]\times\mathbb{R})$ for $N>0$ sufficiently large — note that this notion doesn't depend on the choice of α.

Definition 2.1 Given a contactomorphism $f\colon (Y,\xi)\to(Y,\xi)$, we define its *symplectic lift*

$$\widetilde{f}\colon (SY,\lambda_Y)\to(SY,\lambda_Y),\quad (p,\alpha)\mapsto (f(p),\alpha\circ(df_p)^{-1}). \tag{2-2}$$

One can verify that $\widetilde{f}^*\lambda_Y=\lambda_Y$, so $\widetilde{f}$ is in particular a symplectomorphism. There is a canonical bijection between

(i) contact vector fields on (Y,ξ),

(ii) sections of TY/ξ,

(iii) linear Hamiltonians on the symplectization (recall that H is *linear* if $ZH=H$, where Z denotes the Liouville vector field).

The correspondence between (i) and (ii) is clear; the correspondence between (ii) and (iii) takes a section σ to the Hamiltonian $H(p,\alpha)=\alpha(\sigma(p))$. In particular, the symplectic lift of a (time-dependent) family of contactomorphisms is induced by a (time-dependent) family of linear Hamiltonians; cf [14, Proposition 2.2].

Definition 2.2 Let (Y^+,ξ^+) and (Y^-,ξ^-) be closed co-oriented contact manifolds. An *exact symplectic cobordism* from (Y^+,ξ^+) to (Y^-,ξ^-) is an exact symplectic manifold $(\widehat{X},\widehat{\lambda})$ equipped with embeddings

$$e^+\colon SY^+\to\widehat{X}, \tag{2-3}$$

$$e^-\colon SY^-\to\widehat{X}, \tag{2-4}$$

satisfying the properties

- $(e^\pm)^*\widehat{\lambda}=\lambda_{Y^\pm}$, and
- there exists a neighborhood $U^+\subset SY^+$ of $+\infty$ and a neighborhood $U^-\subset SY^-$ of $-\infty$ such that the restriction of $e^\pm$ to $U^\pm$ is proper, the images $e^+(U^+)$ and $e^-(U^-)$ are disjoint and the complement $\widehat{X}\setminus(e^+(U^+)\cup e^-(U^-))$ is compact.

Definition 2.3 (cf [61, Section 1.3]) Let (Y^+, λ^+) and (Y^-, λ^-) be closed manifolds equipped with contact forms. A (*strict*) *exact symplectic cobordism* from (Y^+, λ^+) to (Y^-, λ^-) is an exact symplectic manifold $(\widehat{X}, \widehat{\lambda})$ equipped with embeddings

$$e^+ : \mathbb{R} \times Y^+ \to \widehat{X}, \tag{2-5}$$

$$e^- : \mathbb{R} \times Y^- \to \widehat{X}, \tag{2-6}$$

satisfying the properties

- $(e^\pm)^* \widehat{\lambda} = \widehat{\lambda}^\pm$, and
- there exists an $N \in \mathbb{R}$ such that the restrictions of e^+ to $[N, \infty) \times Y^+$ and of e^- to $(-\infty, -N] \times Y^-$ are proper and that the images $e^+([N, \infty) \times Y^+)$ and $e^-((-\infty, -N] \times Y^-)$ are disjoint and together cover a neighborhood of infinity (ie the complement of their union is compact).

Notation 2.4 Let $(\widehat{X}, \widehat{\lambda})$ be an exact symplectic cobordism from (Y^+, ξ^+) to (Y^-, ξ^-) in the sense of Definition 2.2. Given any choice of contact forms $\lambda^\pm$ on $(Y^\pm, \xi^\pm)$, one can obtain from $\widehat{X}$ a cobordism from (Y^+, λ^+) to (Y^-, λ^-) in the sense of Definition 2.3 by precomposing the embeddings (2-3)–(2-4) with the canonical identifications $\mathbb{R} \times Y^\pm \to SY^\pm$ induced by $\lambda^\pm$. We will denote this cobordism by $(\widehat{X}, \widehat{\lambda})^{\lambda^+}_{\lambda^-}$, or simply by $\widehat{X}^{\lambda^+}_{\lambda^-}$ when this creates no ambiguity.

Similarly, any cobordism $(\widehat{X}, \widehat{\lambda})$ in the sense of Definition 2.3 can be viewed as a cobordism in the sense of Definition 2.2 as well.

Remark 2.5 In light of the above discussion, Definitions 2.2 and 2.3 are essentially equivalent. However, it will be convenient for us to be able to discuss symplectic cobordisms without fixing a particular choice of contact forms on the ends, so we adopt Definition 2.2 as our main definition moving forward.

Example 2.6 (symplectizations) The symplectization (SY, λ_Y) of a contact manifold (Y, ξ) is canonically endowed with the structure of an exact symplectic cobordism in the sense of Definition 2.2 by letting $e^+ = e^- = \mathrm{id}$. The additional data of a pair of contact forms λ^+ and λ^- for (Y, ξ) endows (SY, λ_Y) with the structure of a strict exact symplectic cobordism in the sense of Definition 2.3, and we write $(SY, \lambda_Y)^{\lambda^+}_{\lambda^-}$.

Definition 2.7 Let $(\widehat{X}^{01}, \widehat{\lambda}^{01})$ and $(\widehat{X}^{12}, \widehat{\lambda}^{12})$ be exact symplectic cobordisms from (Y^0, ξ^0) to (Y^1, ξ^1) and from (Y^1, ξ^1) to (Y^2, ξ^2), respectively. Fix a real number $t \geq 0$ and let $\mu_t : SY^1 \to SY^1$ denote multiplication by e^t. The *t–gluing* of $\widehat{X}^{01}$ and $\widehat{X}^{12}$, denoted by $\widehat{X}^{01} \#_t \widehat{X}^{12}$, is the smooth manifold obtained by gluing $\widehat{X}^{01}$ and $\widehat{X}^{12}$ along the maps

$$\begin{array}{ccc} SY^1 & \xrightarrow{(2\text{-}4)} & \widehat{X}^{01} \\ \big\downarrow{\scriptstyle \mu_t} & & \\ SY^1 & \xrightarrow{(2\text{-}3)} & \widehat{X}^{12} \end{array}$$

Since $\mu_t^* \lambda_{Y^1} = e^t \lambda_{Y^1}$, there is, for any $s \in \mathbb{R}$, a Liouville form on $\widehat{X}^{01} \#_t \widehat{X}^{12}$ which agrees with $e^{t+s}\widehat{\lambda}^{01}$ on $\widehat{X}^{01}$ and with $e^s \widehat{\lambda}^{12}$ on $\widehat{X}^{12}$. We will denote it by $\widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12}$. Note that $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12})$ is canonically equipped with the structure of an exact symplectic cobordism from (Y^0, ξ^0) to (Y^2, ξ^2) via the embeddings

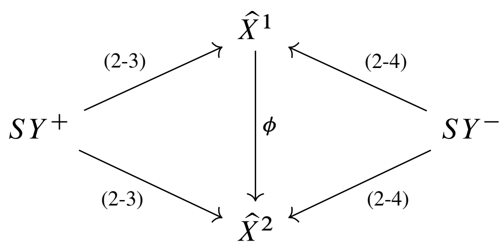
$$SY^0 \xrightarrow{\mu_{-t-s}} SY^0 \xrightarrow{\text{(2-3)}} \widehat{X}^{01} \to \widehat{X}^{01} \#_t \widehat{X}^{12},$$
$$SY^2 \xrightarrow{\mu_{-s}} SY^2 \xrightarrow{\text{(2-4)}} \widehat{X}^{02} \to \widehat{X}^{01} \#_t \widehat{X}^{12}.$$

The precise choice of s doesn't really matter since the forms $\widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12}$ for $s \in \mathbb{R}$ are all constant multiples of each other. When $t = 0$, it is natural to choose $s = 0$, and we will denote the resulting cobordism simply by $(\widehat{X}^{01} \# \widehat{X}^{12}, \widehat{\lambda}^{01} \# \widehat{\lambda}^{12})$. There is no obvious choice for $t > 0$, but for the sake of definiteness we set $\widehat{\lambda}^{01} \#_t \widehat{\lambda}^{12} := \widehat{\lambda}^{01} \#_{t,-t/2} \widehat{\lambda}^{12}$ and will refer to $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{12})$ as "the" t–gluing of $(\widehat{X}^{01}, \widehat{\lambda}^{01})$ and $(\widehat{X}^{12}, \widehat{\lambda}^{12})$.

Remark 2.8 When $t = s = 0$, it follows directly from the definition that the gluing operation is associative: $\big((\widehat{X}^{01} \# \widehat{X}^{12}) \# \widehat{X}^{23}, (\widehat{\lambda}^{01} \# \widehat{\lambda}^{12}) \# \widehat{\lambda}^{23}\big)$ and $\big(\widehat{X}^{01} \# (\widehat{X}^{12} \# \widehat{X}^{23}), \widehat{\lambda}^{01} \# (\widehat{\lambda}^{12} \# \widehat{\lambda}^{23})\big)$ are canonically isomorphic.

Remark 2.9 Multiplication by e^t on SY corresponds to translation by t in the $\mathbb{R}$ coordinate under the identification $SY \cong \mathbb{R} \times Y$ induced by a choice of contact form on Y. Definition 2.7 is therefore consistent with the notion of "t–gluing" in [61, Section 1.5].

Definition 2.10 Let $(\widehat{X}^1, \widehat{\lambda}^1)$ and $(\widehat{X}^2, \widehat{\lambda}^2)$ be cobordisms from (Y^+, ξ^+) to (Y^-, ξ^-). An *isomorphism of exact symplectic cobordisms* $\phi\colon (\widehat{X}^1, \widehat{\lambda}^1) \to (\widehat{X}^2, \widehat{\lambda}^2)$ consists of a diffeomorphism $\phi\colon \widehat{X}^1 \to \widehat{X}^2$ such that $\phi^* \widehat{\lambda}^2 = \widehat{\lambda}^1$ and which is compatible with the ends in the sense that the diagram



commutes.

Example 2.11 Let $(\widehat{X}, \widehat{\lambda})$ be an exact symplectic cobordism from (Y^+, ξ^+) to (Y^-, ξ^-). Then for any $t \geq 0$ and $s \in \mathbb{R}$, the glued cobordisms $(SY^+ \#_t \widehat{X}, \lambda_{Y^+} \#_{t,s} \widehat{\lambda})$ and $(\widehat{X} \#_t SY^-, \widehat{\lambda} \#_{t,s} \lambda_{Y^-})$ are canonically isomorphic to $(\widehat{X}, \widehat{\lambda})$.

Definition 2.12 A *one-parameter family of exact symplectic cobordisms* from (Y^+, ξ^+) to (Y^-, ξ^-) is a manifold $\widehat{X}$ equipped with a family of Liouville forms $\{\widehat{\lambda}^t\}_{t \in I}$ (where $I \subset \mathbb{R}$ is an interval), together with embeddings

$$e_t^+ : SY^+ \to \widehat{X}, \tag{2-7}$$

$$e_t^- : SY^- \to \widehat{X}, \tag{2-8}$$

as in Definition 2.2. We will always assume that the family is fixed at infinity, meaning that for every compact subinterval $[a,b] \subset I$,

- $\{\widehat{\lambda}^t\}_{t\in[a,b]}$ is constant outside of a compact subset of $\widehat{X}$, and
- $\{e_t^+\}_{t\in[a,b]}$ (resp. $\{e_t^-\}_{t\in[a,b]}$) is independent of t on some neighborhood of $+\infty$ in SY^+ (resp. of $-\infty$ in SY^-).

Two cobordisms $(\widehat{X}^0, \widehat{\lambda}^0)$ and $(\widehat{X}^1, \widehat{\lambda}^1)$ are said to be *deformation equivalent* if there exists a one-parameter family $(\widehat{W}, \widehat{\mu}^t)_{t\in[0,1]}$ such that $(\widehat{X}^0, \widehat{\lambda}^0)$ is isomorphic to $(\widehat{W}, \widehat{\mu}^0)$ and $(\widehat{X}^1, \widehat{\lambda}^1)$ is isomorphic to $(\widehat{W}, \widehat{\mu}^1)$. The deformation class of a cobordism $(\widehat{X}, \widehat{\lambda})$ will be denoted by $[\widehat{X}, \widehat{\lambda}]$.

Lemma 2.13 *Given a pair of cobordisms $(\widehat{X}^{01}, \widehat{\lambda}^{01})$ and $(\widehat{X}^{12}, \widehat{\lambda}^{12})$ as in Definition 2.7, the glued cobordisms $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{12})_{t\in[0,\infty)}$ form a one-parameter family.*[5] *Similarly, there is a one-parameter family $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12})_{s\in\mathbb{R}}$ for any fixed $t \geq 0$. In particular, we have that the deformation class $[\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12}]$ is independent of both t and s.*

Proof We will construct a two-parameter family $\phi_{t,s} : \widehat{X}^{01}\#\widehat{X}^{12} \to \widehat{X}^{01} \#_t \widehat{X}^{12}$ of diffeomorphisms, with $\phi_{0,0} = \mathrm{id}$, such that the forms $\phi_{t,s}^*(\widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12})$ agree with $\widehat{\lambda}^{01}\#\widehat{\lambda}^{12}$ outside of a compact set (depending on t, s) and form a smooth family.

In order to simplify the notation, we fix contact forms λ^i on (Y^i, ξ^i) for $i = 0, 1, 2$, so that we can view the symplectization of Y^i as a product $\mathbb{R} \times Y^i$. For $C > 0$ sufficiently large, we can decompose the cobordisms $\widehat{X}^{01}$ and $\widehat{X}^{12}$ as

$$\widehat{X}^{01} = (-\infty, 1] \times Y^1 \cup \overline{X}^{01} \cup [C, \infty) \times Y^0, \tag{2-9}$$

$$\widehat{X}^{12} = (-\infty, -C] \times Y^2 \cup \overline{X}^{12} \cup [-1, \infty) \times Y^1, \tag{2-10}$$

where $\overline{X}^{01} \subset \widehat{X}^{01}$ is a compact submanifold with boundary $\{1\} \times Y^1 \sqcup \{C\} \times Y^0$, and similarly for $\overline{X}^{12}$. This induces a decomposition of $\widehat{X}^{01} \#_t \widehat{X}^{12}$ of the form

$$\widehat{X}^{01} \#_t \widehat{X}^{12} = (-\infty, -C] \times Y^2 \cup \overline{X}^{12} \cup [-1, t+1] \times Y^1 \cup \overline{X}^{01} \cup [C, \infty) \times Y^0 \tag{2-11}$$

for any $t \geq 0$. Hence, in order to define $\phi_{t,s}$, it suffices to make a choice of

- a smooth family of diffeomorphisms $f_t : [-1, 1] \to [-1, t+1]$ which coincide with the identity near -1 and with translation by t near 1,

[5] Strictly speaking, the underlying manifold of $\widehat{X}^{01} \#_t \widehat{X}^{12}$ depends on t, so in order to obtain a family in the sense of Definition 2.12 one needs to choose suitable diffeomorphisms $\widehat{X}^{01} \#_t \widehat{X}^{12} \cong \widehat{X}^{01} \# \widehat{X}^{12}$.

- a smooth family of diffeomorphisms $g_{t,s}\colon [C,\infty) \to [C,\infty)$ which coincide with the identity near C and with translation by $-t-s$ at infinity, and
- a smooth family of diffeomorphisms $h_{t,s}\colon (-\infty,-C] \to (-\infty,-C]$ which coincide with the identity near $-C$ and with translation by $-s$ at infinity.

We of course also require that f_0, $g_{0,0}$ and $h_{0,0}$ be the identity on their respective domains. □

Proposition 2.14 *The deformation class of* $(\hat{X}^{01} \# \hat{X}^{12}, \hat{\lambda}^{01} \# \hat{\lambda}^{12})$ *only depends on the deformation classes of* $(\hat{X}^{01}, \hat{\lambda}^{01})$ *and* $(\hat{X}^{12}, \hat{\lambda}^{12})$.

Proof Let $(\hat{W}^{01}, \hat{\mu}^{01,s})_{s\in[0,1]}$ and $(\hat{W}^{12}, \hat{\mu}^{12,s})_{s\in[0,1]}$ be one-parameter families of exact symplectic cobordisms from (Y^0,ξ^0) to (Y^1,ξ^1) and from (Y^1,ξ^1) to (Y^2,ξ^2), respectively. The negative end (2-8) of $(\hat{W}^{01}, \hat{\mu}^{01,s})$ will be denoted by $e_s^-\colon SY^1 \to \hat{W}^{01}$ and the positive end (2-7) of $(\hat{W}^{12}, \hat{\mu}^{12,s})$ will be denoted by $e_s^+\colon SY^1 \to \hat{W}^{12}$. By definition, we can find a neighborhood $U^+ \subset SY^1$ of $+\infty$ and a neighborhood $U^- \subset SY^1$ such that the restriction of $e_s^\pm$ to $U^\pm$ is independent of s. This common restriction will be denoted by $e^\pm$.

Fix a large $t > 0$ so that the intersection $V := \mu_t^{-1}(U^+) \cap U^-$ is nonempty. Let $\hat{W}^{01} \#_V \hat{W}^{12}$ be the space obtained by gluing $\hat{W}^{01} \setminus e^-(U^- \setminus V)$ and $\hat{W}^{12} \setminus e^+(U^+ \setminus \mu^t(V))$ along the maps

$$\begin{array}{ccc} V & \xrightarrow{\ e^-\ } & \hat{W}^{01} \setminus e^-(U^- \setminus V) \\ \downarrow{\scriptstyle \mu_t} & & \\ U^+ & \xrightarrow{\ e^+\ } & \hat{W}^{12} \setminus e^+(U^+ \setminus \mu^t(V)) \end{array}$$

As a smooth manifold, $\hat{W}^{01} \#_V \hat{W}^{12}$ is canonically identified with $\hat{W}^{01} \#_t \hat{W}^{12}$. Thus we can view $\hat{\mu}^{01,s} \#_t \hat{\mu}^{12,s}$ as a Liouville form on $\hat{W}^{01} \#_V \hat{W}^{12}$ for each s, making $(\hat{W}^{01} \#_V \hat{W}^{12}, \hat{\mu}^{01,s} \#_t \hat{\mu}^{12,s})_{s\in[0,1]}$ into a one-parameter family of cobordisms. In particular, it follows that $(\hat{W}^{01} \#_t \hat{W}^{12}, \hat{\mu}^{01,0} \#_t \hat{\mu}^{12,0})$ and $(\hat{W}^{01} \#_t \hat{W}^{12}, \hat{\mu}^{01,1} \#_t \hat{\mu}^{12,1})$ are deformation equivalent. □

Corollary 2.15 *There is a well-defined gluing operation on deformation classes of exact symplectic cobordisms given by*

$$[\hat{X}^{01}, \hat{\lambda}^{01}] \# [\hat{X}^{12}, \hat{\lambda}^{12}] = [\hat{X}^{01} \#_t \hat{X}^{12}, \hat{\lambda}^{01} \#_{t,s} \hat{\lambda}^{12}] \tag{2-12}$$

for any $t \geq 0$ *and* $s \in \mathbb{R}$.

Proposition 2.16 *The gluing operation* (2-12) *is associative.*

Proof This follows from Remark 2.8. □

We now discuss submanifolds in exact symplectic cobordisms. Let $V \subset (Y,\xi)$ be a contact submanifold. There is a canonical embedding

$$(SV,\lambda_V) \to (SY,\lambda_Y) \tag{2-13}$$

which takes a pair $(p,\alpha_p) \in SV$ to the unique pair $(p,\tilde{\alpha}_p) \in SY$ such that $\tilde{\alpha}_p(w) = \alpha_p(w)$ for some (and hence any) $w \in T_pV \setminus (\xi_p \cap T_pV)$.

Definition 2.17 Let $V^+ \subset (Y^+,\xi^+)$ and $V^- \subset (Y^-,\xi^-)$ be contact submanifolds of the same codimension, and let $(\hat{X},\hat{\lambda})$ be an exact symplectic cobordism from (Y^+,ξ^+) to (Y^-,ξ^-). We say that a smooth submanifold $H \subset \hat{X}$ is *cylindrical* with ends $V^\pm$ if it is closed (as a subset) and there exist neighborhoods $U^\pm \subset SY^\pm$ of $\pm\infty$ such that

$$(e^\pm)^{-1}(H) \cap U^\pm = SV^\pm \cap U^\pm, \tag{2-14}$$

where $e^\pm\colon SY^\pm \to \hat{X}$ are the ends (2-3)–(2-4) of $(\hat{X},\hat{\lambda})$.

If H is a *symplectic* cylindrical submanifold of $(\hat{X},\hat{\lambda})$, then we say that $(\hat{X},\hat{\lambda},H)$ is an *exact relative symplectic cobordism* from (Y^+,ξ^+,V^+) to (Y^-,ξ^-,V^-). Note that in this case, the restrictions of $e^\pm$ to $SV^\pm \cap U^\pm$ endow $(H,\hat{\lambda}|_H)$ with the structure of an exact symplectic cobordism from $(V^+,\xi^+_{V^+})$ to $(V^-,\xi^-_{V^-})$.

Example 2.18 If V is a contact submanifold of (Y,ξ), then, as noted above, SV can be viewed as a symplectic submanifold of (SY,λ_Y), and (SY,λ_Y,SV) is canonically endowed with the structure of an exact relative symplectic cobordism in the sense of Definition 2.17 by letting $e^+ = e^- = \mathrm{id}$.

Notation 2.19 Let $\hat{X}$, $\hat{\lambda}$ and H be as in Definition 2.17. As explained in Notation 2.4, a choice of contact forms $\ker\lambda^\pm = \xi^\pm$ endows $(\hat{X},\hat{\lambda})$ with the structure of a *strict* relative symplectic cobordism. We analogously speak of a *strict* relative exact symplectic cobordism and write $(\hat{X},\hat{\lambda},H)^{\lambda^+}_{\lambda^-}$ when we wish to emphasize that we are fixing contact forms $\lambda^\pm$ on the ends.

Let $\Lambda \subset (Y,\xi)$ be a Legendrian submanifold. The *Lagrangian cone* of Λ is the Lagrangian submanifold

$$L = \{(p,\alpha) \in SY \subset T^*Y \mid p \in \Lambda\} \subset (SY,\lambda_Y). \tag{2-15}$$

Definition 2.20 Let $\Lambda^+ \subset (Y^+,\xi^+)$ and $\Lambda^- \subset (Y^-,\xi^-)$ be Legendrian submanifolds and let $(\hat{X},\hat{\lambda})$ be an exact symplectic cobordism from (Y^+,ξ^+) to (Y^-,ξ^-). We say that a Lagrangian submanifold $L \subset (\hat{X},\hat{\lambda})$ is *cylindrical* with ends $\Lambda^\pm$ if it is closed (as a subset) and there exist neighborhoods $U^\pm \subset SY^\pm$ of $\pm\infty$ such that

$$(e^\pm)^{-1}(L) \cap U^\pm = L^\pm \cap U^\pm, \tag{2-16}$$

where $e^\pm\colon SY^\pm \to \hat{X}$ are the ends (2-3)–(2-4) of $(\hat{X},\hat{\lambda})$ and $L^\pm$ are the Lagrangian cones of $\Lambda^\pm$.

A triple $(\hat{X},\hat{\lambda},L)$ is called an (*exact*) *Lagrangian cobordism* from (Y^+,ξ^+,Λ^+) to (Y^-,ξ^-,Λ^-).

Definition 2.21 The set of equivalence classes of cylindrical codimension 2 submanifolds of $\widehat{X}$ with ends $V^{\pm}$, where two submanifolds are equivalent if they are isotopic via a compactly supported (smooth) isotopy, will be denoted by $\Omega_{2n-2}(\widehat{X}, V^+ \sqcup V^-)$.

Definition 2.22 A contact submanifold $V \subset (Y, \lambda)$ is said to be a *strong contact submanifold* if it is (setwise) invariant under the Reeb flow of λ on Y. We will also say that $(\widehat{X}, \widehat{\lambda}, H)_{\lambda^-}^{\lambda^+}$ is a *strong* relative exact symplectic cobordism if both $V^+ \subset (Y^+, \lambda^+)$ and $V^- \subset (Y^-, \lambda^-)$ are strong contact submanifolds.

Definition 2.23 Let $(\widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01})$ and $(\widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12})$ be exact relative symplectic cobordisms from (Y^0, ξ^0, V^0) to (Y^1, ξ^1, V^1) and from (Y^1, ξ^1, V^1) to (Y^2, ξ^2, V^2), respectively. For any sufficiently large real number $t \geq 0$, $H^{01} \#_t H^{12}$ sits naturally inside $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{12})$ as a symplectic submanifold, and $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{12}, H^{01} \#_t H^{12})$ is a relative cobordism from (Y^0, ξ^0, V^0) to (Y^2, ξ^2, V^2). We will refer to it as the *t–gluing* of $(\widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01})$ and $(\widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12})$.

Definition 2.24 Let $(\widehat{X}^1, \widehat{\lambda}^1, H^1)$ and $(\widehat{X}^2, \widehat{\lambda}^2, H^2)$ be relative cobordisms from (Y^+, ξ^+, V^+) to (Y^-, ξ^-, V^-). An *isomorphism of exact relative symplectic cobordisms* $\phi\colon (\widehat{X}^1, \widehat{\lambda}^1, H^1) \to (\widehat{X}^2, \widehat{\lambda}^2, H^2)$ is an isomorphism $\phi\colon (\widehat{X}^1, \widehat{\lambda}^1) \to (\widehat{X}^2, \widehat{\lambda}^2)$ in the sense of Definition 2.10 which maps H^1 diffeomorphically onto H^2.

Example 2.25 Suppose that $(\widehat{X}, \widehat{\lambda}, H)$ is an exact relative symplectic cobordism from (Y^+, ξ^+, V^+) to (Y^-, ξ^-, V^-). Then for any $t \geq 0$, the glued cobordisms $(SY^+, \lambda_{Y^+}, SV^+) \#_t (\widehat{X}, \widehat{\lambda}, H)$ and $(\widehat{X}, \widehat{\lambda}, H) \#_t (SY^-, \lambda_{Y^-}, SV^-)$ are defined and canonically isomorphic to $(\widehat{X}, \widehat{\lambda}, H)$.

Definition 2.26 A *one-parameter family of exact relative symplectic cobordisms* from (Y^+, ξ^+, V^+) to (Y^-, ξ^-, V^-) is a manifold $\widehat{X}$ equipped with a family of Liouville forms $\{\widehat{\lambda}^t\}_{t \in I}$, a family of symplectic submanifolds $H^t \subset (\widehat{X}, \widehat{\lambda}^t)$, and embeddings

$$e_t^+ \colon SY^+ \to \widehat{X}, \tag{2-17}$$

$$e_t^- \colon SY^- \to \widehat{X}, \tag{2-18}$$

as in Definition 2.17. We will always assume that the family is fixed at infinity, meaning that for every compact subinterval $[a, b] \subset I$,

- $\{\widehat{\lambda}^t\}_{t \in [a,b]}$ and $\{H^t\}_{t \in [a,b]}$ are constant outside of a compact subset of $\widehat{X}$, and
- $\{e_t^+\}_{t \in [a,b]}$ (resp. $\{e_t^-\}_{t \in [a,b]}$) is independent of t on some neighborhood of $+\infty$ in SY^+ (resp. of $-\infty$ in SY^-).

Two relative cobordisms $(\widehat{X}^0, \widehat{\lambda}^0, H^0)$ and $(\widehat{X}^1, \widehat{\lambda}^1, H^1)$ are said to be *deformation equivalent* if there exists a one-parameter family $(\widehat{W}, \widehat{\mu}^t, K^t)_{t \in [0,1]}$ such that $(\widehat{X}^0, \widehat{\lambda}^0, H^0)$ is isomorphic to $(\widehat{W}, \widehat{\mu}^0, K^0)$ and $(\widehat{X}^1, \widehat{\lambda}^1, H^1)$ is isomorphic to $(\widehat{W}, \widehat{\mu}^1, K^1)$. The deformation class of a cobordism $(\widehat{X}, \widehat{\lambda}, H)$ will be denoted by $[\widehat{X}, \widehat{\lambda}, H]$.

Example 2.27 Given $(\widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01})$ and $(\widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12})$ as in Definition 2.23, the glued cobordisms $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{12}, H^{01} \#_t H^{12})_{t \in [N,\infty)}$ form a one-parameter family for $N > 0$ sufficiently large. Similarly, for any fixed $t \gg 0$, $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12}, H^{01} \#_t H^{12})_{s \in \mathbb{R}}$ is a one-parameter family. As in Lemma 2.13, it follows that the deformation class

$$[\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12}, H^{01} \#_t H^{12}] \tag{2-19}$$

is independent of $t \gg 0$ and $s \in \mathbb{R}$.

Proposition 2.28 *The deformation class of* $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{12}, H^{01} \#_t H^{12})$ *only depends on the deformation classes of* $(\widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01})$ *and* $(\widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12})$.

Proof The proof of Proposition 2.14 also works in the relative case as long as $t > 0$ is chosen large enough. □

Corollary 2.29 *There is a well-defined gluing operation on deformation classes of exact relative symplectic cobordisms given by*

$$[\widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01}] \# [\widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12}] = [\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_{t,s} \widehat{\lambda}^{12}, H^{01} \#_t H^{12}] \tag{2-20}$$

for any $t \gg 0$ *and* $s \in \mathbb{R}$.

Proposition 2.30 *The gluing operation* (2-20) *is associative.*

Proof Let $(\widehat{X}^{i,i+1}, \widehat{\lambda}^{i,i+1}, H^{i,i+1})$ be a relative cobordism from (Y^i, ξ^i, V^i) to $(Y^{i+1}, \xi^{i+1}, V^{i+1})$ for $i \in \{0,1,2\}$, and fix $t_1, t_2 \gg 0$. Note that $\big((\widehat{X}^{01} \#_{t_1} \widehat{X}^{12}) \#_{t_2} \widehat{X}^{23}, (H^{01} \#_{t_1} H^{12}) \#_{t_2} H^{23}\big)$ and $\big(\widehat{X}^{01} \#_{t_1} (\widehat{X}^{12} \#_{t_2} \widehat{X}^{23}), H^{01} \#_{t_1} (H^{12} \#_{t_2} H^{23})\big)$ can be canonically identified as pairs of smooth manifolds. Hence, it suffices to show that there exist $s_1, s_2 \in \mathbb{R}$ such that

$$(\widehat{\lambda}^{01} \#_{t_1,s_1} \widehat{\lambda}^{12}) \#_{t_2,s_2} \widehat{\lambda}^{23} = \widehat{\lambda}^{01} \#_{t_1,s_1} (\widehat{\lambda}^{12} \#_{t_2,s_2} \widehat{\lambda}^{23}). \tag{2-21}$$

One can easily see from Definition 2.7 that taking $s_1 = 0$ and $s_2 = -t_2$ works. □

2.2 Homotopy classes of asymptotically cylindrical maps

Definition 2.31 [68, Section 6.1] Suppose that $(\widehat{X}, \widehat{\lambda})$ is an exact symplectic cobordism from (Y^+, λ^+) to (Y^-, λ^-). Given a closed surface Σ and finite subsets $\boldsymbol{p}^+, \boldsymbol{p}^- \subset \Sigma$ (corresponding respectively to positive and negative punctures), a smooth map $u: \Sigma - (\boldsymbol{p}^+ \sqcup \boldsymbol{p}^-) \to \widehat{X}$ is said to be *asymptotically cylindrical* if it converges exponentially near each puncture $z \in \boldsymbol{p}^+ \sqcup \boldsymbol{p}^-$ to a trivial cylinder over a Reeb orbit.

More precisely, given any choice of translation invariant metric on $\mathbb{R} \times Y^\pm$, we require that there exist a choice of holomorphic cylindrical coordinates near each $z \in \boldsymbol{p}^\pm$ such that u takes the form

$$u(s,t) = \exp_{(Ps, \gamma_z(t))} h(s,t) \tag{2-22}$$

for $|s|$ large, where γ_z is a Reeb orbit of period P and $h(s,t)$ is a vector field which decays to zero with all its derivatives as $|s| \to \infty$ (ie these properties hold for $s \gg 0$ if $z \in \boldsymbol{p}^+$ and for $s \ll 0$ if $z \in \boldsymbol{p}^-$).

Remark 2.32 There is also a notion of an asymptotically cylindrical submanifold which will not be needed in this paper.

Definition 2.33 (cf [61, Section 1.2(I)]) Let $(\widehat{X}, \widehat{\lambda})$ be an exact symplectic cobordism from (Y^+, λ^+) to (Y^-, λ^-) and let $\Gamma^\pm$ be a finite set of Reeb orbits in $(Y^\pm, \lambda^\pm)$.

By truncating the ends of $\widehat{X}$, we obtain a compact submanifold $X_0 \subset \widehat{X}$ with boundary $\partial X_0 = Y^+ \sqcup Y^-$. We define the set of homotopy classes $\pi_2(\widehat{X}, \Gamma^+ \sqcup \Gamma^-)$ by

$$\pi_2(\widehat{X}, \Gamma^+ \sqcup \Gamma^-) := [(S, \partial S), (X_0, \Gamma^+ \sqcup \Gamma^-)] / \operatorname{Diff}(S, \partial S), \tag{2-23}$$

where S is a compact connected oriented surface of genus 0 equipped with a homeomorphism $\partial S \to \Gamma^+ \sqcup \Gamma^-$, and $\operatorname{Diff}(S, \partial S)$ is the group of diffeomorphisms of S which fix ∂S pointwise. (The notation $[-,-]$ here stands for *homotopy classes* of maps of pairs.)

Remark 2.34 The right-hand side of (2-23) is independent of the choice of truncation X_0 up to canonical bijection. In the case where $(\widehat{X}, \widehat{\lambda}) = (\mathbb{R} \times Y, e^s \lambda)$ is the symplectization of a contact manifold (Y, λ), we can take $X_0 = \{0\} \times Y$ and (2-23) becomes identical to [61, equation (1.2)].

For any choice of truncation $X_0 \subset \widehat{X}$, there is a canonical retraction $\pi\colon \widehat{X} \to X_0$ induced by quotienting by the Liouville flow (more precisely, one should quotient by the backwards Liouville flow at the positive end and by the forwards Liouville flow at the negative end). If $u\colon \Sigma - (\boldsymbol{p}^+ \sqcup \boldsymbol{p}^-) \to \widehat{X}$ is an asymptotically cylindrical map, then the composition $\pi \circ u$ can be extended to a map

$$\bar{u}\colon (\bar{\Sigma}, \partial\bar{\Sigma}) \to (X_0, \Gamma^+ \sqcup \Gamma^-), \tag{2-24}$$

where $\bar{\Sigma}$ is a compactification of $\Sigma - (\boldsymbol{p}^+ \sqcup \boldsymbol{p}^-)$ obtained by adding one boundary circle for each puncture. The homotopy class $[u] \in \pi_2(\widehat{X}, \Gamma^+ \sqcup \Gamma^-)$ of u is defined to be the equivalence class of (2-24).

Definition 2.35 Let $(\widehat{X}, \widehat{\lambda}, L)$ be an exact Lagrangian cobordism from $(Y^+, \lambda^+, \Lambda^+)$ to $(Y^-, \lambda^-, \Lambda^-)$; see Definition 2.20.

Given a surface with boundary Σ and finite subsets $\boldsymbol{p}^+, \boldsymbol{p}^- \subset \operatorname{int}(\Sigma)$ and $\boldsymbol{c}^+, \boldsymbol{c}^- \in \partial\Sigma$, a smooth map $u\colon \Sigma - (\boldsymbol{p}^\pm \cup \boldsymbol{c}^\pm)$ is said to be cylindrical if it converges asymptotically near each interior puncture to a trivial cylinder over a Reeb orbit, and it converges exponentially near each boundary puncture to a trivial strip over a Reeb chord.

Let $\Gamma^\pm$ be a finite set of Reeb orbits in $(Y^\pm, \lambda^\pm)$ and let $\Gamma_{\Lambda^\pm}$ be a finite *ordered* set of Reeb chords of $\Lambda^\pm \subset (Y^\pm, \lambda^\pm)$. We let $\boldsymbol{p}^\pm$ be a finite set equipped with bijections $\gamma^\pm\colon \boldsymbol{p}^\pm \to \Gamma^\pm$ and we let $\boldsymbol{c} = \boldsymbol{c}^+ \sqcup \boldsymbol{c}^-$ be a finite ordered set equipped with order-preserving bijections $a^\pm\colon \boldsymbol{c}^\pm \to \Gamma_{\Lambda^\pm}$. We let

$$\pi_2(\widehat{X}; \Gamma_{\Lambda^+}, \Gamma_{\Lambda^-}, \Gamma^+, \Gamma^-) \tag{2-25}$$

be the set of equivalence classes of maps from $\Sigma - (\boldsymbol{p}^\pm \cup \boldsymbol{c}^\pm)$ to $\widehat{X}$ which are asymptotic to $\gamma^\pm_p$ at $p \in \boldsymbol{p}^\pm$ (resp. $a^\pm_c$ at $c \in \boldsymbol{c}^\pm$), where two such maps u, v are equivalent if there exists a compactly supported diffeomorphism ϕ of $\Sigma - (\boldsymbol{p}^\pm \cup \boldsymbol{c}^\pm)$ such that u and $v \circ \phi$ are homotopic (through cylindrical maps).

3 Reeb dynamics near a codimension 2 contact submanifold

3.1 The Conley–Zehnder index

A hermitian vector bundle (E, J, ω) is a vector bundle E together with an almost complex structure J and a symplectic structure ω such that J is compatible with ω. An *asymptotic operator* on a hermitian vector bundle (E, J, ω) over S^1 is a real-linear differential operator $\boldsymbol{A}: \Gamma(E) \to \Gamma(E)$ which, in some (and hence any) unitary trivialization, takes the form

$$\boldsymbol{A}: C^\infty(S^1, \mathbb{R}^{2n}) \to C^\infty(S^1, \mathbb{R}^{2n}), \quad \eta \mapsto -J\partial_t \eta - S(t)\eta, \tag{3-1}$$

where $t \in S^1$ and $S: S^1 \to \mathrm{End}(\mathbb{R}^{2n})$ is a loop of symmetric matrices. The asymptotic operator $\boldsymbol{A}$ is said to be nondegenerate if 0 is not an eigenvalue.

Fix a hermitian vector bundle (E, J, ω) over S^1 and a unitary trivialization τ. Given a nondegenerate asymptotic operator $\boldsymbol{A}$, we can obtain a nondegenerate path of symplectic matrices by solving the ordinary differential equation

$$(-J\partial_t - S(t))\Psi(t) = 0, \quad \Psi(0) = \mathrm{id}. \tag{3-2}$$

Conversely, given a nondegenerate path of symplectic matrices, we can recover a nondegenerate asymptotic operator by solving (3-2) for $S(t)$.

The *Conley–Zehnder index* $\mathrm{CZ}(\boldsymbol{A}) = \mathrm{CZ}(\Psi) \in \mathbb{Z}$ is an integer-valued invariant which can be associated equivalently to a nondegenerate asymptotic operator equipped with a unitary trivialization τ or to a nondegenerate path of symplectic matrices. It only depends on τ up to homotopy through unitary trivializations. We refer the reader to [68, Section 3.4] or [33] for a detailed overview of the Conley–Zehnder index.

Definition 3.1 Let $(Y, \xi = \ker\lambda)$ be a contact manifold and let γ be a Reeb orbit of period $P > 0$, parametrized so that $\lambda(\gamma') = P$. Given a choice of $d\lambda$–compatible almost complex structure J on ξ, we can define the asymptotic operator $\boldsymbol{A}_\gamma: \Gamma(\gamma^*\xi) \to \Gamma(\gamma^*\xi)$ by $\boldsymbol{A}_\gamma = -J(\nabla_t - P\nabla R_\lambda)$, where ∇ is some symmetric connection on Y.

The Conley–Zehnder index of a Reeb orbit γ relative to a trivialization τ of $\gamma^*\xi$ will be denoted by $\mathrm{CZ}^\tau(\gamma) := \mathrm{CZ}^\tau(\boldsymbol{A}_\gamma)$.

Let us now consider a contact manifold $(Y^{2n-1}, \xi = \ker\lambda)$ and a strong contact submanifold $(V^{2n-3}, \lambda|_V)$. Observe that the contact distribution splits naturally along V as

$$\xi|_V = \xi|_V^\top \oplus \xi|_V^\perp, \tag{3-3}$$

where $\xi|_V^\top = \xi|_V \cap TV$ and $\xi|_V^\perp$ is the symplectic orthogonal complement of $\xi|_V^\top \subset \xi|_V$ with respect to $d\lambda$. Suppose that J is a $d\lambda$–compatible almost complex structure on ξ which respects the splitting (3-3).

Let $\gamma : S^1 \to V$ be a Reeb orbit. Since J respects the above splitting, then so does the associated asymptotic operator, which we can therefore write as $\boldsymbol{A}_\gamma = \boldsymbol{A}_\gamma^\top \oplus \boldsymbol{A}_\gamma^\perp$. If we choose a unitary trivialization τ of $\xi|_\gamma$ which is also compatible with the splitting, we can define $\mathrm{CZ}_T^\tau(\gamma) := \mathrm{CZ}^\tau(\boldsymbol{A}_\gamma^\top)$ and $\mathrm{CZ}_N^\tau(\gamma) := \mathrm{CZ}^\tau(\boldsymbol{A}_\gamma^\perp)$. We call these respectively the tangential and normal Conley–Zehnder indices of γ with respect to τ.

We define the integers

$$\alpha_N^{\tau;-}(\gamma) := \lfloor \mathrm{CZ}_N^\tau(\gamma)/2 \rfloor, \tag{3-4}$$

$$\alpha_N^{\tau;+}(\gamma) := \lceil \mathrm{CZ}_N^\tau(\gamma)/2 \rceil. \tag{3-5}$$

Let $p_N(\gamma) = \alpha_N^{\tau;+}(\gamma) - \alpha_N^{\tau;-}(\gamma) \in \{0, 1\}$ be the (*normal*) *parity* of γ, and observe that it is independent of the choice of trivialization. We have

$$\mathrm{CZ}_N^\tau(\gamma) = 2\alpha_N^{\tau;-}(\gamma) + p_N(\gamma) = 2\alpha_N^{\tau;+}(\gamma) - p_N(\gamma), \tag{3-6}$$

from which it also follows that $p_N(\gamma) \equiv \mathrm{CZ}_N^\tau(\gamma) \bmod 2$.

There is a canonical isomorphism

$$\xi|_V^\perp \xrightarrow{\sim} N_{Y/V}, \tag{3-7}$$

where $N_{Y/V}$ denotes the normal bundle of $V \subset Y$. If $(V, \xi_V) \subset (Y, \xi)$ are co-oriented and hence oriented, then $N_{Y/V}$ is also oriented and (3-7) is orientation-preserving. If we assume that $N_{Y/V}$ is trivial, then it follows that (3-7) induces a bijection between homotopy classes of trivializations of $N_{Y/V}$ compatible with the orientation and homotopy classes of unitary trivializations on $\xi|_V^\perp$. Since the Conley–Zehnder index only depends on the homotopy class of unitary trivializations, we may define $\mathrm{CZ}_N^\tau(\gamma)$ with respect to any homotopy class of trivializations τ of $N_{Y/V}$ compatible with the orientation.

3.2 Normal dynamics and adapted contact forms

We now state some important definitions, which will be used throughout this paper.

Definition 3.2 A *trivial-normal contact pair* (or just *TN contact pair*) is a datum (Y, ξ, V) consisting of a closed co-oriented contact manifold (Y, ξ) and a co-oriented codimension 2 contact submanifold $V \subset (Y, \xi)$ with trivial normal bundle $N_{Y/V}$.

An important example of a TN contact pair is the binding of a contact open book decomposition; see Definition 3.10. A choice of (homotopy class of) trivialization τ on $N_{Y/V}$ is called a *framing* and we say that $(V, \tau) \subset Y$ is a *framed* codimension 2 submanifold. Note that we do not assume in Definition 3.2 that V and Y are nonempty. For future reference, we let $\xi_\varnothing$ denote the unique contact structure on the empty set.

Definition 3.3 Given a TN contact pair (Y, ξ, V) with V nonempty, let $\mathfrak{R}(Y, \xi, V)$ be the set of triples $\mathfrak{r} = (\alpha_V, \tau, r)$ where

- $\alpha_V \in \Omega^1(V)$ is a nondegenerate contact form for (V, ξ_V),

- τ is a homotopy class of trivializations of $N_{Y/V}$ (ie a framing) which is compatible with the orientation, and
- $r > 0$ is a strictly positive real number, and we have $(1/r)\mathbb{Z} \cap \mathcal{S}(\alpha_V) = \varnothing$, where $\mathcal{S}(\alpha_V)$ is the action spectrum of α_V.

If $V = \varnothing$ (with Y possibly also empty), we define $\mathfrak{R}(Y, \xi, \varnothing) = \{(\alpha_\varnothing, \tau_\varnothing, 0)\}$, where $\alpha_\varnothing$ and $\tau_\varnothing$ are understood as a contact form and normal trivialization on the empty set.

Definition 3.4 Given a TN contact pair (Y, ξ, V) and $\mathfrak{r} = (\alpha_V, \tau, r) \in \mathfrak{R}(Y, \xi, V)$, we say that a contact form $\ker\lambda = \xi$ is *adapted* to $\mathfrak{r}$ if

- λ is nondegenerate,
- $\lambda|_V = \alpha_V$,
- V is a strong contact submanifold of (Y, λ) (see Definition 2.22), and
- we have $\mathrm{CZ}^\tau_N(\gamma) = 1 + 2\lfloor rP_\gamma \rfloor$ for all Reeb orbits $\gamma \subset V$, where P_γ is the period of γ.

In the case that $V = \varnothing$, any contact form λ is considered to be adapted to the unique element $(\alpha_\varnothing, \tau_\varnothing, 0) \in \mathfrak{R}(Y, \xi, \varnothing)$. Given a contactomorphism $f : (Y, \xi, V) \to (Y', \xi', V')$, we write $f_*\mathfrak{r} = (f_*\alpha_V, f_*\tau, r) \in \mathfrak{R}(Y', \xi', V')$. If $\phi_t : V \to Y$ is an isotopy of contact embeddings where ϕ_0 is the tautological embedding $V \xrightarrow{\text{id}} V \subset Y$ and $\phi_1(V) = V'$, then ϕ_t extends to a family of contactomorphisms f_t. We then write $(\phi_1)_*\mathfrak{r} := (f_1)_*\mathfrak{r}$; this is independent of the choice of extension.

We say that λ is *positive elliptic* near $V \neq \varnothing$ if it is adapted to some $\mathfrak{r} = (\alpha_V, \tau, r) \in \mathfrak{R}(Y, \xi, V)$; we refer to $r > 0$ as the *rotation parameter*.

Remark 3.5 Our insistence on allowing the case where $Y = \varnothing$ in the above definitions is explained by the need to treat Liouville manifolds as special cases of Liouville cobordisms in the arguments of Section 7.

We will prove in Proposition 3.9 that adapted contact forms always exist, ie for any TN contact pair (Y, ξ, V) and $\mathfrak{r} = (\alpha_V, \tau, r) \in \mathfrak{R}(Y, \xi, V)$ there exists a contact form adapted to $\mathfrak{r}$. The first step is to construct a suitable local model.

Construction 3.6 For $0 < \epsilon \leq 1$, let $D^2 \subset \mathbb{R}^2$ be the standard disk of radius ϵ (in the sequel, we will often denote this disk by D^2_ϵ). Let (V, α_V) be a contact manifold and let $\phi \colon D^2 \to \mathbb{R}_{>0}$ be a smooth positive function which has a nondegenerate critical point at 0 and satisfies $\phi(0) = 1$. We define

$$\alpha_V^\phi = \frac{1}{\phi}(\alpha_V + \lambda_{D^2}), \tag{3-8}$$

where $\lambda_{D^2} = \frac{1}{2}(x\,dy - y\,dx)$ is the usual Liouville form on D^2. This is a contact form on $V \times D^2$ whose restriction to $V = V \times \{0\}$ coincides with α_V. Its Reeb vector field is given by

$$R_\phi = (\phi - Z_{D^2}\phi)R_V + X_\phi, \tag{3-9}$$

where $Z_{D^2} = \frac{1}{2}(x\partial_x + y\partial_y)$ is the Liouville vector field of λ_{D^2} and $X_\phi = -(\partial_y\phi)\partial_x + (\partial_x\phi)\partial_y$ is the Hamiltonian vector field of ϕ with respect to the symplectic form $\omega_{D^2} = d\lambda_{D^2}$. Our assumptions on ϕ imply that $R_\phi = R_V$ on $V \times \{0\}$, so that (V, α_V) is a strong contact submanifold of $(V \times D^2, \alpha_V^\phi)$. We will let

$$S_\phi = \begin{pmatrix} \partial_{xx}\phi(0) & \partial_{yx}\phi(0) \\ \partial_{xy}\phi(0) & \partial_{yy}\phi(0) \end{pmatrix} \in \mathbb{R}^{2\times 2} \tag{3-10}$$

denote the Hessian of ϕ at the origin. Since S_ϕ is symmetric and nondegenerate, its eigenvalues are real and nonzero, so its signature $\operatorname{Sign}(S_\phi)$ is one of $0, \pm 2$. We will say that ϕ is *hyperbolic* if $\operatorname{Sign}(S_\phi) = 0$, *positive elliptic* if $\operatorname{Sign}(S_\phi) = 2$ and *negative elliptic* if $\operatorname{Sign}(S_\phi) = -2$. In the elliptic case, we define $c_\phi = \sqrt{\det(S_\phi)}/(2\pi)$; this is a positive real number since $\det(S_\phi) > 0$. Finally, we note that the splitting $(\xi_\phi)|_V = (\xi_\phi)|_V^\top \oplus (\xi_\phi)|_V^\perp$ — see equation (3-3) — is given by $(\xi_\phi)|_V^\top = \xi_V$ and $(\xi_\phi)|_V^\perp = T_0 D^2$. We will let τ_ϕ denote the trivialization of $(\xi_\phi)|_V^\perp$ by $\{\partial_x, \partial_y\}$.

We say that α_V^ϕ is *nondegenerate on V* if every Reeb orbit of α_V is nondegenerate when viewed as a Reeb orbit of α_V^ϕ.

Proposition 3.7 *Carrying over the notation of Construction 3.6, suppose that α_V is nondegenerate. If ϕ is elliptic, then α_V^ϕ is nondegenerate on V if and only if $(1/c_\phi)\mathbb{Z} \cap \mathcal{S}(\alpha_V) = \varnothing$, where $\mathcal{S}(\alpha_V)$ denotes the action spectrum of α_V.*

Proof Let γ be a Reeb orbit of period P contained in V. Recall that γ is nondegenerate if and only if its asymptotic operator is nondegenerate. Choose a trivialization τ and an almost complex structure J on $(\xi_\phi)|_\gamma$ which preserve the splitting $(\xi_\phi)|_\gamma = (\xi_V)|_\gamma \oplus T_0 D^2$ and coincide with τ_ϕ and J_0 respectively on $T_0 D^2$, where J_0 denotes the standard almost complex structure on $\mathbb{R}^2 = T_0 D^2$. The asymptotic operator A_γ is compatible with this splitting and can therefore be written as $A_\gamma = A_\gamma^\top \oplus A_\gamma^\perp$. The tangential part $A_\gamma^\top$ is nondegenerate since it coincides with the asymptotic operator of γ as a Reeb orbit in V. The normal part $A_\gamma^\perp$ is given explicitly by

$$A_\gamma^\perp = -J_0\partial_t - P \cdot S_\phi \tag{3-11}$$

(this follows from a short computation using the formula for the Reeb vector field R_ϕ in Construction 3.6). Define a path Ψ of symplectic matrices by $\Psi(t) = \exp(tP \cdot J_0 S_\phi)$. Then $A_\gamma^\perp$ is nondegenerate if and only if $\Psi(1)$ doesn't have 1 as an eigenvalue. If ϕ is elliptic, then the eigenvalues of $\Psi(1)$ are $\exp(\pm iP\sqrt{\det(S_\phi)})$. Hence, γ is nondegenerate if and only if $P\sqrt{\det(S_\phi)}$ is not an integer multiple of 2π, ie $P \notin (1/c_\phi)\mathbb{Z}$. It follows that λ_ϕ is nondegenerate if and only if $(1/c_\phi)\mathbb{Z} \cap \mathcal{S}(\alpha_V) = \varnothing$, as claimed. □

The important feature of Construction 3.6 is that the normal Conley–Zehnder indices of the Reeb orbits in V can be computed explicitly.

Proposition 3.8 *Assume α_V^ϕ is nondegenerate on V. If ϕ is elliptic, then*

$$\mathrm{CZ}_N^{\tau_\phi}(\gamma) = \pm(1 + 2\lfloor c_\phi P_\gamma \rfloor) \tag{3-12}$$

for every Reeb orbit γ contained in V, where $P_\gamma > 0$ denotes the period of γ and the sign is $+$ or $-$ depending on whether ϕ is positive elliptic or negative elliptic.

Proof We have $\mathrm{CZ}_N^{\tau_\phi}(\gamma) = \mathrm{CZ}(\Psi)$, where $\Psi(t) = \exp(tP \cdot J_0 S_\phi)$ is the path of symplectic matrices defined in the proof of Proposition 3.7; see [68, Section 3.4]. Proposition 41 of [33] implies that

$$\mathrm{CZ}(\Psi) = \pm(1 + 2\lfloor c_\phi P \rfloor) \tag{3-13}$$

if $\mathrm{Sign}(S_\phi) = \pm 2$. □

Proposition 3.9 *Fix a TN contact pair (Y, ξ, V) and an element $\mathfrak{r} = (\alpha_V, \tau, r) \in \mathfrak{R}(Y, \xi, V)$. Then there exists a contact form λ on (Y, ξ) which is adapted to $\mathfrak{r}$.*

Proof Let ϕ be as in Construction 3.6. The standard neighborhood theorem for contact submanifolds (see [29, Theorem 2.5.15]) implies that the inclusion map $V \to Y$ extends to a contact embedding $\iota\colon (V \times D_\epsilon^2, \ker(\alpha_V^\phi)) \to (Y, \xi)$ such that $\iota^*\tau$ is homotopic to τ_ϕ, for some $\epsilon > 0$ sufficiently small. Hence there exists a contact form λ for ξ such that $\iota^*\lambda = \alpha_V^\phi$ near V. In addition to choosing ϕ so that λ is adapted to $\mathfrak{r}$, we also need to make sure that λ can be modified away from V so that it becomes nondegenerate. By [2, Theorem 13], this can be achieved by choosing a ϕ such that the following two conditions are satisfied:

- α_V^ϕ is nondegenerate on V.
- All the Reeb orbits of α_V^ϕ in $V \times D_\epsilon^2$ are contained in V.

Let us set $\phi = 1 + \pi r(x^2 + y^2)$. Proposition 3.7 implies that α_V^ϕ is nondegenerate on V since

$$c_\phi = \frac{\sqrt{(2\pi r)^2}}{2\pi} = r.$$

Since $R_\phi = R_V + X_\phi$, every Reeb orbit γ of α_V^ϕ is of the form $\gamma = (\gamma_V, \gamma_\phi)$, where γ_V is an orbit of R_V and γ_ϕ is an orbit of X_ϕ with the same period $P > 0$. From the formula $X_\phi = -2\pi r y \partial_x + 2\pi r x \partial_y$, we see that if γ_ϕ were not constant, we would have $P \in (1/r)\mathbb{Z}$, contradicting our assumption on r; see Definition 3.3. Thus γ is contained in V. □

3.3 Open book decompositions

In this section, we consider normal Reeb dynamics for bindings of open book decompositions. We begin by recalling the definition of an open book decomposition; we refer to [62, Section A.1; 29, Section 4.4.2] for a historically informed survey of this theory.

Definition 3.10 An *open book decomposition* (Y, B, π) of a closed, oriented n–manifold Y consists of the following data:

(i) An oriented, closed, codimension 2 submanifold $B \subset Y$ with trivial normal bundle.

(ii) A fibration $\pi: Y - B \to S^1$ which coincides with the angular coordinate in some neighborhood $B \times \{0\} \subset B \times D^2 = B \times \{(x, y) \mid x^2 + y^2 < 1\}$.

The submanifold $B \subset Y$ is called the *binding* and the fibers of π are called *pages*.

Observe that the data of an open book decomposition induces a natural trivialization of the normal bundle to the binding. We also recall what it means for an open book decomposition to support a contact structure.

Definition 3.11 [31] Given an odd-dimensional manifold Y^{2n-1}, an open book decomposition (Y, B, π) is said to *support* a contact structure ξ if there exists a contact form $\xi = \ker \alpha$ such that the following properties hold:

(i) The restriction of α to B is a contact form.

(ii) The restriction of $d\alpha$ to any page $\pi^{-1}(\theta)$ is a symplectic form.

(iii) The orientation of B induced by α coincides with the orientation of B as the boundary of the symplectic manifold $(P_\theta, d\alpha)$, where $P_\theta = \pi^{-1}(\theta)$ is any page.

Such a contact form is called a *Giroux form* (and is also said in the literature to be *adapted* to the open book decomposition).

Remark 3.12 Condition (ii) in the above definition is equivalent to the Reeb vector field of α being transverse to the pages.

For future convenience, we state the following definition.

Definition 3.13 Let $\mathcal{G}$ be the set of TN contact pairs (Y, ξ, V) having the property that ξ is supported by an open book decomposition $\pi: Y - V \to S^1$ with binding V.

Lemma 3.14 *Let (Y, B, π) be an open book decomposition supporting the contact structure ξ. Let α_B be a contact form for (B, ξ_B) and let $f: [0, 1) \to \mathbb{R}$ be a positive smooth function such that $f(0) = 1$ and $f'(r) < 0$ for $r > 0$. Then there exists a Giroux form α and an embedding $\phi: B \times D^2_\epsilon \to Y$ (for some small $\epsilon > 0$) with the following properties*:

(1) $\alpha|_B = \alpha_B$.

(2) *The projection $\pi \circ \phi$ is given by $(r, \theta) \mapsto \theta$ on $B \times D^2_\epsilon - B \times \{0\}$.*

(3) $\phi^*\alpha = f(r)(\alpha|_B + \lambda_{D^2})$.

Proof The proof is in two steps. First, we will show that there exists a Giroux form $\widetilde{\alpha}$ such that $\widetilde{\alpha}|_B = \alpha_B$. Second, we will construct α by modifying $\widetilde{\alpha}$ so that (1)–(3) are satisfied.

Step 1 Let α' be an arbitrary Giroux form for (Y, B, π). Since $\alpha'|_B$ and α_B define the same contact structure, we can write $\alpha_B = (1+h)(\alpha'|_B)$ for some smooth function $h\colon B \to \mathbb{R}$. We may assume without loss of generality that $h \geq 0$ everywhere, since any positive constant multiple of α' is also a Giroux form.

It is shown in the proof of [18, Proposition 2] that there exists a tubular neighborhood $B \times D^2_\epsilon$ of the binding on which $\pi = \theta$ and $\alpha' = g(\alpha'|_B + \lambda_{D^2})$, where $g\colon B \times D^2_\epsilon \to \mathbb{R}$ is a positive smooth function satisfying $g \equiv 1$ on $B \times \{0\}$, $\lambda_{D^2} = \frac{1}{2}(x\,dy - y\,dx)$ and $\epsilon > 0$ is a suitably small constant. Note that $g(\alpha'|_B + \lambda_{D^2})$ is a Giroux form on $(B \times D^2_\epsilon, B \times \{0\}, \pi = \theta)$ if and only if $\partial g/\partial r < 0$ for $r > 0$.

Let $\sigma\colon [0, \epsilon] \to \mathbb{R}$ be a nonincreasing smooth function such that $\sigma(r) = 1$ for r near 0 and $\sigma(r) = 0$ for r near ϵ. Set $\widetilde{g} := (1+\sigma(r)h)g$. Then $\partial_r \widetilde{g} = \partial_r \sigma \cdot hg + (1+\sigma h)\cdot \partial_r g < 0$. Now we define $\widetilde{\alpha}$ by replacing g with $\widetilde{g}$.

Step 2 By the previous step, we may fix a Giroux form $\widetilde{\alpha}$ so that $\widetilde{\alpha}|_B = \alpha_B$. Appealing again to the proof of [18, Proposition 2], there exists a tubular neighborhood $B \times D^2_{\epsilon'}$ of the binding on which $\pi = \theta$ and $\widetilde{\alpha} = \gamma(\widetilde{\alpha}|_B + \lambda_{D^2})$, where $\gamma\colon B \times D^2_{\epsilon'} \to \mathbb{R}$ is a positive smooth function satisfying $\gamma \equiv 1$ on $B \times \{0\}$, $\lambda_{D^2} = \frac{1}{2}(x\,dy - y\,dx)$ and $\epsilon' > 0$ is a suitably small constant. Again, we have that $\gamma(\widetilde{\alpha}|_B + \lambda_{D^2})$ is a Giroux form on $(B \times D^2_{\epsilon'}, B \times \{0\}, \pi = \theta)$ if and only if $\partial\gamma/\partial r < 0$ for $r > 0$.

Let $\delta\colon B \times D^2_{\epsilon'} \to \mathbb{R}$ be a positive smooth function such that $\delta = f$ near $B \times \{0\}$, $\delta = \gamma$ near $B \times \partial D^2_{\epsilon'}$, and $\partial_r \delta < 0$ for $r > 0$. Let α be the unique contact form on Y which coincides with $\widetilde{\alpha}$ outside the image of ϕ and satisfies $\phi^*\alpha = \delta(\widetilde{\alpha}|_B + \lambda_{D^2})$. Then α is a Giroux form and satisfies conditions (1)–(3). □

Corollary 3.15 *Consider an open book decomposition (Y, B, π) which supports a contact structure ξ and let τ denote the induced trivialization of the normal bundle of $B \subset Y$. Choose an element $\mathfrak{r} = (\alpha_B, \tau, r) \in \mathfrak{R}(Y, \xi, B)$. Then there exists a Giroux form α which is adapted to $\mathfrak{r}$; see Definition 3.4.*

Proof Let $\kappa = \pi r$ and define $f(s) = (1+\kappa s^2)^{-1}$ for $s \in [0, 1)$. Since $f(0) = 1$ and $f'(s) < 0$, it follows that there exists a Giroux form $\widetilde{\alpha}$ satisfying the conditions stated in Lemma 3.14. As we observed in the proof of Proposition 3.9, there exists a neighborhood $\mathcal{U}$ of B with the properties that

- $\widetilde{\alpha}$ is nondegenerate on $\mathcal{U}$, and
- all the Reeb orbits in $\mathcal{U}$ are contained in B.

According to [2, Theorem 13], we can obtain a nondegenerate contact form by multiplying $\widetilde{\alpha}$ by a smooth function $g\colon Y \to \mathbb{R}_+$ with $g \equiv 1$ near B. Moreover, we can assume that $g - 1$ is arbitrarily C^1–small and hence that $g\widetilde{\alpha}$ is still a Giroux form.

Since $g\widetilde{\alpha} = \alpha$ near B, it follows that $(g\widetilde{\alpha})|_B = \alpha_B$ and that B is a strong contact submanifold with respect to $g\widetilde{\alpha}$. Finally, the last point in Definition 3.4 can be verified just as in the proof of Proposition 3.8. □

4 Standard setups and tree categories.

Contact homology is defined in [61] by counting pseudoholomorphic curves (and, more generally, pseudoholomorphic buildings) in four setups. To keep track of the combinatorics of these curves, Pardon introduces certain categories of decorated trees. We briefly review this formalism here, referring the reader to [61, Section 2.1] for details.

4.1 Standard setups

Setup I A datum $\mathcal{D}$ for Setup I consists of a triple (Y, λ, J), where Y is a closed manifold, λ is a nondegenerate contact form on Y and J is a $d\lambda$–compatible almost complex structure on $\xi = \ker\lambda$.

Setup II A datum $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, \hat{X}, \hat{\lambda}, \hat{J})$ for Setup II consists of

- data $\mathcal{D}^\pm = (Y^\pm, \lambda^\pm, J^\pm)$ as in Setup I,
- an exact symplectic cobordism $(\hat{X}, \hat{\lambda})$ with positive end (Y^+, λ^+) and negative end (Y^-, λ^-), and
- a $d\hat{\lambda}$–tame almost complex structure $\hat{J}$ on $\hat{X}$ which agrees with $\hat{J}^\pm$ at infinity.

Setup III A datum $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, (\hat{X}, \hat{\lambda}^t, \hat{J}^t)_{t\in[0,1]})$ for this setting consists of

- data $\mathcal{D}^\pm = (Y^\pm, \lambda^\pm, J^\pm)$ as in Setup I,
- a family of exact symplectic cobordisms $(\hat{X}, \hat{\lambda}^t)_{t\in[0,1]}$ with positive end (Y^+, λ^+) and negative end (Y^-, λ^-), and
- a $d\hat{\lambda}^t$–tame almost complex structure $\hat{J}^t$ on $\hat{X}$ which agrees with $\hat{J}^\pm$ at infinity.

Note that for every $t_0 \in [0,1]$, there is a datum $\mathcal{D}^{t=t_0} = (\mathcal{D}^+, \mathcal{D}^-, \hat{X}, \hat{\lambda}^{t_0}, \hat{J}^{t_0})$ as in Setup II.

Setup IV A datum $\mathcal{D} = (\mathcal{D}^{01}, \mathcal{D}^{12}, (\hat{X}^{02,t}, \hat{\lambda}^{02,t}, \hat{J}^{02,t})_{t\in[0,\infty)})$ for this setting consists of

- data
$$\mathcal{D}^{01} = (\mathcal{D}^0, \mathcal{D}^1, \hat{X}^{01}, \hat{\lambda}^{01}, \hat{J}^{01}) \quad \text{and} \quad \mathcal{D}^{12} = (\mathcal{D}^1, \mathcal{D}^2, \hat{X}^{12}, \hat{\lambda}^{12}, \hat{J}^{12})$$
as in Setup II, where $\mathcal{D}^i = (Y^i, \lambda^i, J^i)$ for $i = 0, 1, 2$,
- a family of exact symplectic cobordisms $(\hat{X}^{02,t}, \hat{\lambda}^{02,t})_{t\in[0,\infty)}$ with positive end (Y^0, λ^0) and negative end (Y^2, λ^2), which for t large coincides with the t–gluing of $(\hat{X}^{01}, \hat{\lambda}^{01})$ and $(\hat{X}^{12}, \hat{\lambda}^{12})$, and
- a $d\hat{\lambda}^{02,t}$–tame almost complex structure $\hat{J}^{02,t}$ on $\hat{X}^{02,t}$ which agrees with $\hat{J}^0$ and $\hat{J}^2$ at infinity, and is induced by $\hat{J}^{01}$ and $\hat{J}^{02}$ for t large.

4.2 Trees

For each setup, Pardon [61, Section 2.1] defines a category $\mathcal{S}_*$ for $* = \mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{IV}$, which depends on some datum $\mathcal{D}$. Each object $T \in \mathcal{S}$ is a decorated tree (or forest) representing a certain class of pseudoholomorphic curves (or more generally of buildings). Geometrically, the vertices correspond to curves, edges correspond to asymptotic orbits, and the decorations keep track of additional information (such as the homology classes of the components, and their "level" in the SFT compactification).

A morphism of trees in $\mathcal{S}_*$ consists of two pieces of data. First, a contraction of some edges, with the important caveat that only certain contractions are allowed, which depend on the decorations on the tree. Second, one specifies some additional data on external edges, which depends on the decorations of the external edges. Geometrically, a morphism of trees correspond to gluing holomorphic curves, and the data which one specifies on the external edges encodes different ways of moving asymptotic markers. For $T \in \mathcal{S}_*$, we let $\mathrm{Aut}(T)$ denote the group of automorphisms of T. Given a morphism $T' \to T$, we let $\mathrm{Aut}(T'/T) \subset \mathrm{Aut}(T')$ be the subgroup of automorphisms of T' which are compatible with $T' \to T$.

In each category $\mathcal{S}_*$, there is an operation called *concatenation* whose input is a collection of trees (satisfying certain conditions, and with additional matching data), and whose output is a single tree. Geometrically, concatenations of trees correspond to "stacking" holomorphic buildings. The precise rules for concatenations are rather involved and depend on the individual setups.

Remark 4.1 A datum $\mathcal{D}$ for Setups II, III, IV determines multiple categories of trees: this is because such a datum itself contains (by definition) data for multiple setups. *We always follow the notation of* [61, Section 2.1] *to denote the resulting tree categories.* So, for example, if $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, \hat{X}, \hat{\lambda}, \hat{J})$ is a datum for Setup II, we write $\mathcal{S}_{\mathrm{II}} := \mathcal{S}_{\mathrm{II}}(\mathcal{D})$, $\mathcal{S}_{\mathrm{I}}^+ := \mathcal{S}_{\mathrm{I}}(\mathcal{D}^+)$ and $\mathcal{S}_{\mathrm{I}}^- := \mathcal{S}_{\mathrm{I}}(\mathcal{D}^-)$. Similarly, a datum for Setup III determines categories $\mathcal{S}_{\mathrm{III}}, \mathcal{S}_{\mathrm{II}}^{t=0}, \mathcal{S}_{\mathrm{II}}^{t=1}, \mathcal{S}_{\mathrm{I}}^{\pm}$, and a datum for Setup IV determines categories $\mathcal{S}_{\mathrm{IV}}, \mathcal{S}_{\mathrm{II}}^{01}, \mathcal{S}_{\mathrm{II}}^{12}, \mathcal{S}_{\mathrm{I}}^{0}, \mathcal{S}_{\mathrm{I}}^{1}, \mathcal{S}_{\mathrm{I}}^{2}$.

4.3 Virtual moduli counts

To an object $T \in \mathcal{S}_*$, we can associate a moduli space $\mathcal{M}(T)$ [61, Section 2.3] which carries an action of $\mathrm{Aut}(T)$ — this action corresponds geometrically to changing asymptotic markers. Note that $T \in \mathcal{S}_*$ has a well-defined notion of index and virtual dimension [61, Definition 2.42]. The compactified moduli space $\overline{\mathcal{M}}(T)$ is defined by (see [61, Definition 2.13])

$$\overline{\mathcal{M}}(T) := \bigsqcup_{T' \to T} \mathcal{M}(T')/\mathrm{Aut}(T'/T). \tag{4-1}$$

Theorem 1.1 in [61] provides a perturbation datum $\theta \in \Theta_*(\mathcal{D})$ and associated virtual moduli counts $\#\overline{\mathcal{M}}(T)_\theta^{\mathrm{vir}} \in \mathbb{Q}$ (which are zero for $\mathrm{vdim}(T) \neq 0$) satisfying the *master equations*

$$0 = \sum_{\mathrm{codim}(T'/T)=1} \frac{1}{|\mathrm{Aut}(T'/T)|} \#\overline{\mathcal{M}}(T')_\theta^{\mathrm{vir}} \tag{4-2}$$

and

$$\#\overline{\mathcal{M}}(\#_i T_i)^{\text{vir}}_{\theta} = \frac{1}{|\text{Aut}(\{T_i\}_i/\#_i T_i)|} \prod_i \#\overline{\mathcal{M}}(T_i)^{\text{vir}}_{\theta}. \tag{4-3}$$

By standard arguments, this can be used to define the various maps involved in the definition of contact homology (such as the differential d) and show that they satisfy the expected relations (such as $d^2 = 0$); see [61, Section 1.7].

5 Intersection theory for punctured holomorphic curves

5.1 Definition of the Siefring intersection number

We will make use in this paper of an intersection theory for asymptotically cylindrical maps and submanifolds. The four-dimensional theory was constructed by Siefring [64] and assigns an integer to a pair of asymptotically cylindrical maps in a 4–dimensional symplectic cobordism; see also the book by Wendl [69]. The higher-dimensional theory, also due to Siefring, assigns an integer to the pairing of a codimension 2 (asymptotically) cylindrical hypersurface with a (asymptotically) cylindrical map. A detailed overview can be found in [54].

Consider a strong exact symplectic cobordism $(\widehat{X}, \widehat{\lambda})$ from (Y^+, λ^+) to (Y^-, λ^-). Let $(V^{\pm}, \lambda^{\pm}|_V) \subset (Y^{\pm}, \lambda^{\pm})$ be strong contact submanifolds and let $H \subset \widehat{X}$ be a codimension 2 submanifold with cylindrical ends $V^+ \sqcup V^-$.

We let τ denote a choice of trivialization of $\xi^{\pm}|^{\perp}_{V^{\pm}}$ along every Reeb orbit in $V^{\pm}$. We require that the trivialization along a multiply covered orbit be pulled back from the chosen trivialization along the underlying simple orbit. Let $u\colon \Sigma - (\boldsymbol{p}_u^+ \sqcup \boldsymbol{p}_u^-) \to \widehat{X}$ be a map which is positively/negatively asymptotic at $z \in \boldsymbol{p}_u^{\pm}$ to the Reeb orbit γ_z. Now set

$$u \bullet_{\tau} H := u^{\tau} \cdot H, \tag{5-1}$$

where u^{τ} is a perturbation of u which is transverse to H and constant with respect to τ at infinity, and $(- \cdot -)$ is the usual algebraic intersection number for transversely intersecting smooth maps. While (5-1) depends on the choice of trivialization τ, Siefring showed that this count can be corrected so as to become independent of τ. This leads to the following definition.

Definition 5.1 [54, Section 2] The *generalized* (*or Siefring*) *intersection number* $u * H \in \mathbb{Z}$ of u and H is defined by

$$u * H = u \bullet_{\tau} H + \sum_{z \in \boldsymbol{p}_u^+} \alpha_N^{\tau;-}(\gamma_z) - \sum_{z \in \boldsymbol{p}_u^-} \alpha_N^{\tau;+}(\gamma_z). \tag{5-2}$$

Proposition 5.2 *The intersection number $u * H$ only depends on the equivalence classes of u in $\pi_2(\widehat{X}, \Gamma^+ \sqcup \Gamma^-)$, and H in $\Omega_{2n-2}(\widehat{X}, V^+ \sqcup V^-)$; see Definition 2.21.*

Proof The intersection number $u^\tau \cdot H$ is clearly invariant under compactly supported isotopies of H.

Given a truncation $X_0 \subset \hat{X}$, we can proceed as in Section 2.2 to associate to u^τ a map $\bar{u}^\tau : \bar{\Sigma} \to X_0$. Let $H_0 = H \cap X_0$. If we choose X_0 sufficiently large (so that H is cylindrical in its complement), then H_0 will be a submanifold with boundary $\partial H_0 = H_0 \cap \partial X_0 = V^+ \sqcup V^-$. Note that $\bar{u}^\tau \cdot H_0$ only depends on $[u] \in \pi_2(\hat{X}, \Gamma^+ \sqcup \Gamma^-)$. Moreover, we have $\bar{u}^\tau \cdot H_0 = u^\tau \cdot H$; indeed, if X_0 is sufficiently large, then the intersections of $\bar{u}^\tau$ with H_0 are exactly the same as those of u^τ with H. □

5.2 Positivity of intersection

We now discuss positivity of intersection for the Siefring intersection number. Given a contact manifold $(Y, \xi = \ker\lambda)$ and an almost complex structure J on ξ, we adopt the usual convention of letting $\hat{J}$ denote the induced almost complex structure on the symplectization. An almost complex structure on a cobordism $(\hat{X}, \hat{\lambda})$ between two contact manifolds $(Y^\pm, \lambda^\pm)$ is called *cylindrical* if it agrees at infinity with $\hat{J}^\pm$ for some choice of $d\lambda^\pm$–compatible almost complex structures $J^\pm$ on $\ker(\lambda^\pm)$.

Proposition 5.3 [54, Corollary 2.3 and Theorem 2.5] *Let $(\hat{X}, \hat{\lambda})$ be an exact symplectic cobordism from (Y^+, λ^+) to (Y^-, λ^-). Let u and H denote an asymptotically cylindrical map and a cylindrical submanifold of codimension 2 in $\hat{X}$, respectively.*

Suppose that u and H are $\hat{J}$–holomorphic for some cylindrical almost complex structure $\hat{J}$ on $\hat{X}$ which is compatible with $d\hat{\lambda}$. If the image of u is not contained in H, then $\mathrm{Im}(u) \cap H$ is a finite set and

$$u * H \geq u \cdot H. \tag{5-3}$$

(Note that, by ordinary positivity of intersection for two pseudoholomorphic submanifolds, this implies that $u * H \geq 0$, and that $\mathrm{Im}(u)$ and H are disjoint if $u * H = 0$.)

When the image of u is contained in H, positivity of intersection does not hold. The following computation, which will be useful to us later, is one example of this. The notation $\hat{\gamma}$ refers to the trivial cylinder $\mathbb{R} \times S^1 \to \hat{Y}$ over the Reeb orbit γ; similarly, $\hat{V} = \mathbb{R} \times V \subset \hat{Y}$ is the cylinder over the strong contact submanifold V.

Corollary 5.4 *Let γ be a Reeb orbit in Y. If γ is contained in V, then*

$$\hat{\gamma} * \hat{V} = -p_N(\gamma). \tag{5-4}$$

Proof By definition,

$$\hat{\gamma} * \hat{V} = \hat{\gamma}^\tau \cdot \hat{V} + \alpha_N^{\tau;-}(\gamma) - \alpha_N^{\tau;+}(\gamma). \tag{5-5}$$

We can choose the perturbation $\hat{\gamma}^\tau$ so that its image is disjoint from $\hat{V}$. The result follows since $\alpha_N^{\tau;+}(\gamma) - \alpha_N^{\tau;-}(\gamma) = p_N(\gamma)$ by definition. □

Remark 5.5 If γ is disjoint from V, then $\hat{\gamma} * \hat{V} = 0$.

Corollary 5.4 shows that positivity of intersection fails for curves contained in $\widehat{V}$. However, we still have a lower bound on the intersection number $u * \widehat{V}$ when $u = \widehat{\gamma}$ is a trivial cylinder, namely, $\widehat{\gamma} * \widehat{V} \geq -1$ since $p_N(\gamma) \in \{0, 1\}$. In the remainder of this section, we show that if (Y, ξ, V) is a TN contact pair and λ is positive elliptic near V, then the intersection number $u * \widehat{V}$ is bounded below for *all* asymptotically cylindrical curves u contained in $\widehat{V}$. We also give an analogous result for cylindrical submanifolds of symplectic cobordisms $H \subset (\widehat{X}, \omega)$ with trivial normal bundle.

Proposition 5.6 *Fix a TN contact pair (Y, ξ, V) and a datum $\mathfrak{r} = (\alpha_V, \tau, r) \in \mathfrak{R}(Y, \xi, V)$. Consider a contact form λ on (Y, ξ) which is adapted to $\mathfrak{r}$, and an almost complex structure J on ξ which is compatible with $d\lambda$ and which preserves ξ_V. Suppose that u is a $\widehat{J}$–holomorphic curve whose image is entirely contained in $\widehat{V}$. If λ is positive elliptic near V, then $u * \widehat{V} \geq 1 - p_u$, where p_u denotes the number of punctures (positive and negative) of u.*

Proof We have by definition that

$$\alpha_N^{\tau;-}(\gamma_z) = \lfloor \mathrm{CZ}_N^{\tau}(\gamma_z)/2 \rfloor = \lfloor r P_z \rfloor \qquad \text{for } z \in \boldsymbol{p}_u^+, \tag{5-6}$$

$$\alpha_N^{\tau;+}(\gamma_z) = \lceil \mathrm{CZ}_N^{\tau}(\gamma_z)/2 \rceil = 1 + \lfloor r P_z \rfloor \quad \text{for } z \in \boldsymbol{p}_u^-, \tag{5-7}$$

where P_z denotes the period of the Reeb orbit γ_z. Using the trivial bounds $x - 1 < \lfloor x \rfloor \leq x$ and the fact that $u^{\tau} \cdot \widehat{V} = 0$, we obtain

$$u * \widehat{V} > \sum_{z \in \boldsymbol{p}_u^+} (r P_z - 1) - \sum_{z \in \boldsymbol{p}_u^-} (1 + r P_z) \geq -p_u + r\Biggl(\sum_{z \in \boldsymbol{p}_u^+} P_z - \sum_{z \in \boldsymbol{p}_u^-} P_z \Biggr). \tag{5-8}$$

The fact that u is $\widehat{J}$–holomorphic implies that $\sum_{z \in \boldsymbol{p}_u^+} P_z - \sum_{z \in \boldsymbol{p}_u^-} P_z$ is nonnegative; see [68, page 60]. Thus $u * \widehat{V} \geq 1 - p_u$, as desired. □

We will need an analog of Proposition 5.6 for cobordisms. Note that if $V \subset Y$ is a codimension 2 contact submanifold, then the normal bundle of $\widehat{V} = \mathbb{R} \times V \subset \mathbb{R} \times Y = \widehat{Y}$ can be identified with the pullback of $\xi|_V^{\perp}$ under the projection $\widehat{V} \to V$. Hence, any trivialization τ of $\xi|_V^{\perp}$ induces a trivialization of the normal bundle of $\widehat{V}$, which we will denote by $\widehat{\tau}$.

Proposition 5.7 *Fix TN contact pairs $(Y^{\pm}, \xi^{\pm}, V^{\pm})$ and elements $\mathfrak{r}^{\pm} = (\alpha_V^{\pm}, \tau^{\pm}, r^{\pm}) \in \mathfrak{R}(Y^{\pm}, \xi^{\pm}, V^{\pm})$. Let $\lambda^{\pm}$ be contact forms on $(Y^{\pm}, \xi^{\pm})$ which are adapted to $\mathfrak{r}^{\pm}$, and let $(\widehat{X}, \widehat{\lambda}, H)_{\lambda^-}^{\lambda^+}$ be a strong relative symplectic cobordism from (Y^+, ξ^+, V^+) to (Y^-, ξ^-, V^-). We assume that there exists a global trivialization τ of the normal bundle of H which coincides with $\widehat{\tau}^{\pm}$ near $\pm\infty$.*

Let $\widehat{J}$ be an almost complex structure on $\widehat{X}$ which is cylindrical and compatible with $d\lambda^{\pm}$ outside a compact set, and such that H is $\widehat{J}$–holomorphic. Let u be an asymptotically cylindrical map in $\widehat{X}$ which is $\widehat{J}$–holomorphic and whose image is entirely contained in H. Following the notation of Proposition 5.6, if λ^+ is positive elliptic near V^+, then $u * H > -p_u + r^+ \sum_{z \in \boldsymbol{p}_u^+} P_z - r^- \sum_{z \in \boldsymbol{p}_u^-} P_z$. In particular, $u * H \geq 0$ if u has no negative puncture.

Proof We have

$$u * H = u^\tau \cdot H + \sum_{z \in \boldsymbol{p}_u^+} \alpha_N^{\tau;-}(\gamma_z) - \sum_{z \in \boldsymbol{p}_u^-} \alpha_N^{\tau;+}(\gamma_z). \tag{5-9}$$

We can choose the perturbation u^τ so that it is disjoint from H, so $u^\tau \cdot H = 0$. We now argue as in Proposition 5.6 to find that

$$u * H > \sum_{z \in \boldsymbol{p}_u^+} (r^+ P_z - 1) - \sum_{z \in \boldsymbol{p}_u^-} (1 + r^- P_z) \geq -p_u + r^+ \sum_{z \in \boldsymbol{p}_u^+} P_z - r^- \sum_{z \in \boldsymbol{p}_u^-} P_z. \tag{5-10}$$ □

5.3 The intersection number for buildings

In this section, we use Siefring's intersection theory to define an intersection number for buildings of asymptotically cylindrical maps and buildings of asymptotically cylindrical codimension 2 submanifolds. Since the differences between $\mathcal{S}_{\mathrm{I}}$, $\mathcal{S}_{\mathrm{II}}$, $\mathcal{S}_{\mathrm{III}}$ and $\mathcal{S}_{\mathrm{IV}}$ don't matter for this purpose, we start by defining a category $\widehat{\mathcal{S}}$ of labeled trees which only keeps track of the information needed for intersection theory; in particular, there are obvious "forgetful" functors $\mathcal{S}_* \to \widehat{\mathcal{S}}$.

The category $\widehat{\mathcal{S}} = \widehat{\mathcal{S}}(\{\widehat{X}^{ij}\}_{ij})$ depends on the following data:

(i) An integer $m \geq 0$ and a collection of $m+1$ co-oriented contact manifolds (Y^i, ξ^i), each equipped with a choice of contact form λ^i for $0 \leq i \leq m$.

(ii) For each pair of integers $0 \leq i \leq j \leq m$, an exact symplectic cobordism $(\widehat{X}^{ij}, \widehat{\lambda}^{ij})$ with positive end (Y^i, ξ^i) and negative end (Y^j, ξ^j). We require that $\widehat{X}^{ii} = SY^i$ be the symplectization of Y^i and that $\widehat{X}^{ik} = \widehat{X}^{ij} \# \widehat{X}^{jk}$ for $i \leq j \leq k$; this makes sense in light of Remark 2.8.

An object $T \in \widehat{\mathcal{S}}$ is a finite directed forest, ie a finite collection of finite directed trees. We require that every vertex has a unique incoming edge. Edges which are adjacent to only one vertex are allowed; we will refer to them as input or output edges depending on whether they are missing a source or a sink. The other edges will be called interior edges. We also have the following decorations:

- For each edge $e \in E(T)$, a symbol $*(e) \in \{0, \dots, m\}$ such that $*(e) = 0$ for input edges and $*(e) = m$ for output edges, together with a Reeb orbit γ_e in $(Y^{*(e)}, \lambda^{*(e)})$.
- For each vertex $v \in V(T)$, a pair $*(v) = (*^+(v), *^-(v)) \in \{0, \dots, m\}^2$ such that $*^+(v) \leq *^-(v)$, and a homotopy class $\beta_v \in \pi_2(\widehat{X}^{*(v)}, \gamma_{e^+(v)} \sqcup \{\gamma_{e^-}\}_{e^- \in E^-(v)})$, where $e^+(v)$ denotes the unique incoming edge of v and $E^-(v)$ denotes its set of outgoing edges. We require that $*(e^+(v)) = *^+(v)$ and $*(e^-) = *^-(v)$ for every $e^- \in E^-(v)$.

Remark 5.8 Geometrically, these decorations specify how different curves and orbits fit together to form a holomorphic building. For example, suppose $m = 1$ and T is a tree with one vertex v and one input edge e. If $*(e) = 0$ and $*(v) = 00$, then T describes a curve in SY^0 with one positive puncture and no negative punctures. If $*(e) = 0$ and $*(v) = 01$, then T describes a curve in $\widehat{X}^{01}$ with one positive puncture and no negative punctures. This labeling scheme of course follows [61, Section 2.1].

We will let Γ_T^+ and Γ_T^- denote the collections of Reeb orbits associated to the input and output edges of an object $T \in \hat{\mathcal{S}}$. In the case where T is a tree, the unique element of Γ_T^+ will be denoted by γ_T^+.

A morphism $\pi : T \to T'$ consists of a contraction of the underlying forests (meaning that T' is identified with the forest obtained by contracting a certain subset of the interior edges of T) subject to the following conditions:

- For every noncontracted edge $e \in E(T)$, we require that $*(\pi(e)) = *(e)$ and $\gamma_{\pi(e)} = \gamma_e$.
- For every vertex $v \in V(T)$, we have $*^+(\pi(v)) \leq *^+(v)$ and $*^-(\pi(v)) \geq *^-(v)$.
- For every vertex $v' \in V(T')$, we require that $\beta_{v'} = \#_{\pi(v)=v'} \beta_v$.

Note that for any morphism $T \to T'$, we have $\Gamma_T^+ = \Gamma_{T'}^+$ and $\Gamma_T^- = \Gamma_{T'}^-$.

Remark 5.9 For every $T \in \hat{\mathcal{S}}$, we get a morphism $T \to T_{\max}$ by contacting all of the interior edges of T. Each component of $T_{\max}$ is a tree with a unique vertex. In the case where T is connected, we will write $\beta_T := \#_v \beta_v \in \pi_2(\hat{X}^{0m}, \gamma_T^+ \sqcup \Gamma_T^-)$ for the homotopy class labeling the unique vertex of $T_{\max}$. Note that for every morphism $T \to T'$, we have $T_{\max} = T'_{\max}$. In particular, if T and T' are trees, then $\beta_T = \beta_{T'}$.

Definition 5.10 Let $T \in \hat{\mathcal{S}}$ and let $\{T_i\}_i$ denote its connected components. The *intersection number* $T * H$ of T with a codimension 2 cylindrical submanifold $H \subset \hat{X}^{0m}$ is defined to be

$$T * H = \sum_i \beta_{T_i} * H. \tag{5-11}$$

By Proposition 5.2, this intersection number only depends on the class of H in $\Omega_{2n-2}(\hat{X}^{0m}, V^0 \sqcup V^m)$. By Remark 5.9, it is "invariant under gluing":

Proposition 5.11 *Let $T, T' \in \hat{\mathcal{S}}$. If there exists a morphism $T \to T'$, then $T * H = T' * H$.*

Suppose now that $m = 0$, so that objects $T \in \hat{\mathcal{S}}$ represent buildings of curves in the symplectization $\hat{Y}$ of a single contact manifold $(Y, \lambda) := (Y^0, \lambda^0)$, and that $H = \hat{V} := \mathbb{R} \times V$ is the trivial cylinder over some strong contact submanifold $V \subset Y$ of codimension 2. In that case, the intersection number $T * H$ can be expressed more explicitly as follows.

Proposition 5.12 *For any $T \in \hat{\mathcal{S}}$, we have*

$$T * \hat{V} = \sum_{v \in V(T)} \beta_v * \hat{V} - \sum_{e \in E^{\text{int}}(T)} \hat{\gamma}_e * \hat{V}. \tag{5-12}$$

Proof The proof will be by induction on the number of interior edges. If this number is zero, then (5-12) is true by definition. Otherwise, pick an edge $e \in E^{\text{int}}(T)$ and contract it to obtain a morphism $\pi : T \to T'$ where T' has one less interior edge than T. We can assume inductively that T' satisfies (5-12). Since $T * \hat{V} = T' * \hat{V}$, it suffices to show that

$$\beta_{v^+} * \hat{V} + \beta_{v^-} * \hat{V} - \hat{\gamma}_e * \hat{V} = \beta_{v'} * \hat{V}, \tag{5-13}$$

where v^+ and v^- are the source and sink of e, respectively, and $v' = \pi(v^+) = \pi(v^-)$.

To do this, start by picking curves $u_\pm\colon \Sigma^\pm \to \widehat{Y}$ representing the classes $\beta_{v^\pm}$. Fix a choice of cylindrical coordinates near the positive puncture of u^- and near the negative puncture of u^+ corresponding to e. We can assume that $u_\pm$ is cylindrical at infinity, so that there exists a constant $C > 0$ such that

$$u_\pm(s,t) = (P_e \cdot s, \gamma_e(t)) \quad \text{for } \mp s \geq C, \tag{5-14}$$

where P_e is the period of the orbit corresponding to e. Now let

$$\overline{\Sigma}^+ = \Sigma^+ \setminus ((-\infty, -3C) \times S^1), \tag{5-15}$$

$$\overline{\Sigma}^- = \Sigma^- \setminus ((3C, \infty) \times S^1), \tag{5-16}$$

and let $\Sigma = \Sigma^+ \# \Sigma^-$ be obtained by identifying $[-3C, -C] \times S^1 \subset \overline{\Sigma}^+$ with $[C, 3C] \times S^1 \subset \Sigma^-$ via translation by $4C$. The curve $u_+ \# u_-\colon \Sigma \to \widehat{Y}$ which is given by $\tau_{2CP_e} \circ u_+$ on $\overline{\Sigma}^+$ and $\tau_{-2CP_e} \circ u_-$ on $\overline{\Sigma}^-$ (where $\tau_s\colon \widehat{Y} \to \widehat{Y}$ denotes translation by s) then represents the homotopy class $\beta_{v^+} \# \beta_{v^-} = \beta_{v'}$. Choose a trivialization τ of $\xi|_V^\perp$ along the relevant Reeb orbits and use it to produce perturbations $u_\pm^\tau$ and $(u_+ \# u_-)^\tau$ as in Section 5.1. We can do this in such a way that $(u_+ \# u_-)^\tau$ is obtained by gluing u_+^τ and u_-^τ. Then

$$(u_+ \# u_-)^\tau \cdot \widehat{V} = u_+^\tau \cdot \widehat{V} + u_-^\tau \cdot \widehat{V}, \tag{5-17}$$

so

$$\begin{aligned}(u_+ \# u_-) * \widehat{V} &= u_+^\tau \cdot \widehat{V} + u_-^\tau \cdot \widehat{V} + \alpha_N^{\tau;-}(\gamma^+) - \sum_{z \in \boldsymbol{p}^-_{u_+ \# u_-}} \alpha_N^{\tau;+}(\gamma_z) \\ &= u_+^\tau \cdot \widehat{V} + u_-^\tau \cdot \widehat{V} + \alpha_N^{\tau;-}(\gamma^+) + \alpha_N^{\tau;+}(\gamma_e) - \sum_{z \in \boldsymbol{p}^-_{u_+}} \alpha_N^{\tau;+}(\gamma_z) - \sum_{z \in \boldsymbol{p}^-_{u_-}} \alpha_N^{\tau;+}(\gamma_z) \\ &= u_+ * \widehat{V} + u_- * \widehat{V} + \alpha_N^{\tau;+}(\gamma_e) - \alpha_N^{\tau;-}(\gamma_e) \\ &= u_+ * \widehat{V} + u_- * \widehat{V} + p_N(\gamma_e).\end{aligned}$$

We have $p(\gamma_e) = -\widehat{\gamma}_e * \widehat{V}$ by Corollary 5.4, so this implies that

$$\beta_{v'} * \widehat{V} = (u_+ \# u_-) * \widehat{V} = u_+ * \widehat{V} + u_- * \widehat{V} - \widehat{\gamma}_e * \widehat{V} = \beta_{v^+} * \widehat{V} + \beta_{v^-} * \widehat{V} - \widehat{\gamma}_e * \widehat{V}, \tag{5-18}$$

as desired. □

Definition 5.13 Given $T \in \widehat{\mathcal{S}}$, we say that T is *representable by a holomorphic building* if there exists a $d\lambda$–compatible almost complex structure J on ξ such that, for every vertex $v \in V(T)$, the homotopy class $\beta_v \in \pi_2(\widehat{Y}, \gamma_{e^+(v)} \sqcup \{\gamma_{e^-}\}_{e^- \in E^-(v)})$ admits a $\widehat{J}$–holomorphic representative. We say that T is representable by a $\widehat{J}$–holomorphic building if we wish to specify $\widehat{J}$.

Corollary 5.14 *Let $T \in \widehat{\mathcal{S}}$. Suppose that there exists a morphism $T' \to T$ and a $d\lambda$–compatible almost complex structure J on ξ such that T' is representable by a $\widehat{J}$–holomorphic building. Suppose also that $\widehat{V}$ is $\widehat{J}$–holomorphic. If λ is positive elliptic near V, then $T * \widehat{V} \geq -\Gamma^-(T, V)$, where $\Gamma^-(T, V)$ denotes the number of output edges e of T such that γ_e is contained in V.*

Proof By Proposition 5.11, $T * \widehat{V} = T' * \widehat{V}$. By Proposition 5.12,

$$T' * \widehat{V} = \sum_{v \in V(T')} \beta_v * \widehat{V} - \sum_{e \in E^{\mathrm{int}}(T')} \widehat{\gamma}_e * \widehat{V}. \tag{5-19}$$

According to Proposition 5.3, we have $\beta_v * \widehat{V} \geq 0$, unless the holomorphic representative of β_v is entirely contained in $\widehat{V}$, in which case Proposition 5.6 tells us that $\beta_v * \widehat{V} \geq -\#E^-(v)$. By Corollary 5.4, we have

$$-\#E^-(v) = \sum_{e \in E^-(v)} \widehat{\gamma}_e * \widehat{V}. \tag{5-20}$$

Given $v \in V(T')$, let us denote by $\Gamma^-(v, V)$ the number of output edges $e \in E^-(v)$ such that $\gamma_e \subset V$. Appealing again to Proposition 5.12, we have

$$\begin{aligned} T'^{*} \widehat{V} &= \sum_{v \in V(T')} \beta_v * \widehat{V} - \sum_{e \in E^{\mathrm{int}}(T')} \widehat{\gamma}_e * \widehat{V} \\ &= \sum_{v \in V(T')} \Bigl(\beta_v * \widehat{V} - \sum_{e \in E^-(v)} \widehat{\gamma}_e * \widehat{V} \Bigr) + \sum_{e \in E^-(T)} \widehat{\gamma}_e * \widehat{V} \\ &= \sum_{v \in V(T')} (\beta_v * \widehat{V} + \Gamma^-(v, V)) - \Gamma^-(T', V) \\ &\geq -\Gamma^-(T, V), \end{aligned} \tag{5-21}$$

where we have used the fact that T and T' have the same exterior edges in the last line. This completes the proof. □

More generally, suppose we are given the following data, where m is now allowed to be any nonnegative integer:

- For each $0 \leq i \leq m$, a strong contact submanifold $V^i \subset Y^i$ of codimension 2.
- For each $0 \leq i \leq j \leq m$, a homotopy class $\eta_{ij} \in \Omega_{2n-2}(\widehat{X}^{ij}, V^i \sqcup V^j)$. We require that $\eta_{ii} := [\widehat{V}_i]$ be the homotopy class of $\widehat{V}^i = \mathbb{R} \times V^i$, and that $\eta_{ik} = \eta_{ij} \# \eta_{jk}$ for any $i \leq j \leq k$.

Let $\eta := \eta_{0m} \in \Omega_{2n-2}(\widehat{X}^{0m}, V^0 \sqcup V^m)$.

Proposition 5.15 *Let $T \in \widehat{\mathcal{S}}$. Then*

$$T * \eta = \sum_{v \in V(T)} \beta_v * \eta_{*(v)} - \sum_{e \in E^{\mathrm{int}}(T)} \widehat{\gamma}_e * \widehat{V}_{*(e)}. \tag{5-22}$$

Proof We will say a vertex $v \in V(T)$ is a *symplectization vertex* if $*(v) = ii$ for some i, and we will call it a *cobordism vertex* otherwise. This induces a partition $E^{\mathrm{int}}(T) = E^{\mathrm{ss}}(T) \sqcup E^{\mathrm{sc}}(T) \sqcup E^{\mathrm{cc}}(T)$ of the set of interior edges according to the types of the vertices they are adjacent to — here the superscripts s and c stand for "symplectization" and "cobordism", respectively. Similarly, the set of exterior edges admits a partition $E^{\mathrm{ext}}(T) = E^{\mathrm{s}}(T) \sqcup E^{\mathrm{c}}(T)$.

We can (and will) assume without loss of generality that $E^{\mathrm{ss}}(T)$ is empty. Indeed, let $T \to T'$ be the morphism obtained by contracting all the edges in $E^{\mathrm{ss}}(T)$. Replacing T with T' doesn't change the left-hand side of (5-22) by Proposition 5.11 and doesn't change the right-hand side by Proposition 5.12.

Let $I = V(T) \sqcup E^{\mathrm{c}}(T) \sqcup E^{\mathrm{cc}}(T)$ and choose a family of curves $\{u_i\}_{i \in I}$ with the following properties:

- For each $v \in V(T)$, u_v is a curve in the homotopy class β_v which is cylindrical at infinity.
- For each $e \in E^{\mathrm{c}}(T) \sqcup E^{\mathrm{cc}}(T)$, $u_e = \hat{\gamma}_e$ is the trivial cylinder over the Reeb orbit γ_e.

For $t > 0$ sufficiently large, we can glue the u_i to obtain a curve u in $X^{00}\#_t X^{01}\#_t X^{11}\#_t \cdots \#_t X^{mm} \cong X^{0m}$ representing T. We can also choose representatives H_{ij} of η_{ij} so that $H_{ii} = \hat{V}_i$ and $H := H_{0m}$ coincides with $H_{00} \#_t H_{01} \#_t \cdots \#_t H_{mm}$.

As in the proof of Proposition 5.12, we can choose perturbations u^τ and $\{u_i^\tau\}$ so that

$$u^\tau \cdot H = \sum_{i \in I} u_i^\tau \cdot H_i, \tag{5-23}$$

where $H_i := H_{*(v)}$ for $i = v \in V(T)$ and $H_i := \hat{V}_{*(e)}$ for $i = e \in E^{\mathrm{c}}(T) \sqcup E^{\mathrm{cc}}(T)$. The difference $\sum_i u_i * H_i - u * H$ is therefore equal to

$$\begin{aligned}\sum_{e \in E^{\mathrm{sc}}(T) \sqcup E^{\mathrm{c}}(T)} \alpha_N^{\tau;-}(\gamma_e) - \alpha_N^{\tau;+}(\gamma_e) + 2 \sum_{e \in E^{\mathrm{cc}}(T)} \alpha_N^{\tau;-}(\gamma_e) - \alpha_N^{\tau;+}(\gamma_e) &= \sum_{e \in E^{\mathrm{sc}}(T) \sqcup E^{\mathrm{c}}(T)} \hat{\gamma}_e * \hat{V}_{*(e)} + 2 \sum_{e \in E^{\mathrm{cc}}(T)} \hat{\gamma}_e * \hat{V}_{*(e)} \\ &= \sum_{e \in E^{\mathrm{int}}(T)} \hat{\gamma}_e * \hat{V}_{*(e)} + \sum_{e \in E^{c}(T) \sqcup E^{\mathrm{cc}}(T)} \hat{\gamma}_e * \hat{V}_{*(e)}.\end{aligned} \tag{5-24}$$

Since $u_i * H_i = \hat{\gamma}_e * \hat{V}_{*(e)}$ for $i = e \in E^{\mathrm{c}}(T) \sqcup E^{\mathrm{cc}}(T)$, we conclude that

$$u * H = \sum_{v \in V(T)} u_v * H_{*(v)} - \sum_{e \in E^{\mathrm{int}}(T)} \hat{\gamma}_e * \hat{V}_{*(e)}, \tag{5-25}$$

which implies (5-22). □

Definition 5.16 Given $T \in \hat{\mathcal{S}}$, we say that T is *representable by a holomorphic building* if for every vertex $v \in V(T)$, there exists an adapted almost complex structure $\hat{J}^v$ on $\hat{X}^{*(v)}$ such that

$$\beta_v \in \pi_2(\hat{X}^{*(v)}, \gamma_{e^+(v)} \sqcup \{\gamma_{e^-}\}_{e^- \in E^-(v)}) \quad \text{and} \quad \eta_{*(v)} \in \Omega_{2n-2}(\hat{X}^{*(v)}, V^{*^+(v)} \sqcup V^{*^-(v)})$$

admit $\hat{J}^v$–holomorphic representatives.

Proposition 5.17 *Let $T \in \hat{\mathcal{S}}$. Suppose that there exists a morphism $T' \to T$ where T' is representable by a holomorphic building. Suppose that λ^i is positive elliptic near V^i for all $0 \le i \le m$, and let $r_i > 0$ be the rotation parameter; see Definition 3.4. If $\beta_v * \eta_{*(v)} \ge -\#\{E^-(v)\}$, then $T * \eta \ge -\Gamma^-(T, V^m)$.* (Recall that $\eta := \eta_{0m}$.)

Proof By Proposition 5.11, we have $T * \eta = T' * \eta$. Given $v \in V(T')$, let us denote by $\Gamma^-(v, V^{*^-(v)})$ the number of output edges $e \in E^-(v)$ such that $\gamma_e \subset V^{*^-(v)}$.

Arguing as in the proof of Corollary 5.14, we obtain from Proposition 5.15 that

$$
\begin{aligned}
T' * \hat{V} &= \sum_{v \in V(T')} \beta_v * \eta_{*(v)} - \sum_{e \in E^{\mathrm{int}}(T')} \hat{\gamma}_e * \hat{V}_{*(e)} \\
&= \sum_{v \in V(T')} \Big(\beta_v * \eta_{*(v)} - \sum_{e \in E^-(v)} \hat{\gamma}_e * \hat{V}_{*(e)} \Big) + \sum_{e \in E^-(T')} \hat{\gamma}_e * \hat{V}_{*(e)} \\
&= \sum_{v \in V(T')} (\beta_v * \hat{V} + \#\Gamma^-(v, V^{*^-(v)})) - \Gamma^-(T', V) \\
&\geq -\Gamma^-(T, V),
\end{aligned}
\tag{5-26}
$$

where we have used the fact that T and T' have the same exterior edges in the last line. This completes the proof. □

5.4 The intersection number for cycles

For future reference, we collect some basic facts about intersection numbers for cycles in oriented manifolds. This subsection takes places entirely in the smooth category and does not involve any contact topology.

Definition 5.18 Let M be an oriented, compact manifold of dimension n, possibly with boundary. Let $S_1, S_2 \subset M$ be disjoint closed embedded submanifolds of M. (We allow the S_i to intersect ∂M, in which case the S_i are required to be embedded submanifolds after enlarging M by a collar). Then we can define a pairing

$$
- \cdot - : H_k(M, S_1; \mathbb{Z}) \times H_{n-k}(M, S_2; \mathbb{Z}) \to \mathbb{Z}, \quad (A, B) \mapsto A \cdot B,
\tag{5-27}
$$

where $A \cdot B$ is a signed count of intersections between cycles representing A and B. More precisely, we represent cycles by C^∞ chains; by general position, these chains may be assumed to intersect transversally after an arbitrarily small perturbation which does not affect their homology class. It is a folklore result, which is beyond the scope of this paper, that the resulting count is well-defined and graded-symmetric; see eg [32, Section 2.3].

We note that the intersection number in Definition 5.18 could be defined under much milder hypotheses, but this is not necessary for our purposes. If A and B are (the pushforward of the fundamental class of) oriented manifolds, then $A \cdot B$ coincides with the usual intersection number for submanifolds. By abuse of notation, we will view the intersection pairing as being defined on both homology classes and oriented submanifolds.

Definition 5.19 Fix a closed manifold Y of dimension $m \geq 3$ and a closed codimension 2 submanifold $V \subset Y$. Suppose that $H_1(Y;\mathbb{Z}) = H_2(Y;\mathbb{Z}) = 0$.

Let $\gamma\colon S^1 \to Y - V$ be a loop with embedded image. The *linking number* of γ with respect to V is denoted by $\mathrm{link}_V(\gamma)$ and defined by

$$\mathrm{link}_V(\gamma) := V \cdot C_\gamma, \tag{5-28}$$

where C_γ is a cycle bounding γ (which exists since $H_1(Y;\mathbb{Z}) = 0$). This is well-defined due to our assumption that $H_2(Y;\mathbb{Z}) = 0$.

Suppose now that $\Lambda \subset Y - V$ is a submanifold with $\pi_0(\Lambda) = \pi_1(\Lambda) = 0$. Let $c\colon [0,1] \to Y - V$ be a path with embedded image having the property that $c(0), c(1) \in \Lambda$. Let $\overline{c}\colon S^1 \to Y - V$ be a loop with embedded image obtained by connecting $c(1)$ to $c(0)$ by a path in Λ. The (*path*) *linking number* of c with respect to V is denoted by $\mathrm{link}_V(c;\Lambda)$ and is defined by setting

$$\mathrm{link}_V(c;\Lambda) := V \cdot C_{\overline{c}}, \tag{5-29}$$

where $C_{\overline{c}}$ is a cycle bounding $\overline{c}$. This is independent of $\overline{c}$ since $\pi_1(\Lambda) = 0$, and independent of $C_{\overline{c}}$ since $H_2(Y;\mathbb{Z}) = 0$.

Remark 5.20 Fix an open book decomposition (Y, B, π) and let $\gamma\colon S^1 \to Y - B$ be a loop. Then it is not hard to show that we have $\mathrm{link}_B(\gamma) = \deg(\pi \circ \gamma)$.

Similarly, suppose $\Lambda \subset Y$ is a submanifold which is contained in a page of (Y, B, π). Let $c\colon [0,1] \to Y - B$ be a path with the property that $c(0), c(1) \in \Lambda$. Then the composition $\pi \circ c\colon [0,1] \to S^1$ induces a map $\overline{c}\colon [0,1]/\{0,1\} \to S^1$. We then have $\mathrm{link}_B(c;\Lambda) = \deg \overline{c}$.

Lemma 5.21 *Let $Y^\pm$ be oriented manifolds with $Y^+ \neq \varnothing$ and let $B^\pm \subset Y^\pm$ be oriented submanifolds. Let W be an oriented, smooth cobordism from Y^+ to Y^- and let $H \subset W$ be an oriented subcobordism from B^+ to B^-, ie H is an embedded submanifold which admits a collar neighborhood near the boundary of W. Suppose that $H_1(Y^\pm;\mathbb{Z}) = H_2(Y^\pm;\mathbb{Z}) = H_2(W, Y^+;\mathbb{Z}) = 0$.*

Let Σ be a Riemann surface with $k+1$ boundary components labeled $\gamma^+, \gamma_1^-, \dots, \gamma_k^-$. Suppose that $u\colon (\Sigma, \partial\Sigma) \to (W, \partial W)$ is a smooth map sending γ^+ into $Y^+ - B^+$ and γ_i^- into $Y^- - B^-$.

Then

$$\mathrm{link}_{B^+}(\gamma^+) - \sum_{i=1}^{k} \mathrm{link}_{B^-}(\gamma_i^-) = H \cdot u(\Sigma), \tag{5-30}$$

where we have identified the boundary components of Σ with the restriction of u to these components.

Proof Choose a 2–chain $B^- \in C_2(Y^-;\mathbb{Z})$ with $\partial B^- = \gamma_1^- \cup \cdots \cup \gamma_k^-$. Glue B^- to $u(\Sigma)$ along the γ_i^- and call the resulting chain $C \in C_2(W;\mathbb{Z})$. We now have $H \cdot C = H \cdot u(\Sigma) + \sum_{i=1}^k \mathrm{link}_{B^-}(\gamma_i^-)$.

By the long exact sequence of the triple (W, Y^+, γ^+) and our assumption that $H_2(W, Y^+;\mathbb{Z}) = 0$, the natural map $H_2(Y, \gamma^+;\mathbb{Z}) \to H_2(W, \gamma^+;\mathbb{Z})$ is surjective. Let $\widetilde{C} \in H_2(Y^+, \gamma^+;\mathbb{Z})$ be a lift of $C \in H_2(W, \gamma^+;\mathbb{Z})$. Then $H \cdot C = H \cdot \widetilde{C} = B^+ \cdot \widetilde{C} = \mathrm{link}_{B^+}(\gamma^+)$. □

Lemma 5.22 *We carry over the setup and notation from Lemma 5.21. In addition to the data considered there, let $\Lambda^\pm \subset Y^\pm - B^\pm$ be an oriented, smooth submanifold and let $\Lambda \subset W$ be an oriented subcobordism from Λ^+ to Λ^- which is disjoint from H. We suppose in addition that $\pi_0(\Lambda^\pm) = \pi_1(\Lambda^\pm) = 0$.*

Let Σ be a closed, oriented surface of genus zero with $s+1$ boundary components labeled $\gamma^*, \gamma_1, \dots, \gamma_n$. For $\sigma \in \mathbb{N}_+$, we place 2σ disjoint marked points on γ^*, thus partitioning γ^* into 2σ subintervals. Let us label these subintervals by the symbols $c^+, b_{01}, c_1^-, b_{12}, c_2^-, \dots, b_{(\sigma-1)\sigma}, c_\sigma^-, b_{\sigma 0}$, in the order induced by the orientation.

Suppose now that $u\colon (\Sigma, \partial\Sigma) \to (W, \partial W \cup \Lambda)$ is a smooth map sending $(c^+, \partial c^+)$ into $(Y^+ - B^+, \Lambda^+)$, sending $(c_i^-, \partial c_i^-)$ into $(Y^- - B^-, \Lambda^-)$, sending $b_{i(i+1)}$ into Λ, and sending the γ_i^- into $Y^- - B^-$.

Then

$$\mathrm{link}_{B^+}(c^+;\Lambda^+) - \sum_{i=1}^{\sigma} \mathrm{link}_{B^-}(c_i^-;\Lambda^-) - \sum_{i=1}^{s} \mathrm{link}_{B^-}(\gamma_i^-) = H \cdot u_*[\Sigma], \tag{5-31}$$

where we have again identified the boundary components of Σ with the restriction of u to these components. □

6 Energy and twisting maps

6.1 Twisting maps

In order to define invariants of codimension 2 contact submanifolds, we will proceed as follows. First, we will use Siefring's intersection theory to define maps $\psi\colon \mathcal{S} \to \mathcal{R}$. Here $\mathcal{R}$ could be any $\mathbb{Q}$–algebra, though we will only use $\mathcal{R} = \mathbb{Q}[U]$ and $\mathcal{R} = \mathbb{Q}$. We will then use these maps to define "twisted" moduli counts

$$\#_\psi \overline{\mathcal{M}}(T)^{\mathrm{vir}} := \#\overline{\mathcal{M}}(T)^{\mathrm{vir}} \cdot \psi(T) \in \mathcal{R}. \tag{6-1}$$

The maps ψ will have the property that

$$\psi(T') = \psi(T) \quad \text{for every morphism } T' \to T \text{, and} \tag{6-2}$$

$$\psi(\#_i T_i) = \prod_i \psi(T_i), \tag{6-3}$$

which implies that the *master equations* (4-2) and (4-3) still hold if $\#\overline{\mathcal{M}}^{\mathrm{vir}}$ is replaced by $\#_\psi \overline{\mathcal{M}}^{\mathrm{vir}}$.

The properties which must be satisfied by the maps ψ in order to obtain twisted counts which are suitable for defining our invariants can be conveniently axiomatized in the notion of a *twisting map*. We now define precisely this notion in each of the four setups.

Setup I Fix a datum $\mathcal{D}$ for Setup I. Let $\mathcal{S}_{\mathrm{I}}^{\neq\varnothing}$ denote the full subcategory of $\mathcal{S}_{\mathrm{I}}$ spanned by objects T for which the moduli space $\overline{\mathcal{M}}(T)$ is nonempty.

Definition 6.1 Let $\mathcal{R}$ be a $\mathbb{Q}$–algebra. The set $\Psi_{\mathrm{I}}(\mathcal{D};\mathcal{R})$ of $\mathcal{R}$–valued *twisting maps* consists of all maps $\psi\colon \mathcal{S}_{\mathrm{I}}^{\neq\varnothing}(\mathcal{D}) \to \mathcal{R}$ satisfying the following two properties:

- For any morphism $T' \to T$, $\psi(T') = \psi(T)$.
- For any concatenation $\{T_i\}_i$, $\psi(\#_i T_i) = \prod_i \psi(T_i)$.

Fix a twisting map $\psi \in \Psi_{\mathrm{I}}(\mathcal{D};\mathcal{R})$. Let

$$CC_\bullet(Y,\xi,\psi)_\lambda := \bigoplus_{n\geq 0} \mathrm{Sym}^n_{\mathcal{R}}\Bigl(\bigoplus_{\gamma\in\mathcal{P}_{\mathrm{good}}} \mathfrak{o}_\gamma\Bigr) \tag{6-4}$$

be the free supercommutative $\mathbb{Z}/2$–graded unital $\mathcal{R}$–algebra generated by the good Reeb orbits. The grading of a Reeb orbit is given by its *parity*, which is defined as

$$|\gamma| = \operatorname{sign}\det(I - A_\gamma) \in \{\pm 1\} = \mathbb{Z}/2, \tag{6-5}$$

where A_γ is the linearized Poincaré return map of ξ along γ; see [61, Section 2.13]. Recall that a Reeb orbit is good if and only if it is not bad; a Reeb orbit γ is bad if it is an even multiple of some simple Reeb orbit γ_s such that γ and γ_s have different parity [61, Definition 2.49].

Theorem 1.1 of [61] provides a set of perturbation data $\Theta_{\mathrm{I}}(\mathcal{D})$ and associated virtual moduli counts $\#\overline{\mathcal{M}}_{\mathrm{I}}(T)^{\mathrm{vir}}_\theta \in \mathbb{Q}$ satisfying (4-2) and (4-3). We define the twisted moduli counts

$$\#_\psi \overline{\mathcal{M}}_{\mathrm{I}}(T)^{\mathrm{vir}}_\theta := \#\overline{\mathcal{M}}_{\mathrm{I}}(T)^{\mathrm{vir}}_\theta \cdot \psi(T) \in \mathcal{R}. \tag{6-6}$$

It follows easily from Definition 6.1 that the twisted moduli counts also satisfy (4-2) and (4-3). We may therefore endow $CC_\bullet(Y,\xi,\psi)_\lambda$ with a differential $d_{\psi,J,\theta}$ which is given by

$$d_{\psi,J,\theta}(\mathfrak{o}_{\gamma^+}) = \sum_{\mu(T)=1} \frac{1}{|\mathrm{Aut}(T)|} \cdot \#_\psi \overline{\mathcal{M}}_{\mathrm{I}}(T)^{\mathrm{vir}}_\theta \mathfrak{o}_{\Gamma^-}, \tag{6-7}$$

where the sum is over all trees $T \in \mathcal{S}_I(\mathcal{D})$ representing curves with positive orbit γ^+ and negative orbits $\Gamma^- \to \mathcal{P}_{\mathrm{good}}$.

The homology of $(CC_\bullet(Y,\xi,\psi)_\lambda, d_{\psi,J,\theta})$ is a supercommutative $\mathbb{Z}/2$–graded unital $\mathcal{R}$–algebra, which is denoted by

$$CH_\bullet(Y,\xi,\psi)_{\lambda,J,\theta}. \tag{6-8}$$

Setup II Fix a datum $\mathcal{D}$ for Setup II. Suppose now we are given a map of $\mathbb{Q}$–algebras $m\colon \mathcal{R}^+ \to \mathcal{R}^-$ and twisting maps $\psi^\pm \in \Psi_{\mathrm{I}}(\mathcal{D}^\pm; \mathcal{R}^\pm)$.

Definition 6.2 The set $\Psi_{\mathrm{II}}(\mathcal{D}; \psi^+, \psi^-)$ consists of all maps $\psi\colon \mathcal{S}_{\mathrm{II}}^{\neq\varnothing}(\mathcal{D}) \to \mathcal{R}^-$ satisfying the following two properties:

- For any morphism $T' \to T$, $\psi(T') = \psi(T)$.
- For any concatenation $\{T_i\}_i$,

$$\psi(\#_i T_i) = \Bigl(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^+} m(\psi^+(T_i))\Bigr)\Bigl(\prod_{T_i \in \mathcal{S}_{\mathrm{II}}} \psi(T_i)\Bigr)\Bigl(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^-} \psi^-(T_i)\Bigr). \tag{6-9}$$

Fix a twisting map $\psi \in \Psi_{\mathrm{II}}(\mathcal{D}; \psi^+, \psi^-)$. Theorem 1.1 of [61] provides a set of perturbation data $\Theta_{\mathrm{II}}(\mathcal{D})$ together with a forgetful map

$$\Theta_{\mathrm{II}}(\mathcal{D}) \to \Theta_{\mathrm{I}}(\mathcal{D}^+) \times \Theta_{\mathrm{I}}(\mathcal{D}^-) \tag{6-10}$$

and associated virtual moduli counts $\#\overline{\mathcal{M}}_{\mathrm{II}}(T)^{\mathrm{vir}}_\theta \in \mathbb{Q}$. We define the twisted moduli counts

$$\#_\psi \overline{\mathcal{M}}_{\mathrm{II}}(T)^{\mathrm{vir}}_\theta := \#\overline{\mathcal{M}}_{\mathrm{II}}(T)^{\mathrm{vir}}_\theta \cdot \psi(T) \in \mathcal{R}^-. \tag{6-11}$$

For any $\theta \in \Theta_{\mathrm{II}}(\mathcal{D})$ mapping to $(\theta^+, \theta^-) \in \Theta_{\mathrm{I}}(\mathcal{D}^+) \times \Theta_{\mathrm{I}}(\mathcal{D}^-)$, we obtain a unital $\mathcal{R}^+$–algebra map

$$\Phi(\hat{X}, \hat{\lambda}, \psi)_{\hat{J},\theta}\colon CC_\bullet(Y^+, \xi^+, \psi^+)_{\lambda^+, J^+, \theta^+} \to CC_\bullet(Y^-, \xi^-, \psi^-)_{\lambda^-, J^-, \theta^-}, \tag{6-12}$$

which maps $\mathfrak{o}_{\gamma^+}$ to

$$\sum_{\mu(T)=0} \frac{1}{|\mathrm{Aut}(T)|} \cdot \#_\psi \overline{\mathcal{M}}_{\mathrm{II}}(T)^{\mathrm{vir}}_\theta \mathfrak{o}_{\Gamma^-}, \tag{6-13}$$

with the sum over all trees $T \in \mathcal{S}_{\mathrm{II}}(\mathcal{D})$ representing curves with positive orbit γ^+ and negative orbits $\Gamma^- \to \mathcal{P}_{\mathrm{good}}(Y^-)$. This is a chain map, since it follows from Definition 6.2 that the twisted moduli counts satisfy (4-2) and (4-3).

Setup III Fix a datum $\mathcal{D}$ for Setup III. There are three types of concatenations $\{T_i\}_i$ in $\mathcal{S}_{\mathrm{III}} = \mathcal{S}_{\mathrm{III}}(\mathcal{D})$:

(1) $\{T_i\} \subset \mathcal{S}_{\mathrm{I}}^+ \sqcup \mathcal{S}_{\mathrm{II}}^{t=0} \sqcup \mathcal{S}_{\mathrm{I}}^-$, in which case $\mathfrak{s}(\#_i T_i) = \{0\}$.

(2) $\{T_i\} \subset \mathcal{S}_{\mathrm{I}}^+ \sqcup \mathcal{S}_{\mathrm{II}}^{t=1} \sqcup \mathcal{S}_{\mathrm{I}}^-$, in which case $\mathfrak{s}(\#_i T_i) = \{1\}$.

(3) $\{T_i\} \subset \mathcal{S}_{\mathrm{I}}^+ \sqcup \mathcal{S}_{\mathrm{III}} \sqcup \mathcal{S}_{\mathrm{I}}^-$ and $T_i \in \mathcal{S}_{\mathrm{III}}$ for a unique $i = i_0$, in which case $\mathfrak{s}(\#_i T_i) = \mathfrak{s}(T_{i_0})$.

(Here $\mathcal{S}_{\mathrm{I}}^\pm, \mathcal{S}_{\mathrm{II}}^{t\in\{0,1\}}, \mathcal{S}_{\mathrm{III}}$ are tree categories determined by $\mathcal{D}$, following the notation of [61, Section 2.1].) Suppose now we are given a map of $\mathbb{Q}$–algebras $m\colon \mathcal{R}^+ \to \mathcal{R}^-$ and twisting maps $\psi^\pm \in \Psi_{\mathrm{I}}(\mathcal{D}^\pm; \mathcal{R}^\pm)$, $\psi^0 \in \Psi_{\mathrm{II}}(\mathcal{D}^{t=0}; \psi^+, \psi^-)$ and $\psi^1 \in \Psi_{\mathrm{II}}(\mathcal{D}^{t=1}; \psi^+, \psi^-)$.

Definition 6.3 The set $\Psi_{\mathrm{III}}(\mathcal{D}; \psi^0, \psi^1)$ consists of all maps $\psi : \mathcal{S}_{\mathrm{III}}^{\neq\varnothing}(\mathcal{D}) \to \mathcal{R}^-$ satisfying the following properties:

- For any morphism $T' \to T$, $\psi(T') = \psi(T)$.
- For any concatenation $\{T_i\}_i$ of the first type,

$$\psi(\#_i T_i) = \Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^+} m(\psi^+(T_i))\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{II}}^{t=0}} \psi^0(T_i)\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^-} \psi^-(T_i)\Big). \tag{6-14}$$

- For any concatenation $\{T_i\}_i$ of the second type,

$$\psi(\#_i T_i) = \Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^+} m(\psi^+(T_i))\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{II}}^{t=1}} \psi^1(T_i)\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^-} \psi^-(T_i)\Big). \tag{6-15}$$

- For any concatenation $\{T_i\}_i$ of the third type,

$$\psi(\#_i T_i) = \Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^+} m(\psi^+(T_i))\Big)\psi(T_{i_0})\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^-} \psi^-(T_i)\Big). \tag{6-16}$$

Fix a twisting map $\psi \in \Psi_{\mathrm{III}}(\mathcal{D}; \psi^0, \psi^1)$. Theorem 1.1 of [61] provides a set of perturbation data $\Theta_{\mathrm{III}}(\mathcal{D})$ together with a forgetful map $\Theta_{\mathrm{III}}(\mathcal{D}) \to \Theta_{\mathrm{II}}(\mathcal{D}^0) \times_{\Theta_{\mathrm{I}}(\mathcal{D}^+) \times \Theta_{\mathrm{I}}(\mathcal{D}^-)} \Theta_{\mathrm{II}}(\mathcal{D}^1)$ and associated virtual moduli counts $\#\overline{\mathcal{M}}_{\mathrm{III}}(T)_\theta^{\mathrm{vir}} \in \mathbb{Q}$; note that the fiber product is defined with respect to (6-10). We define the twisted moduli counts

$$\#_\psi \overline{\mathcal{M}}_{\mathrm{III}}(T)_\theta^{\mathrm{vir}} := \#\overline{\mathcal{M}}_{\mathrm{III}}(T)_\theta^{\mathrm{vir}} \cdot \psi(T) \in \mathcal{R}^-. \tag{6-17}$$

If $(\hat{X}, \hat{\lambda}^t)$ is a family of exact cobordisms, then for any $\theta \in \Theta_{\mathrm{III}}(\mathcal{D})$, we obtain an $\mathcal{R}^+$–linear map

$$K(\hat{X}, \{\lambda^t\}_t, \psi)_{\hat{J}^t,\theta} : CC_\bullet(Y^+, \xi^+, \psi^+)_{\lambda^+, J^+, \theta^+} \to CC_{\bullet+1}(Y^-, \xi^-, \psi^-)_{\lambda^-, J^-, \theta^-} \tag{6-18}$$

which sends the monomial $\prod_{i \in I} \mathfrak{o}_{\gamma_i^+}$ to

$$\sum_{\mathrm{vdim}(\{T_i\}_{i\in I})=0} \frac{1}{|\mathrm{Aut}(\{T_i\}_{i\in I})|} \cdot \#_\psi \overline{\mathcal{M}}_{\mathrm{III}}(\{T_i\}_{i\in I})_\theta^{\mathrm{vir}} \prod_{i\in I} \mathfrak{o}_{\Gamma_i^-}, \tag{6-19}$$

with the sum over trees $T_i \in \mathcal{S}_{\mathrm{III}}(\mathcal{D})$ with positive orbit γ_i^+ and negative orbits $\Gamma_i^- \to \mathcal{P}_{\mathrm{good}}(Y^-)$.

Equations (4-2) and (4-3) applied to the twisted moduli counts imply that this is a chain homotopy between $\Phi(\hat{X}, \hat{\lambda}^0, \psi^0)_{\hat{J}^0,\theta^0}$ and $\Phi(\hat{X}, \hat{\lambda}^1, \psi^1)_{\hat{J}^1,\theta^1}$ and hence that the induced maps on homology

$$CH_\bullet(Y^+, \xi^+, \psi^+)_{\lambda^+, J^+, \theta^+} \underset{\Phi(\hat{X}, \hat{\lambda}^1, \psi^1)_{\hat{J}^1,\theta^1}}{\overset{\Phi(\hat{X}, \hat{\lambda}^0, \psi^0)_{\hat{J}^0,\theta^0}}{\rightrightarrows}} CH_\bullet(Y^-, \xi^-, \psi^-)_{\lambda^-, J^-, \theta^-} \tag{6-20}$$

are equal.

Setup IV Fix a datum $\mathcal{D}$ for Setup IV. There are three types of concatenations $\{T_i\}_i$ in $\mathcal{S}_{\mathrm{IV}} = \mathcal{S}_{\mathrm{IV}}(\mathcal{D})$:

(1) $\{T_i\} \subset \mathcal{S}_{\mathrm{I}}^0 \sqcup \mathcal{S}_{\mathrm{II}}^{02} \sqcup \mathcal{S}_{\mathrm{I}}^2$, in which case $\mathfrak{s}(\#_i T_i) = \{0\}$.

(2) $\{T_i\} \subset \mathcal{S}_{\mathrm{I}}^0 \sqcup \mathcal{S}_{\mathrm{II}}^{01} \sqcup \mathcal{S}_{\mathrm{I}}^1 \sqcup \mathcal{S}_{\mathrm{II}}^{12} \sqcup \mathcal{S}_{\mathrm{I}}^2$. in which case $\mathfrak{s}(\#_i T_i) = \{\infty\}$;

(3) $\{T_i\} \subset \mathcal{S}_{\mathrm{I}}^0 \sqcup \mathcal{S}_{\mathrm{IV}} \sqcup \mathcal{S}_{\mathrm{I}}^2$ and $T_i \in \mathcal{S}_{\mathrm{IV}}$ for a unique $i = i_0$, in which case $\mathfrak{s}(\#_i T_i) = \mathfrak{s}(T_{i_0})$.

(We again follow the notation of [61, Section 2.1] for tree categories determined by $\mathcal{D}$.) Suppose now we are given maps of $\mathbb{Q}$–algebras $m^{01}\colon \mathcal{R}^0 \to \mathcal{R}^1$ and $m^{12}\colon \mathcal{R}^1 \to \mathcal{R}^2$, and twisting maps

$$\begin{aligned} \psi^i &\in \Psi_{\mathrm{I}}(\mathcal{D}^i; \mathcal{R}^i) && \text{for } i = 0,1,2, \\ \psi^{ij} &\in \Psi_{\mathrm{II}}(\mathcal{D}^{ij}; \psi^i, \psi^j) && \text{for } ij = 01, 12, 02. \end{aligned} \tag{6-21}$$

Set $m^{02} = m^{12} \circ m^{01}\colon \mathcal{R}^0 \to \mathcal{R}^2$.

Definition 6.4 The set $\Psi_{\mathrm{IV}}(\mathcal{D}; \{\psi^{ij}\})$ consists of all maps $\psi\colon \mathcal{S}_{\mathrm{IV}}^{\neq\varnothing}(\mathcal{D}) \to \mathcal{R}^2$ satisfying the following properties:

- For any morphism $T' \to T$, $\psi(T') = \psi(T)$.
- For any concatenation $\{T_i\}_i$ of the first type,

$$\psi(\#_i T_i) = \Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^0} m^{02}(\psi^0(T_i))\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{II}}^{02}} \psi^{02}(T_i)\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^2} \psi^2(T_i)\Big). \tag{6-22}$$

- For any concatenation $\{T_i\}_i$ of the second type,

$$\begin{aligned} \psi(\#_i T_i) = \Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^0} m^{02}(\psi^0(T_i))\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{II}}^{01}} m^{12}(\psi^{01}(T_i))\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^1} m^{12}(\psi^1(T_i))\Big) \\ \cdot \Big(\prod_{T_i \in \mathcal{S}_{\mathrm{II}}^{12}} \psi^{12}(T_i)\Big)\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^2} \psi^2(T_i)\Big). \end{aligned} \tag{6-23}$$

- For any concatenation $\{T_i\}_i$ of the third type,

$$\psi(\#_i T_i) = \Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^0} m^{02}(\psi^0(T_i))\Big)\psi(T_{i_0})\Big(\prod_{T_i \in \mathcal{S}_{\mathrm{I}}^2} \psi^2(T_i)\Big). \tag{6-24}$$

Fix a twisting map $\psi \in \Psi_{\mathrm{IV}}(\mathcal{D}; \{\psi^{ij}\})$. Theorem 1.1 of [61] provides a set of perturbation data $\Theta_{\mathrm{IV}}(\mathcal{D})$ together with a forgetful map $\Theta_{\mathrm{IV}}(\mathcal{D}) \to \Theta_{\mathrm{II}}(\mathcal{D}^{02}) \times_{\Theta_{\mathrm{I}}(\mathcal{D}^0) \times \Theta_{\mathrm{I}}(\mathcal{D}^2)} (\Theta_{\mathrm{II}}(\mathcal{D}^{01}) \times_{\Theta_{\mathrm{I}}(\mathcal{D}^1)} \Theta_{\mathrm{II}}(\mathcal{D}^{12}))$ and associated virtual moduli counts $\#\overline{\mathcal{M}}_{\mathrm{IV}}(T)_\theta^{\mathrm{vir}} \in \mathbb{Q}$ — here again, the fiber product is defined using (6-10). We define the twisted moduli counts

$$\#_\psi \overline{\mathcal{M}}_{\mathrm{IV}}(T)_\theta^{\mathrm{vir}} := \#\overline{\mathcal{M}}_{\mathrm{IV}}(T)_\theta^{\mathrm{vir}} \cdot \psi(T) \in \mathcal{R}^2. \tag{6-25}$$

As in the previous section, we obtain an $\mathcal{R}^0$–linear map

$$CC_\bullet(Y^0,\xi^0,\psi^0)_{\lambda^0,J^0,\theta^0} \to CC_{\bullet+1}(Y^2,\xi^2,\psi^2)_{\lambda^2,J^2,\theta^2}, \tag{6-26}$$

which is a chain homotopy between the maps

$$\Phi(\hat{X}^{02},\hat{\lambda}^{02},\psi^{02})_{\hat{J}^{02},\theta^{02}} \quad \text{and} \quad \Phi(\hat{X}^{12},\hat{\lambda}^{12},\psi^{12})_{\hat{J}^{12},\theta^{12}} \circ \Phi(\hat{X}^{01},\hat{\lambda}^{01},\psi^{01})_{\hat{J}^{01},\theta^{01}},$$

so that we get the commuting diagram

$$\begin{array}{ccccc} & \Phi(\hat{X}^{01},\hat{\lambda}^{01},\psi^{01})_{\hat{J}^{01},\theta^{01}} \nearrow & CH_\bullet(Y^1,\xi^1,\psi^1)_{\lambda^1,J^1,\theta^1} & \searrow \Phi(\hat{X}^{12},\hat{\lambda}^{12},\psi^{12})_{\hat{J}^{12},\theta^{12}} & \\ CH_\bullet(Y^0,\xi^0,\psi^0)_{\lambda^0,J^0,\theta^0} & & \xrightarrow[\Phi(\hat{X}^{02},\hat{\lambda}^{02},\psi^{02})_{\hat{J}^{02},\theta^{02}}]{} & & CH_\bullet(Y^2,\xi^2,\psi^2)_{\lambda^2,J^2,\theta^2} \end{array} \tag{6-27}$$

6.2 The energy of a (strict) symplectic cobordism

In this section, we introduce a notion of energy for (families of strict) exact symplectic cobordisms, and for certain classes of almost complex structures.

Notation 6.5 Recall that a strict exact symplectic cobordism from (Y^+,λ^+) to (Y^-,λ^-) is the data of an exact symplectic cobordism $(\hat{X},\hat{\lambda})$ and embeddings

$$e^\pm\colon (\mathbb{R}\times Y^\pm,\hat{\lambda}^\pm) \to (\hat{X},\hat{\lambda}), \tag{6-28}$$

which preserve the Liouville forms and satisfy certain additional properties stated in Definition 2.3. When we consider strict exact symplectic cobordisms in this section, we will routinely abuse notation by identifying subsets of $\mathbb{R}\times Y^\pm$ with their image under $e^\pm$. We hope that this abuse will make this section easier to read without introducing any substantial ambiguities.

We begin with the following definition.

Definition 6.6 Let $(\hat{X},\hat{\lambda})$ be a strict exact symplectic cobordism from (Y^+,λ^+) to (Y^-,λ^-). A *Type A cobordism decomposition* is the data of a pair of hypersurfaces

$$\mathcal{H}_- = \{-C_-\}\times Y^- \quad \text{and} \quad \mathcal{H}_+ = \{C_+\}\times Y^+ \tag{6-29}$$

for $C_\pm \in \mathbb{R}$, such that

$$((-\infty,-C_-)\times Y^-)\cap((C_+,\infty)\times Y^+) = \varnothing. \tag{6-30}$$

The intersection in (6-30) takes place inside $\hat{X}$; if $Y^- = \varnothing$, we set $\mathcal{H}_- = \varnothing$, $C_- = 0$ and we consider that (6-30) is tautologically satisfied. We let $\Sigma(\hat{X},\hat{\lambda}) = \Sigma(\hat{X},\hat{\lambda};\lambda^+,\lambda^-)$ be the set of all such Type A cobordism decompositions.

Remark 6.7 Since we are working with *strict* cobordisms, the real numbers $C_\pm \in \mathbb{R}$ are uniquely determined by the hypersurfaces $\mathcal{H}_\pm$. The data of the pair $(\mathcal{H}_-,\mathcal{H}_+)$ is therefore equivalent to the data of the pair (C_-,C_+).

Definition 6.8 Let $(\widehat{X}, \widehat{\lambda})$ be as in Definition 6.6 and let $\sigma \in \Sigma(\widehat{X}, \widehat{\lambda})$ be a Type A cobordism decomposition. We let $\mathcal{E}(\sigma) := C_- + C_+$ be the *energy* of the decomposition σ. We define

$$\mathcal{E}(\widehat{X}, \widehat{\lambda}) = \mathcal{E}(\widehat{X}, \widehat{\lambda}; \lambda^+, \lambda^-) := \inf_{\sigma \in \Sigma(\widehat{X}, \widehat{\lambda})} \mathcal{E}(\sigma) \in \mathbb{R} \cup \{-\infty\} \tag{6-31}$$

to be the energy of $(\widehat{X}, \widehat{\lambda})$; this is well-defined since a cobordism decomposition clearly always exists. We note that the energy may in general be negative.

Given $C \in \mathbb{R}$, let $\Sigma(\widehat{X}, \widehat{\lambda})_{<C} \subset \Sigma(\widehat{X}, \widehat{\lambda})$ (resp. $\leq C$) denote the subset of cobordism decompositions of energy strictly less than C (resp. at most C).

Lemma 6.9 (energy of a symplectization) *Suppose that $(\widehat{X}, \widehat{\lambda}) = (SY, \lambda_Y)$ is a symplectization which is endowed with the canonical structure of a (strict) exact symplectic cobordism from (Y, λ^+) to (Y, λ^-); see Example 2.6. Fix $f : Y \to \mathbb{R}$ such that $\lambda^+ = e^f \lambda^-$. Then $\mathcal{E}(SY, \lambda_Y) \leq -\min f$.*

Proof Let $e^{\pm} : (\mathbb{R} \times Y, \widehat{\lambda}^{\pm}) \to (SY, \lambda_Y)$ be the canonical identifications induced by $\lambda^{\pm}$. Let $\mathcal{H}_+ = \{0\} \times Y$ in the coordinates induced by e^+. This means that $\mathcal{H}_+ = \{(f(y), y) \mid y \in Y\} \subset \mathbb{R} \times Y$ in the coordinates induced by e^-. Now given any $C_- > -\min f$, we can let $\mathcal{H}_- = \{-C_-\} \times Y$ in the coordinates induced by e^-. It follows that $\mathcal{E}(SY, \widehat{\lambda}_Y) \leq C_-$. Since $C_- > -\min f$ was arbitrary, the claim follows. □

Remark 6.10 If we assume in addition that $\lambda^+ = \lambda^-$, then it is easy to verify that in fact $\mathcal{E}(SY, \lambda_Y) = 0$.

Lemma 6.11 *We have $\mathcal{E}(\widehat{X}, \widehat{\lambda}) = -\infty$ if and only if $Y^- = \varnothing$.*

Proof Suppose that Y^- is nonempty and choose a cobordism decomposition σ for $(\widehat{X}, \widehat{\lambda})$ given by a pair of hypersurfaces $\mathcal{H}_-, \mathcal{H}_+ \subset \widehat{X}$. Let (X, λ) be the truncated Liouville cobordism with negative boundary $\mathcal{H}_-$ and positive boundary $\mathcal{H}_+$. Observe that the image of the negative boundary under the Liouville flow must touch the positive boundary in some finite time $T < \infty$ — indeed, this follows from the fact that (X, λ) has finite volume. Given any other cobordism decomposition σ', we now have $\mathcal{E}(\sigma') \geq \mathcal{E}(\sigma) - T$.

Suppose now that Y^- is empty. Then (6-30) is a vacuous condition. Since the backwards Liouville flow of any slice $\{C_+\} \times Y^+$ is defined for all time, it follows that we can find a cobordism decomposition of arbitrarily negative energy. □

Lemma 6.12 *Fix a strict exact symplectic cobordism $(\widehat{X}, \widehat{\lambda})$ from (Y^+, λ^+) to (Y^-, λ^-). Then $\Sigma(\widehat{X}, \widehat{\lambda})_{<C} \subset \Sigma(\widehat{X}, \widehat{\lambda})$ is*

(a) *nonempty for $C > \mathcal{E}(\widehat{X}, \widehat{\lambda})$,*

(b) *path-connected for all $C \in \mathbb{R}$.* (Note that the empty set is path-connected.)

If moreover $(\widehat{X}, \widehat{\lambda}) = (SY, \lambda_Y)$ is a symplectization and $\lambda^+ = \lambda^-$, then $\Sigma(\widehat{X}, \widehat{\lambda})_{\leq 0}$ is nonempty and path-connected.

Proof Note first that (a) is tautologically true. Next, note that (b) is obvious when $Y^- = \varnothing$. It therefore remains to prove (b) under the assumption that $Y^- \neq \varnothing$.

Let us consider a pair of cobordism decompositions $\sigma, \sigma' \in \Sigma(\widehat{X}, \widehat{\lambda})_{<C}$. By definition, σ and σ' are entirely determined by the constants $-C_-, C_+ \in \mathbb{R}$ (resp. $-C'_-, C'_+$), where we are following the notation of Definition 6.6.

Suppose first that $-C_- = -C'_-$. Up to relabeling σ and σ', we can assume that $C_+ \le C'_+$. Now just translate C'_+ in the negative direction until $C'_+ = C_+$. This translation defines a one-parameter family of cobordism decompositions taking σ' to σ, whose energy is clearly bounded by $\max(\mathcal{E}(\sigma), \mathcal{E}(\sigma')) = \mathcal{E}(\sigma') < C$. An analogous argument works if we now suppose $C_+ = C'_+$ and $-C_- \neq -C'_-$.

Suppose finally that $-C_- \neq -C'_-$ and $C_+ \neq C'_+$. Up to relabeling σ and σ', we can assume that $C_+ < C'_+$. If $-C'_- < -C_-$, then we translate $-C'_-$ in the positive direction until $-C'_- = -C_-$. If instead $-C_- < -C'_-$, then we simultaneously translate $-C_-$ and C_+ in the positive direction until either $-C_- = -C'_-$ or $C_+ = C'_+$. This takes us back to the case treated in the previous paragraph.

Finally, if $(\widehat{X}, \widehat{\lambda}) = (SY, \lambda_Y)$ is a symplectization with $\lambda^+ = \lambda^-$, then any Type A cobordism decomposition $\sigma \in \Sigma(\widehat{X}, \widehat{\lambda})_{\le 0}$ has vanishing energy (Remark 6.10) and is equivalent to a choice of hypersurface $\mathcal{H} = \mathcal{H}_- = \mathcal{H}_+ = \{\widetilde{C} \times Y\}$. The space of such choices is in natural bijection with $\mathbb{R}$, so it is in particular nonempty and connected. □

Definition 6.13 Let $(\widehat{X}, \widehat{\lambda}_t)_{t \in [0,1]}$ be a one-parameter family of (strict) exact symplectic cobordisms; cf Definition 2.12. A one-parameter family of Type A cobordism decompositions is just the data of a family of hypersurfaces

$$\mathcal{H}_-(t) = \{-C_-(t)\} \times Y^- \quad \text{and} \quad \mathcal{H}_+(t) = \{C_+(t)\} \times Y^+ \tag{6-32}$$

such that

$$((-\infty, -C_-(t)) \times Y^-) \cap ((C_+(t), \infty) \times Y^+) = \varnothing. \tag{6-33}$$

(If $Y^- = \varnothing$, we again set $\mathcal{H}_-(t) = \varnothing$, $C_-(t) = 0$ and we consider that (6-33) is tautologically satisfied.) We let $\Sigma(\widehat{X}, \widehat{\lambda}_t)_{t \in [0,1]}$ be the set of all such families of cobordism decompositions. (Note that $\Sigma(\widehat{X}, \widehat{\lambda}_{t_0})$ is a Type A cobordism decomposition for each fixed choice of t_0.)

Definition 6.14 With the notation as above, with define the energy of a family of Type A cobordism decompositions $\sigma \in \Sigma(\widehat{X}, \widehat{\lambda}_t)_{t \in [0,1]}$ to be $\mathcal{E}(\sigma) := \sup_t (C_-(t) + C_+(t))$.

Let $(\widehat{X}^{01}, \widehat{\lambda}^{01})$ (resp. $(\widehat{X}^{12}, \widehat{\lambda}^{12})$) be a strict exact symplectic cobordism from (Y^0, λ^0) to (Y^1, λ^1) (resp. from (Y^1, λ^1) to (Y^2, λ^2)). Let $(\widehat{X}, \widehat{\lambda}_t)_{t \in [0,\infty)}$ be a one-parameter family of strict exact symplectic cobordisms which agrees for $t \geq a$ large enough with the t–gluing $(\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{02})_{t \in [a,\infty)}$; see Definition 2.7. For $t \geq a$, note that there are canonical Liouville embeddings

$$\iota_{0,t} \colon (\widehat{X}^{01}, \widehat{\lambda}^{01}) \xrightarrow{\mu_{t/2}} (\widehat{X}^{01}, e^{t/2}\widehat{\lambda}^{01}) \to (\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{02}),$$
$$\iota_{2,t} \colon (\widehat{X}^{12}, \widehat{\lambda}^{12}) \xrightarrow{\mu_{-t/2}} (\widehat{X}^{12}, e^{-t/2}\widehat{\lambda}^{12}) \to (\widehat{X}^{01} \#_t \widehat{X}^{12}, \widehat{\lambda}^{01} \#_t \widehat{\lambda}^{02}).$$

Definition 6.15 A *Type B cobordism decomposition* of $(\widehat{X}, \widehat{\lambda}_t)_{t \in [0,\infty)}$ is the data of a family of hypersurfaces

$$\mathcal{H}_2(t) = \{-C_2(t)\} \times Y^2 \quad \text{and} \quad \mathcal{H}_0(t) = \{C_0(t)\} \times Y^0, \tag{6-34}$$

and a Liouville embedding $([-C_1(t), \widetilde{C}_1(t)] \times Y^1, e^s\lambda_1) \hookrightarrow (\widehat{X}, \widehat{\lambda}_t)$ such that

$$(-\infty, -C_2(t)) \times Y^2, \quad (-C_1(t), \widetilde{C}_1(t)) \times Y^1 \quad \text{and} \quad (C_0(t), \infty) \times Y^0 \tag{6-35}$$

are pairwise disjoint. (In the case that $Y^- = \varnothing$, we set $\mathcal{H}_2(t) = \varnothing$ and $C_2(t) = 0$, and replace (6-35) by the condition that $(-C_1(t), \widetilde{C}_1(t)) \times Y^1$ and $(C_0(t), \infty) \times Y^0$ are pairwise disjoint.)

We let

$$\mathcal{H}_1(t) = \{-C_1(t)\} \times Y^1 \quad \text{and} \quad \widetilde{\mathcal{H}}_1(t) = \{\widetilde{C}_1(t)\} \times Y^1. \tag{6-36}$$

This data is required to satisfy the following hypotheses:

(1) $\widetilde{C}_1(0) = -C_1(0)$.

(2) For t large enough, $\mathcal{H}_0$ and $\widetilde{\mathcal{H}}_1(t)$ (resp. $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$) are in the image of the canonical embedding $\iota_{0,t}$ (resp. $\iota_{2,t}$). Moreover, their preimages define a Type A decomposition on $(\widehat{X}^{01}, \widehat{\lambda}^{01})$ (resp. on $(\widehat{X}^{12}, \widehat{\lambda}^{12})$) *which is independent of* t.

We let $\Sigma_B((\widehat{X}, \widehat{\lambda}_t)_{t \in [0,\infty)})$ denote the set of all such cobordism decompositions. We will write $\Sigma(-)$ instead of $\Sigma_B(-)$ when the subscript is understood from the context. As in Remark 6.7, note that the data of the hypersurfaces $\mathcal{H}_2(t)$, $\mathcal{H}_1(t)$, $\widetilde{\mathcal{H}}_1'(t)$ and $\mathcal{H}_0(t)$ is equivalent to the data of the constants $C_2(t)$, $C_1(t)$, $\widetilde{C}_1(t)$ and $C_0(t)$.

Definition 6.16 It follows from property (1) of Definition 6.15 that a Type B cobordism decomposition $\sigma^{02} \in \Sigma_B((\widehat{X}, \widehat{\lambda}_t)_{t \in [0,\infty)})$ induces a Type A cobordism decomposition $\sigma \in \Sigma_A(\widehat{X}, \widehat{\lambda}_0)$ by taking $\mathcal{H}_- = \mathcal{H}_2(0)$ and $\mathcal{H}_+ = \mathcal{H}_0(0)$. We say that σ is *induced at zero* by σ^{02}.

Similarly, it follows from property (2) of Definition 6.15 that a Type B cobordism decomposition $\sigma^{02} \in \Sigma_B((\widehat{X}, \widehat{\lambda}_t)_{t \in [0,\infty)})$ induces a pair of Type A decompositions $\sigma^{01} \in \Sigma_A(\widehat{X}^{01}, \widehat{\lambda}^{01})$ and $\sigma^{12} \in \Sigma_A(\widehat{X}^{02}, \widehat{\lambda}^{02})$. We say that the pair $(\sigma^{01}, \sigma^{12})$ is *induced at infinity* by σ^{02}.

Definition 6.17 With the notation as above, we define the energy of a Type B cobordism decomposition $\sigma \in \Sigma_B(\widehat{X}, \widehat{\lambda}_t)$ to be $\mathcal{E}(\sigma) := \sup_t (C_2(t) + C_0(t) - C_1(t) - \widetilde{C}_1(t))$. We let

$$\mathcal{E}(\widehat{X}, \widehat{\lambda}_t) := \inf_{\sigma \in \Sigma_B(\widehat{X}, \widehat{\lambda}_t)} \mathcal{E}(\sigma) \in \mathbb{R} \cup \{-\infty\}. \tag{6-37}$$

Given $C \in \mathbb{R}$, let $\Sigma(\widehat{X}, \widehat{\lambda}_t)_{<C} \subset \Sigma(\widehat{X}, \widehat{\lambda}_t)$ (resp. $\leq C$) denote the subset of Type B cobordism decompositions of energy strictly less than C (resp. at most C).

The following lemma asserts that our notions of energy for Type A and Type B decompositions are compatible with the map which associates to a Type B decomposition the Type A decomposition induced at zero or infinity. It will be used implicitly in the sequel.

Lemma 6.18 *Let σ^{02} be a Type B cobordism decomposition. Suppose that σ is induced at zero by σ^{02} and that $(\sigma^{01}, \sigma^{12})$ is induced at infinity. Then $\mathcal{E}(\sigma) \leq \mathcal{E}(\sigma^{02})$ and $\mathcal{E}(\sigma^{01}) + \mathcal{E}(\sigma^{12}) \leq \mathcal{E}(\sigma^{02})$.*

Proof The first claim follows from (1) in Definition 6.15 and the definition of energy for Type A and Type B cobordism decomposition. The second claim follows similarly from (2) in Definition 6.15. □

Corollary 6.19 *We have $\mathcal{E}(\widehat{X}, \widehat{\lambda}_t) = -\infty$ if and only if $Y^2 = \varnothing$.*

Proof One direction follows from Lemmas 6.11 and 6.18. The other one can be checked by inspection, using the backwards Liouville flow as in the proof of the corresponding statement in Lemma 6.11. □

Definition 6.20 Suppose that $(\widehat{X}, \widehat{\lambda})$ and $(\widehat{X}', \widehat{\lambda}')$ are exact symplectic cobordisms. For $C \in \mathbb{R}$, let $\Sigma_A((\widehat{X}, \widehat{\lambda}), (\widehat{X}', \widehat{\lambda}'))_{<C} \subset \Sigma_A(\widehat{X}, \widehat{\lambda}) \times \Sigma_A(\widehat{X}, \widehat{\lambda})$ (resp. $(-)_{\leq C}$) be the subspace of pairs (σ, σ') of Type A cobordism decompositions such that $\mathcal{E}(\sigma) + \mathcal{E}(\sigma') < C$ (resp. $\leq C$).

Lemma 6.21 *Given $C \in \mathbb{R}$ such that $\Sigma(X^{02,t}, \widehat{\lambda}^{02,t})_{<C}$ is nonempty, the map which associates to a decomposition $\sigma^{02} \in \Sigma(X^{02,t}, \widehat{\lambda}^{02,t})_{<C}$ the pair $(\sigma^{01}, \sigma^{12}) \in \Sigma_A((\widehat{X}^{01}, \widehat{\lambda}^{01}), (\widehat{X}^{12}, \widehat{\lambda}^{12}))_{<C}$ induced by σ^{02} at infinity is surjective. If $\lambda^0 = \lambda^1 = \lambda^2$ and $(\widehat{X}^{01}, \widehat{\lambda}^{01}), (\widehat{X}^{12}, \widehat{\lambda}^{12})$ are symplectizations, the same statement holds with $\leq$ in place of $<$.*

Proof Choose a Type B decomposition $\widetilde{\sigma}^{02}$. Let $(\widetilde{\sigma}^{01}, \widetilde{\sigma}^{12})$ be the Type A decompositions induced by $\widetilde{\sigma}^{02}$ at infinity. According to Definition 6.15, this means that there exists a $T > 0$ so that for $t \geq T$, we have that $\mathcal{H}_0(t)$ and $\widetilde{\mathcal{H}}_1(t)$ are independent of t after pulling back via the canonical embedding ι_{01} (and similarly $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ are independent of t after pulling back by ι_{12}). By a routine modification of the arguments of Lemma 6.12(b), one can now construct a Type B decomposition σ^{02} so that $\sigma^{02}_t = \widetilde{\sigma}^{02}_t$ for $t \in [0, T]$, $\mathcal{E}(\sigma^{02}) \leq \mathcal{E}(\widetilde{\sigma}^{02})$ and σ^{02} induces the pair $(\sigma^{01}, \sigma^{12})$. □

Lemma 6.22 *Suppose that $(\widehat{X}^{02,t},\widehat{\lambda}^{02,t})$ is the $(t+T)$–gluing of two exact symplectic cobordisms $(\widehat{X}^{01},\widehat{\lambda}^{01})$ and $(\widehat{X}^{12},\widehat{\lambda}^{12})$, for $T\geq 0$ an arbitrary fixed constant and $t\in[0,\infty)$ a parameter; see Definition 2.7 for the definition of this gluing and Lemma 2.13 for the parametric version. Suppose that either $(\widehat{X}^{01},\widehat{\lambda}^{01})$ or $(\widehat{X}^{12},\widehat{\lambda}^{12})$ is a symplectization; see Example 2.6. Then $\mathcal{E}(\widehat{X}^{02,t},\widehat{\lambda}^{02,t})=\mathcal{E}(\widehat{X}^{01},\widehat{\lambda}^{01})+\mathcal{E}(\widehat{X}^{12},\widehat{\lambda}^{12})$.*

Proof By Lemma 6.11 and Corollary 6.19, we may assume that $\widehat{X}^{12}$ has a nonempty negative end.

We only treat the case where $(\widehat{X}^{01},\widehat{\lambda}^{01})$ is a symplectization and $T=0$, since the other cases are analogous.

Choose σ^{01} so that $\mathcal{E}(\sigma^{01})\leq\mathcal{E}(\widehat{X}^{01},\widehat{\lambda}^{01})+\epsilon$, and choose σ^{12} so that $\mathcal{E}(\sigma^{12})\leq\mathcal{E}(\widehat{X}^{12},\widehat{\lambda}^{12})+\epsilon$. Let $\widetilde{X}^{01}\subset\widehat{X}^{01}$ and $\widetilde{X}^{12}\subset\widehat{X}^{12}$ be the Liouville subdomains which determine the Type A decompositions σ^{01} and σ^{12}, respectively.

Note that $\widehat{X}^{02,t}$ comes equipped with tautological embeddings $\iota_{0,t}\colon\widehat{X}^{01}\to\widehat{X}^{02,t}$ and $\iota_{2,t}\colon\widehat{X}^{12}\to\widehat{X}^{02,t}$; see Definition 2.7. For T' large enough and $t\geq T'$, note that $\iota_{0,t}(\mathcal{H}^0_-)$ is in the image of $\iota_{2,t}(\mathcal{H}^2_+)$ under the Liouville flow. These hypersurfaces therefore bound Liouville domains $([-C_1(t),\widetilde{C}_1(t)]\times Y^1,e^s\lambda_1)$.

Let $f\colon[0,\infty)\to\mathbb{R}$ be a function which equals $-(C_1(T')+\widetilde{C}_1(T'))$ on $[0,T']$, is nondecreasing on $[T',T'+1]$ and is zero on $[T'+1,\infty)$. Let $\tau_f\colon\widehat{X}^{01}\times[0,\infty)\to\widehat{X}^{01}$ be defined by $\tau_f(x,t)=\phi^{01}_{f(t)+T'+1-t}(x)$, where ϕ^{01}_t is the time-t Liouville flow on $\widehat{X}^{01}$. Now define a map $\overline{\iota}_{0,t}\colon\widehat{X}^{01}\to\widehat{X}^{02,t}$ by letting $\overline{\iota}_{0,t}(x)=\iota_{0,t}\circ\tau_f(x,t)$.

We now define the data of a Type B cobordism decomposition by letting

$$\mathcal{H}_2(t)=\iota_{2,t}(\mathcal{H}^2_-),\quad \mathcal{H}_1(t)=\iota_{2,t}(\mathcal{H}^2_+),\quad \widetilde{\mathcal{H}}_1=\overline{\iota}_{0,t}(\mathcal{H}^0_-(t)),\quad \mathcal{H}_0=\overline{\iota}_{0,t}(\mathcal{H}^0_+(t)). \tag{6-38}$$

One can check that this data indeed defines a Type B cobordism decomposition, which has energy precisely equal to $\mathcal{E}(\sigma^{01})+\mathcal{E}(\sigma^{12})\leq\mathcal{E}(\widehat{X}^{01},\widehat{\lambda}^{01})+\mathcal{E}(\widehat{X}^{12},\widehat{\lambda}^{12})+2\epsilon$. Since ϵ was arbitrary, we conclude that $\mathcal{E}(\widehat{X}^{02,t},\widehat{\lambda}^{02,t})\leq\mathcal{E}(\widehat{X}^{01},\widehat{\lambda}^{01})+\mathcal{E}(\widehat{X}^{12},\widehat{\lambda}^{12})$. □

We now discuss almost complex structures for Setups II–IV.

Setup II Fix a datum $\mathcal{D}=(\mathcal{D}^+,\mathcal{D}^-,\widehat{X},\widehat{\lambda},\widehat{J})$ for Setup II, with $\mathcal{D}^\pm=(Y^\pm,\lambda^\pm,J^\pm)$. Let $(V^\pm,\tau^\pm)\subset(Y^\pm,\xi^\pm)$ be framed codimension 2 contact submanifolds; let $\alpha^\pm:=\lambda^\pm|_{V^\pm}$ and assume that $V^\pm$ is a strong contact submanifold with respect to $\lambda^\pm$. Let $J^\pm$ be $d\lambda^\pm$–compatible almost complex structures on $\xi^\pm\subset TY^\pm$ which preserve $\xi^\pm\cap TV^\pm$.

Let $H\subset\widehat{X}$ be a codimension 2 symplectic submanifold such that $(\widehat{X},\widehat{\lambda},H)$ is an exact relative symplectic cobordism from (Y^+,ξ^+,V^+) to (Y^-,ξ^-,V^-). We will also consider (see Notation 2.4) the strict symplectic cobordisms $(\widehat{X},\widehat{\lambda})^{\lambda^+}_{\lambda^-}$ and $(H,\widehat{\lambda}|_H)^{\alpha^+}_{\alpha^-}$.

Definition 6.23 Fix a Type A cobordism decomposition $\sigma \in \Sigma(H, \widehat{\lambda}|_H)$, which is specified by a pair of hypersurfaces $\mathcal{H}_- = \{-C_-\} \times V^-$ and $\mathcal{H}_+ = \{C_+\} \times V^+$. We say that an almost complex structure $\widehat{J}$ on $\widehat{X}$ is *adapted* to σ if the following properties hold:

- $\widehat{J}$ is compatible with $d\widehat{\lambda}$,
- $\widehat{J}$ coincides with $\widehat{J}^\pm$ near the ends (where $\widehat{J}^\pm$ is the canonical cylindrical almost complex structure induced on $(\widehat{Y}, \widehat{\lambda}^\pm)$ by $J^\pm$),
- $H \subset \widehat{X}$ is a $\widehat{J}$–complex hypersurface,
- $\widehat{J}$ preserves $\ker \alpha^+ \subset TV^+$ on $[C_2, \infty) \times V^+$ (resp. preserves $\ker \alpha^- \subset TV^-$ on $(-\infty, -C_1] \times V^-$) and the induced almost complex structure is $d\alpha^+$–compatible (resp. $d\alpha^-$–compatible).

In the case that $V^- = \varnothing$, the conditions involving V^- are considered to be vacuously satisfied.

Definition 6.24 Given an almost complex structure $\widehat{J}$ on $(\widehat{X}, \widehat{\lambda})$ we define its energy

$$\mathcal{E}(\widehat{J}) := \inf\{\mathcal{E}(\sigma) \mid \sigma \in \Sigma(H, \widehat{\lambda}|_H),\ \widehat{J} \text{ is adapted to } \sigma\} \in \mathbb{R} \cup \{\pm\infty\}. \tag{6-39}$$

We define $\mathcal{E}(\widehat{J}) = \infty$ if $\widehat{J}$ is not adapted to any cobordism decomposition.

Let $\mathcal{J}(\widehat{X}, \widehat{\lambda}, H)_{<C}$ (resp. $\leq C$) be the set of almost complex structures of energy less than C (resp. at most C). Let $\mathcal{J}(\widehat{X}, \widehat{\lambda}, H) := \mathcal{J}(\widehat{X}, \widehat{\lambda}, H)_{<\infty}$ be the set of almost complex structures adapted to some decomposition $\sigma \in \Sigma(\widehat{H}, \widehat{\lambda}|_H)$.

Lemma 6.25 *The set $\mathcal{J}(\widehat{X}, \widehat{\lambda}, H)_{<C}$ is*

(a) *nonempty for $C > \mathcal{E}(H, \widehat{\lambda}|_H)$, and*

(b) *path-connected for all $C \in \mathbb{R}$.* (Note that the empty set is path-connected.)

If moreover $(H, \widehat{\lambda}|_H) = (SV, \lambda_V)$ is a symplectization and $\alpha^+ = \alpha^-$, then $\mathcal{J}(\widehat{X}, \widehat{\lambda}, H)_{\leq 0}$ is nonempty and path-connected.

Proof To prove (a), it is enough to show that given any cobordism decomposition σ, there exists an almost complex structure adapted to it, ie meeting the conditions of Definition 6.23. To prove (b), it follows from Lemma 6.12 that it is enough to prove a similar statement in families: namely, if $\{\sigma_t\}_{t\in[a,b]}$ is a family of cobordism decompositions and J_a and J_b are almost complex structures adapted to σ_a and σ_b, respectively, then there is a family $\{J_t\}_{t\in[a,b]}$ adapted to σ_t. All of these statements can be proved by standard arguments, using the fact that the space of almost complex structures compatible with a given symplectic structure can be viewed as the space of sections of a bundle with contractible fibers; see eg [52, Proposition 2.6.4].

If $\alpha^+ = \alpha^-$ and $(H, \widehat{\lambda}|_H)$ is a symplectization, then Lemma 6.12 implies that $\Sigma(H, \widehat{\lambda}|_H)_{\leq 0}$ is nonempty and path-connected. Hence the same arguments involving extensions of almost complex structures imply that $\mathcal{J}(\widehat{X}, \widehat{\lambda}, H)_{\leq 0}$ is nonempty and path-connected. □

Setup IV Fix a datum $\mathcal{D}$ for Setup IV. We write $\mathcal{D} = (\mathcal{D}^{01}, \mathcal{D}^{12}, (\hat{X}^{02,t}, \hat{\lambda}^{02,t}, \hat{J}^{02,t})_{t\in[0,\infty)})$, where

$$\begin{aligned}\mathcal{D}^{01} &= (\mathcal{D}^0, \mathcal{D}^1, \hat{X}^{01}, \hat{\lambda}^{01}, \hat{J}^{01}),\\ \mathcal{D}^{12} &= (\mathcal{D}^1, \mathcal{D}^2, \hat{X}^{12}, \hat{\lambda}^{12}, \hat{J}^{12}),\\ \mathcal{D}^i &= (Y^i, \lambda^i, J^i) \quad \text{for } i = 0, 1, 2.\end{aligned}$$

Let $(V^i, \tau^i) \subset (Y^i, \xi)$ be framed codimension 2 contact submanifolds; set $\alpha_{V^i} = \lambda^i|_{V^i}$ and assume that V^i are strong contact submanifolds with respect to λ^i. Let

$$H^{01} \subset \hat{X}^{01}, \quad H^{12} \subset \hat{X}^{12} \quad \text{and} \quad (H^{02,t} \subset \hat{X}^{02,t})_{t\in[0,\infty)}$$

be cylindrical symplectic submanifolds such that $(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})_{t\in[0,\infty)}$ is a family of relative symplectic cobordisms that agrees for t large with the t–gluing of the relative symplectic cobordisms $(\hat{X}^{01}, \hat{\lambda}^{01}, H^{01})$ and $(\hat{X}^{12}, \hat{\lambda}^{12}, H^{12})$. Note that $\{H^{02,t}\}$ forms a family of Liouville manifolds with respect to (the restriction of) $\hat{\lambda}^{02,t}$.

Definition 6.26 Fix a Type B cobordism decomposition $\sigma^{02} \in \Sigma_B(\hat{H}^{02,t}, \hat{\lambda}^{02,t}_{H^{02,t}})$. Recall that σ^{02} consists of the data of hypersurfaces

$$\mathcal{H}_2(t) = \{-C_2(t)\}\times V^2, \quad \mathcal{H}_1(t) = \{-C_1(t)\}\times V^1, \quad \tilde{\mathcal{H}}_1(t) = \{\tilde{C}_1(t)\}\times V^1, \quad \mathcal{H}_0(t) = \{C_0(t)\}\times V^0.$$

We say that an almost complex structure $\hat{J}^{02,t}$ is *adapted* to σ^{02} if the following properties hold:

- $\hat{J}^{02,t}$ is compatible with $d\hat{\lambda}^{02,t}$.
- $\hat{J}^{02,t}$ coincides with $\hat{J}^0$ (resp. $\hat{J}^2$) near the positive (resp. negative) end.
- $H^{02,t}$ is a $\hat{J}^{02,t}$–complex hypersurface. $\hat{J}^{02,t}$ is compatible with the restriction of $d\hat{\lambda}^{02,t}$ to $H^{02,t}$, and
- $\hat{J}^{02,t}$ preserves $\ker\alpha_0$ on $[C_0(t), \infty) \times V^0$, and preserves $\ker\alpha_1$ on $[-C_1(t), \tilde{C}_1(t)] \times V_1$, and preserves $\ker\alpha_2$ on $(-\infty, -C_2(t)] \times V_2)$. Moreover, the induced almost complex structure is $d\alpha_0$–compatible, $d\alpha_1$–compatible, and $d\alpha_2$–compatible.

In the case that $V^2 = \varnothing$, all conditions involving V^2 are considered to be vacuously satisfied.

Definition 6.27 Given a family of almost complex structures $\hat{J}_t$, we define its energy by

$$\mathcal{E}(\hat{J}_t) := \inf\{\mathcal{E}(\sigma) \mid \sigma \in \Sigma(\hat{X}^{02,t}, \hat{\lambda}^{02,t}),\ \hat{J}_t \text{ is adapted to } \sigma\} \in \mathbb{R}\cup\{\pm\infty\}. \tag{6-40}$$

If $\hat{J}_t$ is not adapted to any cobordism decomposition, we set $\mathcal{E}(\hat{J}_t) = \infty$.

Let $\mathcal{J}(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})$ be the set of almost complex structures adapted to some Type B decomposition $\sigma \in \Sigma_B(H^{02,t}, \hat{\lambda}^{02,t}|_{H^{02,t}})$. For $C \in \mathbb{R}$, let $\mathcal{J}(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})_{<C}$ (resp. $\leq c$) be the set of all such decompositions having energy less than C (resp. at most C).

Let $\mathcal{J}((\hat{X}^{01}, \hat{\lambda}^{01}), (\hat{X}^{12}, \hat{\lambda}^{12}))_{<C} \subset \mathcal{J}(\hat{X}^{01}, \hat{\lambda}^{01}) \times \mathcal{J}(\hat{X}^{12}, \hat{\lambda}^{12})$ (resp. $\leq C$) be the subspace of pairs (J, J') with the property that $\mathcal{E}(J) + \mathcal{E}(J') < C$ (resp. $\leq C$).

The following lemma is an analog of Lemma 6.25 and can be proved by similar arguments.

Lemma 6.28 *The set $\mathcal{J}(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})_{<C}$ is nonempty for $C > \mathcal{E}(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})$. If moreover $\alpha_0 = \alpha_1 = \alpha_2$ and $(H^{01}, \hat{\lambda}^{01}|_{H^{01}})$, $(H^{12}, \hat{\lambda}^{12}|_{H^{12}})$ and $(H^{02,t}, \hat{\lambda}^{02,t}|_{H^{02,t}})$ are symplectizations, then $\mathcal{J}(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})_{\leq 0}$ is nonempty.* □

We will also need the following lemma, which follows from Lemma 6.21 and standard arguments for extending compatible almost complex structures.

Lemma 6.29 *Suppose that $\mathcal{J}(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})_{<C}$ is nonempty. The map that associates to an almost complex structure $\hat{J}_t \in \mathcal{J}(\hat{X}^{02,t}, \hat{\lambda}^{02,t}, H^{02,t})_{<C}$ the pair $(\hat{J}^{01}, \hat{J}^{12}) \in \mathcal{J}((\hat{X}^{01}, \hat{\lambda}^{01}), (\hat{X}^{12}, \hat{\lambda}^{12}))_{<C}$ is surjective for all $C > 0$.*

If moreover $\alpha_0 = \alpha_1 = \alpha_2$, and $(H^{01}, \hat{\lambda}^{01}|_{H^{01}})$, $(H^{12}, \hat{\lambda}^{12}|_{H^{12}})$ and $(H^{02,t}, \hat{\lambda}^{02,t}|_{H^{02,t}})$ are symplectizations, then the same statement holds for $C = 0$ with $\leq$ in place of $<$. □

7 Enriched setups and twisted moduli counts

7.1 Enriched setups

The construction of invariants of codimension 2 contact submanifolds in this paper follows the same general scheme as Pardon's construction of contact homology. However, we work with a class of "enriched" Setups I*–IV*, which contain more information than the standard Setups I–IV considered by Pardon and reviewed in Section 4.1.

We will show in Section 7.2 that the data associated to our enriched setups give rise to twisting maps. These twisting maps are constructed using Siefring's intersection theory, and will be used to define "twisted" moduli counts, following the construction of Section 6.1.

Given a datum $\mathcal{D}$ for any of Setups I*–IV*, there is a "forgetful functor" which allows one to view $\mathcal{D}$ as a datum of Setups I–IV. However, it is not the case that every datum of Setups I–IV admits an enrichment. Nevertheless, we will show in Section 8.1 that the class of enriched data is large enough for the purpose of defining invariants in the spirit of contact homology.

Setup I* A datum $\mathcal{D} = ((Y, \xi, V), \mathfrak{r}, \lambda, J)$ for Setup I* consists of

- a TN contact pair (Y, ξ, V),
- an element $\mathfrak{r} = (\alpha_V, \tau, r) \in \mathfrak{R}(Y, \xi, V)$,
- a contact form $\ker \lambda = \xi$ which is adapted to $\mathfrak{r}$, and
- an almost complex structure J which is compatible with $d\lambda$ and preserves ξ_V.

Observe that there is a "forgetful functor" from Setup I* to Setup I which remembers (Y, λ, J) but forgets V and $\mathfrak{r}$. One has analogous forgetful functors for the other setups.

Setup II* A datum $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, \widehat{X}, \widehat{\lambda}, H, \widehat{J})$ for Setup II* consists of

- data $\mathcal{D}^\pm = ((Y^\pm, \xi^\pm, V^\pm), \mathfrak{r}^\pm, \lambda^\pm, J^\pm)$ for Setup I*, where we write $\mathfrak{r}^\pm = (\alpha_V^\pm, \tau^\pm, r^\pm)$,
- an exact relative symplectic cobordism $(\widehat{X}, \widehat{\lambda}, H)$ with positive end (Y^+, λ^+, V^+) and negative end $(Y^-, \lambda^-, V^\pm)$, and
- an $d\widehat{\lambda}$–tame almost complex structure $\widehat{J}$ on $\widehat{X}$ which agrees with $\widehat{J}^\pm$ at infinity.

This datum is moreover subject to the conditions that

- there exists a trivialization of the normal bundle of H which restricts to τ^+ (resp. τ^-) on the positive (resp. negative) end, and
- $\mathcal{E}(\widehat{J}) < \infty$ and $r^+ \geq e^{\mathcal{E}(\widehat{J})} r^-$.

Setup III* A datum $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, \widehat{X}, \widehat{\lambda}^t, H^t, \widehat{J}^t)_{t \in [0,1]}$ for Setup III* consists of

- data $\mathcal{D}^\pm = ((Y^\pm, \xi^\pm, V^\pm), \mathfrak{r}^\pm, \lambda^\pm, J^\pm)$ for Setup I*,
- a family of exact relative symplectic cobordisms $(\widehat{X}, \widehat{\lambda}^t, \widehat{H}^t)$ for $t \in [0, 1]$, with positive end (Y^+, λ^+, V^+) and negative end (Y^-, λ^-, V^-), and
- a family $d\widehat{\lambda}^t$–tame almost complex structures $\widehat{J}^t$ on $\widehat{X}$, which agree with $\widehat{J}^\pm$ at infinity.

This datum is moreover subject to the conditions that

- there exists a trivialization of the normal bundle of H which restricts to τ^+ (resp. τ^-) on the positive (resp. negative) end, and
- $\mathcal{E}(\widehat{J}^t) < \infty$ and $r^+ \geq e^{\mathcal{E}(\widehat{J}^t)} r^-$.

Setup IV* A datum $\mathcal{D} = (\mathcal{D}^{01}, \mathcal{D}^{12}, (\widehat{X}^{02,t}, \widehat{\lambda}^{02,t}, H^{02,t}, \widehat{J}^{02,t})_{t \in [0,\infty)})$ for Setup IV* consists of

- data $\mathcal{D}^i = ((Y^i, \xi^i, V^i); \mathfrak{r}^i, \lambda^i, J^i)$ for Setup I* for $i = 0, 1, 2$,
- a datum $\mathcal{D}^{01} = (\mathcal{D}^0, \mathcal{D}^1, \widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01}, \widehat{J}^{01})$ for Setup II*,
- a datum $\mathcal{D}^{12} = (\mathcal{D}^1, \mathcal{D}^2, \widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12}, \widehat{J}^{12})$ for Setup II*, and
- a family of cylindrical symplectic submanifolds $H^{02,t} \subset \widehat{X}^{02,t}$ for $t \in [0, \infty)$, such that
$$(\widehat{X}^{02,t}, \widehat{\lambda}^{02,t}, H^{02,t})_{t \in [0,\infty)}$$
is a family of exact relative symplectic cobordisms that agrees for t large with the t–gluing of the relative symplectic cobordisms $(\widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01})$ and $(\widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12})$.

This datum is moreover subject to the conditions that

- there exists a trivialization of the normal bundle of $H^{02,t}$ which restricts to τ^0 (resp. τ^2) on the positive (resp. negative) end,
- there exists a trivialization of the normal bundle of H^{01} which restricts to τ^0 (resp. τ^1) on the positive (resp. negative) end,

- there exists a trivialization of the normal bundle of H^{12} which restricts to τ^1 (resp. τ^2) on the positive (resp. negative) end,
- $\mathcal{E}(\hat{J}^{02,t}) < \infty$, $\mathcal{E}(\hat{J}^{01}) < \infty$ and $\mathcal{E}(\hat{J}^{12}) < \infty$, and
- $r^0 \geq e^{\mathcal{E}(\hat{J}^{02,t})} r^2$; $r^0 \geq e^{\mathcal{E}(\hat{J}^{01})} r^1$; and $r^1 \geq e^{\mathcal{E}(\hat{J}^{12})} r^2$.

We note that the requirement in Setups II*–IV* that the almost complex structures have finite energy is of course vacuous if r^-, r_1 and r_2 are nonzero (or equivalently, V^-, V_1 and V_2 are nonempty).

7.2 Twisting maps associated to enriched setups

In this section, we construct twisting maps on the contact homology algebra. These maps depend on geometric data involving codimension 2 contact submanifolds and relative symplectic cobordisms.

Setup I* Let $\mathcal{D} = ((Y, \xi, V), \mathfrak{r}, \lambda, J)$ be a datum for Setup I*, where $\mathfrak{r} = (\alpha_V, \tau, r)$. There is an obvious functor from $\mathcal{S}_{\mathrm{I}}(\mathcal{D})$ to the category $\hat{\mathcal{S}}(\hat{Y})$ defined in Section 5.3. We therefore have a well-defined intersection number $T * \hat{V}$ for $T \in \mathcal{S}_{\mathrm{I}}(\mathcal{D})$. We now introduce twisting maps associated to the above setup.

Definition 7.1 We define a map $\psi_V(T) \colon \mathcal{S}_{\mathrm{I}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}[U]$ by

$$\psi_V(T) = U^{T*\hat{V} + \Gamma^-(T,V)}, \tag{7-1}$$

where $\Gamma^-(T, V)$ denotes the number of output edges e of T such that the corresponding Reeb orbit γ_e is contained in V. Corollary 5.14 ensures that the exponents appearing in these definitions are nonnegative.

Remark 7.2 Corollary 5.14 only applies to trees T such that $\overline{\mathcal{M}}(T) \neq \varnothing$. This is why the definition of twisting maps only requires them to be defined on $\mathcal{S}^{\neq\varnothing}$ and not on the whole category $\mathcal{S}$.

Definition 7.3 We define a map $\tilde{\psi}_V \colon \mathcal{S}_{\mathrm{I}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}$ by

$$\tilde{\psi}_V(T) = \begin{cases} 1 & \text{if } T * \hat{V} = 0 \text{ and } |\gamma_e| \cap V = \varnothing \text{ for every } e \in E(T), \\ 0 & \text{otherwise.} \end{cases} \tag{7-2}$$

We must now check that the maps in Definitions 7.1 and 7.3 satisfy the axioms of Definition 6.1.

Proposition 7.4 *The map ψ_V introduced in Definition 7.1 is a twisting map.*

Proof It follows from Proposition 5.11 that $\psi_V(T) = \psi_V(T')$ for any morphism $T \to T'$.

Let $\{T_i\}_i$ be a concatenation in $\mathcal{S}_{\mathrm{I}}^{\neq\varnothing}$. We need to show that $\psi_V(\#_i T_i) = \prod_i \psi_V(T_i)$, ie

$$(\#_i T_i) * \hat{V} + \Gamma^-(\#_i T_i, V) = \sum_i T_i * \hat{V} + \Gamma^-(T_i, V). \tag{7-3}$$

We assume $V \neq \varnothing$ (otherwise there is nothing to say). Since the contact form λ is positive-elliptic near V, we have $p_N(\gamma) = 1$ for every Reeb orbit γ contained in V by Proposition 3.8. Remark 5.5 and Corollary 5.4 therefore imply that $\widehat{\gamma} * \widehat{V}$ is equal to -1 if γ is contained in V and 0 otherwise. By Proposition 5.12, this means that

$$T * \widehat{V} = \sum_{v \in V(T)} \beta_v * \widehat{V} + \Gamma^{\mathrm{int}}(T, V) \tag{7-4}$$

for all $T \in \mathcal{S}_{\mathrm{I}}$, where $\Gamma^{\mathrm{int}}(T, V)$ denotes the number of edges $e \in E^{\mathrm{int}}(T)$ such that γ_e is contained in V. Equation (7-3) is therefore equivalent to

$$\Gamma^{\mathrm{int}}(\#_i T_i, V) + \Gamma^{-}(\#_i T_i, V) = \sum_i (\Gamma^{\mathrm{int}}(T_i, V) + \Gamma^{-}(T_i, V)). \tag{7-5}$$

The result now follows from the observation that there is a (canonical) label-preserving bijection between $E^{\mathrm{int}}(\#_i T_i) \cup E^{-}(\#_i T_i)$ and $\bigcup_i (E^{\mathrm{int}}(T_i) \cup E^{-}(T_i))$ — this is an immediate consequence of the definition: every interior edge of T_j corresponds to an interior edge of $\#_i T_i$, and every output edge of T_j corresponds either to an interior or an output edge of $\#_i T_i$ depending on whether it is identified with another edge in the concatenation or not. □

It will be convenient to introduce the following definition.

Definition 7.5 Given a tree $T \in \mathcal{S}_{\mathrm{I}}$, a vertex $v \in V(T)$ is *mean* if it is an interior vertex and $|\gamma_e| \subset V$ for all $e \in e^{+}(v) \sqcup E^{-}(v)$. All other vertices are said to be *nice*. These sets are denoted by $V_m(T) \subset V(T)$ and $V_n(T) \subset V(T)$, respectively.

Remark 7.6 This notion of *nice/mean* vertices is purely auxiliary (and has nothing to do with good/bad Reeb orbits!). Geometrically, mean vertices correspond to holomorphic buildings which have intermediate orbits intersecting $\widehat{V}$. Nice orbits do not affect the intersection number of the building, but mean orbits do affect it and must therefore be treated carefully (hence the adjective).

Proposition 7.7 *The map $\widetilde{\psi}_V$ introduced in Definition 7.3 is a twisting map.*

Proof Fix a tree $T \in \mathcal{S}_{\mathrm{I}}^{\neq \varnothing}$. We first show that $\widetilde{\psi}_V(T') = \widetilde{\psi}_V(T)$ for any tree $T' \in \mathcal{S}_{\mathrm{I}}^{\neq \varnothing}$ admitting a morphism $T' \to T$. Observe that we may assume without loss of generality that T' is representable by a $\widehat{J}$–holomorphic building; see Definition 5.13. Indeed, since $T, T' \in \mathcal{S}_{\mathrm{I}}^{\neq \varnothing}$, there exists $T'' \to T' \to T$ such that T'' is representable by a $\widehat{J}$–holomorphic building. So we may as well prove that $\widetilde{\psi}_V(T'') = \widetilde{\psi}_V(T')$ and $\widetilde{\psi}_V(T'') = \widetilde{\psi}_V(T)$.

Let us therefore fix $T' \in \mathcal{S}_{\mathrm{I}}^{\neq \varnothing}$ such that T' is representable by a $\widehat{J}$–holomorphic building, and a morphism $T' \to T$. It follows from Proposition 5.11 that $T' * \widehat{V} = T * \widehat{V}$. Note that T' and T have the same exterior edges. If one of these edges is contained in V, then $\widetilde{\psi}_V(T') = \widetilde{\psi}_V(T) = 0$. So we can assume that the exterior edges of T' and T are not contained in V.

Suppose now that T' has an interior edge contained in V. For $i = 0, 1, 2$, let $X_i \geq 0$ be the number of edges $e \in E(T')$ such that $|\gamma_e| \subset V$ and e is adjacent to exactly i mean vertices. By assumption, we have $X_2 + X_1 + X_0 \geq 1$. According to Proposition 5.12, we have

$$T' * \widehat{V} = \sum_{v \in V_n(T')} \beta_v * \widehat{V} + \sum_{v \in V_m(T')} \beta_v * \widehat{V} + X_2 + X_1 + X_0. \tag{7-6}$$

According to Proposition 5.3, we also have that $\sum_{v \in V_n(T')} \beta_v * \widehat{V} \geq 0$ — here we use the fact that T' is representable by a $\widehat{J}$–holomorphic building. If there are no mean vertices, then we have that $\sum_{v \in V_m(T')} \beta_v * \widehat{V} = 0$, $X_1 = X_2 = 0$ and $X_0 \geq 1$. So $T' * \widehat{V} > 0$. If there exists at least one mean vertex, observe that we have $X_2 \leq \# V_m(T') - 1$. Moreover, given $v \in V_m(T')$, Proposition 5.6 together with the fact that T' is representable by a $\widehat{J}$–holomorphic building imply that $\beta_v * \widehat{V} \geq 1 - p_v$, where p_v is the number of edges adjacent to v. It follows that $\sum_{v \in V_m(T')} \beta_v * \widehat{V} + X_2 + X_1 \geq (\# V_m(T') - X_1 - 2X_2) + X_2 + X_1 = \# V_m(T') - X_2 \geq 1$. It thus follows again that $T' * \widehat{V} > 0$. We conclude that $\widetilde{\psi}_H(T') = \widetilde{\psi}_H(T) = 0$ if T' has an interior edge contained in V.

We are left with the case where T' and hence T have no edges contained in V. It is then immediate that $\widetilde{\psi}_V(T') = \widetilde{\psi}_V(T)$.

We now show that any concatenation $\{T_i\}_i$ satisfies $\widetilde{\psi}_V(\#_i T_i) = \prod_i \widetilde{\psi}_V(T_i)$. If one of the T_i has an edge contained in V, then $\#_i T_i$ also has an edge contained in V and we have $\widetilde{\psi}_V(\#_i T_i) = \prod_i \widetilde{\psi}_V(T_i) = 0$. If none of the T_i have an edge contained in V, then the same is true for $\#_i T_i$. Hence Proposition 5.12 implies that $\#_i T_i * \widehat{V} = \sum_i T_i * \widehat{V}$. By positivity of intersection (Proposition 5.3), $\sum_i T_i * \widehat{V} = 0$ if and only if $T_i * \widehat{V} = 0$ for all i. It then follows that $\widetilde{\psi}_V(\#_i T_i) = \prod_i \widetilde{\psi}_V(T_i)$. □

Setup II* Fix a datum $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, \widehat{X}, H, \widehat{\lambda}, \widehat{J})$ for Setup II*, where we write

$$\mathcal{D}^\pm = ((Y^\pm, \xi^\pm, V^\pm), \mathfrak{r}^\pm, \lambda^\pm, J^\pm) \quad \text{and} \quad \mathfrak{r}^\pm = (\alpha^\pm, \tau^\pm, r^\pm).$$

We now introduce the following twisting maps.

Definition 7.8 We define a map $\psi_H : \mathcal{S}_{\mathrm{II}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}[U]$ by

$$\psi_H(T) = U^{T * H + \Gamma^-(T, V^-)}. \tag{7-7}$$

Definition 7.9 We define a map $\widetilde{\psi}_H : \mathcal{S}_{\mathrm{II}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}$ by

$$\widetilde{\psi}_H(T) = \begin{cases} 1 & \text{if } T * H = 0 \text{ and } |\gamma_e| \cap V^\pm = \varnothing \text{ for every } e \in E(T), \\ 0 & \text{otherwise.} \end{cases} \tag{7-8}$$

We need to verify that the above definitions satisfy the axioms of twisting maps. The first step is to prove that the $\psi_H(T)$ are nonnegative powers of U. This is the content of Corollary 7.12, whose proof requires some preparatory lemmas. (In the next two lemmas, $\dot{\Sigma}$ always denotes an arbitrary punctured Riemann surface.)

Lemma 7.10 *For $n \geq 1$, suppose that $\beta \in \pi_2\big(\widehat{X}, \gamma^+ \sqcup \big(\bigcup_{i=1}^n \gamma_i^-\big)\big)$ is represented by a $\widehat{J}$–holomorphic curve $u\colon \dot{\Sigma} \to \widehat{X}$ which is contained in H. Then $P^+ - e^{-\mathcal{E}(\widehat{J})}\big(\sum_{i=1}^n P_i^-\big) \geq 0$, where P^+ (resp. P_i^-) is the period of $\gamma^+ \subset (Y^+, \lambda^+)$ (resp. the period of $\gamma_i^- \subset (Y^-, \lambda^-)$).*

Proof The claim is trivial if $\mathcal{E}(\widehat{J}) = \infty$, so let us assume that $\widehat{J} \in \mathcal{J}(\widehat{X}, \widehat{\lambda}, H)$. We may therefore fix a Type A decomposition σ of (H, λ_H), which is specified by a pair of hypersurfaces $\mathcal{H}_- = \{-C_1\} \times V^-$ and $\mathcal{H}_+ = \{C_+\} \times V^+$.

It will be convenient to define the regions $R^- := (-\infty, -C_1] \times V^-$, $R^+ := [C_2 \times \infty) \times V^+$ and $\widetilde{H} = H - (\operatorname{int}(R^-) \cup \operatorname{int}(R^+))$. Let us first assume that u is transverse to the boundary of $\widetilde{H}$. Consider now the sum

$$\int_{u^{-1}(R^-)} e^{-C_1} u^* d\alpha^- + \int_{u^{-1}(\widetilde{H})} u^* d\widehat{\lambda} + e^{C_2} \int_{u^{-1}(R^+)} u^* d\alpha^+. \tag{7-9}$$

Each summand is nonnegative due to the fact that u is $\widehat{J}$–holomorphic and that $\widehat{J}$ is adapted to σ. By Stokes' theorem, the sum of the integrals is $e^{C_2} P^+ - e^{-C_1}\big(\sum_{i=1}^n P_i^-\big) \geq 0$. This implies that $P^+ \geq e^{-\mathcal{E}(\sigma)}\big(\sum_{i=1}^n P_i^-\big)$.

If u is not transverse to the boundary of $\widetilde{H}$, observe by Sard's theorem that transversality can be achieved for a sequence of domains $\widetilde{H}^n := \widetilde{H} \cup [-C_1^n, -C_1] \cup [C_2, C_2^n]$, where $\{C_i^n\}_{n=0}^\infty$ is monotonically decreasing and $C_i^n \to C_i$. It is easy to verify that $\widehat{J}$ is still adapted to the Type A decompositions induced by the boundary of $\widetilde{H}^n$, so the above argument goes through and passing to the limit gives $P^+ \geq e^{-\mathcal{E}(\sigma)}\big(\sum_{i=1}^n P_i^-\big)$.

The lemma now follows from the definition of $\mathcal{E}(\widehat{J})$. □

Lemma 7.11 *For $n \geq 0$, suppose that $\beta \in \pi_2(\widehat{X}, \gamma^+ \sqcup (\bigcup_{i=1}^n \gamma_i^-))$ is represented by a $\widehat{J}$–holomorphic curve $u\colon \dot{\Sigma} \to \widehat{X}$. (Note that unlike in Lemma 7.10, we allow $n = 0$ in which case the union is interpreted as being empty.) Then $\beta * H \geq -n_u$, where n_u is the total number of negative punctures of u contained in $V^- \subset Y^-$.*

Proof According to Proposition 5.3, we only need to consider the case where the image of u is contained in H. By definition of a datum for Setup II*, the trivializations $\tau^\pm$ extend to a global trivialization τ of the normal bundle of H, which implies that $u^\tau \cdot H = 0$. Using the fact that $u^\tau \cdot H = 0$, we have (see Definition 5.1 and the proof of Proposition 5.6)

$$\begin{aligned} u * H = \alpha_N^{\tau;-}(\gamma^+) - \sum_{i=1}^n \alpha_N^{\tau;+}(\gamma_i^-) &= \lfloor \mathrm{CZ}_N^\tau(\gamma^+)/2 \rfloor - \sum_{i=1}^n \lceil \mathrm{CZ}_N^\tau(\gamma_i^-)/2 \rceil \\ &= \lfloor r^+ P^+ \rfloor - \sum_{i=1}^n \lfloor r^- P_i \rfloor, \end{aligned} \tag{7-10}$$

where the sum is interpreted as zero if u has no negative punctures.

We may assume that $n \geq 1$ and $r^- > 0$ (otherwise the lemma is automatic). Let $p_u = n_u + 1$ be the total number of punctures (positive and negative) of u contained in $V^\pm \subset Y^\pm$. Using the trivial bounds $x - 1 < \lfloor x \rfloor \leq x$, we obtain

$$u * H > (r^+ P^+ - 1) - \sum_{i=1}^{n} (1 + r^- P_i^-) = -p_u + r^+ P^+ - r^- \sum_{i=1}^{n} P_i^-. \tag{7-11}$$

Using now Lemma 7.10 and the fact that $r^+ \geq e^{\mathcal{E}(\hat{J})} r^-$, we have

$$-p_u + r^+ P^+ - r^- \sum_{i=1}^{n} P_i^- \geq -p_u + r^+ \epsilon^{-\mathcal{E}(\hat{J})} \sum_{i=1}^{n} P_i^- - r^- \sum_{i=1}^{n} P_i^- \geq -p_u. \tag{7-12}$$

The claim follows. □

Corollary 7.12 *We have $T * H \geq -\Gamma^-(T, V^-)$ for any $T \in \mathcal{S}_{\mathrm{II}}^{\neq\varnothing}(\mathcal{D})$. Hence $\psi_H(T) \in \mathbb{Q}[U]$.*

Proof Since $T \in \mathcal{S}_{\mathrm{II}}^{\neq\varnothing}(\mathcal{D})$, there exists $T' \to T$ such that T' is representable by a holomorphic building. Since the Siefring number is invariant under gluing (Proposition 5.11), we may assume that T is representable by a holomorphic building. We now apply Proposition 5.17: it therefore suffices to check that for each $v \in T$, the intersection number $\beta_v * \eta_{*(v)}$ is bounded below by $-\#\{E^-(v)\}$. In the case that $*(v) = 01$, this follows from Lemma 7.11. In the case that $*(v) = 00$ or $*(v) = 11$, this follows either from Lemma 7.11 or (more directly) from Proposition 5.6. □

Proposition 7.13 *Let $\{T_i\}_i$ be a concatenation in $\mathcal{S}_{\mathrm{II}}$. Then we have*

$$\begin{aligned}(\#_i\, T_i) * H &+ \Gamma^-(\#_i T_i, V^-) \\ &= \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^+} (T_i * \hat{V}^+ + \Gamma^-(T_i, V^+)) + \sum_{T_i \in \mathcal{S}_{\mathrm{II}}} (T_i * H + \Gamma^-(T_i, V^-)) + \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^-} (T_i * \hat{V}^- + \Gamma^-(T_i, V^-)).\end{aligned}$$

Proof As in the proof of Proposition 7.4, our assumptions imply that $\hat{\gamma} * V^\pm = -1$ if γ is contained in $V^\pm$ and 0 otherwise. By Proposition 5.15, we have

$$T * H = \sum_{\substack{v \in V(T) \\ *(v)=00}} \beta_v * \hat{V}^+ + \sum_{\substack{v \in V(T) \\ *(v)=01}} \beta_v * H + \sum_{\substack{v \in V(T) \\ *(v)=11}} \beta_v * \hat{V}^- + \Gamma^{\mathrm{int}}(T, V^+) + \Gamma^{\mathrm{int}}(T, V^-) \tag{7-13}$$

for all $T \in \mathcal{S}_{\mathrm{II}}$. By applying this formula to $T = \#_i\, T_i$ (and also using (7-4)), we see that it suffices to prove that

$$\begin{aligned}\Gamma^{\mathrm{int}}(\#_i T_i, V^+) &+ \Gamma^{\mathrm{int}}(\#_i T_i, V^-) + \Gamma^-(\#_i T_i, V^-) \\ &= \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^+} \Gamma^{\mathrm{int}}(T_i, V^+) + \Gamma^-(T_i, V^+) + \sum_{T_i \in \mathcal{S}_{\mathrm{II}}} \Gamma^{\mathrm{int}}(T_i, V^+) + \Gamma^{\mathrm{int}}(T_i, V^-) + \Gamma^-(T_i, V^-) \\ &\quad + \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^-} \Gamma^{\mathrm{int}}(T_i, V^-) + \Gamma^-(T_i, V^-).\end{aligned} \tag{7-14}$$

As in the proof of Proposition 7.4, this is just a matter of understanding how the edges of $\#_i T_i$ are obtained from the edges of the T_i, following the discussion in [61, Section 2.2]. More precisely, let us check that every edge counted on the right-hand side of (7-14) is also counted on the left-hand side. Note that under concatenation, interior edges remain interior edges. Output edges either remain output edges, or they become interior edges. The output edges corresponding to $\Gamma^-(T_i, V^+)$ for $T_i \in \mathcal{S}_{\mathrm{I}}^+$ must all become interior edges of $\#_i T_i$: indeed, any such output edge has label $*(e) = 0$, but the output edges of $\#_i T_i$ have label $*(e) = 1$. These output edges are thus counted in $\Gamma^{\mathrm{int}}(\#_i T_i, V^+)$.

The output edges corresponding to $\Gamma^-(T_i, V^-)$ for $T_i \in \mathcal{S}_{\mathrm{II}}$ may either become interior edges of $\#_i T_i$ (in which case they are counted in $\Gamma^{\mathrm{int}}(\#_i T_i, V^-)$), or remain output edges (in which case they are counted in $\Gamma^-(\#_i T_i, V^-)$). Similarly, the output edges corresponding to $\Gamma^-(T_i, V^-)$ for $T_i \in \mathcal{S}_{\mathrm{I}}^-$ may either become interior edges of $\#_i T_i$ (counted in $\Gamma^{\mathrm{int}}(\#_i T_i, V^-)$) or remain output edges (counted in $\Gamma^-(\#_i T_i, V^-)$). □

Corollary 7.14 *Under the assumptions of Proposition 7.13, $\psi_H \in \Psi_{\mathrm{II}}(\mathcal{D}; \psi_{V^+}, \psi_{V^-})$.*

Proof Proposition 5.11 implies that $\psi_H(T) = \psi_H(T')$ for any morphism $T \to T'$. Proposition 7.13 implies that ψ_H acts correctly on concatenations. □

We now want to show that $\widetilde{\psi}_H$ is a twisting map. We will need the following definition.

Definition 7.15 Given a tree $T \in \mathcal{S}_{\mathrm{II}}$, a vertex $v \in V(T)$ is *mean* if it is an interior vertex and $|\gamma_e| \subset V^\pm$ for all $e \in e^+(v) \sqcup E^-(v)$. All other vertices are said to be *nice*. These sets are denoted by $V_m(T) \subset V(T)$ and $V_n(T) \subset V(T)$, respectively.

Proposition 7.16 *Under the assumptions of Definition 7.9, $\widetilde{\psi}_H \in \Psi_{\mathrm{II}}(\mathcal{D}; \widetilde{\psi}_{V^+}, \widetilde{\psi}_{V^-})$.*

Proof Consider a tree $T' \in \mathcal{S}_{\mathrm{II}}^{\neq\varnothing}$ with a morphism $T' \to T$. We wish to show that $\widetilde{\psi}_H(T') = \widetilde{\psi}_H(T)$. As in the proof of Proposition 7.7, we may assume that T' is representable by a building; see Definition 5.16.

It follows from Proposition 5.11 that $T' * H = T * H$. Note that T' and T have the same exterior edges. If one of these edges is contained in $V^\pm$, then $\widetilde{\psi}_H(T') = \widetilde{\psi}_H(T) = 0$. So we can assume that the exterior edges of T' and T are not contained in $V^\pm$.

Suppose now that T' has an interior edge contained in $V^\pm$. Arguing as in the proof of Proposition 7.7, let $X_i \geq 0$ for $i = 0, 1, 2$ denote the number of edges $e \in E(T')$ such that $|\gamma_e| \subset V^\pm$ and e is adjacent to exactly i mean vertices. By assumption $X_0 + X_1 + X_2 \geq 1$. By Proposition 5.15, we have

$$T' * H = \sum_{v \in V_n(T')} \beta_v * H_v + \sum_{v \in V_m(T')} \beta_v * H_v + X_2 + X_1 + X_0, \tag{7-15}$$

where we write $H_v = \widehat{V}^+$ if $*(v) = 00$, $H_v = H$ if $*(v) = 01$ and $H_v = \widehat{V}^-$ if $*(v) = 11$. According to Proposition 5.6 and the fact that T' is representable by a building, we have that $\sum_{v \in V_n(T')} \beta_v * H_v \geq 0$.

If there are no mean vertices, then $\sum_{v\in V_m(T')}\beta_v * H_v = X_1 = X_2 = 0$ and $X_0 \geq 1$. Hence $T' * H \geq 1$. If there exists at least one mean vertex, observe that $X_2 \leq \#V_m(T') - 1$. According to Lemma 7.11 and the fact that T' is representable by a building, we have that $\sum_{v\in V_m(T)}\beta_v * H_v + X_2 + X_1 \geq \#V_m(T') - X_1 - 2X_2 + X_2 + X_1 = \#V_m(T') - X_2 \geq 1$. It thus follows again that $T' * \widehat{V} \geq 1$. We conclude that $\widetilde{\psi}_V(T') = \widetilde{\psi}_V(T) = 0$ if T' has an interior edge contained in $V^{\pm}$.

We are left with the case where T' and hence T have no edges contained in $V^{\pm}$. It's then immediate that $\widetilde{\psi}_H(T') = \widetilde{\psi}_H(T)$.

If $\{T_i\}_i$ is a concatenation, then the argument is the same as in the proof of Proposition 7.7 (using Proposition 5.15 instead of Proposition 5.12). □

Setup III* Fix a datum $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, \widehat{X}, \widehat{\lambda}^t, H^t, \widehat{J}^t)_{t\in[0,1]}$ for Setup III*, where

$$\mathcal{D}^{\pm} = ((Y^{\pm}, \xi^{\pm}, V^{\pm}), \mathfrak{r}^{\pm}, \lambda^{\pm}, J^{\pm}).$$

We now introduce the following twisting maps.

Definition 7.17 We define a map $\psi_{H^t} : \mathcal{S}_{\mathrm{III}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}[U]$ by

$$\psi_{H^t}(T) = U^{T*H^t+\Gamma^-(T,V^-)}. \tag{7-16}$$

Definition 7.18 We define a map $\widetilde{\psi}_{H^t} : \mathcal{S}_{\mathrm{III}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}$ by

$$\widetilde{\psi}_{H^t}(T) = \begin{cases} 1 & \text{if } T * H^t = 0 \text{ and } |\gamma_e| \cap V^{\pm} = \varnothing \text{ for every } e \in E(T), \\ 0 & \text{otherwise.} \end{cases} \tag{7-17}$$

There is no difference between $\mathcal{S}_{\mathrm{III}}$ and $\mathcal{S}_{\mathrm{II}}$ from the point of view of the intersection theory defined in Section 5.3. It can therefore be shown by essentially the same arguments as in the previous section that the above definitions do indeed satisfy the axioms for twisting maps.

Corollary 7.19 *We have* $\psi_{H^t} \in \Psi_{\mathrm{III}}(\mathcal{D}; \psi_{H^0}, \psi_{H^1})$ *and* $\widetilde{\psi}_{H^t} \in \Psi_{\mathrm{III}}(\mathcal{D}; \widetilde{\psi}_{H^0}, \widetilde{\psi}_{H^1})$.

Setup IV* Fix datum $\mathcal{D} = (\mathcal{D}^{01}, \mathcal{D}^{12}, (\widehat{X}^{02,t}, \widehat{\lambda}^{02,t}, H^{02,t}, \widehat{J}^{02,t})_{t\in[0,\infty)})$ for Setup IV*. Here,

- $\mathcal{D}^{01} = (\mathcal{D}^0, \mathcal{D}^1, \widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01}, \widehat{J}^{01})$ is a datum for Setup II*,
- $\mathcal{D}^{12} = (\mathcal{D}^1, \mathcal{D}^2, \widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12}, \widehat{J}^{12})$ is a datum for Setup II*, and
- $\mathcal{D}^i = ((Y^i, \xi^i, V^i); \mathfrak{r}^i, \lambda^i, J^i)$ is a datum for Setup I* for $i = 0, 1, 2$.

We introduce the following twisting maps.

Definition 7.20 We define $\psi_{H^{02,t}} : \mathcal{S}_{\mathrm{IV}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}[U]$ by

$$\psi_{H^{02,t}}(T) = U^{T*\eta+\Gamma^-(T,V^2)}. \tag{7-18}$$

Definition 7.21 We define $\widetilde{\psi}_{H^{02,t}} : \mathcal{S}_{\mathrm{IV}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}$ by

$$\widetilde{\psi}_{H^{02,t}}(T) = \begin{cases} 1 & \text{if } T * \eta = 0 \text{ and } |\gamma_e| \cap V^i = \varnothing \text{ for all } e \in E(T) \text{ and } i \in \{0,1,2\}, \\ 0 & \text{otherwise.} \end{cases} \tag{7-19}$$

We need to show that the powers of U appearing in Definition 7.20 are nonnegative. This will be the content of Corollary 7.24, which requires some preparatory lemmas.

Lemma 7.22 *For $n \geq 1$, suppose that $u : \dot{\Sigma} \to \widehat{X}^{02,t}$ is $\widehat{J}^{02,t}$–holomorphic with positive orbit γ^+ and negative orbits $\bigcup_{i=1}^n \gamma_i^-$. Then we have $P^+ - e^{-\mathcal{E}(\widehat{J}^{02,t})} \sum_i P_i^- \geq 0$, where P^+ (resp. P_i^-) is the period of $\gamma^+ \subset V^0$ (resp. $\gamma_i^- \subset V^2$).*

Proof The proof is analogous to that of Lemma 7.10. If $\mathcal{E}(\widehat{J}^{02,t}) = \infty$, the result is trivial. Hence we may assume that $\widehat{J}^{02,t} \in \mathcal{J}(\widehat{X}^{02,t}, \widehat{\lambda}^{02,t}, H^{02,t})$ and fix a Type B cobordism decomposition $\sigma^{02,t}$ of $(H^{02,t}, \widehat{\lambda}^{02,t}|_{H^{02,t}})$ to which $\widehat{J}^{02,t}$ is adapted. The decomposition $\sigma^{02,t}$ is specified by a family of hypersurfaces

$$\mathcal{H}_2(t) = \{-C_2(t)\} \times V^2, \quad \mathcal{H}_1(t) = \{-C_1(t)\} \times V^1, \quad \widetilde{\mathcal{H}}_1(t) = \{\widetilde{C}_1(t)\} \times V^1, \quad \mathcal{H}_0(t) = \{C_0(t)\} \times V^0.$$

It will be convenient to define the regions $R_2(t) = (-\infty, -C_2(t)] \times V^2$, $R_0(t) = [C_0(t), \infty) \times V^0$ and $R_1(t) = [-C_1(t), \widetilde{C}_1(t)] \times V^1$.

Suppose first that $R_1(t)$ is empty. Then $\widetilde{C}_1(t) + C_1(t) = 0$, and hence $\mathcal{E}(\sigma^{02,t}) = C_0(t) + C_2(t)$. Hence $\mathcal{E}(\sigma^{02,t})$ coincides with the energy of $\sigma^{02,t}$ if it is viewed as a Type A cobordism decomposition (Definition 6.8) by forgetting C_1 and $\widetilde{C}_1$. Hence, when $R_1(t)$ is empty, the claim reduces to Lemma 7.10.

We now assume that $R_1(t)$ is nonempty. We suppose that $\widetilde{H}_{12}^{02,t} \sqcup \widetilde{H}_{01}^{02,t}$ are the connected components of $X^{02,t} - \operatorname{int}\big(R_2(t) \cup R_1(t) \cup R_0(t)\big)$. Let us first assume that the image of u intersects the boundaries of $\widetilde{H}_{01}^{02,t}$ and $\widetilde{H}_{12}^{02,t}$ transversally. We then have the following computations:

- $\displaystyle\int_{u^{-1}(R_2(t))} u^*\alpha_2 = \int_{u^{-1}(\mathcal{H}_2(t))} u^*\alpha_2 - \sum_{i=1}^n P_i^- \geq 0,$
- $\displaystyle\int_{u^{-1}(H_{21}^{02,t})} u^* d(e^s \alpha_2) = e^{-C_1(t)} \int_{u^{-1}(\mathcal{H}_1(t))} u^*\alpha_1 - e^{-C_2(t)} \int_{u^{-1}(\mathcal{H}_2(t))} u^*\alpha_2 \geq 0,$
- $\displaystyle\int_{u^{-1}(R_1(t))} u^*\alpha_1 = \int_{u^{-1}(\widetilde{\mathcal{H}}_1(t))} u^*\alpha_1 - \int_{u^{-1}(\mathcal{H}_1(t))} u^*\alpha_1 \geq 0,$
- $\displaystyle\int_{u^{-1}(\widetilde{H}_{01}^{02,t})} u^* d(e^s \alpha_1) = e^{C_0(t)} \int_{u^{-1}(\mathcal{H}_0(t))} u^*\alpha_0 - e^{\widetilde{C}_1(t)} \int_{u^{-1}(\widetilde{\mathcal{H}}_1(t))} u^*\alpha_1 \geq 0,$
- $\displaystyle\int_{u^{-1}(R_0(t))} u^*\alpha_0 = P_i^+ - \int_{u^{-1}(\mathcal{H}_0(t))} u^*\alpha_0 \geq 0.$

After appropriate rescalings, these terms form a telescoping sum. We find that

$$P^+ - e^{C_2(t)+C_0(t)-C_1(t)-\widetilde{C}_1(t)} \sum_i P_i^- = P^+ - e^{-\mathcal{E}(\sigma^{02,t})} \sum_i P_i^- \geq 0.$$

Suppose now that the image of u does not intersect the boundaries of $\widetilde{H}_{01}^{02,t}$ and $\widetilde{H}_{12}^{02,t}$ transversally. For $\epsilon_n \downarrow 0$, set

$$R_2^{(n)} := (-\infty, -C_2(t)-\epsilon_n] \times V^2, \quad R_0^{(n)} = [C_0(t)+\epsilon_n, \infty) \times V^0, \quad R_1^{(n)} = [-C_1(t)+\epsilon_n, \widetilde{C}_1(t)-\epsilon_n] \times V^1.$$

By Sard's theorem, we may assume by choosing ϵ_n appropriately that u intersects the boundary of the $R_i^{(n)}$ transversally. Now repeat the above argument with $R_i^{(n)}$ in place of $R_i(t)$. This yields the inequality $P^+ - e^{C_2(t)+C_0(t)-C_1(t)-\widetilde{C}_1(t)+4\epsilon_n} \sum_i P_i^- \geq 0$. The claim now follows by passing to the limit. □

Lemma 7.23 *For $n \geq 0$, suppose that $u \colon \dot{\Sigma} \to \widehat{X}^{02,t}$ is a $\widehat{J}^{02,t}$–holomorphic curve in the homotopy class $\beta \in \pi_2\big(\widehat{X}^{02,t}, \gamma^+ \sqcup \big(\bigcup_{i=1}^n \gamma_i^-\big)\big)$ for $t < \infty$. Then $\beta * [H^{02,t}] \geq -n_u$, where n_u is the total number of negative punctures.* (Note that unlike in Lemma 7.22, we allow $n = 0$ here, in which case the union is interpreted as being empty.)

Proof We argue as in the proof of Lemma 7.11. It is enough to consider the case where the image of u is contained in $H^{02,t}$. The trivialization τ extends to a global trivialization along $H^{02,t}$, implying that $\beta \cdot_\tau H^{02,t} = 0$.

We thus have

$$\begin{aligned} u * H^{02,t} = \alpha_N^{\tau;-}(\gamma^+) - \sum_{i=1}^n \alpha_N^{\tau;+}(\gamma_i^-) &= \lfloor \mathrm{CZ}_N^\tau(\gamma^+)/2 \rfloor - \sum_{i=1}^n \lceil \mathrm{CZ}_N^\tau(\gamma_i^-)/2 \rceil \\ &= \lfloor r_0 P^+ \rfloor - \sum_{i=1}^n + \lfloor r P_i \rfloor, \end{aligned}$$

where the sum is interpreted as zero if u has no negative punctures. Thus the lemma is verified if $n = 0$ or $r_2 = 0$. It remains only to consider the case where $n \geq 1$ and $r_2 > 0$.

Using the trivial bounds $x - 1 < \lfloor x \rfloor \leq x$, we obtain

$$u * \widehat{V} > \sum_{z \in \boldsymbol{p}_u^+} (r_0 P_z - 1) - \sum_{z \in \boldsymbol{p}_u^-} (1 + r_2 P_z) \geq -p_u + e^{\mathcal{E}(\sigma^{02})} r_2 \bigg(\sum_{z \in \boldsymbol{p}_u^+} P_z \bigg) - r_2 \sum_{z \in \boldsymbol{p}_u^-} P_z.$$

It follows from Lemma 7.22 that $e^{\widehat{J}^{02,t}} r_2 \big(\sum_{z \in \boldsymbol{p}_u^+} P_z\big) - r_2 \sum_{z \in \boldsymbol{p}_u^-} P_z \geq 0$. The claim follows. □

Corollary 7.24 *We have that $T * \eta + \Gamma^-(T, V^2) \geq 0$.*

Proof We need to consider two cases. If $\mathfrak{s}(T) \in [0, \infty)$, then the claim follows by combining Proposition 5.17 and Lemma 7.23. If $\mathfrak{s}(T) = \{\infty\}$, then the argument is the same as in the proof of Corollary 7.12. □

Proposition 7.25 *Let $\{T_i\}_i$ be a concatenation in $\mathcal{S}_{\mathrm{IV}}$ of type* (2) *in Setup IV on page 45. Then we have*

$$\begin{aligned}(\#_i\, T_i) * \eta + \Gamma^-(T, V^2) &= \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^0} (T_i * \hat{V}^0 + \Gamma^-(T_i, V^0)) + \sum_{T_i \in \mathcal{S}_{\mathrm{II}}^{01}} (T_i * H^{01} + \Gamma^-(T_i, V^1)) + \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^1} (T_i * \hat{V}^1 + \Gamma^-(T_i, V^1)) \\ &\quad + \sum_{T_i \in \mathcal{S}_{\mathrm{II}}^{12}} (T_i * H^{12} + \Gamma^-(T_i, V^2)) + \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^2} (T_i * \hat{V}^2 + \Gamma^-(T_i, V^2)).\end{aligned}$$

Proof As in the proof of Proposition 7.4, our assumptions imply that $\hat{\gamma} * V^j = -1$ if γ is contained in V^j and 0 otherwise. Proposition 5.15 implies that

$$\begin{aligned}(\#_i T_i) * \eta &= \sum_{\substack{v \in V(\#_i T_i) \\ *(v)=00}} \beta_v * \hat{V}^0 + \sum_{\substack{v \in V(\#_i T_i) \\ *(v)=01}} \beta_v * H^{01} + \sum_{\substack{v \in V(\#_i T_i) \\ *(v)=11}} \beta_v * \hat{V}^1 + \sum_{\substack{v \in V(\#_i T_i) \\ *(v)=12}} \beta_v * H^{12} + \sum_{\substack{v \in V(\#_i T_i) \\ *(v)=22}} \beta_v * \hat{V}^2 \\ &\quad + \Gamma^{\mathrm{int}}(\#_i T_i, V^0) + \Gamma^{\mathrm{int}}(\#_i T_i, V^1) + \Gamma^{\mathrm{int}}(\#_i T_i, V^2).\end{aligned}$$

As in the proof of Proposition 7.13, it follows that the result is equivalent to

$$\begin{aligned}&\Gamma^{\mathrm{int}}(\#_i T_i, V^0) + \Gamma^{\mathrm{int}}(\#_i T_i, V^1) + \Gamma^{\mathrm{int}}(\#_i T_i, V^2) + \Gamma^-(T, V^2) \\ &\quad = \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^0} \Gamma^{\mathrm{int}}(T_i, V^0) + \Gamma^-(T_i, V^0) + \sum_{T_i \in \mathcal{S}_{\mathrm{II}}^{01}} \Gamma^{\mathrm{int}}(T_i, V^0) + \Gamma^{\mathrm{int}}(T_i, V^1) + \Gamma^-(T_i, V^1) \\ &\qquad + \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^1} \Gamma^{\mathrm{int}}(T_i, V^1) + \Gamma^-(T_i, V^1) + \sum_{T_i \in \mathcal{S}_{\mathrm{II}}^{12}} \Gamma^{\mathrm{int}}(T_i, V^1) + \Gamma^{\mathrm{int}}(T_i, V^2) + \Gamma^-(T_i, V^2) \\ &\qquad + \sum_{T_i \in \mathcal{S}_{\mathrm{I}}^2} \Gamma^{\mathrm{int}}(T_i, V^2) + \Gamma^-(T_i, V^2),\end{aligned}$$

which is a consequence of the way the edges of $\#_i T_i$ are obtained from the edges of the T_i. □

Corollary 7.26 *We have $\psi_{H^{02,t}} \in \Psi_{\mathrm{IV}}(\mathcal{D}; \psi_{H^{01}}, \psi_{H^{12}}, \psi_{H^{02,0}})$ and we have that*

$$\tilde{\psi}_{H^{02,t}} \in \Psi_{\mathrm{IV}}(\mathcal{D}; \tilde{\psi}_{H^{01}}, \tilde{\psi}_{H^{12}}, \tilde{\psi}_{H^{02,0}}).$$

Proof Proposition 7.25 shows that $\psi_{H^{02,t}}$ acts correctly on concatenations of type (2); the proof that $\psi_{H^{02,t}}$ behaves well with respect to the other two types of concatenation is virtually identical. Proposition 5.11 implies that $\psi_{H^{02,t}}(T) = \psi_{H^{02,t}}(T')$ for any morphism $T \to T'$. The argument that $\tilde{\psi}_{H^{02,t}}(T) = \tilde{\psi}_{H^{02,t}}(T')$ is essentially the same as the proof of Proposition 7.16, except that we appeal to Lemma 7.23 instead of Lemma 7.11. □

The results from the previous sections can be conveniently packaged into the following theorem.

Theorem 7.27 (cf [61, Theorem 1.1]) *Let $\mathcal{D}$ be a datum for any one of Setups I*–IV*. Then there exists a set of perturbation data $\Theta(\mathcal{D})$ and twisted moduli counts*

$$\#_{\psi}\overline{\mathcal{M}}(T)^{\mathrm{vir}}_{\theta} \in \mathbb{Q}[U] \quad \textit{and} \quad \#_{\widetilde{\psi}}\overline{\mathcal{M}}(T)^{\mathrm{vir}}_{\theta} \in \mathbb{Q}$$

for $\theta \in \Theta(\mathcal{D})$ and $T \in \mathcal{S}_*(\mathcal{D})$ for $* = \mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{IV}$, satisfying the obvious analogs of (i)–(v) *in Theorem* 1.1 *in* [61].

Proof There is a forgetful functor taking a datum for the enriched setups I*–IV* (Section 7.1) to a datum for the standard Setups I–IV (Section 4.1). So the set of perturbation data is furnished by [61, Theorem 1.1]. We showed in Section 7.2 a datum for Setups I*–IV* gives rise to twisting maps, from which we may define our twisted moduli counts as in Section 6.1. The properties (i)–(iv) are tautological and (v) is a consequence of the axioms of twisting maps, as explained in Section 6.1. □

8 Construction of the main invariants

In this section, we construct the invariants which are the central objects of this paper. To the data of a TN contact pair (Y, ξ, V) and an element $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$, we associate a unital, $\mathbb{Z}/2$–graded $\mathbb{Q}[U]$–algebra

$$CH_\bullet(Y, \xi, V; \mathfrak{r}). \tag{8-1}$$

There is a natural map to ordinary contact homology $CH_\bullet(Y, \xi, V; \mathfrak{r}) \to CH_\bullet(Y, \xi)$ given by setting $U = 1$.

A contactomorphism $f\colon (Y, \xi, V) \to (Y', \xi', V')$ induces an identification

$$CH_\bullet(Y, \xi, V; \mathfrak{r}) = CH_\bullet(Y', \xi', V'; f_*\mathfrak{r}).$$

An exact relative symplectic cobordism $(\hat{X}, \hat{\lambda}, H)$ from (Y^+, ξ^+, V^+) to (Y^-, ξ^-, V^-) satisfying an energy condition induces a map $CH_\bullet(Y^+, \xi^+, V^+; \mathfrak{r}^+) \to CH_\bullet(Y^-, \xi^-, V^-; \mathfrak{r}^-)$. Unfortunately, our notions of energy are not well behaved under compositions of arbitrary relative symplectic cobordisms, so the composition of maps is not always defined.

We also define a reduced version of (8-1), which only counts Reeb orbits in the complement of a codimension 2 submanifold, and certain "asymptotic invariants" which have good functoriality properties.

8.1 Construction and basic properties of the invariants

The following subsection is entirely parallel to [61, Section 1.7]. More precisely, Pardon constructs (ordinary) contact homology by applying [61, Theorem 1.1.] to data from Setups I–IV. We construct our new invariants by applying Theorem 7.27 to data from Setups I*–IV*.

Setup I* Fix a TN contact pair (Y, ξ, V) and an element $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$. According to Proposition 3.9, we may choose a contact form $\xi = \ker \lambda$ which is adapted to $\mathfrak{r}$. Let $J : \xi \to \xi$ be a $d\lambda$–compatible almost complex structure which preserves ξ_V. We therefore obtain a datum $\mathcal{D}$ for Setup I*. Theorem 7.27 applied to $\mathcal{D}$ furnishes a $\mathbb{Z}/2$–graded, unital $\mathbb{Q}[U]$–algebra

$$CH_\bullet(Y, \xi, V; \mathfrak{r})_{\lambda, J, \theta} \tag{8-2}$$

for any choice of perturbation datum $\theta \in \Theta_{\mathrm{I}}(\mathcal{D})$; cf (6-7) and (6-8).

Setup II* Fix pairs $\mathcal{D}^\pm = ((Y^\pm, \xi^\pm, V^\pm), \mathfrak{r}^\pm, \lambda^\pm, J^\pm)$ of data for Setup I*, where we write $\mathfrak{r}^\pm = (\alpha^\pm, \tau^\pm, r^\pm)$. Let $(\hat{X}, \hat{\lambda}, H)$ be an exact relative symplectic cobordism with positive end (Y^+, λ^+, V^+) and negative end (Y^-, λ^-, V^-), and suppose that there exists a trivialization of the normal bundle of H which restricts to $\tau^\pm$ on the positive/negative end.

Proposition 8.1 *Suppose that $r^+ > e^{\mathcal{E}(H, \hat{\lambda}|_H)} r^-$. Then there is an induced map on homology*

$$\Phi(\hat{X}, \hat{\lambda}, H)_{\hat{J}, \theta} : CH_\bullet(Y^+, \xi^+, V^+; \mathfrak{r}^+)_{\lambda^+, J^+, \theta^+} \to CH_\bullet(Y^-, \xi^-, V^-; \mathfrak{r}^-)_{\lambda^-, J^-, \theta^-}. \tag{8-3}$$

If $\alpha^+ = \alpha^-$ and $(H, \hat{\lambda}|_H)$ is a symplectization, then the same conclusion holds provided that $r^+ \geq r^-$.

Proof According to Lemma 6.25, we can choose an almost complex structure $\hat{J}$ on $\hat{X}$ which is $d\hat{\lambda}$–compatible and agrees with $\hat{J}^\pm$ at infinity, and is such that $r^+ \geq e^{\mathcal{E}(\hat{J})} r^-$. We thus obtain a datum $\mathcal{D} = (\mathcal{D}^+, \mathcal{D}^-, \hat{X}, H, \hat{\lambda}, \hat{J})$ for Setup II*.

Given $(\theta^+, \theta^-) \in \Theta_{\mathrm{I}}(\mathcal{D}^+) \times \Theta_{\mathrm{I}}(\mathcal{D}^-)$, Theorem 7.27 thus provides a perturbation datum $\theta \in \Theta_{\mathrm{II}}(\mathcal{D})$ with $\theta \mapsto (\theta^+, \theta^-)$, and twisted moduli counts which give rise to the map (8-3); cf (6-12). □

Setup III* We have the following proposition.

Proposition 8.2 *Under the assumptions of Proposition 8.1, the map* (8-3) *is independent of the pair* $(\hat{J}, \theta)$.

Proof Let $(\hat{J}_0, \theta_0)$ and $(\hat{J}_1, \theta_1)$ be two possible choices of such pairs. Let us first treat the case where $(H, \hat{\lambda}|_H)$ is not a symplectization. For any $\epsilon > 0$, Lemma 6.25 provides an interpolating family of almost complex structures $\{\hat{J}_t\}_{t \in [0,1]}$ such that $\mathcal{E}(\hat{J}_t) \leq \max(\mathcal{E}(\hat{J}_0), \mathcal{E}(\hat{J}_1)) + \epsilon$. Choosing ϵ small enough so that $r^+ > e^{\mathcal{E}(\hat{J}_t)} r^-$, we thus get a datum $\mathcal{D}$ for Setup III*.

Theorem 7.27 now provides perturbation data $\theta \in \Theta_{\mathrm{III}}(\mathcal{D})$ mapping to (θ_0, θ_1), and a chain homotopy between the maps $\Phi(\hat{X}, \hat{\lambda}, H)_{\hat{J}_0, \theta_0}$ and $\Phi(\hat{X}, \hat{\lambda}, H)_{\hat{J}_1, \theta_1}$; cf (6-18) and (6-20).

If $\alpha^+ = \alpha^-$ and $(H, \hat{\lambda}|_H)$ is a symplectization, then Lemma 6.25 implies that we may repeat the above argument for a family of almost complex structures $\hat{J}_t$ which have vanishing energy. Tracing through the proof, it is straightforward to check that the desired conclusion goes through provided that $r^+ \geq r^-$. □

Setup IV* Let us consider data $\widetilde{\mathcal{D}} = (\widetilde{\mathcal{D}}^{01}, \widetilde{\mathcal{D}}^{12}, (\widehat{X}^{02,t}, \widehat{\lambda}^{02,t})_{t\in[0,\infty)})$, where

$$\begin{aligned}
\widetilde{\mathcal{D}}^{01} &= (\mathcal{D}^0, \mathcal{D}^1, \widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01}),\\
\widetilde{\mathcal{D}}^{12} &= (\mathcal{D}^1, \mathcal{D}^2, \widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12}),\\
\mathcal{D}^i &= ((Y^i, \xi^i, V^i), \mathfrak{r}^i, \lambda^i, J^i) \quad \text{for } i = 0, 1, 2.
\end{aligned}$$

Here $\widetilde{\mathcal{D}}^{01}$, $\widetilde{\mathcal{D}}^{12}$ and $\widetilde{\mathcal{D}}$ are "partial data" for Setups II* and IV*, since they do not contain any information about almost complex structures. These "partial data" are assumed to obey all the axioms stated in Section 7.1 which do not involve complex structures.

The $\mathcal{D}^i$ are (ordinary) data for Setup I*.

Proposition 8.3 *Suppose that the following conditions hold*:

- $r_0 > e^{\mathcal{E}(H^{02,t}, \widehat{\lambda}^{02,t}|_{H^{02,t}})} r_2$,
- $r_0 > e^{\mathcal{E}(H^{01}, \widehat{\lambda}^{01}|_{H^{01}})} r_1$,
- $r_1 > e^{\mathcal{E}(H^{12}, \widehat{\lambda}^{12}|_{H^{12}})} r_2$.

Then the following diagram commutes:

$$\begin{array}{ccccc}
 & & CH_\bullet(Y^1, \xi^1, V^1; \mathfrak{r}^1)_{\lambda^1, J^1, \theta^1} & & \\
 & \nearrow^{\Phi(\widehat{X}^{01}, \widehat{\lambda}^{01}, H^{01})} & & \searrow^{\Phi(\widehat{X}^{12}, \widehat{\lambda}^{12}, H^{12})} & \\
CH_\bullet(Y^0, \xi^0, V^0; \mathfrak{r}^0)_{\lambda^0, J^0, \theta^0} & & \xrightarrow{\Phi(\widehat{X}^{02}, \widehat{\lambda}^{02}, H^{02})} & & CH_\bullet(Y^2, \xi^2, V^2; \mathfrak{r}^2)_{\lambda^2, J^2, \theta^2}
\end{array}$$

If $\alpha_0 = \alpha_1 = \alpha_2$ and if $(H^{01}, \widehat{\lambda}^{01}|_{H^{01}})$, $(H^{12}, \widehat{\lambda}^{12}|_{H^{12}})$ and $(H^{02,t}, \widehat{\lambda}^{02,t}|_{H^{02,t}})$ are symplectizations, then the conclusion still holds if we only assume that $r_i \geq r_j$ for $i \geq j$.

Proof According to Lemma 6.28, one can choose a family of almost complex structures $\widehat{J}^{02,t}$ so that $r_0 > e^{\mathcal{E}(\widehat{J}^{02,t})} r_2$. Moreover, by Lemma 6.29, one may also assume that $r_0 > e^{\mathcal{E}(\widehat{J}^{01})} r_1$ and $r_1 > e^{\mathcal{E}(\widehat{J}^{12})} r_2$, where J^{01} and J^{12} are the almost complex structures induced at infinity by $J^{02,t}$. We therefore obtain a datum for Setup IV* by considering $\mathcal{D}^{01} = (\widetilde{\mathcal{D}}^{01}, \widehat{J}^{01})$, $\mathcal{D}^{12} = (\widetilde{\mathcal{D}}^{12}, \widehat{J}^{12})$ and $\mathcal{D} = (\widetilde{\mathcal{D}}, \widehat{J}^{02,t})$. Theorem 7.27 applied to $\mathcal{D}$ now implies the commutativity of the above diagram; cf (6-27).

Under the additional hypotheses that $\alpha_0 = \alpha_1 = \alpha_2$ and that the relevant cobordisms are symplectizations, Lemmas 6.28 and 6.29 allow us to work with (families of) almost complex structures with vanishing energy. Retracing through the above argument, we find that the desired conclusion follows if $r_i \geq r_j$ for $i \geq j$. □

Proposition 8.4 *Let $(\widehat{X}, \widehat{\lambda}, H)$ be a relative symplectic cobordism from (Y^+, ξ^+, V^+) to (Y^-, ξ^-, V^-). Let $(\widehat{V}^\pm, \widehat{\lambda}^\pm_{\widehat{V}})$ be the Liouville structure induced on $\widehat{V}^\pm$ from the canonical Liouville structure of the symplectization $(\widehat{Y}^\pm, \lambda_{Y^\pm})$. For $i \in \{1, 2\}$, consider elements $\mathfrak{r}_i^\pm \in \mathfrak{R}(Y^\pm, \xi^\pm, V^\pm)$ and let $\lambda_i^\pm$ be a contact form on $Y^\pm$ which is adapted to $\mathfrak{r}_i^\pm$. Suppose finally that we have*:

(1) $r_i^+ > e^{\mathcal{E}(H,\hat{\lambda}|_H)} r_i^-$.

(2) $r_1^+ > e^{\mathcal{E}(\hat{V}^+,\hat{\lambda}^+_{\hat{V}})} r_2^+$ and $r_1^- > e^{\mathcal{E}(\hat{V}^-,\hat{\lambda}^-_{\hat{V}})} r_2^-$.

Then the following diagram commutes:

$$\begin{array}{ccc} CH_\bullet(Y^+,\xi^+,V^+;\mathfrak{r}_1^+)_{\lambda_1^+,J^+,\theta^+} & \xrightarrow{\Phi(\hat{Y}^+,\hat{V}^+)} & CH_\bullet(Y^+,\xi^+,V^+;\mathfrak{r}_2^+)_{\lambda_2^+,J^+,\theta^+} \\ \big\downarrow \Phi(\hat{X},\hat{\lambda},H) & & \big\downarrow \Phi(\hat{X},\hat{\lambda},H) \\ CH_\bullet(Y^-,\xi^-,V^-;\mathfrak{r}_1^-)_{\lambda_1^-,J^-,\theta^-} & \xrightarrow{\Phi(\hat{Y}^-,\hat{V}^-)} & CH_\bullet(Y^-,\xi^-,V^-;\mathfrak{r}_2^-)_{\lambda_2^-,J^-,\theta^-} \end{array}$$

As usual, if $\alpha_1^+ = \alpha_1^- = \alpha_2^+ = \alpha_2^-$ and $(H,\hat{\lambda}|_H)$ is a symplectization, then it is enough to assume that $r_i^+ \geq r_i^-$, $r_1^+ \geq r_2^+$ and $r_1^- \geq r_2^-$.

Proof Observe first that the conditions (1) and (2) along with Proposition 8.1 ensure that the maps appearing in the commutative diagram are well-defined. Let us now consider the strict exact symplectic cobordisms $(\hat{X},\hat{\lambda},H)^{\lambda_1^+}_{\lambda_1^-}$ and $(\hat{Y}^-,\lambda_2^-,\hat{V}^-)^{\lambda_1^-}_{\lambda_2^-}$. For $t \in [0,\infty)$ and $T_0 > 0$ large enough, we can consider their $(t+T_0)$–gluing $(\hat{X}^t,\hat{\lambda}^t,H^t)$; cf Definition 2.23. According to Lemma 6.22, we have that (cf Notation 2.19)

$$\mathcal{E}((H^t,\hat{\lambda}^t|_{H^t})^{\alpha_1^+}_{\alpha_2^-}) = \mathcal{E}((H,\hat{\lambda}|_H)^{\alpha_1^+}_{\alpha_1^-}) + \mathcal{E}((\hat{V}^-,\hat{\lambda}^-_{\hat{V}})^{\alpha_1^-}_{\alpha_2^-}). \tag{8-4}$$

It then follows from (1) and (2) that $r_1^+ > e^{\mathcal{E}((H^t,\hat{\lambda}^t|_{H^t})^{\alpha_1^+}_{\alpha_2^-})} r_2^-$.

We can now appeal to Proposition 8.3, which implies that the composition $\Phi(\hat{Y},\hat{V}^-) \circ \Phi(\hat{X},\hat{\lambda},H)$ agrees with the map induced by $(\hat{X}^0,\hat{\lambda}^0,H^0) = (\hat{X},\hat{\lambda},H)^{\lambda_1^+}_{\lambda_2^-}$; see Example 2.11. The same argument shows that composition along the upper right-hand side of the diagram agrees with the map induced by $(\hat{X},\hat{\lambda},H)^{\lambda_1^+}_{\lambda_2^-}$. This proves the claim. □

We obtain the following corollary by putting together the results of the previous section.

Corollary 8.5 *Consider a TN contact pair (Y,ξ,V) and fix an element $\mathfrak{r} \in \mathfrak{R}(Y,\xi,V)$. Let $\mathcal{D}^\pm = (Y,\xi,V),\mathfrak{r},\lambda^\pm,J^\pm)$ be a pair of data for Setup I*, and fix $\theta^\pm \in \Theta_{\mathrm{I}}(Y^\pm,\lambda^\pm,J^\pm)$.*

The map

$$\Phi(\hat{Y},\hat{\lambda},\hat{V})\colon CH_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda^+,J^+,\theta^+} \to CH_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda^-,J^-,\theta^-} \tag{8-5}$$

defined in Proposition 8.1 is an isomorphism.

Proof In light of Proposition 8.4 and Lemma 6.9, it's enough to consider the case $\lambda^+ = \lambda^- = \lambda$ and $J^+ = J^- = J$. Let $\theta \in \Theta_{\mathrm{II}}(\hat{Y},\hat{\lambda},\hat{J})$ be a lift of (θ^+,θ^-) under the forgetful map $\Theta_{\mathrm{II}}(\hat{Y},\hat{\lambda},\hat{J}) \to$

$\Theta_{\mathrm{I}}(Y,\lambda,J)\times\Theta_{\mathrm{I}}(Y,\lambda,J)$. The proof of [61, Lemma 1.2] can be adapted to show that the map

$$\Phi(\hat{Y},\hat{\lambda},\hat{V})_{\hat{J},\theta}\colon CC_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda,J,\theta^+}\to CC_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda,J,\theta^-} \tag{8-6}$$

is an isomorphism of chain complexes: one simply needs to observe that the twisted counts of trivial cylinders coincide with the usual counts. □

We now arrive at the definition of our main invariants.

Definition 8.6 (full invariant) Consider a TN contact pair (Y,ξ,V) and choose an element $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. Let

$$CH_\bullet(Y,\xi,V;\mathfrak{r}) \tag{8-7}$$

be the limit (or equivalently the colimit) of $\{CH_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda,J,\theta}\}_{\lambda,J,\theta}$ along the maps (8-5). Proposition 8.4 and Corollary 8.5 imply that $CH_\bullet(Y,\xi,V;\mathfrak{r})$ is canonically isomorphic to $CH_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda,J,\theta}$ for any admissible choice of (λ,J,θ).

Given $s\in\mathbb{Q}$, define

$$CH_\bullet^{U=s}(Y,\xi,V;\mathfrak{r}) := CH_\bullet(Y,\xi,V;\mathfrak{r})\otimes_{\mathbb{Q}[U]}\mathbb{Q}, \tag{8-8}$$

where the map $\mathbb{Q}[U]\to\mathbb{Q}$ sends $U\mapsto s$. There is a natural evaluation morphism of $\mathbb{Q}[U]$–algebras

$$\mathrm{ev}_{U=s}\colon CH_\bullet(Y,\xi,V;\mathfrak{r})\to CH_\bullet^{U=s}(Y,\xi,V;\mathfrak{r}). \tag{8-9}$$

It follows tautologically from the construction that $CH_\bullet^{U=1}(Y,\xi,V;\mathfrak{r}) = CH_\bullet(Y,\xi)$. The invariant $CH_\bullet(Y,\xi,V;\mathfrak{r})$ therefore admits a $\mathbb{Q}[U]$ algebra morphism to ordinary contact homology (which is viewed as a $\mathbb{Q}[U]$–algebra by letting U act by the identity).

We can also define a "reduced" variant of the invariants (8-7), which are based on the twisting map $\tilde{\psi}$. These invariants are naturally $\mathbb{Q}$–algebras (as opposed to $\mathbb{Q}[U]$–algebras) and only take into account Reeb orbits in the complement of the codimension 2 submanifold.

More precisely, given a datum $((Y,\xi,V),\mathfrak{r},\lambda,J)$ for Setup I*, we may proceed as in Setup I* in Section 8.1 and let

$$(\widetilde{CC}_\bullet(Y,\xi,V;\mathfrak{r})_\lambda, d_{\tilde{\psi},J,\theta}) \tag{8-10}$$

be the complex generated by the (good) Reeb orbits *not contained in* $V\subset Y$, for some perturbation datum $\theta\in\Theta_{\mathrm{I}}(\mathcal{D})$. By repeating the above arguments with the twisting maps $\tilde{\psi}_-$ in place of the twisting maps ψ_-, one can establish the obvious analogs of Proposition 8.4 and Corollary 8.5. In particular, given choices of data (λ^+,J^+,θ^+) and (λ^-,J^-,θ^-) as in Corollary 8.5, there is an isomorphism

$$\Phi(\hat{Y},\hat{\lambda},\hat{V})\colon \widetilde{CH}_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda^+,J^+,\theta^+}\to \widetilde{CH}_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda^+,J^+,\theta^+}. \tag{8-11}$$

Definition 8.7 (reduced invariant) Consider a TN contact pair (Y,ξ,V) and fix an element $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. Let

$$\widetilde{CH}_\bullet(Y,\xi,V;\mathfrak{r}) \tag{8-12}$$

be the limit (or equivalently the colimit) of the algebras $\{\widetilde{CH}_\bullet(Y,\xi,V;\mathfrak{r})_{\lambda,J,\theta}\}_{\lambda,J,\theta}$ along the maps (8-11).

For future reference, we record the following corollary of the above discussion.

Corollary 8.8 *Let $(Y^\pm,\xi^\pm,V^\pm)$ be TN contact pairs, and choose elements $\mathfrak{r}^\pm=(\alpha^\pm,\tau^\pm,r^\pm)\in\mathfrak{R}(Y^\pm,\xi^\pm,V^\pm)$. Let $(\widehat{X},\widehat{\lambda},H)$ be an exact relative symplectic cobordism with positive end (Y^+,ξ^+,V^+) and negative end (Y^-,ξ^-,V^-), and suppose that τ^+ and τ^- extend to a global trivialization of the normal bundle of H. If $r^+\geq e^{\mathcal{E}((H,\widehat{\lambda}|_H)^{\alpha^+}_{\alpha^-})}r^-$, then there is an induced map*

$$\Phi(\widehat{X},\widehat{\lambda},H)\colon CH_\bullet(Y^+,\xi^+,V^+;\mathfrak{r}^+)\to CH_\bullet(Y^-,\xi^-,V^-;\mathfrak{r}^-). \tag{8-13}$$

Similarly, suppose that $(\widehat{X},\widehat{\lambda}^t,H^t)_{t\in[0,1]}$ is a family of exact relative symplectic cobordisms with ends $(V^\pm,\xi^\pm,V^\pm)$ and such that $\tau^\pm$ extends to a global trivialization of the normal bundle of H^t. If $r^+\geq e^{\mathcal{E}(H^t,(\widehat{\lambda}^t)|_{H^t})}r^-$, then

$$\Phi(\widehat{X},\widehat{\lambda}^0,H^0)=\Phi(\widehat{X},\widehat{\lambda}^1,H^1). \tag{8-14}$$

The analogous statement holds for the reduced invariants $\widetilde{CH}_\bullet(-)$.

8.2 (Bi)gradings

The $\mathbb{Z}/2$–grading by parity on the deformed invariants $CH_\bullet(-),\widetilde{CH}_\bullet(-)$ shall be referred to as the *homological grading*. As in the case of (ordinary) contact homology, the homological grading can be lifted to a $\mathbb{Z}$–grading under certain topological assumptions. We will also to refer to this $\mathbb{Z}$–grading as the homological grading when it exists.

Definition 8.9 [61, Section 1.8] Let (Y^{2n-1},ξ,V) be a TN contact pair and choose $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. Suppose that $H_1(Y;\mathbb{Z})=0$ and $c_1(\xi)=0$. Then the homological $\mathbb{Z}/2$–grading lifts to a canonical $\mathbb{Z}$–grading defined on generators by

$$|\gamma|=\mathrm{CZ}^\tau(\gamma)+n-3, \tag{8-15}$$

where τ is any trivialization of the contact distribution along γ — this is independent of τ due to our assumption that $c_1(\xi)=0$.

Remark 8.10 In Definition 8.9, our assumption that $c_1(\xi)=0$ is equivalent to the statement that the canonical bundle $\Lambda^{n-1}_{\mathbb{C}}\xi$ is trivial. The grading in general depends on a trivialization of the canonical bundle; however, our assumption that $H_1(Y;\mathbb{Z})=0$ along with the universal coefficients theorem implies that $H^1(Y;\mathbb{Z})=0$. Hence the canonical bundle admits a unique trivialization.

Lemma 8.11 (see [61, equation (2.50)]) *With the notation of Corollary 8.8, suppose that* $H_1(Y^{\pm};\mathbb{Z})=0$ *and that* $c_1(\xi^{\pm}) = c_1(TX) = 0$. *Then the cobordism maps described in Corollary 8.8 preserve the homological* $\mathbb{Z}$*–grading.* □

Under certain topological assumptions, the reduced invariant $\widetilde{CH}_{\bullet}(-;-)$ admits an additional $\mathbb{Z}$–grading, which we will refer to as the *linking number grading*.

Definition 8.12 Let (Y,ξ,V) be a TN contact pair and choose $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. Suppose that $H_1(Y;\mathbb{Z}) = H_2(Y;\mathbb{Z}) = 0$. Then the *linking number grading* $|\cdot|_{\mathrm{link}}$ on $\widetilde{CH}_{\bullet}(Y,\xi,V;\mathfrak{r})$ is given on generators by

$$|\gamma|_{\mathrm{link}} = \mathrm{link}_V(\gamma). \tag{8-16}$$

(See Definition 5.19.) The linking number grading of a word of generators is then defined to be the sum of the linking number grading of each letter. One can verify using Lemma 5.21 that this grading is well-defined.

We let

$$\widetilde{CH}_{\bullet,\bullet}(Y,\xi,V;\mathfrak{r}) \tag{8-17}$$

be the (super)commutative bigraded $\mathbb{Q}$–algebra, where

- the first bullet refers to the homological $\mathbb{Z}$–grading (which exists in view of our topological assumption and the universal coefficients theorem, see Definition 8.9);
- the second bullet refers to the linking number $\mathbb{Z}$–grading.

We sometimes drop the second grading in our notation, so the reader should keep in mind that the notation $\widetilde{CH}_{\bullet}(-;-)$ always refers to the homological grading.

We have the following lemma as a consequence of Lemmas 5.21 and 8.11.

Lemma 8.13 *With the notation of Corollary 8.8, suppose that* $H_1(Y^{\pm};\mathbb{Z}) = H_2(X,Y^{+};\mathbb{Z}) = 0$. *Then the cobordism maps described in Corollary 8.8 preserve the linking number* $\mathbb{Z}$*–grading. If we also have that* $c_1(\xi^{\pm}) = c_1(TX) = 0$, *then the cobordism maps preserve the* $(\mathbb{Z}\times\mathbb{Z})$*–bigrading* (8-17). □

8.3 Asymptotic invariants

Given a TN contact pair (Y,ξ,V) and a trivialization τ of the normal bundle $N_{Y/V}$, let

$$\mathfrak{R}^{\tau}(Y,\xi,V) = \{\mathfrak{r}=(\alpha,\tau',r')\in\mathfrak{R}(Y,\xi,V) \mid \tau'=\tau\} \subset \mathfrak{R}(Y,\xi,V). \tag{8-18}$$

We equip $\mathfrak{R}^{\tau}(Y,\xi,V)$ with a preorder[6] $\preceq$ defined by setting $(\alpha_V^-,\tau,r^-)\preceq(\alpha_V^+,\tau,r^+)$ if $r^+ \geq e^{-\min f} r^-$, where $\alpha_V^+ = e^f\alpha_V^-$. We let $\preceq^{\mathrm{op}}$ denote the opposite preorder.

[6] A preorder on a set is a binary relation which is reflexive and transitive. Equivalently, a preordered set is a category with at most one morphism from any object x to any other object y.

We now define a functor $\mathcal{F}(Y,\xi,V)$ from the preordered set $(\mathfrak{R}^\tau(Y,\xi,V),\preceq^{\mathrm{op}})$ to the category of $\mathbb{Q}[U]$–algebras. On objects, the functor takes $\mathfrak{r}$ to $CH_\bullet(Y,\xi,V;\mathfrak{r})$. It remains to define the functor on morphisms.

Given elements $\mathfrak{r}^\pm=(\alpha_V^\pm,\tau,r^\pm)\in\mathfrak{R}^\tau(Y,\xi,V)$, let $\lambda^\pm$ be a contact form on Y which is adapted to $\mathfrak{r}^\pm$. Consider the symplectization $(\hat{Y},\hat{\lambda},\hat{V})_{\lambda^-}^{\lambda^+}$. If $\mathfrak{r}^-\preceq\mathfrak{r}^+$, then Lemma 6.9 and Propositions 8.1 and 8.4 imply that there is a map

$$\Phi(\hat{Y},\hat{\lambda},\hat{V})\colon CH_\bullet(Y,\xi,V;\mathfrak{r}^+)\to CH_\bullet(Y,\xi,V;\mathfrak{r}^-). \tag{8-19}$$

This defines $\mathcal{F}(Y,\xi,V)$ on morphisms. One can check using Proposition 8.4 that $\mathcal{F}(Y,\xi,V)$ is indeed a functor.

We can similarly define a functor $\mathcal{F}_+(Y,\xi,V)$ from $(\mathfrak{R}^\tau(Y,\xi,V),\preceq^{\mathrm{op}})$ to the category of $\mathbb{Q}$–algebras using $\widetilde{CH}_\bullet(-)$.

Definition 8.14 (asymptotic invariants) Noting that the category of $\mathbb{Q}[U]$–algebras is complete and cocomplete, we denote by

$$\underleftarrow{CH}_\bullet(Y,\xi,V;\tau)\text{ and }\underrightarrow{CH}_\bullet(Y,\xi,V;\tau) \tag{8-20}$$

the limit (resp. colimit) of the $\mathbb{Q}[U]$–algebras $\{CH_\bullet(Y,\xi,V;\mathfrak{r})\}$ over the preordered set $(\mathfrak{R}^\tau(Y,\xi,V),\preceq^{\mathrm{op}})$. We let $\underleftarrow{\widetilde{CH}}_\bullet(Y,\xi,V;\tau)$ and $\underrightarrow{\widetilde{CH}}_\bullet(Y,\xi,V;\tau)$ be defined similarly over the category of $\mathbb{Q}$–algebras.

It's easy to check that $(\mathfrak{R}^\tau(Y,\xi,V),\preceq^{\mathrm{op}})$ is a filtered preordered set. In particular, (co)limits can be computed by restricting to (co)final subsets. In contrast to the invariants defined in Section 8.1, the asymptotic invariants are fully functorial under compositions of arbitrary relative symplectic cobordisms which respect normal trivializations. (A verification of this is tedious and essentially consists of repeating the arguments of Section 8.1 — the key is that the energy conditions can always be satisfied by sending the rotation parameter to zero or infinity.)

8.4 Mixed morphisms

Consider a TN contact pair (Y,ξ,V) and elements $\mathfrak{r}^\pm=(\alpha^\pm,\tau^\pm,r^\pm)\in\mathfrak{R}(Y,\xi,V)$. In this section, we exhibit a $\mathbb{Q}$–algebra map

$$CH_\bullet^{U=0}(Y,\xi,V;\mathfrak{r}^+)\to\widetilde{CH}_\bullet(Y,\xi,V;\mathfrak{r}^-) \tag{8-21}$$

under certain assumptions on r^+ and r^-. Precomposing with (8-9) gives a $\mathbb{Q}$–algebra map

$$CH_\bullet(Y,\xi,V;\mathfrak{r}^+)\to\widetilde{CH}_\bullet(Y,\xi,V;\mathfrak{r}^-). \tag{8-22}$$

Let us begin by considering a datum $\mathcal{D}=(\mathcal{D}^+,\mathcal{D}^-,\hat{X},H,\hat{\lambda},\hat{J})$ for Setup II*, where we let $\mathcal{D}^\pm=((Y,\xi,V),\mathfrak{r}^\pm,\lambda^\pm,J^\pm)$.

Definition 8.15 We define a map $\psi_{\text{mix}}\colon \mathcal{S}_{\text{II}}^{\neq\varnothing}(\mathcal{D}) \to \mathbb{Q}$ by

$$\psi_{\text{mix}}(T) = \begin{cases} 1 & \text{if } T * \hat{V} = 0 \text{ and } |\gamma_e| \cap V^- = \varnothing \text{ for every } e \in E(T), \\ 0 & \text{otherwise.} \end{cases} \tag{8-23}$$

Proposition 8.16 *The map $\psi_{\text{mix}}(-)$ is a twisting map provided that the following properties hold*:

(i) $r^+ \geq 2e^{\mathcal{E}(\hat{J})} r^-$, *and*

(ii) $r^+ > 2/R_\alpha^{\min}$, *where $R_\alpha^{\min}$ denotes the smallest action of all Reeb orbits of α.*

We need the following lemma for proving Proposition 8.16.

Lemma 8.17 *Suppose that $\beta \in \pi_2\big(\hat{Y}, \gamma^+ \sqcup \big(\bigcup_{i=1}^n \gamma_i^-\big)\big)$ is represented by a $\hat{J}$–holomorphic curve $u\colon \dot{\Sigma} \to \hat{X}$ contained in $\hat{V}$. Suppose also that r^+ and r^- satisfy assumptions* (i) *and* (ii) *in Definition 8.15. Then $\beta * \hat{V} \geq 2 - p_u$, where p_u is the total number of punctures* (*positive and negative*) *of u contained in $V^\pm$.*

Proof By Lemma 7.10, we have $r^+P^+ - r^- \sum_{i=1}^n P_i^- \geq r^+P^+ - r^-P^+e^{\mathcal{E}(\hat{J})} \geq P^+(r^+ - r^-e^{\mathcal{E}(\hat{J})}) \geq R_\alpha^{\min}((r^+/2 - r^-e^{\mathcal{E}(\hat{J})}) + r^+/2) \geq R_\alpha^{\min} r^+/2 \geq 1$. The lemma now follows from Proposition 5.7. □

Definition 8.18 Given a tree $T \in \mathcal{S}_{\text{II}}^{\neq\varnothing}$, we say that a vertex $v \in V(T)$ is *mean* if all adjacent edges are contained in $V^\pm$. Otherwise, we say that $v \in V(T)$ is *nice*. We denote by $V_b(T)$ (resp. $V_n(T)$) the set of mean (resp. nice) vertices of T.

Proof of Proposition 8.16 Choose a tree $T' \in \mathcal{S}_{\text{II}}^{\neq\varnothing}$. Let $T' \to T$ be a morphism. It follows from Proposition 5.11 that $T' * \hat{V} = T * \hat{V}$. If T' has no edges contained in V^-, then neither does T' and we see that $\psi_{\text{mix}}(T') = \psi_{\text{mix}}(T)$.

Let us now suppose that T' has an edge contained in V^-. Note that T' and T have the same exterior edges. If one of these edges is contained in V^-, then $\psi_{\text{mix}}(T') = \psi_{\text{mix}}(T) = 0$. Let us therefore assume that the exterior edges of T' and T are not contained in V^-.

We are left with the case where T' has at least one interior edge contained in V^-. If T' had no mean vertices, then it would follow from Proposition 5.15 that there are no interior edges contained in $V^\pm$, which is a contradiction. It follows that T' has at least one mean vertex. Let $E_b^{\text{int}}(T') \subset E^{\text{int}}(T')$ be the set of interior edges which occur as an outgoing edge of some mean vertex. According to Proposition 5.15, we have

$$\begin{aligned} T' * \hat{V} &= \sum_{v \in V_n(T')} \beta_v * \hat{V} + \sum_{v \in V_m(T')} \beta_v * \hat{V} + |E^{\text{int}}(T') - E_b^{\text{int}}(T')| + |E_b^{\text{int}}(T')| \\ &\geq \sum_{v \in V_m(T')} \beta_v * \hat{V} + |E_b^{\text{int}}(T')| \\ &= \sum_{v \in V_m(T')} (\beta_v * \hat{V} + p_v^-), \end{aligned}$$

where p_v^- denotes the number of outgoing edges of v. (Here, we have used the fact that the outgoing edges of a mean vertex are all interior edges, which follows from our assumption that the exterior edges of T' and T are not contained in V^-.) It now follows from Lemma 8.17 and the fact that T' has at least one mean vertex that $\sum_{v \in V_m(T')}(\beta_v * \hat{V} + p_v) \geq \sum_{v \in V_m(T')}(2 - p_v + p_v^-) \geq \sum_{v \in V_m(T')} 1 \geq 1$. Hence $\psi_{\mathrm{mix}}(T') = \psi_{\mathrm{mix}}(T) = 0$. This completes the proof that $\psi_{\mathrm{mix}}(T') = \psi_{\mathrm{mix}}(T)$.

If $\{T_i\}_i$ is a concatenation, then the argument is the same as in the proof of Proposition 7.13 since every edge in $\#_i T_i$ appears in at least one of the T_i. □

Proposition 8.19 *Consider a TN contact pair (Y, ξ, V) and elements $\mathfrak{r}^{\pm} = (\alpha^{\pm}, \tau^{\pm}, r^{\pm}) \in \mathfrak{R}(Y, \xi, V)$. If $r' > 2e^{\mathcal{E}(\hat{V}, \hat{\lambda}|_{\hat{V}})} r^-$ and $r^+ > 2/R_\alpha^{\min}$, then there is a map of $\mathbb{Q}$–algebras*

$$CH_\bullet^{U=0}(Y, \xi, V; \mathfrak{r}^+) \to \widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r}^-). \tag{8-24}$$

Proof The argument is essentially the same as the proof of Proposition 8.1. Choose data of Type I* $\mathcal{D}^{\pm} = ((Y, \xi, V), \mathfrak{r}^{\pm}, \lambda^{\pm}, J^{\pm})$. Now consider the symplectization $(\hat{Y}, \hat{\lambda}, \hat{V})$. Lemma 6.25 furnishes an almost complex structure $\hat{J}$ on $\hat{Y}$ which is $d\hat{\lambda}$–compatible and agrees with $\hat{J}^{\pm}$ at infinity, and such that $r^+ \geq e^{\mathcal{E}(\hat{J})} r^-$. It now follows as in Setup II of Section 6.1 that we have a $\mathbb{Q}$–algebra chain map

$$\Phi(\hat{Y}, \hat{\lambda}, \psi_{\mathrm{mix}})_{\hat{J}, \Theta} \colon CC_\bullet^{U=0}(Y, \xi, V; \mathfrak{r}')_{J^+, \theta^+} \to \widetilde{CC}_\bullet(Y, \xi, V; \mathfrak{r})_{J^-, \theta^-} \tag{8-25}$$

for perturbation data $\Theta \mapsto (\theta^+, \theta^-)$. □

Remark 8.20 The proof of Proposition 8.19 does not show that (8-24) is independent of auxiliary choices (ie $\hat{J}, J^{\pm}, \Theta, \theta^{\pm}$). To show this, one needs to extend the definition of the twisting map ψ_{mix} to Setups III* and IV*. One can then prove analogs of Theorem 7.27 and Corollary 8.8 for ψ_{mix}. All of the ingredients for this are already in place, but we omit the details since Proposition 8.19 is sufficient for our applications.

Corollary 8.21 *Suppose that $\widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r}) \neq 0$ for some $\mathfrak{r} = (\alpha_V, \tau, r) \in \mathfrak{R}(Y, \xi, V)$. Letting $\mathfrak{r}' = (\alpha_V', \tau', r')$, we have $CH_\bullet(Y, \xi, V; \mathfrak{r}') \neq 0$ if r' is large enough. In particular, $\underleftarrow{CH}_\bullet(Y, \xi, V; \tau) \neq 0$.* □

9 Augmentations and linearized invariants

9.1 Differential graded algebras

Let R be a commutative ring containing $\mathbb{Q}$. The following two categories occur naturally in contact topology: the category **cdga** of unital $\mathbb{Z}$–graded (super)commutative dg R–algebras; the category **dga** of unital $\mathbb{Z}$–graded associative dg R–algebras. When we speak of a *dg algebra*, we mean an object of either one of these categories.

Let us say that a dg algebra is *action-filtered* if the underlying graded algebra is the free algebra on a free graded module U having the following property: U admits a basis $\{x_\alpha \mid \alpha \in \mathcal{A}\}$ for some well-ordered set $\mathcal{A}$ such that dx_α is a sum of words in the letters x_β for $\beta < \alpha$. *Note that the dg algebras which arise in Symplectic Field Theory are naturally action-filtered by Reeb length.*

The categories **cdga** and **dga** each carry a model structure described by Hinich [34] (see also Section 1.11 of [66]) with the following properties:

(i) the weak equivalences are the quasi-isomorphisms, and

(ii) all objects are fibrant and the set of cofibrant objects includes all objects which are *action-filtered*.[7]

Note that in any model category, there is a notion of two maps being left (resp. right) homotopic, defined in terms of cylinder (resp. path) objects [35, Section 7.3.1]. These notions coincide on objects which are fibrant and cofibrant.

We let **hcdga** and **hdga** be the associated homotopy categories: the objects are those objects of **cdga** (resp. **dga**) which are fibrant and cofibrant, and the morphisms are the homotopy classes of morphisms in **cdga** (resp. **dga**). *In particular*, **hcdga** *and* **hdga** *contain all action-filtered dg algebras.*

An *augmentation* of a dg R–algebra A is a morphism of dg algebras $\epsilon\colon A \to R$, where R is viewed as a dg algebra concentrated in degree 0. A dg algebra equipped with an augmentation is said to be augmented. We let $\mathbf{cdga}_{/R}$ and $\mathbf{dga}_{/R}$ be the overcategory of augmented objects, which naturally inherit model structures. We let $\mathbf{hcdga}_{/R}$ (resp. $\mathbf{hdga}_{/R}$) be the associated homotopy categories.

Definition 9.1 Given an augmented dg algebra $\epsilon\colon (A, d) \to R$, we consider the graded R–module $A_\epsilon := \ker\epsilon/(\ker\epsilon)^2$. The differential d descends to a differential d_ϵ on A_ϵ. The resulting differential graded module (A_ϵ, d_ϵ) is called the *linearization*[8] of (A, d) at the augmentation ϵ.

It follows from the definition that linearization defines a functor from $\mathbf{cdga}_{/R}$ (resp. $\mathbf{dga}_{/R}$) to the category of chain complexes of R–modules. It is an important fact that this functor is left Quillen [51, Proposition 12.1.7], and therefore induces a functor of homotopy categories. We state this as a corollary:

Corollary 9.2 *Linearization defines a functor from* $\mathbf{hcdga}_{/R}$ *(resp.* $\mathbf{hdga}_{/R}$*) to the homotopy category of chain complexes of R–modules.* □

[7] Both [34] and [51] work in the setting of dg algebras over operads. The categories **cdga** and **dga** are, respectively, the category of dg algebras over the operads uComm and uAss (u stands for *unital*), so they can be treated on equal footing. A good reference for the material in this section, which mostly avoids operadic language (but only treats the case of **cdga**), is [47].

[8] The linearization is sometimes also called the "indecomposable quotient" in the rational homotopy theory literature. In the contact topology literature, one also often encounters an equivalent construction of the linearization in which one twists the differential by the augmentation; see [57, Remark 2.8].

In particular, Corollary 9.2 implies that homotopic augmentations of action-filtered dg algebras induce isomorphic linearizations. Special cases of this statement have already appeared in the Legendrian contact homology literature; see eg [58, Section 5.3.2].

Remark 9.3 Let (A, d) be a dg algebra, where A is the free R–algebra generated by the set $\{x_\alpha \mid \alpha \in \mathcal{A}\}$. Let $\epsilon\colon A \to R$ be the unique R–algebra map sending the generators x_α to zero. Then ϵ is an augmentation if and only if dx_α is contained in the ideal $(x_\beta \mid \beta \in \mathcal{A})$ for all $\alpha \in \mathcal{A}$ (or equivalently, if and only if the differential has no constant term). If ϵ is an augmentation, it is called the *zero augmentation*.

Suppose now that (A, d) is the (possibly deformed) contact algebra of some contact manifold, ie (A, d) is the commutative R–algebra generated by good Reeb orbits (for $R = \mathbb{Q}$ or $\mathbb{Q}[U]$) and the differential is defined as in Section 6.1. Suppose that $\epsilon\colon A \to R$ is the zero augmentation. Then $\ker \epsilon$ is the free R–module generated by (good) Reeb orbits, and d_ϵ counts curves with one input and one output (ie d_ϵ is defined as in (6-7), where the sum is restricted to curves with $|\Gamma^-| = 1$). It follows that the homology of the complex (A_ϵ, d_ϵ) can be interpreted as the (possibly deformed) *cylindrical* contact homology. This latter invariant admits a rigorous definition for contact structures under certain assumptions, such as the existence of a contact form with no contractible Reeb orbits; see [4; 42] in dimension 3 and [61, Section 1.8] in general.

9.2 Cyclic homology

Definition 9.4 (cf [17]) Let S be a countable, well-ordered set equipped with a map $|\cdot|\colon S \to \mathbb{Z}$. Let $A = R\langle S\rangle$ be the free $\mathbb{Z}$–graded R–algebra generated by S, where the $\mathbb{Z}$–grading is induced by extending $|\cdot|$ multiplicatively.

Let $d\colon A \to A$ be a differential of degree -1. Let $\bar{A} := A/R$, and consider the cyclic permutation map $\tau\colon \bar{A} \to \bar{A}$ which is defined on monomials by $\tau(\gamma_1 \cdots \gamma_l) = (-1)^{|\gamma_1|(|\gamma_2|+\cdots+|\gamma_l|)}\gamma_2 \cdots \gamma_l \gamma_1$ and extended R–linearly. Let $\bar{A}^\tau := \bar{A}/(1-\tau)$ be the $\mathbb{Z}$–graded R–module of coinvariants. Observe that d passes to the quotient. We denote the induced differential by d^τ.

We now define

$$\overline{HC}_\bullet(A) := H_\bullet(\bar{A}^\tau, d^\tau), \tag{9-1}$$

and refer to this invariant as the *reduced cyclic homology* of the dg algebra (A, d).

Remark 9.5 Definition 9.4 agrees with other definitions of reduced cyclic homology of dg algebras (such as [50, Section 5.3]) which may be more familiar to the reader, when both are defined. We adopt the present definition for consistency with [7].

In the special case where A is the Chekanov–Eliashberg dg algebra of a Legendrian knot in a contact manifold satisfying the assumptions of [7, Section 4.1], the algebraic invariants considered in [7, Section 4] can be translated as follows: $L\mathbb{H}^{\mathrm{cyc}}(A) = \overline{HC}_\bullet(A)$, $L\mathbb{H}^{\mathrm{Ho}+}(A) = \overline{HH}_\bullet(A)$ and $L\mathbb{H}^{\mathrm{Ho}}(A) = HH_\bullet(A)$. Here $HH_\bullet(-)$ and $\overline{HH}_\bullet(-)$ denote, respectively, Hochschild homology and reduced Hochschild homology.

We record the following computation which will be useful to us later on.

Lemma 9.6 *Under the assumptions of Definition 9.4, if (A,d) is acyclic, then*

$$\overline{HC}_k(A) = \begin{cases} R & \text{if } k \text{ is odd and positive,} \\ 0 & \text{otherwise.} \end{cases} \tag{9-2}$$

Proof Let us first prove the lemma under the assumption that S is a finite set. Note first that the Hochschild homology of an acyclic finitely generated dg algebra vanishes identically. Moreover, we have an exact triangle

$$R[0] \to HH_\bullet(A) \to \overline{HH}_\bullet(A) \xrightarrow{[-1]}, \tag{9-3}$$

which implies that $\overline{HH}_\bullet(A)$ is just a copy of R concentrated in degree 1.

We now consider the following Gysin-type exact triangle (see [7, Proposition 4.9]):

$$\overline{HC}_\bullet(A) \xrightarrow{[-2]} \overline{HC}_\bullet(A) \xrightarrow{[+1]} \overline{HH}_\bullet(A) \xrightarrow{[0]} . \tag{9-4}$$

The desired result now follows immediately by induction, using (9-4) and the fact that $\overline{HC}_\bullet(A)$ vanishes in sufficiently large positive and negative degrees due to our finiteness hypotheses. We remark that (9-4) is constructed from a spectral sequence, whose convergence can only be verified under finiteness assumptions.

Let us now drop our assumption that S is a finite set. We instead consider an exhaustion of S by finite subsets $S^{(1)} \subset S^{(2)} \subset \cdots \subset S$. Let $(A^{(k)}, d) \subset (A, d)$ be the dg subalgebra generated by $S^{(k)}$. One can readily verify that

$$\varinjlim \overline{HC}_\bullet(A^{(k)}) = \overline{HC}_\bullet(\varinjlim A^{(k)}) = \overline{HC}_\bullet(A). \tag{9-5}$$

Observe that $(A^{(k)}, d)$ is acyclic for k large enough and satisfies the assumption of Definition 9.4. Since we have already proved the lemma under the assumption that S is finite, it is enough to prove that the natural maps $\overline{HC}_\bullet(A^{(k)}) \to \overline{HC}_\bullet(A^{(k+1)})$ are isomorphisms for k large enough.

To this end, note that the exact triangles (9-3) and (9-4) can be shown to be functorial under morphisms of bounded dg algebras. Since quasi-isomorphisms induce isomorphisms on Hochschild homology, it follows from (9-3) that the natural map $\overline{HH}_\bullet(A^{(k)}) \to \overline{HH}_\bullet(A^{(k+1)})$ is an isomorphism. Since $\overline{HH}_\bullet(A^{(k)}) = \overline{HH}_\bullet(A^{(k+1)})$ is concentrated in degree 1, and since $\overline{HC}_i(A^{(k)})$ and $\overline{HC}_i(A^{(k+1)})$ vanish for $|i|$ sufficiently large, the desired claim can be checked by inductively applying the five lemma; cf [50, Section 2.2.3]. □

9.3 Augmentations from relative fillings

According to the philosophy of [24], contact homology is supposed to be well-defined as a differential graded $\mathbb{Q}$–algebra up to strict isomorphism. However, [61] only proves that contact homology is well-defined as a graded $\mathbb{Q}$–algebra, ie after passing to homology. Similarly, the invariants introduced in

Section 8 are merely graded R–algebras for $R=\mathbb{Q}$ or $\mathbb{Q}[U]$. For the purpose of linearizing the invariants of Section 8, we saw in Section 9.1 that the following intermediate assumption — weaker than what is conjectured in [24] but stronger than what is proved in [61] — is sufficient:

Assumption 9.7 The constructions of Section 8 can be lifted to the category **hcdga**. (That is, cobordisms induce a unique morphism of commutative dg algebras up to homotopy, and composition of cobordism is functorial up to homotopy.)

The rest of this section depends on the unproved Assumption 9.7; all statements which depend on this assumption are therefore starred, in accordance with the convention stated in the introduction.

Definition* 9.8 Fix a TN contact pair (Y,ξ,V) and an element $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. Let

$$\mathcal{A}(Y,\xi,V;\mathfrak{r})\in\mathbb{Q}[U]\text{–}\mathbf{hcdga} \tag{9-6}$$

be the limit (or equivalently the colimit) of the dg algebras $\{(CC_\bullet(Y,\xi,V;\mathfrak{r})_\lambda,d_{\psi,J,\theta})\}_{\lambda,J,\theta}$ under the lifts of the maps (8-5) which are furnished by Assumption 9.7. We also analogously define

$$\widetilde{\mathcal{A}}(Y,\xi,V;\mathfrak{r})\in\mathbb{Q}\text{–}\mathbf{hcdga}. \tag{9-7}$$

Remark 9.9 (bigradings) With the notation of Definition* 9.8, suppose that $H_1(Y;\mathbb{Z})=H_2(Y;\mathbb{Z})=0$. Combining Assumption 9.7 with the discussion of Section 8.2, it then follows that $\widetilde{\mathcal{A}}(Y,\xi,V;\mathfrak{r})$ is a $(\mathbb{Z}\times\mathbb{Z})$–bigraded differential algebra, where the differential has bidegree $(-1,0)$.

Definition* 9.10 Given an augmentation $\epsilon\colon\mathcal{A}(Y,\xi,V;\mathfrak{r})\to\mathbb{Q}[U]$, we let $\mathcal{A}^\epsilon(Y,\xi,V;\mathfrak{r})$ be the linearized chain complex (in the sense of Definition 9.1) with respect to ϵ and let $CH^\epsilon_\bullet(Y,\xi,V;\mathfrak{r})$ be the resulting homology.

We have analogous invariants in the reduced case, which are denoted by $\widetilde{\mathcal{A}}^{\widetilde{\epsilon}}(Y,\xi,V;\mathfrak{r})$ and $\widetilde{CH}^{\widetilde{\epsilon}}_\bullet(Y,\xi,V;\mathfrak{r})$ for an augmentation $\widetilde{\epsilon}\colon\widetilde{\mathcal{A}}(Y,\xi,V;\mathfrak{r})\to\mathbb{Q}$.

Definition 9.11 Given a contact manifold (Y,ξ) and a codimension 2 contact submanifold V, a *relative filling* $(\widehat{X},\widehat{\lambda},H)$ is a relative symplectic cobordism from (Y,ξ,V) to the empty set.

Let $(\widehat{X},\widehat{\lambda},H)$ be a relative filling of (Y,ξ,V) and fix $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. Suppose τ extends to a normal trivialization of H. Then Lemma 8.11 and Assumption 9.7 furnish an augmentation $\epsilon(\widehat{X},\widehat{\lambda},H)\colon\mathcal{A}(Y,\xi,V;\mathfrak{r})\to\mathbb{Q}[U]$. Similarly, we have an augmentation $\widetilde{\epsilon}(\widehat{X},\widehat{\lambda},H)\colon\widetilde{\mathcal{A}}(Y,\xi,V;\mathfrak{r})\to\mathbb{Q}$.

If we suppose that $H_1(Y;\mathbb{Z})=H_2(Y;\mathbb{Z})=H_2(X,Y;\mathbb{Z})=0$ and $c_1(TX)=0$, then Lemma 8.13 and Assumption 9.7 imply that $\widetilde{\epsilon}(\widehat{X},\widehat{\lambda},H)$ preserves the $(\mathbb{Z}\times\mathbb{Z})$–bigrading defined in Remark 9.9. It follows that the linearized complex

$$\widetilde{\mathcal{A}}^{\widetilde{\epsilon}}(Y,\xi,V;\mathfrak{r}) \tag{9-8}$$

inherits a $(\mathbb{Z}\times\mathbb{Z})$–bigrading with differential of bidegree $(-1,0)$. Hence

$$\widetilde{CH}^{\epsilon}_{\bullet}(Y,\xi,V;\mathfrak{r}) \tag{9-9}$$

is a $(\mathbb{Z}\times\mathbb{Z})$–bigraded $\mathbb{Q}$–vector space.

We end this section by collecting some lemmas which will be useful later. The reader is referred to Section 3.3 for a review of open book decompositions.

Lemma* 9.12 *Suppose that $(\widehat{W},\widehat{\lambda},H)$ is a relative filling of (Y,ξ,V). Suppose that V is the binding of an open book decomposition (Y,V,π) which supports ξ. Fix an element $\mathfrak{r}=(\alpha_V,\tau,r)\in\mathfrak{R}(Y,\xi,V)$, where τ is the canonical trivialization induced by the open book.*

Suppose that H admits a normal trivialization which restricts to τ. Suppose also that $H_1(Y;\mathbb{Z})=H_2(Y;\mathbb{Z})=H_2(W,Y;\mathbb{Z})=0$ and that $c_1(TW)=0$. Then the augmentation

$$\widetilde{\epsilon}\colon \widetilde{\mathcal{A}}(Y,\xi,V;\mathfrak{r})\to\mathbb{Q} \tag{9-10}$$

is the zero augmentation. In particular, it depends only on $(\widehat{W},\widehat{\lambda})$ and not on H.

Proof It is shown in Corollary 3.15 that there exists a nondegenerate contact form α for (Y,ξ) which is adapted to $\mathfrak{r}$ and has the property that all Reeb orbits are transverse to the pages of the open book decomposition. Given auxiliary choices of almost complex structures and perturbation data, the augmentation $\widetilde{\epsilon}$ counts (possibly broken) holomorphic planes u which are asymptotic to a Reeb orbit γ disjoint from H, and such that $[u]*H=0$. However, our topological assumptions and Lemma 5.21 implies that $[u]*H$ is precisely the linking number of γ with the binding V, which is strictly positive by assumption; see Remark 5.20. □

Lemma* 9.13 *Let $(\widehat{X},\widehat{\lambda},H)$ and $(\widehat{X}',\widehat{\lambda}',H')$ be relative symplectic fillings of (Y,ξ,V) and (Y',ξ',V'). Let $f:(Y,\xi,V)\to(Y',\xi',V')$ be a contactomorphism. Suppose that there exists a symplectomorphism $\phi\colon(\widehat{X},\widehat{\lambda})\to(\widehat{X}',\widehat{\lambda}')$ which coincides near infinity with the induced map $\widetilde{f}:SY\to SY'$.*

Given any $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$, the following diagram commutes:

$$\begin{array}{ccc} CH_\bullet(Y,\xi,V;\mathfrak{r}) & \xrightarrow{\Phi(\widehat{X},\widehat{\lambda},H)} & \mathbb{Q}[U] \\ \downarrow{=} & & \downarrow \\ CH_\bullet(Y',\xi',V';f_*\mathfrak{r}) & \xrightarrow{\Phi(\widehat{X}',\phi_*\widehat{\lambda},\phi(H))} & \mathbb{Q}[U] \\ \downarrow{=} & & \downarrow \\ CH_\bullet(Y',\xi',V';f_*\mathfrak{r}) & \xrightarrow{\Phi(\widehat{X}',\widehat{\lambda}',\phi(H))} & \mathbb{Q}[U] \end{array} \tag{9-11}$$

The analogous statement holds for $\widetilde{CH}_\bullet(-)$, with $\mathbb{Q}$ in place of $\mathbb{Q}[U]$. In addition, if $H_1(Y;\mathbb{Z})=H_2(Y;\mathbb{Z})=0$ and $c_1(TX)=0$, then all arrows can be assumed to preserve the bigrading in Definition 8.12.

Proof The commutativity of the top square is essentially tautological; more precisely, it follows from the functoriality of the moduli counts in Theorem 7.27. The commutativity of the bottom square follows from the observation that ϕ preserves the Liouville form outside a compact set. Hence $(\hat{X}', \phi_* \hat{\lambda})$ and $(\hat{X}', \hat{\lambda}')$ are deformation equivalent. It follows by Corollary 8.8 that they induce the same morphism on homology. The fact that the maps preserve the bigradings on $\widetilde{CH}(-;-)$ (under the above topological assumptions) is a consequence on Lemma 8.13. □

Corollary* 9.14 *Let $(\hat{X}, \hat{\lambda}, H)$ (resp. $(\hat{X}, \hat{\lambda}, H')$) be relative fillings for (Y, ξ, V) (resp. (Y, ξ, V')). Suppose that V is the binding of an open book decomposition of (Y, ξ) and fix $\mathfrak{r} = (\alpha, \tau, r) \in \mathfrak{R}(Y, \xi, V)$, where τ is induced by the open book. Let $f : (Y, \xi, V) \to (Y, \xi, V')$ be a contactomorphism.*

Suppose that $H_1(Y;\mathbb{Z}) = H_2(Y;\mathbb{Z}) = H_2(X, Y;\mathbb{Z}) = 0$ and that $c_1(TX) = 0$. Then

$$\widetilde{CH}^{\tilde{\epsilon}}_{\bullet,\bullet}(Y, \xi, V; \mathfrak{r}) = \widetilde{CH}^{\tilde{\epsilon}'}_{\bullet,\bullet}(Y, \xi, V'; f_* \mathfrak{r}), \tag{9-12}$$

where both augmentations are induced by the relative fillings and the $(\mathbb{Z}\times\mathbb{Z})$–bigrading is defined in Definition 8.12.

Proof Indeed, since the lift of a contactomorphism to the symplectization is a Hamiltonian symplectomorphism, it is easy to construct a symplectic automorphism of $(X, d\hat{\lambda})$ satisfying the conditions of Lemma* 9.13; see eg [14, Section 3.2]. The claim now follows from Lemma* 9.12. □

10 Invariants of Legendrian submanifolds

10.1 Invariants of contact pushoffs

Definition 10.1 (see Definition 3.1 in [10]) Let (Y, ξ) be a contact manifold and let $\Lambda \hookrightarrow Y$ be a Legendrian embedding. By the Weinstein neighborhood theorem, the map extends to an embedding $\mathrm{Op}(\Lambda) \subset (J^1\Lambda, \xi_{\mathrm{std}}) \to (Y, \xi)$, where $\mathrm{Op}(\Lambda) \subset (J^1\Lambda, \xi_{\mathrm{std}})$ denotes an open neighborhood of the zero section.

Let $\tau(\Lambda)$ be the induced codimension 2 contact embedding

$$\tau(\Lambda) \colon \partial(D^*_{\epsilon,g}\Lambda) = \partial(D^*_{\epsilon,g}\Lambda) \times 0 \subset T^*\Lambda \times \mathbb{R} = J^1\Lambda \hookrightarrow (Y, \xi). \tag{10-1}$$

Here $D^*_{\epsilon,g}\Lambda$ is the sphere bundle of covectors of length ϵ with respect to some metric g, which is a contact manifold with respect to the restriction of the canonical 1–form on $T^*\Lambda$. We refer to $\tau(\Lambda)$ as the *contact pushoff* of $\Lambda \hookrightarrow Y$.

Standard arguments establish that the contact pushoff is canonical up to isotopy through codimension 2 contact embeddings. By abuse of notation, we will routinely identify $\tau(\Lambda)$ with its image. Observe that it follows that $CH_\bullet(Y, \xi, \tau(\Lambda); \mathfrak{r})$ and $\widetilde{CH}_\bullet(Y, \xi, \tau(\Lambda); \mathfrak{r})$ can be viewed as invariants of Λ.

10.2 Deformations of the Chekanov–Eliashberg dg algebra

In the spirit of the previous sections, we now consider deformations of the Chekanov–Eliashberg dg algebra of a Legendrian induced by a codimension 2 contact submanifold. We begin with some preliminary definitions.

Definition 10.2 Let $\Lambda \subset (Y, \xi)$ be a Legendrian submanifold. Given a contact form $\ker \alpha = \xi$, consider a Reeb chord $c\colon [0, R] \to Y$. The linearized Reeb flow defines a path of symplectomorphisms $\mathcal{P}_r\colon \xi|_{c(0)} \to \xi|_{c(r)}$. We say that the Reeb chord c is nondegenerate if $\mathcal{P}_R(T_{c(0)}\Lambda) \cap T_{c(R)}\Lambda = \{0\}$.

Definition 10.3 (cf Section 2.1 in [7]) With the notation of Definition 10.2, let $\bigwedge_{\mathbb{C}}^{n-1}(\xi, d\alpha)$ be the canonical bundle of ξ and suppose that it admits a trivialization σ. Let $\Lambda_1, \dots, \Lambda_k$ be an enumeration of the components of Λ. Suppose that each Λ_i has vanishing Maslov class.

Suppose first of all that $k = 1$ (ie Λ is connected). Given a nondegenerate Reeb chord c, pick a path c_- in Λ connecting $c(R)$ to $c(0)$. Observe that $\bigwedge^{n-1} T_{c_-}\Lambda_i \subset \bigwedge_{\mathbb{C}}^{n-1}(\xi, d\alpha)$ is a path of Lagrangian subspaces along c_-. We call this path L_{c_-}. The map $\mathcal{P}_r$ also defines a path of Lagrangian subspaces $\bigwedge^{n-1} \mathcal{P}_r(T_{c(0)}\Lambda_i) \subset \bigwedge_{\mathbb{C}}^{n-1}(\xi, d\alpha)$ along c. We call this path L_c.

Let $\tilde{c} = c_- * c$ be obtained by concatenating c_- and c (the concatenation is from left to right). Now consider the path of Lagrangian subspaces $L_{\tilde{c}} = L_{c^-} * L_c * \boldsymbol{P}^+$, where $\boldsymbol{P}^+$ is a positive rotation from $\mathcal{P}_R(T_{c(0)}\Lambda)$ to $T_{c(R)}\Lambda$ (this is well-defined by our assumption that c is nondegenerate).

The *Conley–Zehnder index for chords* of c with respect to σ is denoted by $\mathrm{CZ}^{+,\sigma}(c)$ and defined by

$$\mathrm{CZ}^{+,\sigma}(c) = \mu^\sigma(L_{\tilde{c}}), \tag{10-2}$$

where $\mu^\sigma(-)$ is the Maslov index with respect to σ; see [52, Theorem 2.3.7]. This definition is independent of the choice of c_- due to our assumption that Λ has vanishing Maslov class. Note also that the resulting index depends on σ, but its parity does not.

In the case that $k > 1$, the definition of the Conley–Zehnder index for chords is more complicated, and depends on additional choices. We refer the reader to [7, Section 2.1] — we warn the reader that there is a typo in the formula stated there: the correct formula for the Conley–Zehnder index for chords should read $\mathrm{CZ}^{+,\sigma}(c) = |c| - 1 = (\phi_- - \phi_\Lambda(x_1))/\pi + (n-1)/2$.

Remark 10.4 It may happen that a Reeb orbit can also be viewed as a Reeb chord with same starting and end point. In this case, we have in general that $\mathrm{CZ}^+(c) \neq \mathrm{CZ}(c)$.

Let us now consider a TN contact pair (Y, ξ, V) and a Legendrian submanifold $\Lambda \subset (Y - V, \xi)$. We let $\Lambda_1, \dots, \Lambda_k$ be an enumeration of the connected components of Λ. As in Definition 10.3, we assume that the Λ_i have vanishing Maslov class.

Definition/Assumption* 10.5 (cf Proposition 2.8.2 in [24]) Fix $\mathfrak{r} \in \mathfrak{R}(Y,\xi,V)$. Let us also choose the following additional data:

- A contact form $\lambda \in \Omega^1(Y)$ which is adapted to $\mathfrak{r}$ and has the property that all Reeb orbits and Λ–Reeb chords are nondegenerate.
- A $d\lambda$–compatible almost complex structure J on ξ which preserves TV.

Given a class $\beta \in \pi_2(\widehat{Y}; c^+, \Gamma_\Lambda^-, \Gamma^-)$, we let

$$\mathcal{M}(c^+, \Gamma_\Lambda^-, \Gamma^-; \beta)_J \tag{10-3}$$

be the moduli space of connected $\widehat{J}$–holomorphic curves, modulo $\mathbb{R}$–translation representing the class β. (Here we follow the notation of Section 2.2, where c^+ is a Reeb chord of $\Lambda \subset (Y,\lambda)$, $\Gamma_\Lambda^- = \{c_1^-, \dots, c_\sigma^-\}$ is an ordered collections of (not necessarily distinct) Reeb chords, and Γ^- is a collection of Reeb orbits.) Since $V \subset (Y-\Lambda,\lambda)$ is a strong contact submanifold, a straightforward extension of Siefring's intersection theory defines an intersection number $\widehat{V} * \beta \in \mathbb{Z}$.

Let us now consider the semisimple ring

$$\mathcal{R} = \bigoplus_{i=1}^{k} \mathbb{Q}[U], \tag{10-4}$$

and let $e_1, \dots, e_k$ be the idempotents corresponding to the unit in each summand.

Let $CL_\bullet(Y,\xi,V,\Lambda;\mathfrak{r})_\lambda$ be the free $\mathcal{R}$ algebra generated by (good) Reeb orbits of (Y,α) and Reeb chords of Λ, subject to the following relations:

- $\gamma_1\gamma_2 = (-1)^{|\gamma_1||\gamma_2|}\gamma_2\gamma_1$ for Reeb orbits a and b.
- If c_{ij} is a Reeb chord from Λ_i to Λ_j, then $e_k c_{ij} e_l = \delta_{jk} c_{ij} \delta_{il}$.

We assume that there exists a suitable virtual perturbation framework compatible with [61], so that we can define a differential d_J (squaring to zero) on generators as follows:

- For a Reeb chord c^+, we let

$$d_J(c^+) = \sum \frac{1}{|\mathrm{Aut}|} \#\overline{\mathcal{M}}(c^+, \Gamma_\Lambda^-, \Gamma^-; \beta)_J U^{\widehat{V} * \beta} c_1^- \cdots c_\sigma^- \gamma_1 \cdots \gamma_s, \tag{10-5}$$

  where the sum is over choices of $\beta \in \pi_2(\widehat{Y}; c^+, \Gamma_\Lambda^-, \Gamma^-)$ for all possible choices of Γ_Λ^- and Γ^-.
- For a Reeb orbit γ, we let $d_J(\gamma)$ be the usual deformed contact homology differential, as described in Section 8.1.

We assume that $(CL_\bullet(Y,\xi,V,\Lambda;\mathfrak{r})_\lambda, d_J)$ is independent of λ and J up to canonical isomorphism in $\mathbb{Q}[U]$–**hdga**. We denote the resulting object by

$$\mathcal{L}(Y,\xi,V,\Lambda;\mathfrak{r}) \tag{10-6}$$

and we let $CH_\bullet(Y,\xi,V,\Lambda;\mathfrak{r})$ be its homology. We assume that $\mathcal{L}(Y,\xi,V,\Lambda;\mathfrak{r})$ satisfies the limited functoriality described in Proposition* 10.7.

Definition/Assumption* 10.6 (cf Proposition 2.8.2 in [24]) Carrying over the hypotheses and notation from Definition/Assumption* 10.5, let us consider the semisimple ring

$$\widetilde{\mathcal{R}} = \bigoplus_{i=1}^{k} \mathbb{Q}, \tag{10-7}$$

where we again let $e_1, \dots, e_k$ be the idempotents corresponding to the unit in each summand.

We let $\widetilde{CL}_\bullet(Y, \xi, V, \Lambda; \mathfrak{r})_\lambda$ be the free $\widetilde{\mathcal{R}}$ algebra generated by (good) Reeb orbits of (Y, α) *which are not contained in* V and Λ Reeb chords, subject to the following relations:

- $\gamma_1 \gamma_2 = (-1)^{|\gamma_1||\gamma_2|} \gamma_2 \gamma_1$ for Reeb orbits a and b.
- If c_{ij} is a Reeb chord from Λ_i to Λ_j, then $e_k c_{ij} e_l = \delta_{jk} c_{ij} \delta_{il}$.

This algebra is again $\mathbb{Z}/2$–graded in general, and $\mathbb{Z}$–graded when the canonical bundle is trivialized. We assume again that there exists a suitable virtual perturbation framework so that one can define a differential $\widetilde{d}_L$ (squaring to zero) on generators as follows:

- For a Reeb chord c^+, we let

$$d_J(c^+) = \sum \frac{1}{|\mathrm{Aut}|} \# \overline{\mathcal{M}}(c^+, \Gamma_\Lambda^-, \Gamma^-; \beta)_J \delta(\widehat{V} * \beta) c_1^- \cdots c_\sigma^- \gamma_1 \cdots \gamma_s, \tag{10-8}$$

  where $\delta \colon \mathbb{R} \to \{0, 1\}$ satisfies $\delta(0) = 1$ and $\delta(s) = 0$ for $s \neq 0$ and the sum is over all possible choices of homotopy classes as in Definition/Assumption* 10.5.

- For a Reeb orbit γ, we let $\widetilde{d}_J(\gamma)$ be the reduced contact homology differential associated to the twisted moduli counts $\#_{\widetilde{\psi}} \mathcal{M}$, which only counts curves disjoint from $\widehat{V}$.

We assume that $(\widetilde{CL}_\bullet(Y, \xi, V, \Lambda; \mathfrak{r})_\lambda, \widetilde{d}_J)$ is independent of λ, J up to canonical isomorphism in $\mathbb{Q}$–**hdga**. We denote the resulting object by

$$\widetilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r}) \tag{10-9}$$

and we let $CH_\bullet(Y, \xi, V, \Lambda; \mathfrak{r})$ be its homology. We also assume that $\widetilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r})$ satisfies the limited functoriality described in Proposition* 10.7.

Proposition* 10.7 (cf Corollary 8.8) *Let $(Y^\pm, \xi^\pm, V^\pm)$ be TN contact pairs and choose elements $\mathfrak{r}^\pm = (\alpha^\pm, \tau^\pm, r^\pm) \in \mathfrak{R}(Y^\pm, \xi^\pm, V^\pm)$. Consider an exact relative symplectic cobordism $(\widehat{X}, \widehat{\lambda}, H)$ with positive end (Y^+, ξ^+, V^+) and negative end (Y^-, ξ^-, V^-), and suppose that τ^+ and τ^- extend to a global trivialization of the normal bundle of H.*

Suppose that $L \subset (\widehat{X}, \widehat{\lambda}, H)$ is a cylindrical Lagrangian submanifold which is disjoint from H, with ends $\Lambda^\pm \subset (Y^\pm - V^\pm, \xi^\pm)$.

If $r^+ \geq e^{\mathcal{E}((H,\hat{\lambda}_H)_{\alpha^-}^{\alpha^+})} r^-$*, then there is an induced map*

$$\Phi(\hat{X}, \hat{\lambda}, H, L)\colon \mathcal{L}(Y^+, \xi^+, V^+; \mathfrak{r}^+) \to \mathcal{L}(Y^-, \xi^-, V^-; \mathfrak{r}^-). \tag{10-10}$$

The analogous statement holds for the reduced invariants $\tilde{\mathcal{L}}(-)$.

Definition* 10.8 Let $\epsilon\colon \mathcal{A}(Y, \xi, V; \mathfrak{r}) \to \mathbb{Q}[U]$ be an augmentation. Then we let

$$\mathcal{L}^{\epsilon}(Y, \xi, V, \Lambda; \mathfrak{r}) := \mathcal{L}(Y, \xi, V, \Lambda; \mathfrak{r}) \otimes_{\mathcal{A}(Y,\xi,V;\mathfrak{r})} \mathbb{Q}[U], \tag{10-11}$$

with differential $d_L \otimes 1$. The structure maps implicit in forming the tensor product (10-11) are, respectively, furnished by the inclusion $\mathcal{A}(Y, \xi, V; \mathfrak{r}) \subset \mathcal{L}(Y, \xi, V, \Lambda; \mathfrak{r})$ and the augmentation $\epsilon\colon \mathcal{A}(Y, \xi, V; \mathfrak{r}) \to \mathbb{Q}[U]$. The resulting tensor product is naturally also a differential graded $\mathbb{Q}[U]$ algebra.

We similarly define

$$\tilde{\mathcal{L}}^{\tilde{\epsilon}}(Y, \xi, V, \Lambda; \mathfrak{r}) := \tilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r}) \otimes_{\tilde{\mathcal{A}}(Y,\xi,V;\mathfrak{r})} \mathbb{Q}, \tag{10-12}$$

using the maps $\tilde{\mathcal{A}}(Y, \xi, V; \mathfrak{r}) \subset \tilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r})$ and $\tilde{\epsilon}\colon \tilde{\mathcal{A}}(Y, \xi, V; \mathfrak{r}) \to \mathbb{Q}$, which is naturally a differential graded $\mathbb{Q}$–algebra.

Remark 10.9 The algebra $\mathcal{L}^{\epsilon}(Y, \xi, V, \Lambda; \mathfrak{r})$ is the twisted analog of the Legendrian homology dg algebra (or Chekanov–Eliashberg dg algebra) described in [7, Section 4.1].

We now discuss gradings on the above Legendrian invariants.

Definition* 10.10 Let (Y, ξ, V) be a TN contact pair and choose $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$. Let $\Lambda \subset (Y - V, \xi)$ be a Legendrian submanifold. Suppose that $H_1(Y; \mathbb{Z}) = 0$ and that $c_1(\xi) = 0$. Then the Legendrian homological $\mathbb{Z}/2$–grading of $\mathcal{L}(Y, \xi, V, \Lambda; \mathfrak{r})$ (resp. $\tilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r})$) lifts to a canonical $\mathbb{Z}$–grading given on orbits by (8-15) and given on chords by

$$|c| = \mathrm{CZ}^{+,\tau}(c) - 1, \tag{10-13}$$

which is well-defined due to our topological assumptions.

The invariants $\mathcal{L}^{\epsilon}(Y, \xi, V, \Lambda; \mathfrak{r})$, $\overline{HC}(\mathcal{L}^{\epsilon}(Y, \xi, V, \Lambda; \mathfrak{r}))$, $\tilde{\mathcal{L}}^{\epsilon}(Y, \xi, V, \Lambda; \mathfrak{r})$ and $\overline{HC}(\tilde{\mathcal{L}}^{\epsilon}(Y, \xi, V, \Lambda; \mathfrak{r}))$ inherit a $\mathbb{Z}$–grading, which we also refer to as the homological grading.

Lemma* 10.11 *With the notation of Proposition* 10.7, suppose that* $H_1(Y^{\pm}; \mathbb{Z}) = 0$ *and that* $w_2(L) = c_1(\xi^{\pm}) = c_1(TX) = 0$. *Then the cobordism maps described in Proposition* 10.7 preserve the Legendrian homological* $\mathbb{Z}$*–grading.* □

As in Section 8.2, there is also a linking number $\mathbb{Z}$–grading on the reduced Legendrian invariants under certain topological assumptions.

Definition* 10.12 Let (Y,ξ,V) be a TN contact pair and choose $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. Let $\Lambda\subset(Y,\xi)$ be a Legendrian submanifold. Suppose that $H_1(Y;\mathbb{Z})=H_2(Y;\mathbb{Z})=\pi_0(\Lambda)=\pi_1(\Lambda)=0$. Then the *linking number grading* $|\cdot|_{\mathrm{link}}$ on $\widetilde{\mathcal{L}}(Y,\xi,V,\Lambda;\mathfrak{r})$ is given on Reeb chords by

$$|c|_{\mathrm{link}}=\mathrm{link}_V(c;\Lambda). \tag{10-14}$$

It is given on Reeb orbits by (8-16). The grading is extended to arbitrary words by defining the grading of a word to be the sum of the gradings of its letters. One can verify using Lemma 5.22 that this grading is well-defined.

We let

$$\widetilde{\mathcal{L}}_{\bullet,\bullet}(Y,\xi,V,\Lambda;\mathfrak{r}) \tag{10-15}$$

be the bigraded differential $\mathbb{Q}$–algebra of bidegree $(-1,0)$, where

- the first bullet refers to the (Legendrian) homological $\mathbb{Z}$–grading (which is well-defined in view of our topological assumptions and the universal coefficients theorem, see Definition* 10.10), and
- the second bullet refers to the (Legendrian) linking number grading.

We also have the following lemma, which follows from Lemma* 10.11 and Lemma 5.22.

Lemma* 10.13 *With the notation of Proposition* 10.7, suppose that* $H_1(Y^\pm;\mathbb{Z})=H_2(Y^\pm;\mathbb{Z})=H_2(X,Y^+;\mathbb{Z})=0$ *and* $\pi_0(\Lambda^\pm)=\pi_1(\Lambda^\pm)=0$. *Then the cobordism maps described in Proposition* 10.7 preserve the linking number* $\mathbb{Z}$*–grading. In the case that we also have* $w_2(\Lambda^\pm)=c_1(\xi^\pm)=c_1(TX)=0$, *then the cobordism maps preserve the* $(\mathbb{Z}\times\mathbb{Z})$*–bigrading* (10-15). ◻

Corollary* 10.14 *Consider a TN contact pair* (Y,ξ,V) *and an element* $\mathfrak{r}\in\mathfrak{R}(Y,\xi,V)$. *Let* $\Lambda\subset(Y,\xi)$ *be a Legendrian submanifold. Let* (W,λ,H) *be a relative filling for* (Y,ξ,V) *and let* $\widetilde{\epsilon}\colon\widetilde{\mathcal{A}}(Y,\xi,V;\mathfrak{r})\to\mathbb{Q}$ *be the induced augmentation. Suppose that* $H_1(Y;\mathbb{Z})=H_2(Y;\mathbb{Z})=H_2(W,Y;\mathbb{Z})=0$, *that* $\pi_0(\Lambda)=\pi_1(\Lambda)=0$ *and that* $w_2(\Lambda)=c_1(\xi)=c_2(TW)=0$.

Then

$$\widetilde{\mathcal{L}}^{\widetilde{\epsilon}}_{\bullet,\bullet}(Y,\xi,V;\mathfrak{r}) \tag{10-16}$$

inherits the structure of a $(\mathbb{Z}\times\mathbb{Z})$*–bigraded* $\mathbb{Q}$*–algebra with differential of bidegree* $(-1,0)$. *Moreover,*

$$\overline{HC}_{\bullet,\bullet}(\widetilde{\mathcal{L}}^{\widetilde{\epsilon}}(Y,\xi,V;\mathfrak{r})) \tag{10-17}$$

inherits the structures of a $(\mathbb{Z}\times\mathbb{Z})$*–bigraded* $\mathbb{Q}$*–vector space.*

Proof According to Lemma 5.21 and our topological hypotheses, the augmentation $\widetilde{\epsilon}$ preserves the linking number. The first claim follows. For the second claim, note that both the homological grading and linking number grading are preserved by the cyclic permutation operator τ, and hence pass to reduced cyclic homology; see Definition 9.4. ◻

10.3 The effect of Legendrian surgery

The familiar procedure of attaching a handle in differential topology can be performed in the symplectic category. There are various essentially equivalent approaches to doing this in the literature. For concreteness, we exclusively follow in this paper the construction described in [46, Section 3.1], which we now summarize.

Construction 10.15 (attaching a handle) Let (X_0^{2n}, λ_0) be a Liouville cobordism with positive boundary $(Y_0^{2n-1}, \xi_0 = \ker(\lambda_0|_{Y_0}))$. Let $\Lambda \subset (Y_0 - V, \xi_0)$ be an isotropic sphere with trivialized conformal symplectic normal bundle (the latter condition is vacuous if Λ is a Legendrian). Choose an arbitrary open neighborhood $\mathcal{U}$ of Λ, which we refer to as the *attaching region*.

We may now glue a model handle H along Y_0 inside $\mathcal{U}$, following the detailed construction given in [46, Section 3.1]. The gluing is carried out by identifying the Liouville flow near Λ with the flow on H. We note that this gluing procedure involves some auxiliary choices, which we do not state here.

The outcome of the procedure (for any of the above auxiliary choices) is a Liouville cobordism (X, λ) with positive boundary $(Y, \xi = \ker(\lambda|_Y))$. We say that this domain is obtained from (X_0, λ_0) by *attaching a handle along* Λ, or *Legendrian surgery on* Λ. As is well-known from differential topology, Y differs from Y_0 by surgery along Λ.

In [7], Bourgeois, Ekholm and Eliashberg study the effect of handle attachment on various flavors of symplectic and contact homology. In particular, they describe conjectural exact sequences which govern the change in these invariants and describe the moduli spaces of holomorphic curves which should underlie the existence of these exact sequences. In Theorem/Assumption* 10.16 below (see also Remark 10.17), we state an analog of the surgery exact sequence for linearized contact homology in [7, Theorem 5.1]. The proofs sketched in [7, Section 6] also apply *mutatis mutandis* in the present setting, so will not be repeated. As discussed in Section 1.3, we expect that if [7, Theorem 5.1] can be made rigorous in the setting of Pardon's VFC package, it should pose no substantial additional difficulties to also establish Theorem/Assumption* 10.16.

To set the stage for Theorem/Assumption* 10.16, let (Y_0^{2n-1}, ξ_0) be a contact manifold and let $V \subset (Y_0, \xi_0)$ be a codimension 2 contact submanifold with trivial normal bundle. Let (X_0, λ_0) be a Liouville domain with positive boundary $(Y_0, \xi = \ker \lambda_0)$ and let $H_0 \subset (X, \lambda_0)$ be a symplectic submanifold which is preserved setwise by the Liouville flow near $\partial X_0 = Y_0$ and such that $\partial H = V$.

Let $\Lambda \subset (Y_0 - V, \xi_0)$ be an isotropic sphere with trivialized conformal symplectic normal bundle (the latter condition is vacuous if Λ is a Legendrian). Let (X, λ) be the Liouville domain obtained by attaching a Weinstein handle along Λ according to Construction 10.15 and let $(Y, \xi = \ker \lambda)$ be the positive boundary. We may assume that the attaching region is disjoint from $V \subset Y_0$. By abuse of notation, we therefore view V as a codimension 2 contact submanifold of (Y_0, ξ_0) and (Y, ξ) and view H_0 as a submanifold of X_0 and X. We also identify $\mathfrak{R}(Y_0, \xi_0, V) = \mathfrak{R}(Y, \xi, V)$.

We let $(\widehat{X}_0, \widehat{\lambda}_0, H)$ be the completion of (X_0, λ_0, H_0), and let $(\widehat{X}, \widehat{\lambda}, H)$ be the completion of (X, λ, H_0). There are natural (strict) markings $e_0\colon \mathbb{R} \times Y_0 \to \widehat{X}_0$ taking $(t, y_0) \mapsto \psi^0_t(y_0)$ and $e\colon \mathbb{R} \times Y \to \widehat{X}$ taking $(t, y) \mapsto \psi_t(y_0)$, where ψ^0 (resp. ψ) is the Liouville flow in $\widehat{X}_0$ (resp. in $\widehat{X}$).

Finally, in order to have well-defined homological $\mathbb{Z}$–gradings, we assume that $H_1(Y_0; \mathbb{Z}) = H_1(Y; \mathbb{Z}) = 0$ and that $c_1(TX_0) = c_1(TX) = 0$.

Theorem/Assumption* 10.16 (cf Theorem 5.1 in [7]) *With the above setup and $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$, consider the augmentations $\widetilde{\epsilon}((\widehat{X}_0, \widehat{\lambda}_0, H), e_0)\colon \widetilde{\mathcal{A}}(Y, \xi, V) \to \mathbb{Q}$ and $\widetilde{\epsilon}((\widehat{X}, \widehat{\lambda}, H), e)\colon \widetilde{\mathcal{A}}(Y, \xi, V) \to \mathbb{Q}$. If Λ is a Legendrian sphere, we have the following exact triangle, where the top horizontal arrow is the natural map induced by an exact relative symplectic cobordism*:

$$\begin{array}{ccc} \widetilde{CH}^{\widetilde{\epsilon}}_{\bullet-(n-3)}(Y, \xi, V; \mathfrak{r}) & \longrightarrow & \widetilde{CH}^{\widetilde{\epsilon}_0}_{\bullet-(n-3)}(Y_0, \xi_0, V; \mathfrak{r}) \\ & \nwarrow \qquad \swarrow_{[-1]} & \\ & \overline{HC}_\bullet(\widetilde{\mathcal{L}}^{\epsilon_0}(Y_0, \xi_0, V, \Lambda; \mathfrak{r})) & \end{array} \tag{10-18}$$

If Λ is an isotropic sphere of dimension $k \le n-2$, then we have that

$$H_*\big(\mathrm{Cone}\big(\widetilde{CH}^{\widetilde{\epsilon}}_{\bullet-(n-3)}(Y, \xi, V; \mathfrak{r}) \to \widetilde{CH}^{\widetilde{\epsilon}_0}_{\bullet-(n-3)}(Y_0, \xi_0, V; \mathfrak{r})\big)\big) = \begin{cases} \mathbb{Q} & \textit{if } * = n-k+2\mathbb{N}, \\ 0 & \textit{otherwise.} \end{cases} \tag{10-19}$$

Remark 10.17 There is a natural analog of Theorem/Assumption* 10.16 involving the invariants $CH^{\epsilon_0}_{\bullet-(n-3)}(Y_0, \xi_0, V; \mathfrak{r})$, $CH^{\epsilon}_{\bullet-(n-3)}(Y, \xi, V; \mathfrak{r})$ and $\overline{HC}_\bullet(\mathcal{L}^{\epsilon_0}(Y_0, \xi_0, V, \Lambda; \mathfrak{r}))$ which we also expect to hold. We do not state it here since we do not use it in this paper.

Remark 10.18 With the setup of Theorem/Assumption* 10.16, let us in addition assume that $H_2(Y_0; \mathbb{Z}) = H_2(X_0, Y_0) = H_2(Y; \mathbb{Z}) = H_2(X, Y; \mathbb{Z}) = 0$. Then Lemma 8.13 and Corollary* 10.14 provide an additional linking number $\mathbb{Z}$–grading on the invariants appearing in the surgery exact sequences.

The resulting $(\mathbb{Z} \times \mathbb{Z})$–bigrading is preserved by the maps in the surgery exact sequence. Indeed, Lemma 8.13 ensures the top horizontal map preserves the linking number $\mathbb{Z}$–grading. The bottom right map counts holomorphic disks with one positive interior puncture, k negative boundary punctures, and with boundary mapping to $S\Lambda$ (the relevant moduli space is described in [7, Section 2.6]). Hence one can readily verify (cf Lemma 5.22) that this map also preserves the linking number grading. Finally, the bottom left map is defined algebraically as the connecting map in the long exact sequence. Since the internal differentials of the relevant chain complexes preserve the linking number grading, this connecting map does too.

11 Some computations

11.1 Vanishing results

Recall that a contact manifold (Y^{2n-1}, ξ) is said to be *overtwisted* if it contains an overtwisted disk; see [5, Section 1]. In general, if (Y, ξ) is overtwisted and $C \subset Y$ is a closed subset, then $(Y - C, \xi)$ may not be overtwisted.

Theorem 11.1 *Suppose that (Y, ξ, V) is a TN contact pair such that $(Y - V, \xi)$ is overtwisted. Given any element $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$, we have*

$$CH_\bullet(Y, \xi, V; \mathfrak{r}) = \widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r}) = 0. \tag{11-1}$$

We collect some definitions which will be useful in proving Theorem 11.1. Let $\mathfrak{alm}_{\mathrm{U}(n-1)}(S^{2n-1})$ be the set of almost-contact structures on S^{2n-1}; see Definition 12.1. It follows by the main theorem of [5] that $\mathfrak{alm}_{\mathrm{U}(n-1)}(S^{2n-1})$ is in canonical correspondence with the set of overtwisted contact structures on the sphere, a fact which will be used implicitly in the proof of Theorem 11.1.

A folklore result in contact topology (see eg [12, Section 6]) states that for any fixed element $\beta \in \mathfrak{alm}_{\mathrm{U}(n-1)}(S^{2n-1})$, the operation of connected sum endows $\mathfrak{alm}_{\mathrm{U}(n-1)}(S^{2n-1})$ with a group structure with identity element β. The isomorphism class of the resulting group is moreover independent of β. Since these facts are not to our knowledge available in the literature, we have provided careful proofs in the appendix.

For the remainder of this section, we fix $\beta \in \mathfrak{alm}_{\mathrm{U}(n-1)}(S^{2n-1})$ to be the almost-contact structure induced by the standard contact structure on the sphere. Given a pair of contact manifolds (M_1, α_1) and (M_2, α_2), one can also consider their connected sum $(M_1 \# M_2, \alpha_1 \# \alpha_2)$, which is obtained by gluing-in a neck along Darboux balls in M_1, M_2. This operation is discussed in Remark A.10. As noted there, the two a priori different notions of a connected sum of (almost-)contact manifolds commute with the forgetful map from contact manifolds to almost-contact manifolds.

Proof of Theorem 11.1 It is enough to prove that the invariants vanish for a particular choice of nondegenerate contact form $\widetilde{\alpha}$ on Y which is adapted to $\mathfrak{r}$. To construct such a form, we follow arguments of Bourgeois and Van Koert in [8, Section 6.2].

Using Construction 3.6, we define an auxiliary contact form α in a small neighborhood $\mathcal{N}$ of V with the property that V is a strong contact submanifold and that α is adapted to $\mathfrak{r}$. After possibly shrinking $\mathcal{N}$, we can assume that $(Y - \mathcal{N}, \xi)$ is overtwisted. We now extend α arbitrarily to a globally defined, nondegenerate contact form on (Y, ξ). (Since α_V is nondegenerate, Construction 3.6 produces a nondegenerate contact form on $\mathcal{N}$, so it extends unproblematically to a global nondegenerate contact form.)

Choose a Darboux ball $B \subset Y$ whose closure is disjoint from $\mathcal{N}$. Let $B' \subset B$ be a smaller Darboux ball and let $A = B - \bar{B}'$. Let β_0 denote the almost-contact structure on B obtained by restricting ξ. Let $\mathfrak{alm}_{\mathrm{U}(n-1)}(B, A; \beta_0)$ be the set of almost-contact structures on B agreeing with β_0 near A. The group $\mathfrak{alm}_{\mathrm{U}(n-1)}(S^{2n-1})$ acts on $\mathfrak{alm}_{\mathrm{U}(n-1)}(B, A; \beta_0)$ by connect-summing with an almost-contact sphere along a disk whose closure is disjoint from A.

We now appeal to work of Bourgeois and van Koert: in [8, Section 2.2], they construct a special contact form α_L on the sphere S^{2n-1} (this form turns out to be overtwisted by [12], although we won't need this). They prove [8, Sections 2–5] that α_L admits a Reeb orbit γ which bounds a single, transversally cut-out J–holomorphic plane, for some suitable J on the symplectization.

We now form the connected sum of (S^{2n-1}, α_L) with (Y, α), where we assume that the gluing happens entirely inside of B'. Bourgeois and van Koert (following earlier work of Ustilovsky [65]) explain how to perform this connected sum so that the orbit γ survives, and still has the property that it bounds a single transversally cut-out plane (basically, one can suitably adjust the neck to ensure that the plane cannot cross the neck; see [8, Section 6.2]).

Finally, we connect sum with another overtwisted contact sphere (S^{2n-1}, α'_L) so that $(B, \alpha \# \alpha_L \# \alpha'_L|_B)$ is formally contact isotopic to $(B, \alpha|_B)$, through a contact isotopy fixed near A. Note that we may freely assume that the two connected sums happen in disjoint regions of B', so they do not interfere with each other. Unwinding the definitions, this means that there exists a diffeomorphism $\psi : B \to B \# S^{2n-1} \# S^{2n-1}$ fixed near the boundary and a formal contact isotopy from $\ker \psi^*(\alpha \# \alpha_L \# \alpha'_L)|_B$ to $\xi_B = \ker \alpha|_B$, which is fixed near A.

If we extend ψ to a diffeomorphism $Y \to Y \# S^{2n-1} \# S^{2n-1}$ by letting it be the identity outside of B, we observe that $\ker \psi^*(\alpha \# \alpha_L \# \alpha'_L)$ is formally isotopic to $\xi = \ker \alpha$. Moreover, these contact structures agree on $Y - B \supset \mathcal{N}$. Since $(Y - \mathcal{N}, \xi)$ is overtwisted, it follows from the relative h-principle for overtwisted contact structures (see [5, Theorem 1.2]) that there is a smooth isotopy ϕ_t fixed on $\mathcal{N}$ so that $\tilde{\alpha} := \phi_1^* \psi^*(\alpha \# \alpha_L \# \alpha'_L)$ is a contact form for (Y, ξ). By construction, $\tilde{\alpha} = \alpha$ on $\mathcal{N}$, so $\tilde{\alpha}$ is adapted to $\mathfrak{r}$. Finally, it follows from the above discussion that $CH_\bullet(Y, \xi, V; \mathfrak{r})$ vanishes when we compute it using the form $\tilde{\alpha}$ (since γ bounds a rigid plane). An analogous argument shows that $\widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r})$ vanishes as well. □

We also state a vanishing result for the deformed Chekanov–Eliashberg dg algebra of certain loose Legendrians. To set the notation, let us now assume that (Y, ξ, V) is an arbitrary TN contact pair and fix $\mathfrak{r} \in \mathfrak{R}(Y, \xi, V)$.

Proposition* 11.2 *Suppose that $\Lambda \subset (Y - V, \xi)$ is a loose Legendrian submanifold. Then $\mathcal{L}(Y, \xi, V, \Lambda; \mathfrak{r})$ and $\widetilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r})$ are acyclic. Given augmentations $\epsilon : \mathcal{A}(Y, \xi, V; \mathfrak{r}) \to \mathbb{Q}[U]$ and $\tilde{\epsilon} : \widetilde{\mathcal{A}}(Y, \xi, V; \mathfrak{r}) \to \mathbb{Q}$, the invariants $\mathcal{L}^\epsilon(Y, \xi, V, \Lambda; \mathfrak{r})$ and $\widetilde{\mathcal{L}}^\epsilon(Y, \xi, V, \Lambda; \mathfrak{r})$ are also acyclic.*

Proof The argument is the same as that which shows that the (undeformed) Chekanov–Eliashberg dg algebra of a loose Legendrian is acyclic (see eg [55, Section 5]): up to Legendrian isotopy in $Y - V$, we can find a chord c of arbitrarily small action which bounds a single half-disk. This disk can be assumed to stay in a small ball disjoint from V for action reasons. Hence we have $d(c) = 1$. □

11.2 Nonvanishing results: bindings of open books

The following theorem is the main result of this section.

Theorem 11.3 *Consider a TN contact pair (Y, ξ, V). Suppose that Y admits an open book decomposition (Y, B, π) which supports the contact structure ξ and realizes $V = B$ as its binding. Viewing (B, τ) as a framed contact submanifold, where τ denote the trivialization of $B \subset Y$ induced by the open book decomposition, we have*

$$\widetilde{CH}_\bullet(Y, \xi, B; \mathfrak{r}) \neq 0 \tag{11-2}$$

for any $\mathfrak{r} = (\alpha_B, \tau, r) \in \mathfrak{R}(Y, \xi, B)$.

By combining Theorem 11.3 with Corollary 8.21, we obtain the following result.

Corollary 11.4 *Under the hypotheses of Theorem 11.3, if r' is large enough and we write $\mathfrak{r}' = (\alpha_B, \tau, r')$, then*

$$CH_\bullet(Y, \xi, B; \mathfrak{r}') \neq 0. \tag{11-3}$$

Proof of Theorem 11.3 According to Corollary 3.15, the open book decomposition (Y, B, π) supports a nondegenerate Giroux form α which is adapted to $\mathfrak{r}$ for any $\mathfrak{r} = (\alpha_B, \tau, r) \in \mathfrak{R}(Y, \xi, B)$.

Consider the algebra $\widetilde{CC}_\bullet(Y, \xi, B; \mathfrak{r})$ generated by (good) Reeb orbits of α not contained in B. After fixing an almost complex structure $J : \xi \to \xi$ which is compatible with $d\alpha$ and preserves $\xi|_B$, and a choice of perturbation data $\theta \in \Theta_{\mathrm{I}}((Y, \xi, B), \alpha, J)$, we get a differential $d_J = d(\widetilde{\psi}_B, J, \theta)$ and the homology of the resulting chain complex is (canonically isomorphic to) $\widetilde{CH}_\bullet(Y, \xi, B; \mathfrak{r})$.

Let us suppose for contradiction that $\widetilde{CH}_\bullet(Y, \xi, B; \mathfrak{r}) = 0$. This means that 1 is in the image of the differential. By the Leibniz rule, this implies that there exists some good Reeb orbit $\gamma : S^1 \to Y$ and a relative homotopy class $\beta \in \pi_2(Y, \gamma)$ such that the twisted moduli count of planes positively asymptotic to γ in the homotopy class β is nonzero. To state this more formally in the language of Section 5.3, let $T \in \mathcal{S}_{\mathrm{I}}((Y, \xi, B), \alpha, J)$ be the tree with a single input edge e and a single vertex v, where e is decorated with the Reeb orbit γ and v is decorated with the $\beta \in \pi_2(Y, \gamma)$. Then we have that $\widetilde{\psi}_B(T) \neq 0$.

In particular, this implies that $\overline{\mathcal{M}}(T) \neq \varnothing$. Hence there exists $T' \to T$ such that T' is representable by a J–holomorphic building. The proof of Proposition 7.7 shows that we may assume that T' does not have any edges contained in B (since otherwise we would have $\widetilde{\psi}_B(T') = \widetilde{\psi}_B(T) = 0$).

It follows by Proposition 5.12 that $T' * \widehat{B} = \sum_{v \in V(T')} \beta_v * \widehat{B}$, and Corollary 5.14 implies that all the terms on the right-hand side are nonnegative. Since $\widetilde{\psi}_B(T') = \widetilde{\psi}_B(T) \neq 0$, it follows by definition of the reduced twisting maps that $T' * \widehat{V} = 0$. Hence $\beta_v * \widehat{B} = 0$ for all $v \in V(T')$.

For topological reasons, there exists $\widetilde{v} \in V(T')$ such that $\widetilde{v}$ has a single incoming edge and no outgoing edges. Hence $\widetilde{v}$ is represented by a J–holomorphic plane u which is asymptotic to some Reeb orbit $\widetilde{\gamma}$. By positivity of intersection (see Proposition 5.3) and the fact that $\beta_{\widetilde{v}} * \widehat{B} = 0$, the image of u is contained in $\mathbb{R} \times (Y \setminus B)$. Thus $\widetilde{\gamma}$ is contractible in $Y \setminus B$, which implies that the composition $\pi \circ \widetilde{\gamma} \colon S^1 \to Y \setminus B \to S^1$ has degree 0. This is a contradiction: since α is a Giroux form, $\pi \circ \widetilde{\gamma}$ must be an immersion by Remark 3.12, and hence have nonzero degree. □

We also state a vanishing result for the (reduced) Chekanov–Eliashberg dg algebra introduced in Section 10.2.

Theorem* 11.5 *Let (Y, ξ, V) be a TN contact pair and let $\Lambda \subset (Y - V, \xi)$ be a Legendrian submanifold. Suppose that ξ supports an open book decomposition π with binding $B = V$ such that Λ is contained in a single page. Let τ be the trivialization of $N_{Y/V}$ induced by the open book. Then we have*

$$\widetilde{\mathcal{L}}(Y, \xi, V, \Lambda; \mathfrak{r}) \neq 0. \tag{11-4}$$

Proof The proof is identical to that of Theorem 11.3; namely, one argues that the image of any Reeb orbit or chord under the differential cannot contain a term of degree zero, which immediately implies the claim. □

We note that is was proved by Honda and Huang [39, Corollary 1.3.3] that any Legendrian Λ in a contact manifold (Y, ξ) is contained in the page of some compatible open book decomposition. Hence it follows from Theorem* 11.5 that every Legendrian is tight in the complement of some codimension 2 contact submanifold.

11.3 Explicit computations in open books

We now perform certain explicit computations in open book decompositions which will be used in applications in the next sections. We assume throughout this section that $n \geq 4$. This assumption is needed for the purpose of obtaining a $(\mathbb{Z} \times \mathbb{Z})$–bigrading on $\widetilde{CH}(-)$; see Lemma 11.6 and Corollary 12.17.

Let us endow S^{n-1} with a Riemannian metric h having the property that all geodesics are nondegenerate. Such metrics, which are typically referred to as "bumpy" in the literature, are generic in the space of Riemannian metrics; see [1] or [44, 3.3.9]. It can be shown [1] that any manifold endowed with a bumpy metric admits a closed geodesic of minimal length. We let $\rho > 0$ be the length of the shortest geodesic of (S^{n-1}, h).

To set the stage for this section, it will be useful to recall some general facts about coordinate systems. Given a system of local coordinates $(q_1,\dots,q_m)$ on some manifold M, the *dual coordinates* $(p_1,\dots,p_m)$ in the fibers of the T^*M are characterized by the property that

$$T^*M \ni (\boldsymbol{q},\boldsymbol{p}) = (q_1,\dots,q_m,p_1,\dots,p_m) = \sum_{i=1}^{m} p_i\,dq_i. \tag{11-5}$$

Unless otherwise indicated, a pair $(\boldsymbol{q},\boldsymbol{p})$ refers to a system of local coordinates in the cotangent bundle of a manifold, where $\boldsymbol{p}$ is dual to $\boldsymbol{q}$.

It will sometimes also be useful to work with Riemannian normal coordinates. Recall that on a Riemannian manifold (M,g), a system of normal coordinates $(x_1,\dots,x_m)$ has the property that for any vector $\boldsymbol{a}\in T_xM$, the path $\theta\mapsto \boldsymbol{a}t$ is a geodesic. If $(q_1,\dots,q_n)$ is a system of Riemannian normal coordinates, then the path $t\mapsto(\gamma(t),\dot\gamma^\flat)$ can be written in coordinates $(\boldsymbol{q},\boldsymbol{p})$ as

$$t\mapsto(\boldsymbol{a}t,\boldsymbol{a})\in T^*M. \tag{11-6}$$

We now introduce a Liouville manifold which will be studied throughout the remainder of this section. For $a>0$, define

$$(\widehat{W}_0,\hat\lambda^a) = \Big(D^2\times T^*S^{n-1},\ \hat\lambda^a := \frac{1}{a}s^2\,d\theta+\lambda_{\mathrm{std}}\Big), \tag{11-7}$$

where we have chosen local coordinates $(s,\theta,\boldsymbol{q},\boldsymbol{p})$. We emphasize that the Liouville structure depends on the parameter $a>0$.

Let $\phi\colon \widehat{W}_0\to\mathbb{R}$ be the function

$$\phi(s,\theta,\boldsymbol{q},\boldsymbol{p}) = s^2+\|\boldsymbol{p}\|^2. \tag{11-8}$$

We consider the Liouville domain

$$(W_0,\lambda^a) = (\{\phi\le 1\},\ \lambda^a := \hat\lambda^a|_{W_0}) \tag{11-9}$$

and its contact-type boundary

$$(Y_0,\xi_0) = (\{\phi=1\},\ \xi=\ker\lambda_0), \tag{11-10}$$

where $\lambda_0 = (\hat\lambda^a)|_{Y_0}$ is the induced contact form. Consider also the codimension 2 contact submanifold

$$V = \{\phi=1, s=0\}\subset(Y_0,\xi_0) \tag{11-11}$$

and the Legendrian

$$\Lambda := \{\phi=1,\ \theta=\text{constant},\ s=1,\ \|p\|=0\}. \tag{11-12}$$

We define $\alpha := (\lambda_0)|_V$ and let τ be the trivialization of $N_{Y_0/V}$, which is unique by Lemma 11.6. We set

$$\mathfrak{r} = (\alpha,\tau,a)\in\mathfrak{R}(Y_0,\xi_0,V). \tag{11-13}$$

Finally, we let $H=\{0\}\times T^*S^{n-1}\subset\widehat{W}_0$.

Observe that $\mathfrak{r}$ *depends on our choice of* $a>0$. More generally, the contact form $\lambda_0=(\hat{\lambda}^a)|_{Y_0}$ on (Y_0,ξ_0) obviously depends on $a>0$. The plan is now to study the Reeb dynamics on (Y_0,ξ_0) with respect to this contact form. By taking $a\gg 0$ large enough, we will be able to obtain a sufficiently good understanding of the Reeb dynamics to compute the invariant $\widetilde{CH}(Y_0,\xi_0,V;\mathfrak{r})$ in low degrees; see Proposition* 11.17.

Lemma 11.6 *The manifolds* W_0 *and* Y_0 *have vanishing first and second homology and cohomology with* $\mathbb{Z}$*–coefficients. In addition, we have* $H^1(V;\mathbb{Z})=0$ *and* $w_2(\Lambda)=0$.

Proof The first claim is proved in Corollary 12.17. To compute $H^1(V;\mathbb{Z})$, note that V is the sphere bundle associated to T^*S^{n-1}. Hence, we have a fibration $S^{n-2}\hookrightarrow V\to S^{n-1}$ giving rise to a Gysin sequence

$$\cdots\to H^k(S^{n-1};\mathbb{Z})\to H^k(V;\mathbb{Z})\to H^{k-(n-2)}(S^{n-1};\mathbb{Z})\to\cdots. \tag{11-14}$$

Taking $k=1$ immediately gives the desired result since $n\geq 4$. Finally, note that $\Lambda=S^{n-1}$, which has vanishing homology (with any coefficients) in degrees $1\leq i\leq n-2$. Hence $w_2(\Lambda)=0$ for $n\geq 4$. □

Observe that there is a natural marking

$$e_0\colon \mathbb{R}\times Y_0\to(\widehat{W}_0,\hat{\lambda}^a,H),\quad (t,y)\mapsto\psi_t(y), \tag{11-15}$$

where $\psi_{(-)}$ is the Liouville flow associated to $\hat{\lambda}^a$.

This endows $(\widehat{W}_0,\hat{\lambda}^a,H)$ with the structure of a (strict) relative exact symplectic cobordism. We thus obtain an augmentation

$$\tilde{\epsilon}_0\colon\widetilde{\mathcal{A}}(Y_0,\xi_0,V;\mathfrak{r})\to\mathbb{Q}. \tag{11-16}$$

It follows from Lemma 11.6 and the discussion following Definition* 9.10 that $\widetilde{\mathcal{A}}(Y_0,\xi_0,V;\mathfrak{r})$ and $\widetilde{\mathcal{A}}^{\tilde{\epsilon}_0}(Y_0,\xi_0,V;\mathfrak{r})$ admit a $(\mathbb{Z}\times\mathbb{Z})$–bigrading.

We now analyze the structure of (Y_0,λ_0) in more detail. First, observe that $(Y_0-V,\lambda_0|_{Y_0-V})$ is strictly contactomorphic to

$$\Big(S^1\times D^*S^{n-1},\alpha_V:=\frac{1}{a}(1-\|\boldsymbol{p}\|^2)\,d\theta+\lambda_{\mathrm{std}}\Big) \tag{11-17}$$

via the map

$$S^1\times D^*S^{n-1}\to Y_0-V,\quad (\theta,\boldsymbol{q},\boldsymbol{p})\mapsto\big(\sqrt{1-\|\boldsymbol{p}\|^2},\theta,\boldsymbol{q},\boldsymbol{p}\big), \tag{11-18}$$

where $D^*S^{n-1}=\{(\boldsymbol{q},\boldsymbol{p})\in T^*S^{n-1}\mid\|\boldsymbol{p}\|<1\}$. We let $\mathcal{N}\subset Y_0$ denote the image of $S^1\times S^{n-1}$ under this map; equivalently, $\mathcal{N}=\{\|\boldsymbol{p}\|=0\}$. The complement $(Y_0-\mathcal{N},\lambda_0|_{Y_0-\mathcal{N}})$ is strictly contactomorphic to

$$\Big(B\times U,\alpha_{\mathcal{N}}:=\frac{1}{a}(x\,dy-y\,dx)+\sqrt{1-x^2-y^2}\alpha_U\Big),$$

where $B \subset \mathbb{R}^2$ denotes the open unit disk and (U, α_U) denotes the unit cotangent bundle of (S^{n-1}, h), equipped with the contact form $\alpha_U := \lambda_{\text{std}}$ induced by the canonical Liouville form on T^*S^{n-1}. A contactomorphism is given by

$$B \times U \to Y_0 - \mathcal{N}, \quad (x, y, \boldsymbol{q}, \boldsymbol{p}) \mapsto \big(x, y, \boldsymbol{q}, \sqrt{1-x^2-y^2}\,\boldsymbol{p}\big). \tag{11-19}$$

Observe that the size of the tubular neighborhood $B \times U$ depends on our choice of $a > 0$.

Our first task is to study the Reeb orbits of λ_0 which are in the complement of $\mathcal{N}$. In particular, we wish to show that they are nondegenerate for a generic choice of a, and moreover that their Conley–Zehnder indices depend linearly on a. This is the content of Proposition 11.7 and Corollary 11.8 below.

Proposition 11.7 *Let $\gamma_U : \mathbb{R}/\mathbb{Z} \to U$ be a Reeb orbit of α_U of period P_U. Then:*

(A) *The map*

$$\gamma_1 : \mathbb{R}/\mathbb{Z} \to B \times U, \quad t \mapsto (0, 0, \gamma_U(t)), \tag{11-20}$$

is a Reeb orbit of $\alpha_{\mathcal{N}}$ of period $P_1 := P_U$.

(B) *Given any $r_0 \in (0,1)$ and integers $m, n > 0$ such that*

$$\frac{a P_U}{4\pi\sqrt{1-r_0^2}} = \frac{m}{n}, \tag{11-21}$$

the map

$$\gamma_2 : \mathbb{R}/\mathbb{Z} \to B \times U, \quad t \mapsto (r_0\cos(2\pi m t), r_0 \sin(2\pi m t), \gamma_U(nt)), \tag{11-22}$$

is a Reeb orbit of $\alpha_{\mathcal{N}}$ of period

$$P_2 := (2-r_0^2)\frac{2\pi m}{a} = (2-r_0^2)\frac{n P_U}{2\sqrt{1-r_0^2}}.$$

Every Reeb orbit of $\alpha_{\mathcal{N}}$ is of the form (A) *or* (B) *for some choice of γ_U, r_0, m and n.*

If α_U is nondegenerate and a satisfies

$$a^{-1} \notin \bigcup_{q \in \mathbb{Q}_{>0}} \frac{1}{4\pi\sqrt{q}} \mathcal{S}(\alpha_U), \tag{11-23}$$

where $\mathcal{S}(\alpha_U) \subset \mathbb{R}$ is the action spectrum of α_U, then $\alpha_{\mathcal{N}}$ is nondegenerate. Moreover, given any trivialization τ_0 of $\gamma_U^* \ker(\alpha_U)$, there exist trivializations τ_i of $\gamma_i^* \ker(\alpha_{\mathcal{N}})$ for $i = 1, 2$ such that

$$\mathrm{CZ}^{\tau_1}(\gamma_1) = 1 + 2\left\lfloor \frac{P_1 a}{4\pi} \right\rfloor + \mathrm{CZ}^{\tau_0}(\gamma_U), \tag{11-24}$$

$$\mathrm{CZ}^{\tau_2}(\gamma_2) = 1 + 2\left\lfloor \frac{P_2 a}{\pi(2-r_0^2)^2} \right\rfloor + \mathrm{CZ}^{\tau_0}(\gamma_U^n). \tag{11-25}$$

If τ_0 extends to a disk spanning γ_U, then τ_i extends to a disk spanning γ_i.

Proof The Reeb vector field of $\alpha_{\mathcal{N}}$ is given by

$$R_{\alpha_{\mathcal{N}}} = \frac{1}{2-x^2-y^2}\left(a(x\partial_y - y\partial_x) + 2\sqrt{1-x^2-y^2}\,R_U\right), \tag{11-26}$$

where R_U denotes the Reeb vector field of α_U (recall that $a > 0$ is a constant fixed above). A simple computation shows that γ_1 and γ_2 are Reeb orbits with periods as claimed, and that there are no other orbits.

Note that the contact structure $\xi = \ker(\alpha_{\mathcal{N}})$ splits as

$$\xi = \langle e_1, e_2\rangle \oplus \ker(\alpha_U), \tag{11-27}$$

where e_1 and e_2 are the vector fields on $B \times U$ defined by

$$e_1 = \partial_x + \frac{y}{a\sqrt{1-x^2-y^2}}R_U, \tag{11-28}$$

$$e_2 = \partial_y - \frac{x}{a\sqrt{1-x^2-y^2}}R_U. \tag{11-29}$$

In particular, given a trivialization τ_0 of $\gamma_U^* \ker(\alpha_U)$, we get trivializations $\tau_1 = \langle \gamma_1^* e_1, \gamma_1^* e_2\rangle \oplus \tau_0$ and $\tau_2 = \langle \gamma_2^* e_1, \gamma_2^* e_2\rangle \oplus \tau_0^n$ of $\gamma_1^*\xi$ and $\gamma_2^*\xi$, where τ_0^n denotes the trivialization of $(\gamma_U^n)^* \ker(\alpha_U)$ induced by τ_0.

We have

$$\begin{aligned}\mathcal{L}_{e_1} R_{\alpha_{\mathcal{N}}} &= -\partial_x\left(\frac{ay}{2-x^2-y^2}\right)e_1 + \partial_x\left(\frac{ax}{2-x^2-y^2}\right)e_2 \\ &= \frac{a}{(2-x^2-y^2)^2}\left(-2xye_1 + (2+x^2-y^2)e_2\right),\end{aligned} \tag{11-30}$$

$$\begin{aligned}\mathcal{L}_{e_2} R_{\alpha_{\mathcal{N}}} &= \partial_y\left(\frac{ay}{2-x^2-y^2}\right)e_1 - \partial_y\left(\frac{ax}{2-x^2-y^2}\right)e_2 \\ &= \frac{a}{(2-x^2-y^2)^2}\left(-(2-x^2+y^2)e_1 + 2xye_2\right).\end{aligned} \tag{11-31}$$

Moreover, for any vector field X on U such that $X \in \ker(\alpha_U)$, we have

$$\mathcal{L}_X R_{\alpha_{\mathcal{N}}} = \frac{2\sqrt{1-x^2-y^2}}{2-x^2-y^2}\mathcal{L}_X R_U. \tag{11-32}$$

Hence, if $\Psi_i(t)\colon \xi_{\gamma_i(0)} \to \xi_{\gamma_i(t)}$ denotes the linearized Reeb flow along γ_i (viewed as a matrix via the trivialization τ_i) for $i = 1, 2$, then $\Psi_i'(t) = S_i(t)\Psi_i(t)$ with

$$S_1(t) = \frac{aP_U}{2}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} \oplus P_U S_U(t), \tag{11-33}$$

$$S_2(t) = \frac{2\pi m}{2-r_0^2}\begin{pmatrix}-r_0^2\sin(4\pi mt) & -2+r_0^2\cos(4\pi mt)\\ 2+r_0^2\cos(4\pi mt) & r_0^2\sin(4\pi mt)\end{pmatrix} \oplus nP_U S_U(nt), \tag{11-34}$$

where $S_U(t)$ is the matrix such that the linearized Reeb flow $\Psi_U : (\xi_U)_{\gamma_U(0)} \to (\xi_U)_{\gamma_U(t)}$ of R_U along γ_U satisfies $\Psi_U'(t) = P_U S_U(t)\Psi_U(t)$.

It follows that $\mathrm{CZ}^{\tau_1}(\gamma_1)=\mathrm{CZ}(\psi_1)+\mathrm{CZ}^{\tau_0}(\gamma_U)$ and $\mathrm{CZ}^{\tau_2}(\gamma_2)=\mathrm{CZ}(\psi_2)+\mathrm{CZ}^{\tau_0}(\gamma_U^n)$, where ψ_1 and ψ_2 are paths of 2×2 matrices given by $\psi_i(t)=\exp(P_i(t))$ with

$$P_1(t)=t\frac{aP_U}{2}\begin{pmatrix}0&-1\\1&0\end{pmatrix}, \tag{11-35}$$

$$\begin{aligned}P_2(t)&=\int_0^t\frac{2\pi m}{2-r_0^2}\begin{pmatrix}-r_0^2\sin(4\pi ms)&-2+r_0^2\cos(4\pi ms)\\2+r_0^2\cos(4\pi ms)&r_0^2\sin(4\pi ms)\end{pmatrix}ds,\\&=\frac{2\pi m}{2-r_0^2}\begin{pmatrix}\frac{r_0^2}{4\pi m}(\cos(4\pi mt)-1)&-2t+\frac{r_0^2}{4\pi m}\sin(4\pi mt)\\2t+\frac{r_0^2}{4\pi m}\sin(4\pi mt)&-\frac{r_0^2}{4\pi m}(\cos(4\pi mt)-1)\end{pmatrix}.\end{aligned} \tag{11-36}$$

Note that $P_1(t)$ and $P_2(t)$ are diagonalizable with eigenvalues $\pm 2\pi i\lambda_1(t)$ and $\pm 2\pi i\lambda_2(t)$, respectively, where

$$\lambda_1(t)=t\frac{aP_U}{4\pi}, \tag{11-37}$$

$$\lambda_2(t)=\frac{1}{2-r_0^2}\sqrt{4m^2t^2-\frac{r_0^4}{8\pi^2}(1-\cos(4\pi mt))}. \tag{11-38}$$

It follows that $\ker(\psi_i(t)-\mathrm{Id})$ is either $\mathbb{R}^2$ or 0, depending on whether $\lambda_i(t)$ is an integer or not. Assumption (11-23) implies that $\lambda_i(1)$ is not an integer and hence that $\psi_i(1)$ doesn't have 1 as an eigenvalue, ie ψ_i is nondegenerate. This is clear for $\lambda_1(1)$; to check this for $\lambda_2(1)$, note that it follows from (11-21) that

$$\lambda_2(1)=\frac{2m}{2-r_0^2}=\frac{2m}{1+\frac{n^2a^2P_U^2}{(4\pi)^2m^2}}=\frac{2m^3(4\pi)^2}{(4\pi)^2m^2+n^2a^2P_U^2}. \tag{11-39}$$

If this expression were integral, then the reciprocal would be rational; hence $n^2a^2P_U^2/(2m^3(4\pi)^2)$ would be rational, contradicting Assumption (11-23).

Since $-J_0P_i'(t)$ is positive-definite for all t, it follows from [33, Proposition 52] that

$$\mathrm{CZ}(\psi_i)=1+2\,\#\{t\in(0,1)\mid\lambda_i(t)\in\mathbb{Z}\}. \tag{11-40}$$

Since λ_i is strictly increasing with $\lambda_i(0)=0$ and $\lambda_i(1)\notin\mathbb{Z}$, the right-hand side is equal to $1+2\lfloor\lambda_i(1)\rfloor$. Thus

$$\mathrm{CZ}(\psi_1)=1+2\left\lfloor\frac{aP_U}{4\pi}\right\rfloor=1+2\left\lfloor\frac{aP_1}{4\pi}\right\rfloor, \tag{11-41}$$

$$\mathrm{CZ}(\psi_2)=1+2\left\lfloor\frac{2m}{2-r_0^2}\right\rfloor=1+2\left\lfloor\frac{aP_2}{\pi(2-r_0^2)^2}\right\rfloor, \tag{11-42}$$

as desired. □

Corollary 11.8 *Suppose that γ is a closed Reeb orbits of $(Y_0, \xi = \ker \lambda_0)$ which is contained in the complement of $\mathcal{N} \subset Y_0$. Then*

$$\mathrm{CZ}^{\tau}(\gamma) > \left\lfloor \frac{a\rho}{\pi} \right\rfloor, \tag{11-43}$$

where τ is a trivialization which extends to a spanning disk and $\rho > 0$ is as on page 92.

Proof It is well-known that the Reeb orbits on U correspond bijectively to geodesics on (S^{n-1}, h); with our normalization, the action of a closed Reeb orbit equals twice the length of the corresponding unit-speed geodesic; see eg [29, Section 1.5]. Moreover, according to [24, Proposition 1.7.3], given a Reeb orbit $\widetilde{\gamma}$ which corresponds to a geodesic γ, we have

$$\mu_M(\gamma) = \mathrm{CZ}^{\tau}(\widetilde{\gamma}), \tag{11-44}$$

where μ_M is the Morse index of the geodesic and τ extends to a spanning disk; see Remark 11.9. Since the Morse index of a geodesic is nonnegative by definition, the corollary follows from Proposition 11.7. □

Remark 11.9 The trivialization considered in [24, Proposition 1.7.3] is in fact constructed as follows. Choose a spanning disk $\widetilde{v}\colon D^2 \to U \subset T^*S^{n-1}$ for $\widetilde{\gamma}$ and let $v := \pi \circ \widetilde{v}$, where $\pi\colon T^*S^{n-1} \to S^{n-1}$ is the projection. Let $\{\sigma^1, \dots, \sigma^{n-1}\}$ be a trivialization of v^*TS^{n-1}. For points $\pi\colon \widetilde{x} \mapsto x$, let $Q_{x;\widetilde{x}}\colon \pi^{-1}(x) \to T_{\widetilde{x}}(\pi^{-1}(x))$ be the canonical identification. Now define $\widetilde{\sigma}^i_p = Q_{v(p);\widetilde{v}(p)}\sigma_p$ for $p \in D^2$. Then $\{\widetilde{\sigma}^1, \dots, \widetilde{\sigma}^{n-1}\}$ defines a Lagrangian subbundle of the symplectic vector bundle $(\widetilde{v}^*(\xi), d\lambda_0)$. Hence it induces a unique trivialization of $\widetilde{v}^*\xi$, which restricts on the boundary to a trivialization of $\widetilde{\gamma}^*\xi$.

We now turn to the Reeb dynamics near $\mathcal{N}$. Recall from page 94 that $\mathcal{N}$ is contained in $(Y_0 - V, \lambda_0)$, which is strictly contactomorphic to $(S^1 \times D^*S^{n-1}, \alpha_V)$, where $\alpha_V = (1/a)(1 - \|\boldsymbol{p}\|^2)\, d\theta + \lambda_{\mathrm{std}}$.

Lemma 11.10 *Let $\boldsymbol{q} = (q_1, \dots, q_{n-1})$ be Riemannian normal coordinates in some open set $\mathcal{U} \subset (S^{n-1}, h)$ and let $\boldsymbol{p} = (p_1, \dots, p_n)$ be the dual coordinates. The Reeb vector field of α_V is given by*

$$R_{\alpha_V} = \frac{1}{1 + \|p\|^2}\left(a\partial_\theta + 2\sum_{i,j} h^{ij} p_i \partial_{q_j} - \sum_{i,j,k} p_i p_j \partial_k h^{ij} \partial_{p_k}\right) \tag{11-45}$$

on $S^1 \times D^*\mathcal{U}$. (Here we follow the convention of using superscripts $(h^{ij}) = (h_{ij})^{-1}$ to denote the coefficients of the metric induced by h on T^*S^{n-1}.)

Proof A direct computation using the formulas

$$\alpha_V = \frac{1}{a}\left(1 - \sum_{i,j} p_i p_j h^{ij}\right) d\theta + \sum_i p_i\, dq_i, \tag{11-46}$$

$$d\alpha_V = -\frac{2}{a}\sum_{i,j} h^{ij} p_i\, dp_j \wedge d\theta - \frac{1}{a}\sum_{i,j,k} p_i p_j \partial_k h^{ij}\, dq_k \wedge d\theta + \sum_i dp_i \wedge dq_i \tag{11-47}$$

shows that $\alpha_V(R_{\alpha_V}) = 1$ and $d\alpha_V(R_{\alpha_V}, -) = 0$. □

Lemma 11.11 *Consider the map $\pi\colon Y_0 - V \to S^1$ given by $\pi(s,\theta,q,p) = \theta$. Then the pair (V,π) defines an open book decomposition of Y. Moreover, λ_0 is a Giroux form for the contact structure $\xi_0 = \ker\lambda_0$.*

Proof It is clear that (V,π) defines an open book decomposition of Y. To verify that λ_0 is a Giroux form, observe by Lemma 11.10 that the Reeb vector field is transverse to the pages of π. The claim then follows by combining Definition 3.11 and Remark 3.12. □

By Lemma 11.10, the map $\gamma_0\colon \mathbb{R}/\mathbb{Z} \to S^1 \times D^*\mathcal{U}$ given by the formula $\gamma_0(t) = (2\pi t, 0, 0)$ defines a simple Reeb orbit in Y_0. Let γ_0^k denote its k–fold cover. There is an obvious trivialization τ_0 of $\xi|_{\gamma_0^k}$ given by

$$\tau_0 = \{\partial_{p_1}, \dots, \partial_{p_{n-1}}, \partial_{q_1}, \dots, \partial_{q_{n-1}}\}. \tag{11-48}$$

Let τ be the trivialization of $\xi|_{\gamma_0^k}$ defined as

$$\tau = \{\sin(2\pi kt)\partial_{q_1} + \cos(2\pi kt)\partial_{p_1}, \partial_{p_2}, \dots \partial_{p_n}, \cos(2\pi kt)\partial_{q_1} - \sin(2\pi kt)\partial_{p_1}, \partial_{q_2}, \dots, \partial_{q_n}\}. \tag{11-49}$$

Observe that τ extends to a disk spanning γ_0 in Y_0.

Lemma 11.12 *With respect to the trivialization τ_0, the linearized Reeb flow along γ_0^k is given by the matrix*

$$\begin{pmatrix} 1 & 0 \\ 2t & 1 \end{pmatrix}, \tag{11-50}$$

where each entry of this matrix should be viewed as an $(n-1) \times (n-1)$ diagonal matrix.

Proof Note that τ_0 can be extended to a trivialization $\tilde{\tau}_0$ of $\ker(\alpha_V)$ over $S^1 \times D^*\mathcal{U}$, where

$$\tilde{\tau}_0 = \left\{ \partial_{p_1}, \dots, \partial_{p_{n-1}}, \partial_{q_1} - \frac{ap_1}{1 - \|\boldsymbol{p}\|^2}\partial_\theta, \dots, \partial_{q_{n-1}} - \frac{ap_{n-1}}{1 - \|\boldsymbol{p}\|^2}\partial_\theta \right\}. \tag{11-51}$$

Using the formula for R_{α_V} given in Lemma 11.10, one can easily compute

$$\mathcal{L}_{\partial_{p_i}} R_{\alpha_V}|_{\boldsymbol{p}=0,\boldsymbol{q}=0} = 2\partial_{q_i}, \tag{11-52}$$

$$\mathcal{L}_{\partial_{q_i} - (ap_i/(1-\|\boldsymbol{p}\|^2))\,\partial_\theta} R_{\alpha_V}|_{\boldsymbol{p}=0,\boldsymbol{q}=0} = 0. \tag{11-53}$$

Hence, the matrix $A(t)$ representing the linearized Reeb flow $\xi_{\gamma_0^k(0)} \to \xi_{\gamma_0^k(t)}$ with respect to the trivialization τ_0 is given by

$$A(t) = \exp\left(t \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 \\ 2t & 1 \end{pmatrix}, \tag{11-54}$$

where each entry should be interpreted as a multiple of the $(n-1) \times (n-1)$ identity matrix. □

Corollary 11.13 *The Robbin–Salamon index satisfies*

$$\mu_{\mathrm{RS}}^{\tau_0}(\gamma_0^k) = \tfrac{1}{2}(n-1). \tag{11-55}$$

Hence,

$$\mu_{\mathrm{RS}}^{\tau}(\gamma_0^k) = \tfrac{1}{2}(n-1) + 2k. \tag{11-56}$$

Proof The first computation follows from [33, Proposition 54]; there is a sign change due to the fact that the matrix we are considering is the transpose of that considered in [33, Proposition 54], but the proof is entirely analogous. The second computation follows from the fact (see the proof of Lemma 57 in [33]) that the Robbin–Salamon index satisfies the so-called "loop property", ie given a path of symplectic matrices $\phi\colon [0,1] \to \mathrm{Sp}(2n,\mathbb{R})$ with $\phi(0) = \phi(1) = \mathrm{id}$, and given a path $\psi\colon [0,1] \to \mathrm{Sp}(2n,\mathbb{R})$, we have

$$\mu_{\mathrm{RS}}(\phi\psi) = \mu_{\mathrm{RS}}(\psi) + 2\mu(\phi), \tag{11-57}$$

where ϕ is the Maslov index of the path. □

By Lemma 11.10, $\mathcal{N} = \{\|p\| = 0\}$ is preserved by the Reeb flow and is foliated by Reeb orbits in a Morse–Bott family.

Given $\epsilon > 0$ which will be fixed later, let $\mathcal{U}_\epsilon = \{\|p\| < \epsilon\} \cap Y_0$. This is a neighborhood of $\mathcal{N}$, which we identify with $S^1 \times D_\epsilon^* S^{n-1}$ via the contactomorphism defined on page 94. Let $f\colon \mathcal{U}_\epsilon \to \mathbb{R}$ be the function corresponding under this identification to

$$S^1 \times D_\epsilon^* S^{n-1} \to \mathbb{R}, \quad (\theta, \boldsymbol{q}, \boldsymbol{p}) \mapsto \rho(\|\boldsymbol{p}\|) g(\boldsymbol{q}), \tag{11-58}$$

where g is a perfect Morse function on S^{n-1} and $\rho\colon \mathbb{R} \to [0,1]$ is a smooth bump function with $\rho(x) = 1$ for x near 0 and $\rho(x) = 0$ for $x > \epsilon/2$; cf [6, Section 2.2].

Lemma 11.14 *Fix $P > 0$. If ϵ is small enough, all closed Reeb orbits of (Y_0, λ_0) which are contained in $\mathcal{U}_\epsilon - \mathcal{N}$ have action at least P.* □

We now consider a perturbed contact form $\lambda_\delta := (1 + \delta f)\lambda_0$. Since f is compactly supported in $\mathcal{U}_\epsilon$, the form λ_δ can be viewed as a contact form both on $\mathcal{U}_\epsilon$ and on Y_0.

Lemma 11.15 *Fix $P > 0$. If ϵ and δ are small enough, then there are exactly two simple Reeb orbits in $\mathcal{U}_\epsilon$ with action $< P$. We label them γ_a and γ_b, and they correspond respectively to the minimum and maximum of f.*

Proof Combine Lemma 11.14 with the argument of [6, Lemma 2.3]. □

Lemma 11.16 *Let $P > 0$ be as in Lemma 11.15. After possibly further shrinking ϵ and δ, we may assume that any Reeb orbit of (Y_0, λ_δ) contained in $\mathcal{U}_\epsilon$ and having Conley–Zehnder index (measured with respect to a trivialization which extends to a spanning disk) less than P/a is a multiple of γ_a or γ_b. In addition, we have*

$$\mathrm{CZ}^\tau(\gamma_a^k) = \mu_{\mathrm{RS}}^\tau(\gamma_0^k) - \tfrac{1}{2}(n-1) + \mathrm{ind}_a(\delta f) = 2k, \tag{11-59}$$

$$\mathrm{CZ}^\tau(\gamma_b^k) = \mu_{\mathrm{RS}}^\tau(\gamma_0^k) - \tfrac{1}{2}(n-1) + \mathrm{ind}_b(\delta f) = (n-1) + 2k. \tag{11-60}$$

Proof First of all, observe by Lemma 11.10 that the boundary of $\mathcal{U}_\epsilon$ is preserved by the Reeb flow of λ_0. It follows that the Reeb flow of λ_0 has "bounded return time", in the terminology of [6, Definition 2.5].

Next, it follows from (11-56) that the Robbin–Salamon index of any Reeb orbit γ contained in the Morse–Bott submanifold $\mathcal{N} = \{\|p\|=0\} \subset Y_0$ satisfies

$$\mu_{\mathrm{RS}}(\gamma) = \tfrac{1}{2}(n-1) + 2\,\mathrm{wind}(\gamma) = \tfrac{1}{2}(n-1) + 2T_\gamma a, \tag{11-61}$$

where P_γ is the length of γ. It follows that these orbits satisfy "index positivity" (with constant $2/a$), in the terminology of [6, Definition 2.6].

The first claim now follows from [6, Lemma 2.7]. The index computations follow by combining (11-56) with [6, Lemma 2.4]. □

We now put together the above results. For any integer $N > 0$, let us define

$$\Sigma_N^1 = \{k \in \mathbb{Z} \mid 0 < k < N,\ k \text{ even}\}, \tag{11-62}$$

and let

$$\Sigma_N^2 = \{k \in \mathbb{Z} \mid k < N,\ k = n-1+2j \text{ for } j \geq 1\}. \tag{11-63}$$

Proposition* 11.17 *Given any $N > 0$, there exists $A > 0$ such that*

$$CH^{U=0}_{k-(n-3)}(Y_0, \xi_0, V; \mathfrak{r}) = \widetilde{CH}_{k-(n-3)}(Y_0, \xi_0, V; \mathfrak{r}) = \begin{cases} \mathbb{Q} \oplus \mathbb{Q} & \text{if } k \in \Sigma_N^1 \cap \Sigma_N^2, \\ \mathbb{Q} & \text{if } k \in \Sigma_N^1 \cup \Sigma_N^2 - (\Sigma_N^1 \cap \Sigma_N^2), \\ 0 & \text{if } k \notin \Sigma_N^1 \cup \Sigma_N^2 \text{ and } k < N, \end{cases} \tag{11-64}$$

whenever $a > A$. (Recall from (11-13) that $\mathfrak{r}$ depends on $a > 0$ and hence on $N > 0$.)

Proof According to Corollary 11.8, we may fix $A > 0$ large enough so that all Reeb orbits for (Y_0, λ_0) in the complement of $\mathcal{N} \subset Y_0$ have index at least N. We now choose ϵ and δ small enough so that the conclusions of Lemma 11.16 hold with $P = N$. Since f_δ is compactly supported in $\mathcal{U}_\epsilon$, we find that the only Reeb orbits of (Y_0, λ_δ) having index less than N are multiples of γ_a and γ_b.

According to (11-59) and (11-60), it is now enough to check that the differential vanishes on the set

$$\Omega_N = \{\gamma_a^{k_a}, \gamma_b^{k_b} \mid \mathrm{CZ}^\tau(\gamma_a^{k_a}) < N, \mathrm{CZ}^\tau(\gamma_b^{k_b}) < N\}.$$

To see this, observe that for $\gamma \in \Omega_N$ we have

$$\mathrm{CZ}^\tau(\gamma) = 2\,\mathrm{wind}(\gamma) \mod (n-1). \tag{11-65}$$

Suppose that there exists a homotopy class β of curves of index 1 with $\hat{V} * \beta = 0$. Then the linking number of the positive puncture equals the sum of the linking numbers of the negative punctures. Hence, by (11-65), the index of the positive puncture equals the sum of the indices of the negative punctures mod $(n-1)$. Since β has index 1, this means that $1 = 0 \mod (n-1)$. This is a contradiction since $n > 2$. □

Corollary* 11.18 *Let $N > 0$ be as in Proposition* 11.17. Then for all integers $k < N$ we have*

$$\widetilde{CH}^{\tilde{\epsilon}_0}_{k-(n-3)}(Y_0, \xi_0, V; \mathfrak{r}) = \widetilde{CH}_{k-(n-3)}(Y_0, \xi_0, V; \mathfrak{r}), \tag{11-66}$$

where the right-hand side was computed in Proposition 11.17.* □

We now turn out attention to computing certain Legendrian invariants. Let $\Lambda \subset (Y_0, \xi_0)$ be defined as above; see (11-12). Recall that the relative symplectic filling $(\hat{W}_0, \hat{\lambda}_0, H)$ gives an augmentation $\tilde{\epsilon}_0 : \tilde{\mathcal{A}}(Y_0, \xi_0, V; \mathfrak{r}) \to \mathbb{Q}$.

It follows from Corollary* 10.14 and Lemma 11.6 that $\tilde{\mathcal{L}}^{\tilde{\epsilon}_0}(Y_0, \xi_0, V, \Lambda; \mathfrak{r})$ is a $(\mathbb{Z}\times\mathbb{Z})$–bigraded algebra with differential of bidegree $(-1, 0)$, and that $\overline{HC}_{\bullet,\bullet}(\tilde{\mathcal{L}}^{\epsilon}(Y, \xi, V; \mathfrak{r}))$ is a $(\mathbb{Z}\times\mathbb{Z})$–bigraded $\mathbb{Q}$–vector space.

We now have the following computation.

Proposition* 11.19 *Given a positive integer $N \gg 1$, let*

$$\bigoplus_{j \le N} \tilde{\mathcal{L}}^{\tilde{\epsilon}_0}_{\bullet,j}(Y_0, \xi_0, V, \Lambda; \mathfrak{r}) \subset \tilde{\mathcal{L}}^{\tilde{\epsilon}_0}(Y_0, \xi_0, V, \Lambda; \mathfrak{r}) \tag{11-67}$$

be the bigraded submodule of elements of winding number at most N. Then this submodule can be generated by products of total winding number $\le N$ of Reeb chords $\{a_k\}_{k\in\mathbb{N}_+}$ and $\{b_k\}_{k\in\mathbb{N}_+}$, where $|a_k| = 2k-1$ and $|b_k| = n-2+2k$. (Note that we do not say anything about the differential.)

Proof Since $(Y_0 - V, \lambda_0)$ is strictly contactomorphic to $(S^1 \times D^*S^{n-1}, \alpha_V)$, we have that $(Y_0 - V, \lambda_\delta)$ is strictly contactomorphic to $(S^1 \times D^*S^{n-1}, \alpha_\delta := (1 + \delta f)\alpha_V)$.

Recall that f depends on a positive real parameter $\epsilon > 0$, which can be taken to be arbitrarily small. Moreover, the restriction of f to the Legendrian $\Lambda = \{0\} \times S^{n-1}$ is equal to g, a Morse function with exactly two critical points: one minimum a and one maximum b.

Let c denote either a or b. As in Lemma 11.16, we let γ_c denote the simple Reeb chord (which is also a Reeb orbit) passing through c, and let γ_c^k denote its k–fold cover. Observe first of all that all Reeb chords for Λ are contained in $\mathcal{U}_\epsilon$ – this follows from the fact that f is compactly supported in $\mathcal{U}_\epsilon$; see page 100. Hence, as in Lemma 11.14, if we assume that $\epsilon > 0$ is small enough, then there exists $P > 0$ large enough so that all Reeb chords of action greater than P have winding number greater than N. By a routine adaptation of Lemma 11.15 (or rather the proof of [6, Lemma 2.3]), one concludes that the only Reeb chords of winding number less than or equal to N are the γ_a^k and γ_b^k.

We can assume without loss of generality that there are normal coordinates $\boldsymbol{q} = (q_1, \dots, q_{n-1})$ defined in a neighborhood $U_c \subset \Lambda$ of c in which g is given by

$$g = g(c) + \epsilon \sum_{i=1}^{n-1} q_i^2, \tag{11-68}$$

where $\epsilon = 1$ if $c = a$, and $\epsilon = -1$ if $c = b$. The Reeb vector field of α_δ is given by

$$R_{\alpha_\delta} = \frac{1}{1+\delta f} R_\alpha + \frac{2\epsilon\delta}{(1+\delta f)^2} \sum_i \biggl(q_i - \frac{2p_i}{1+\|\boldsymbol{p}\|^2} \sum_{j,k} h^{jk} p_k q_j \biggr) \partial_{p_i} \tag{11-69}$$

on $S^1 \times D^* U_c$ for $\|\boldsymbol{p}\|$ sufficiently small, ie satisfying $\rho(\|\boldsymbol{p}\|) = 1$. We will now show that for every $k \geq 1$, the indices of γ_a^k and γ_b^k as Reeb chords are given by

$$\mathrm{CZ}^+(\gamma_a^k) = 2k \quad \text{and} \quad \mathrm{CZ}^+(\gamma_b^k) = 2k + n - 1. \tag{11-70}$$

Hence, setting $a_k = \gamma_a^k$ and $b_k = \gamma_b^k$, we have

$$\begin{aligned} |a_k| &= \mathrm{CZ}^+(a_k) - 1 = 2k - 1, \\ |b_k| &= \mathrm{CZ}^+(b_k) - 1 = 2k + n - 2, \end{aligned}$$

as desired.

To compute $\mathrm{CZ}^+(\gamma_c^k)$, we start by computing the linearized Reeb flow along γ_c^k with respect to the trivialization τ_0; see (11-48). We proceed as in Lemma 11.12: we have

$$\mathcal{L}_{\partial_{p_i}} R_{\alpha_\delta} \big|_{\boldsymbol{p}=0, \boldsymbol{q}=0} = \frac{2}{1+\delta g(c)} \partial_{q_i}, \tag{11-71}$$

$$\mathcal{L}_{\partial_{q_i} - (a p_i/(1-\|\boldsymbol{p}\|^2)) \partial_\theta} R_{\alpha_\delta} \big|_{\boldsymbol{p}=0, \boldsymbol{q}=0} = \epsilon \frac{2\delta}{(1+\delta g(c))^2} \partial_{p_i}. \tag{11-72}$$

Hence, the matrix $A(t)$ representing the linearized Reeb flow $\xi_{\gamma_c^k(0)} \to \xi_{\gamma_c^k(t)}$ with respect to the trivialization τ_0 satisfies $A'(t) = S A(t)$ with

$$S = \begin{pmatrix} 0 & \epsilon \dfrac{2\delta}{(1+\delta g(c))^2} \\ \dfrac{2}{1+\delta g(c)} & 0 \end{pmatrix}, \tag{11-73}$$

where each entry should be interpreted as a multiple of the $(n-1) \times (n-1)$ identity matrix.

Setting $\mu = 2\delta/(1+\delta g(c))^2$ and $\nu = 2/(1+\delta g(c))$ for ease of notation, it follows that

$$A(t) = \exp(tS) = \begin{cases} \begin{pmatrix} \cosh(t\sqrt{\mu\nu}) & \sqrt{\mu/\nu}\sinh(t\sqrt{\mu\nu}) \\ \sqrt{\nu/\mu}\sinh(t\sqrt{\mu\nu}) & \cosh(t\sqrt{\mu\nu}) \end{pmatrix} & \text{if } \epsilon = 1, \\ \begin{pmatrix} \cos(t\sqrt{\mu\nu}) & -\mu/\nu\sin(t\sqrt{\mu\nu}) \\ \nu/\mu\sin(t\sqrt{\mu\nu}) & \cos(t\sqrt{\mu\nu}) \end{pmatrix} & \text{if } \epsilon = -1. \end{cases} \tag{11-74}$$

(Note that $\mu, \nu > 0$ if δ is sufficiently small.)

Let $L(t) \subset \xi_{\gamma_c^k(t)}$ be the path of Lagrangian subspaces obtained by applying the linearized Reeb flow to the tangent space $T_c\Lambda \subset \xi_c$ and let $\widetilde{L}(t)$ be the loop obtained by closing up $L(t)$ by a positive rotation. Since $T_c\Lambda$ is represented by

$$\begin{pmatrix} 0 \\ I_{n-1} \end{pmatrix}$$

in the trivialization τ_0, $L(t)$ is represented by

$$A(t)\begin{pmatrix} 0 \\ I_{n-1} \end{pmatrix}.$$

In the two-dimensional case (ie $n-1=1$), one can easily deduce (eg using the standard properties of the Maslov index stated in [52, Theorem 2.3.7]) that

$$\mu^{\tau_0}(\widetilde{L}(t)) = \begin{cases} 0 & \text{if } \epsilon = 1, \\ 1 & \text{if } \epsilon = -1. \end{cases} \tag{11-75}$$

In general, $L(t)$ splits as a direct sum of $n-1$ copies of the two-dimensional case, so the additivity property of the Maslov index [52, Theorem 2.3.7] implies that

$$\mu^{\tau_0}(\widetilde{L}(t)) = \begin{cases} 0 & \text{if } \epsilon = 1, \\ n-1 & \text{if } \epsilon = -1. \end{cases} \tag{11-76}$$

According to Definition 10.3 and the definition of the Maslov index [52, Theorem 2.3.7], we have $\mathrm{CZ}^+(\gamma_c^k) = \mu^\tau(\Lambda_{\mathbb{C}}^{n-1}\widetilde{\Lambda}) = \mu^\tau(\widetilde{L}(t))$, where τ is a trivialization of the contact structure along γ_c^k which extends to a spanning disk. For example, we can take τ to be the trivialization defined in equation (11-49). The difference $\mu^\tau(\widetilde{L}(t)) - \mu^{\tau_0}(\widetilde{L}(t))$ is equal to twice the Maslov index of the loop of symplectic matrices relating τ and τ_0, ie

$$\mu^\tau(\widetilde{L}(t)) - \mu^{\tau_0}(\widetilde{L}(t)) = 2\mu\begin{pmatrix} \cos(2\pi kt) & -\sin(2\pi kt) \\ \sin(2\pi kt) & \cos(2\pi kt) \end{pmatrix} = 2k. \tag{11-77}$$

It follows that

$$\mathrm{CZ}^+(\gamma_a^k) = 2k \quad \text{and} \quad \mathrm{CZ}^+(\gamma_b^k) = 2k+n-1, \tag{11-78}$$

as desired. □

It will be useful to record the following consequence of the above computation.

Corollary* 11.20 *Suppose that* $n \geq 4$ *is* **even***. Then we have*

$$\operatorname{rk} \overline{HC}_{2n,2}(\widetilde{\mathcal{L}}^{\epsilon_0}(Y_0, \xi_0, V, \Lambda; \mathfrak{r})) = 1. \tag{11-79}$$

Proof Indeed, note that the generators described in Proposition* 11.19 satisfy $\operatorname{link}(a_k) = \operatorname{link}(b_k) = k$. It thus follows that

$$\overline{CC}_{2n-1,2}(\widetilde{\mathcal{L}}^{\epsilon}(Y_0, \xi_0, V; \mathfrak{r})) = \overline{CC}_{2n+1,2}(\widetilde{\mathcal{L}}^{\epsilon}(Y_0, \xi_0, V; \mathfrak{r})) = 0. \tag{11-80}$$

On the other hand, $\overline{CC}_{2n,2}(\widetilde{\mathcal{L}}^{\epsilon}(Y_0, \xi_0, V; \mathfrak{r}))$ is generated by the word $b_1 b_1$. □

12 Applications to contact topology

12.1 Contact and Legendrian embeddings

We begin by introducing some standard definitions in the theory of contact and Legendrian embeddings.

Definition 12.1 Given a smooth manifold Y^{2n-1}, a *formal contact structure* (or *almost-contact structure*) is the data of a pair (η, ω), where $\eta \subset TY$ is a codimension 1 distribution and $\omega \in \Omega^2(Y)$ is a 2–form whose restriction to η is nondegenerate. A formal contact structure is said to be *genuine* if it is induced by a contact structure.

If Y^{2n-1} is orientable, then a formal contact structure is the same thing as a lift of the classifying map $Y \to B\mathrm{SO}(2n+1)$ to a map $Y \to B(U(n) \times \mathrm{id}) = BU(n)$.

Definition 12.2 (see Definition 2.2 in [10]) Let $(Y^{2n-1}, \xi = \ker\alpha)$ be a contact manifold. Given a formal contact manifold (V^{2m-1}, η, ω) where $1 \leq m \leq n-1$, a formal (iso)contact embedding is a pair (f, F_s) where

- F_s is a fiberwise injective bundle map $TV \to TY$ defined for $s \in [0,1]$,
- $f : V \to Y$ is a smooth map and $df = F_0$, and
- F_1 defines a fiberwise conformally symplectic map $(\eta, \omega) \to (\xi, d\alpha)$.

Observe that the above properties are independent of the choice of contact form α.

Two formal contact embeddings $i_0, i_1 : (V, \zeta, \omega) \to (Y, \xi)$ are said to be formally isotopic if they can be connected by a family $\{i_t\}_{t \in [0,1]}$ of formal contact embeddings.

A (genuine) contact embedding $(V, \zeta) \to (Y, \xi)$ is simply a smooth embedding $\phi : V \to Y$ such that $\phi_*(\zeta) = \xi_{\phi(V)}$. In particular, every contact embedding induces a formal contact embedding by taking $F_s = F_0 = df$.

Definition 12.3 (see Definition 2.1 in [10]) Let (Y^{2n-1}, ξ) be a contact manifold. Given a smooth n–dimensional manifold Λ, a formal Legendrian embedding is a pair (f, F_s) where

- F_s is a fiberwise injective bundle map $TV \to TY$ defined for $s \in [0,1]$,
- $df = F_0$, and
- $\mathrm{im}(F_1) \subset \xi$.

Two formal Legendrian embeddings are said to be formally isotopic if they can be connected by a family of Legendrian embeddings. A (genuine) Legendrian embedding $\Lambda \to (Y, \xi)$ is a smooth embedding $\phi\colon \Lambda \to Y$ such that $d\phi(T\Lambda) \subset \xi \subset TY$. In particular, a Legendrian embedding canonically induces a formal Legendrian embedding.

We now review some foundational facts about loose Legendrians. Recall that a Legendrian Λ in a (possibly noncompact) contact manifold (Y, ξ) of dimension at least five is defined to be loose if it admits a loose chart. For concreteness, we adopt as our definition of a loose chart the one given in [16, Section 7.7].

Loose Legendrians satisfy the following h-principle due to Murphy [55, Theorem 1.2]: given a pair of loose Legendrian embeddings $f_0, f_1\colon \Lambda \to (Y, \xi)$ which are formally isotopic, then f_0 and f_1 are genuinely isotopic, ie isotopic through Legendrian embeddings.

Given an arbitrary Legendrian submanifold Λ_0 in a contact manifold (Y_0, ξ_0) of dimension at least five, one can perform a local modification, called *stabilization*, which makes Λ_0 loose without changing the formal isotopy class of the tautological embedding $\Lambda_0 \xrightarrow{\mathrm{id}} \Lambda_0$. This modification can be realized in multiple essentially equivalent ways. In this paper, we will take as our definition of stabilization any construction which satisfies the properties stated in the following lemma.

Lemma 12.4 *Given a Legendrian submanifold $\Lambda \subset (Y_0, \xi_0)$ and an open set $U \subset Y_0$ such that $U \cap \Lambda_0$ is nonempty, there exists a Lagrangian embedding $f_1\colon \Lambda_0 \to Y$ which is formally isotopic to the tautological embedding $\Lambda_0 \xrightarrow{\mathrm{id}} \Lambda_0$ via a family of formal Legendrian embeddings $\{(f_t, F_s^t)\}_{t\in[0,1]}$ which are independent of t on $\Lambda_0 \cap (Y - U)$. We put $\Lambda := f_1(\Lambda_0)$ and say that Λ is the* ***stabilization*** *of Λ_0 inside U.*

Proof To construct Λ, we follow the procedure described in [16, Section 7.4]. As the reader may verify, this construction can be assumed to happen entirely inside a suitably chosen Darboux chart $\mathcal{U} \subset U$. In addition, the construction depends on the choice of a function f; using that Y has dimension at least five, we may (and do) assume that $\chi(\{f \geq 1\}) = 0$. To construct the formal isotopy, we simply follow the proof of [16, Proposition 7.23] (using the assumption that $\chi(\{f \geq 1\}) = 0$). The argument there is entirely local, so that the isotopy can be assumed to be fixed outside of $\mathcal{U}$ (and in particular, outside of U). □

12.2 Embeddings into overtwisted contact manifolds

We begin with the following proposition.

Proposition 12.5 *Suppose that (Y,ξ) is an overtwisted contact manifold and let*

$$i:(V,\zeta)\to(Y,\xi) \tag{12-1}$$

be a formal contact embedding. Then there exists an open subset $\Omega\subset Y$ such that $Y-\overline{\Omega}$ is overtwisted, and a ***genuine*** *contact embedding*

$$j:(V,\zeta)\to\Omega\subset(Y,\xi) \tag{12-2}$$

such that i and j are formally contact isotopic in (Y,ξ).

Proof We will assume for simplicity that V is connected, but the proof can easily be generalized. Let $D_{\mathrm{ot}}\subset(Y,\xi)$ be an overtwisted disk. Let f_t be a family of formal contact embeddings such that f_0 is the underlying smooth map induced by i, and $\mathrm{Im}(f_1)\cap D_{\mathrm{ot}}=\varnothing$. Let $\Omega\subset Y$ be a connected open subset such that $\mathrm{Im}(f_1)\subset\Omega\subset\overline{\Omega}\subset Y-D_{\mathrm{ot}}$. According to [5, Proposition 3.8], we can assume by choosing Ω large enough that (Ω,ξ) is overtwisted.

For purely algebrotopological reasons, there exists a family ξ_t of formal contact structures on Y with the following properties:

- $\xi_0=\xi$,
- ξ_t is constant in the complement of $\overline{\Omega}$, and
- ξ_1 is a genuine contact structure in a neighborhood $\mathcal{V}\subset\Omega$ of $\mathrm{Im}(f_1)$, and f_1 is a genuine contact embedding with respect to ξ_1.

Since ξ_1 is genuine on $\mathcal{V}\cup(Y-\overline{\Omega})$, it follows from the relative h-principle for overtwisted contact structures [5, Theorem 1.2] that ξ_1 is homotopic to a genuine overtwisted contact structure through a homotopy fixed on $\mathcal{V}\cup(Y-\overline{\Omega})$. Thus we may as well assume in the third property above that ξ_1 is genuine everywhere.

Since (Ω,ξ) is overtwisted, it follows from [5, Theorem 1.2] that there exists a homotopy $\tilde{\xi}_t$ of genuine contact structures such that $\tilde{\xi}_0=\xi_0=\xi$, $\tilde{\xi}_1=\xi_1$ and $\tilde{\xi}_t$ is independent of t on $Y-\overline{\Omega}$.

By Gray's theorem, there is an ambient isotopy $\psi_t:Y\to Y$ which is fixed on $Y-\overline{\Omega}$ and has the property that $\psi_t^*\tilde{\xi}_t=\tilde{\xi}_0=\xi_0$. The composition $\psi_1^{-1}\circ f_1$ is in the same class of formal contact embeddings as f_1, and gives the desired genuine embedding. □

We now describe a procedure for constructing pairs of codimension 2 contact embeddings in overtwisted manifolds which are formally isotopic but fail to be isotopic as contact embeddings.

Construction 12.6 Let (Y_0^{2n-1}, ξ_0) be a closed, overtwisted contact manifold and let (Y_0, B, π) be an open book decomposition which supports ξ_0. Let $i_0\colon (B, (\xi_0)_B) \to (Y_0, \xi_0)$ be the tautological embedding of the binding and let $j_0\colon (B, (\xi_0)_B) \to (Y_0, \xi_0)$ be a contact embedding with overtwisted complement which is formally isotopic to i_0 (the existence of such an embedding follows from Proposition 12.5). Let $D_{\mathrm{ot}} \subset Y_0$ be an overtwisted disk which is disjoint from $j_0(B)$.

Choose an open subset $\mathcal{U} \subset Y_0$ whose closure is disjoint from $i_0(B) \cup j_0(B) \cup D_{\mathrm{ot}}$, and such that i_0 and j_0 are formally isotopic in the complement of $\overline{\mathcal{U}}$. Now let (Y, ξ) be obtained by attaching handles of arbitrary index along isotropic submanifolds contained inside $\mathcal{U}$; see Construction 10.15. We let $(\widehat{X}, \widehat{\lambda})$ denote the resulting Weinstein cobordism with positive end (Y, ξ) and negative end (Y_0, ξ).

Observe that (Y, ξ) is still overtwisted and that i_0 and j_0 can also be viewed as codimension 2 contact embeddings into (Y, ξ). We denote these latter embeddings by $i, j\colon (B, (\xi_0)_B) \to (Y, \xi)$. By construction, the embeddings i and j are formally isotopic.

Theorem 12.7 *The embeddings i and j which arise from Construction 12.6 are not genuinely isotopic. In fact, i is not genuinely isotopic to any reparametrization of j in the source, meaning that the codimension 2 submanifolds $(i(B), \xi_{i(B)})$, $(j(B), \xi_{j(B)})$ are not contact isotopic.*

Proof According to Corollary 8.8, the cobordism $(\widehat{X}, \widehat{\lambda})$ induces a map of unital $\mathbb{Q}$–algebras

$$\widetilde{CH}_\bullet(Y, \xi, B; \mathfrak{r}) \to \widetilde{CH}_\bullet(Y_0, \xi_0, B; \mathfrak{r})$$

for any element $\mathfrak{r} \in \mathfrak{R}(Y_0, \xi_0, B) \equiv \mathfrak{R}(Y, \xi, B)$. Moreover, Theorem 11.3 implies that $\widetilde{CH}_\bullet(Y_0, \xi_0, B; \mathfrak{r}) \neq 0$ for appropriate $\mathfrak{r} \in \mathfrak{R}(Y, \xi, B)$. It follows that $\widetilde{CH}_\bullet(Y, \xi, B; \mathfrak{r}) \neq 0$.

If we assume that $(i(B), \xi_{i(B)})$, $(j(B), \xi_{j(B)})$ are isotopic as codimension 2 contact submanifolds, then $\widetilde{CH}_\bullet(Y, \xi, B; \mathfrak{r}) = \widetilde{CH}_\bullet(Y, \xi, j(B); \mathfrak{r}')$ for some datum $\mathfrak{r}' \in \mathfrak{R}(Y, \xi, j(B))$. On the other hand, observe that $(Y - j(B), \xi)$ is overtwisted by construction. Hence Theorem 11.1 implies $\widetilde{CH}_\bullet(Y, \xi, j(B); \mathfrak{r}') = 0$. □

Example 12.8 By a well-known theorem of Giroux and Mohsen [29, Theorem 7.3.5], any contact manifold (Y, ξ) admits an open book decomposition (Y, B, π) which supports ξ. Hence Construction 12.6 and Theorem 12.7 can be applied to any overtwisted contact manifold.

We also consider the following modification of Construction 12.6.

Construction 12.9 Let (Y_0^{2n-1}, ξ_0) be a closed, overtwisted contact manifold and let (Y_0, B, π) be an open book decomposition which supports ξ_0. Suppose that there exists a Legendrian submanifold $\Lambda \subset Y_0$ such that $B = \tau(\Lambda)$ is a contact pushoff of Λ. Let $D_{\mathrm{ot}} \subset Y_0$ be an overtwisted disk.

Let $\mathcal{U}_1 \subset Y_0 - B - D_{\mathrm{ot}}$ be an open ball which intersects Λ. Let $\Lambda' \subset (Y_0, \xi_0)$ be obtained by stabilizing Λ inside $\mathcal{U}_1$; see Lemma 12.4. Let $\mathcal{U}_2$ be the union of $\mathcal{U}_1$ with a tubular neighborhood of Λ. Let $V' = \tau(\Lambda') \subset \mathcal{U}_2$ be a choice of contact pushoff for Λ'.

Let $i_0\colon (B,(\xi_0)_B)\to(Y_0,\xi_0)$ be the tautological embedding. By [10, Lemma 3.4], i_0 is formally isotopic to some codimension 2 contact embedding $j_0\colon (B,(\xi_0)_B)\to(Y_0,\xi_0)$, where $j_0(B)=B'$. Choose such a formal isotopy and let $\mathcal{T}\subset Y_0$ be its trace.

Let $\mathcal{U}\subset Y_0$ be an open set whose closure is disjoint from $\mathcal{T}\cup\mathcal{U}_2\cup D_{\mathrm{ot}}$. As in Construction 12.6, let (Y,ξ) be obtained by attaching handles of arbitrary index along some collection of isotropics inside $\mathcal{U}$. Let $(\widehat{X},\widehat{\lambda})$ denote the resulting Weinstein cobordism with positive end (Y,ξ) and negative end (Y_0,ξ).

It follows from our choice of $\mathcal{U}$ that (Y,ξ) is overtwisted and that Λ, Λ', V and V' can be viewed as submanifolds of (Y,ξ). It also follows that Λ' is the stabilization of Λ as submanifolds of (Y,ξ), and that V (resp. V') is the contact pushoff of Λ (resp. Λ').

Corollary 12.10 *The submanifolds (V,ξ_V) and $(V',\xi_{V'})$ are not isotopic through codimension 2 contact submanifolds. Hence the Legendrian submanifolds $\Lambda,\Lambda'\subset(Y,\xi)$ are not isotopic through Legendrian submanifolds.*

Proof The proof of the first statement is identical to that of Theorem 12.7. The second statement follows from the fact that V and V' are, respectively, the contact pushoff of Λ and Λ'. □

Example 12.11 Let $(Y_0,\xi_0)=\mathrm{obd}(T^*S^{n-1},\tau^{-1})$, where τ^{-1} is a left-handed Dehn twist. Note by [12, Theorem 1.1] that (Y_0,ξ_0) is overtwisted — in fact, (Y,ξ) is contactomorphic to $(S^{2n-1},\xi_{\mathrm{ot}})$. Let $P=T^*S^{n-1}\subset Y_0$ be a page of the open book and let $\Lambda\subset(Y_0,\xi_0)$ be the Legendrian which corresponds to the zero section of P. Then the binding of the open book decomposition is also a contact pushoff of Λ. We may therefore apply Construction 12.9 to this data.

Remark 12.12 Consider the special case of Constructions 12.6 and 12.9 where $\mathcal{U}$ is empty, ie one does not attach any handles. In this case, Theorem 12.7 and Corollary 12.10 are essentially equivalent to the statement that the binding of an open book decomposition is tight, ie must intersect any overtwisted disk. This statement was proved in dimension 3 by Etnyre and Vela-Vick [28, Theorem 1.2]; the higher-dimensional case follows from work of Klukas [45, Corollary 3], who proved (following an outline of Wendl [67, Remark 4.1]) the stronger fact that a local filling obstruction (such as an overtwisted disk) in a closed contact manifold must intersect the binding of any supporting open book.

12.3 Contact embeddings into the standard contact sphere

In this section, we exhibit examples of pairs of codimension 2 contact embeddings into tight contact manifolds which are formally isotopic but are not isotopic through genuine contact embeddings. We begin with the following construction.

Construction 12.13 Let (Y_0^{2n-1},ξ_0) be a contact manifold for $n\geq 3$. Let $V\subset(Y_0,\xi_0)$ be a codimension 2 contact submanifold and let $\Lambda\subset(Y_0,\xi_0)$ be a loose Legendrian such that $\Lambda\cap V=\varnothing$.

Choose an open ball $\mathcal{U} \subset Y_0$ such that $(\mathcal{U}, \mathcal{U} \cap \Lambda)$ is a loose chart for Λ. Next, choose an open ball $\mathcal{O} \subset Y_0 - V - \mathcal{U}$. (By definition of a loose chart, $\mathcal{U} \cap \Lambda$ is a proper subset of Λ, so it is clear that such choices exist.)

Let Λ' be obtained by stabilizing Λ inside $\mathcal{O}$. It follows from Lemma 12.4 that Λ and Λ' are formally isotopic via a formal isotopy fixed outside of $\mathcal{O}$.

According to Lemma 12.14 below, we can (and do) fix a contactomorphism $f : (Y_0, \xi_0) \to (Y_0, \xi_0)$ with the following properties:

(1) f is isotopic to the identity,

(2) $f(\Lambda) = \Lambda'$, and

(3) the tautological contact embedding $i'_0 : (V, (\xi_0)_V) \to (Y_0 - \Lambda', \xi_0)$ is formally isotopic to the embedding $i'_1 := f \circ i'_0 : (V, (\xi_0)_V) \to (Y_0 - \Lambda', \xi_0)$. (We emphasize here that the formal isotopy is contained in the open contact manifold $(Y_0 - \Lambda', \xi_0)$.)

Finally, let (Y, ξ) be obtained by attaching a Weinstein n–handle along $\Lambda' \subset (Y_0, \xi_0)$ as described in Construction 10.15. We assume without loss of generality that the attaching region $\Lambda' \subset \mathcal{V}$ disjoint from V and $f(V)$, and that i'_0 and i'_1 are formally isotopic in $Y_0 - \mathcal{V}$. We let $\iota : Y_0 - \mathcal{V} \hookrightarrow Y$ be the canonical inclusion.

Let

(12-3) $$i_0 = \iota \circ i'_0 : (V, \xi_V) \to (Y, \xi)$$

be the tautological contact embedding and define

(12-4) $$i_1 := \iota \circ i'_1 : (V, \xi_V) \to (Y, \xi).$$

It is an immediate consequence of (3) and our choice of $\mathcal{V}$ that i_0 and i_1 are formally isotopic.

Lemma 12.14 *With the notation of Construction 12.13, there exists a contactomorphism* $f : (Y_0, \xi_0) \to (Y_0, \xi_0)$ *satisfying the properties* (1)–(3) *stated in Construction 12.13.*

Proof Recall that $\mathcal{U}$ is disjoint from $\mathcal{O}$. Recall also that $(\mathcal{U}, \mathcal{U} \cap \Lambda)$ is a loose chart for Λ, which means in particular that $\mathcal{U}$ deformation retracts onto $\mathcal{U} \cap \Lambda$. Using these two facts, it is not hard to verify that there exists a family of *formal* contact embeddings $j_t : (V, (\xi_0)_V) \to (Y_0, \xi_0)$ for $t \in [0, 1]$, with the following properties:

- $j_0 = i'_0$,
- $j_t(V)$ is disjoint from $\mathcal{O} \cup \Lambda$ for all $t \in [0, 1]$, and
- $j_1(V)$ is disjoint from $\mathcal{U} \cup \mathcal{O} \cup \Lambda$.

By the h–principle for loose Legendrian embeddings [55, Theorem 1.2], there exists a global contact isotopy ϕ_t for $t \in [0,1]$ such that $\phi_0 = \mathrm{Id}$ and $\phi_1(\Lambda) = \Lambda'$. By the Legendrian isotopy extension theorem [29, Theorem 2.6.2], this isotopy can be assumed to be compactly supported and constant in a neighborhood $\mathcal{W}$ of $j_1(V)$, where $\mathcal{W}$ is disjoint from $\mathcal{U} \cup \mathcal{O} \cup \Lambda$.

Let $f := \phi_1$ and observe that f satisfies (1)–(2). Observe that $f \circ j_t$ defines a formal contact isotopy from $f \circ i_0' = i_1'$ to j_1 in the complement of $\Lambda' = f(\Lambda)$. Since i_0' is formally isotopic to j_1 in the complement of $\Lambda' \subset \Lambda \cup \mathcal{O}$, we find that f satisfies (3). □

It will be useful to record the following basic observation, which is a consequence of the fact stated in Definition 2.1 that an isotopy of contactomorphisms induces a Hamiltonian isotopy of symplectizations.

Lemma 12.15 (cf Definition 2.1) *Let $(\widehat{X}, \widehat{\lambda})$ be a relative cobordism from (Y^+, λ^+) to (Y^-, λ^-). Given contactomorphisms $f^\pm : (Y^\pm, \lambda^\pm) \to (Y^\pm, \lambda^\pm)$ which are contact isotopic to the identity, there is a symplectomorphism $F : (\widehat{X}, \widehat{\lambda}) \to (\widehat{X}, \widehat{\lambda})$ which agrees near infinity with the lifts $\widetilde{f}^\pm : (SY^\pm, \lambda_{Y^\pm}) \to (SY^\pm, \lambda_{Y^\pm})$.* □

Let us now return to the geometric setup considered in Section 11.3. In particular, we let

$$(\widehat{W}_0, \widehat{\lambda}^a) := \Big(D^2 \times T^* S^{n-1}, \frac{1}{a} r^2 d\theta + \lambda_{\mathrm{std}}\Big), \tag{12-5}$$

where $a > 0$ is a constant which will be fixed later; see (11-7).

We let (W_0, λ^a), $(Y_0, \xi_0 = \ker \lambda_0)$, $V \subset Y_0$, $\Lambda \subset Y_0$ and $H = \{0\} \times T^* S^{n-1}$ be defined as in Section 11.3. Note that Λ is a loose Legendrian according to [11, Proposition 2.9].

Construction 12.13 applied to the above data produces a contactomorphism $f : (Y_0, \xi_0) \to (Y_0, \xi_0)$, a Liouville domain (X, λ) with positive contact boundary $(Y, \xi = \ker \lambda)$, and a pair of formally isotopic contact embeddings $i_0, i_1 : (V, \xi_V) \to (Y, \xi)$.

Let $\mathfrak{r} = (\alpha, \tau, a) \in \mathfrak{R}(Y_0, \xi_0, V)$, where $\alpha := (\lambda_0)|_V$ and the trivialization τ is unique since $H^1(V; \mathbb{Z}) = 0$; see Corollary 12.17. We let $\mathfrak{r}' = ((i_1')_* \alpha, \tau, a) \in \mathfrak{R}(Y_0, \xi_0, V)$, where i_1' is defined as in Construction 12.13 and τ is again unique. Since the surgery resulting from Construction 12.13 is away from V and $i_1'(V)$, we may identify $\mathfrak{R}(Y, \xi, V) = \mathfrak{R}(Y_0, \xi_0, V)$ and $\mathfrak{R}(Y, \xi, i_1(V)) = \mathfrak{R}(Y_0, \xi_0, i_1'(V))$.

As in Section 11.3, let $e_0 : \mathbb{R} \times Y_0 \to (\widehat{W}_0, \widehat{\lambda}_0, H)$ be the canonical marking furnished by the Liouville flow and let $\widetilde{\epsilon}_0 : \widetilde{\mathcal{A}}(Y_0, \xi_0, V; \mathfrak{r}) \to \mathbb{Q}$ be the associated augmentation.

By Lemma 12.15, there is a symplectomorphism $\psi : (\widehat{W}_0, \widehat{\lambda}_a) \to (\widehat{W}_0, \widehat{\lambda}_a)$ which coincides near infinity with the lift $\widetilde{f} : SY_0 \to SY_0$. Let $H' \subset \widehat{W}_0$ be a symplectic submanifold which is cylindrical at infinity and coincides with the symplectization of $f(V) = i_1'(V)$ on $[0, \infty) \times Y_0$. Such a surface can be constructed by taking the backwards Liouville flow of $\psi(H)$.

Let $\widetilde{\epsilon}_0' : \widetilde{\mathcal{A}}(Y_0, \xi_0, i_1'(V); \mathfrak{r}') \to \mathbb{Q}$ be the augmentation induced by the relative symplectic cobordism $((\widehat{W}_0, \widehat{\lambda}_a, H'), \widetilde{\epsilon}_0)$.

Observe that $(Y_0, \xi_0, V) \in \mathcal{G}$ and hence also $(Y_0, f_*\xi_0, f(V)) = (Y_0, \xi_0, i_1'(V)) \in \mathcal{G}$; see Definition 3.13. The following lemma shows that we also have $(Y, \xi, i_1(V)) \in \mathcal{G}$.

Lemma 12.16 *Up to contactomorphism,* $(Y, \xi) = \mathrm{ob}(T^*S^{n-1}, \tau_S) = (S^{2n-1}, \xi_{\mathrm{std}})$, *where* τ_S *denotes a right-handed Dehn twist. Moreover, the first contactomorphism can be assumed to take* $i_1(V)$ *to the binding of the open book decomposition* $\mathrm{ob}(T^*S^{n-1}, \tau_S)$.

Proof By construction, there is an open book decomposition of (Y_0, ξ_0) agreeing (up to contactomorphism) with $\mathrm{ob}(T^*S^{n-1}, \mathrm{id})$, such that $i_1(V)$ is the binding and Λ' is the zero section of a page. Note now that attaching a handle to the zero section of a page of $(Y_0, \xi_0) = \mathrm{ob}(T^*S^{n-1}, \mathrm{id})$ simply changes the monodromy of the open book by a positive Dehn twist [46, Theorem 4.6]. Hence, $i_1(V)$ is the binding of $\mathrm{ob}(T^*S^{n-1}, \tau_S) = (S^{2n-1}, \xi_{\mathrm{std}})$. □

Corollary 12.17 *The manifolds* Y_0, W_0, Y *and* W *have vanishing first and second homology and cohomology with* $\mathbb{Z}$*–coefficients. Hence the same is also true for the pairs* (W_0, Y_0) *and* (W, Y). *Finally, we have* $H^1(V; \mathbb{Z}) = 0$.

Proof By construction, W is obtained by attaching a handle of index n to W_0. The union of the core and co-core of this handle has codimension n. Hence, for $i \le n-2$, we have $H_i(W_0; \mathbb{Z}) = H_i(W; \mathbb{Z})$ and $H_i(Y_0; \mathbb{Z}) = H_i(Y; \mathbb{Z})$. Now, W_0 is homotopy equivalent to S^{n-1} by definition, while Y is homeomorphic to S^{2n-1} by Lemma 12.16. Since $n \ge 4$, it follows that Y_0, W_0, Y and W have vanishing first and second homology. The vanishing of cohomology in the same degrees follows by the universal coefficients theorem for cohomology. The vanishing of $H^1(V; \mathbb{Z})$ was proved in Lemma 11.6. □

As a result of Corollary 12.17, Definition 8.12, Lemma 8.13 and Definition* 10.12, the invariants considered in the proof of Theorem* 12.18 below, as well as the maps between these invariants, are all canonically $(\mathbb{Z}\times\mathbb{Z})$–bigraded.

Theorem* 12.18 *For* $n \ge 4$ ***even*** *and* $a \gg 0$ *large enough, the contact embeddings* $i_0, i_1 \colon (V, \xi_V) \to (Y, \xi) = (S^{2n-1}, \xi_{\mathrm{std}})$ *are not isotopic through contact embeddings.*

Proof We suppose for contradiction that i_0 and i_1 are isotopic through contact embeddings. This means that there exists a contactomorphism $g \colon (Y, \xi, V) \to (Y, \xi, i_1(V))$. It follows by Lemma 12.16 that $(Y, \xi, V) \in \mathcal{G}$.

According to Corollary* 11.18 (and the description of the generators in Proposition* 11.17), we may (and do) fix $a \gg 0$ large enough that

$$\widetilde{CH}^{\tilde{\epsilon}_0}_{k-(n-3),2}(Y_0, \xi_0, V; \mathfrak{r}) = \begin{cases} \mathbb{Z} & \text{if } k = 4, 4+(n-1), \\ 0 & \text{otherwise.} \end{cases} \tag{12-6}$$

In particular, since $n \ge 4$, we have

$$\widetilde{CH}^{\tilde{\epsilon}_0}_{2n-(n-3),2}(Y_0, \xi_0, V; \mathfrak{r}) = \widetilde{CH}^{\tilde{\epsilon}_0}_{2n+1-(n-3),2}(Y_0, \xi_0, V; \mathfrak{r}) = 0. \tag{12-7}$$

Since $(Y_0, \xi_0, V) \in \mathcal{G}$, it follows by Corollary* 9.14 that

$$\widetilde{CH}^{\tilde{\epsilon}_0}_{\bullet,\bullet}(Y_0, \xi_0, V; \mathfrak{r}) = \widetilde{CH}^{\tilde{\epsilon}'_0}_{\bullet,\bullet}(Y_0, \xi_0, i_1(V); \mathfrak{r}'). \tag{12-8}$$

Similarly, it follows by Definition/Assumption* 10.6 that

$$\widetilde{\mathcal{L}}^{\tilde{\epsilon}_0}_{\bullet,\bullet}(Y_0, \xi_0, V, \Lambda; \mathfrak{r}) = \widetilde{\mathcal{L}}^{\tilde{\epsilon}'_0}_{\bullet,\bullet}(Y_0, \xi_0, i_1(V), \Lambda'; \mathfrak{r}'). \tag{12-9}$$

Let $e\colon \mathbb{R} \times Y \to \widehat{W}$ be the canonical marking and consider the resulting relative filling $((\widehat{W}, \hat{\lambda}, H), e)$. Let $\tilde{\epsilon}\colon \widetilde{\mathcal{A}}(Y, \xi, V; \mathfrak{r}) \to \mathbb{Q}$ be the induced augmentation. Let $\phi\colon (\widehat{W}, \hat{\lambda}, H) \to (\widehat{W}, \hat{\lambda}, H)$ be a symplectomorphism which agrees with the lift of g near infinity. Let $\tilde{\epsilon}'\colon \widetilde{\mathcal{A}}(Y, \xi, i_1(V); \mathfrak{r}') \to \mathbb{Q}$ be the augmentation induced by $((\widehat{W}, \hat{\lambda}, H'), e)$. Then according to Lemma* 9.13 and Corollary* 9.14, we have

$$\widetilde{CH}^{\tilde{\epsilon}}_{\bullet,\bullet}(Y, \xi, V; \mathfrak{r}) = \widetilde{CH}^{\tilde{\epsilon}'}_{\bullet,\bullet}(Y, \xi, i_1(V); \mathfrak{r}'). \tag{12-10}$$

It then follows by Lemma* 12.19 that $\widetilde{CH}^{\tilde{\epsilon}}_{2n-(n-3),2}(Y, \xi, V; \mathfrak{r}) \neq 0$. Hence Lemma* 12.20 implies that

$$\widetilde{CH}^{\tilde{\epsilon}_0}_{2n-(n-3),2}(Y_0, \xi_0, V; \mathfrak{r}) \neq 0. \tag{12-11}$$

This contradicts (12-7). □

Lemma* 12.19 *We have*

$$\widetilde{CH}^{\tilde{\epsilon}'}_{2n-(n-3),2}(Y, \xi, i_1(V); \mathfrak{r}') \neq 0. \tag{12-12}$$

Proof On the one hand, Corollary* 11.20 and (12-9) imply that

$$\operatorname{rk} \overline{HC}_{2n,2}(\widetilde{\mathcal{L}}^{\tilde{\epsilon}'_0}(Y_0, \xi_0, i_1(V), \Lambda'; \mathfrak{r}')) = 1. \tag{12-13}$$

On the other hand, by (12-7) and (12-8), we have that

$$\widetilde{CH}^{\tilde{\epsilon}'_0}_{2n-(n-3),2}(Y_0, \xi_0, i_1(V); \mathfrak{r}') = \widetilde{CH}^{\tilde{\epsilon}'_0}_{2n+1-(n-3),2}(Y_0, \xi_0, i_1(V); \mathfrak{r}') = 0. \tag{12-14}$$

It then follows by Theorem/Assumption* 10.16 and Remark 10.18 that

$$\widetilde{CH}^{\tilde{\epsilon}'}_{2n-(n-3),2}(Y, \xi, i_1(V); \mathfrak{r}') \simeq \overline{HC}_{2n,2}(\widetilde{\mathcal{L}}^{\tilde{\epsilon}'_0}(Y_0, \xi_0, i_1(V), \Lambda'; \mathfrak{r}')). \tag{12-15}$$

This proves the claim. □

Lemma* 12.20 *The natural map*

$$\widetilde{CH}^{\tilde{\epsilon}}_{2n-(n-3),2}(Y, \xi, V; \mathfrak{r}) \to \widetilde{CH}^{\tilde{\epsilon}_0}_{2n-(n-3),2}(Y_0, \xi_0, V; \mathfrak{r}) \tag{12-16}$$

is injective.

Proof Since Λ' is loose in $Y_0 - V$, it follows by Proposition* 11.2 and Lemma 9.6 that

$$\overline{HC}_{2k}(\mathcal{L}^{\tilde{\epsilon}_0}(Y_0, \xi_0, V, \Lambda'; \mathfrak{r})) = 0 \tag{12-17}$$

for all $k \in \mathbb{Z}$. The lemma thus follows from Theorem/Assumption* 10.16 and Remark 10.18. □

Remark 12.21 One can slightly tweak Construction 12.13 so that Λ and Λ' are disjoint and $\Lambda \cup \Lambda'$ is a loose Legendrian link. One can then upgrade Lemma 12.14 to require that $f(\Lambda) = \Lambda'$ and $f(\Lambda') = \Lambda$ in (2) of Construction 12.13. In particular, this means that Λ' is a stabilization of Λ, and Λ is a stabilization of Λ'.

Let us apply this tweaked construction to the setup considered in Construction 12.13, where $(Y_0, \xi_0) = \mathrm{ob}(T^*S^{n-1}, \mathrm{id})$ for $n \geq 4$, $V \subset (Y_0, \xi_0)$ is the binding and $\Lambda \subset (Y_0, \xi_0)$ is the zero section of a page. It is well known that the zero section of a page in $\mathrm{ob}(T^*S^{n-1}, \mathrm{id})$ is the standard Legendrian unknot. Hence Lemma 12.16 implies that $i_1(V)$ is the pushoff of the standard unknot. By construction, it now also follows that $i_0(V)$ is the contact pushoff of a stabilization of the unknot. Theorem* 12.18 thus provides an alternative way to distinguish (for $n \geq 4$ even) the basic example considered by Casals and Etnyre in [10, Section 5].

12.4 Relative symplectic and Lagrangian cobordisms

In this final section, we exhibit some constraints on relative symplectic and Lagrangian cobordisms. In particular, we prove the results which were advertised in Section 1.5 of the introduction.

Proof of Theorem 1.9 Suppose for contradiction that such a relative symplectic cobordism exists. According to Theorem 11.3, we have $\widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r}) \neq 0$ for some $\mathfrak{r} = (\alpha, \tau, r) \in \mathfrak{R}(Y, \xi, V)$, which we now view as fixed. According to Theorem 11.1, we also have $\widetilde{CH}_\bullet(Y, \xi, V'; \mathfrak{r}') = 0$ for all $\mathfrak{r}' = (\alpha', \tau', r') \in \mathfrak{R}(Y, \xi, V')$. Choose $\mathfrak{r}'$ depending on our previous choice of $\mathfrak{r}$ so that $r' \geq e^{\mathcal{E}((H, \lambda_H)_\alpha^{\alpha'})} r$. Then Corollary 8.8 along with our topological assumptions on H furnishes a unital $\mathbb{Q}$–algebra map

$$\widetilde{CH}_\bullet(Y, \xi, V'; \mathfrak{r}') \to \widetilde{CH}_\bullet(Y, \xi, V; \mathfrak{r}). \tag{12-18}$$

This gives the desired contradiction. □

In contrast to Theorem 1.9, we expect that one could prove that V is concordant to V' by adapting work of Eliashberg and Murphy [26], but we do not pursue this here. Note that we could also have proved Theorem 1.9 using the full invariant $CH_\bullet(-;-)$ instead of its reduced counterpart.

For Lagrangian cobordisms, we have the following result.

Proposition 12.22 *Let (Y, ξ) be a contact manifold. Let Λ and Λ' be Legendrian knots such that $H^1(\tau(\Lambda'); \mathbb{Z}) = H^2(\tau(\Lambda'); \mathbb{Z}) = 0$. Suppose that (SY, λ_Y, L) is a Lagrangian concordance from Λ' to Λ. Given $\mathfrak{r} = (\alpha, \tau, r) \in \mathfrak{R}(Y, \xi, \tau(\Lambda))$, there is a map of $\mathbb{Q}$–algebras*

$$\widetilde{CH}_\bullet(Y, \xi, \tau(\Lambda'); \mathfrak{r}') \to \widetilde{CH}_\bullet(Y, \xi, \tau(\Lambda); \mathfrak{r}) \tag{12-19}$$

for some $\mathfrak{r}' = (\alpha', \tau', r') \in \mathfrak{R}(Y, \xi, \tau(\Lambda'))$. (A similar statement holds for the nonreduced invariants $CH_\bullet(-)$.)

Proof Observe that the trivial Lagrangian cobordism $L = \mathbb{R} \times \Lambda \subset \mathbb{R} \times Y$ admits a "symplectic push-off" $\tau(L) := \mathbb{R} \times \tau(\Lambda) \subset \mathbb{R} \times Y$. It follows by the Lagrangian neighborhood theorem that any Lagrangian concordance (SY, λ_Y, L) also admits a symplectic push-off (SY, λ_Y, H), which is a relative symplectic cobordism from $(Y, \xi, \tau(\Lambda'))$ to $(Y, \xi, \tau(\Lambda))$. Fix α' arbitrarily and choose r' so that $r' \geq e^{\mathcal{E}((H,\lambda_H)_\alpha^{\alpha'})} r$ (note that τ' is unique since $H^1(\tau(\Lambda'); \mathbb{Z}) = 0$). The claim now follows from Corollary 8.8. □

Proof of Theorem 1.10 Suppose for contradiction that Λ' is concordant to Λ. As in the proof of Corollary 12.10, we have

$$\widetilde{CH}_\bullet(Y, \xi, \tau(\Lambda); \mathfrak{r}) \neq 0 \tag{12-20}$$

for a suitable choice $\mathfrak{r} \in \mathfrak{R}(Y, \xi, \tau(\Lambda))$. On the other hand, we have $\widetilde{CH}_\bullet(Y, \xi, \tau(\Lambda'); \mathfrak{r}') = 0$ for all $\mathfrak{r}' \in \mathfrak{R}(Y, \xi, \tau(\Lambda'))$. This gives a contradiction in view of Proposition 12.22. □

We end by exhibiting examples of Lagrangian cobordisms which cannot be displaced from a codimension 2 symplectic submanifold.

Construction 12.23 Let $(Y_0, \xi_0) = \mathrm{obd}(T^*S^{n-1}, \mathrm{id})$ for $n \geq 3$. Let $V \subset (Y_0, \xi_0)$ be the binding and let $\Lambda \subset (Y_0, \xi_0)$ be the zero section of a page, which is a loose Legendrian by [11, Proposition 2.9]. Let $\mathcal{U}_1 \subset Y_0 - V$ be a small ball which intersects Λ in an $(n-1)$–ball and let Λ' be obtained by stabilizing Λ inside $\mathcal{U}_1$.

Let $\mathcal{U}_2 \subset Y_0 - (V \cup \Lambda \cup \mathcal{U}_1)$ be an open subdomain. Let (Y, ξ) be obtained by attaching a sequence of handles along isotropics contained in $\mathcal{U}_2$. Observe that V, Λ and Λ' can be viewed as submanifolds of both Y_0 and Y; we will not distinguish these embeddings in our notation. We let $(\widehat{X}, \hat{\lambda}, \widehat{V})$ be the associated relative symplectic cobordism from (Y, ξ, V) to (Y_0, ξ_0, V).

Proof of Theorem* 1.12 We can identify $\mathfrak{R}(Y_0, \xi_0, V) = \mathfrak{R}(Y, \xi, V)$. According to Theorem* 11.5, $\mathcal{L}(Y_0, \xi_0, V, \Lambda; \mathfrak{r}) \neq 0$ for suitable $\mathfrak{r} \in \mathfrak{R}(Y_0, \xi_0, V)$. In contrast, $\mathcal{L}(Y, \xi, V, \Lambda'; \mathfrak{r}) = 0$ by Proposition* 11.2, since Λ' is loose in $Y - V$ by construction. By Proposition* 10.7, the existence of a concordance from Λ' to Λ which doesn't intersect $\widehat{V}$ would imply that there is a unital map of $\mathbb{Q}[U]$–algebras

$$\mathcal{L}(Y, \xi, V, \Lambda'; \mathfrak{r}) \to \mathcal{L}(Y_0, \xi_0, V, \Lambda; \mathfrak{r}). \tag{12-21}$$

This gives a contradiction. □

We remark that Construction 12.23 could be generalized in various directions without affecting the validity of Theorem* 1.12, but we do not pursue this here.

Appendix Connected sums of almost-contact manifolds

Let G be a connected[9] subgroup of $\mathrm{SO}(n)$. An *almost G–structure* on a smooth oriented manifold M is a homotopy class of maps $M \to BG$ lifting the classifying map of the tangent bundle of M:

$$\begin{array}{ccc} & & BG \\ & \nearrow & \downarrow \\ M & \xrightarrow[TM]{} & B\mathrm{SO}(n) \end{array}$$

An *almost G–manifold* is a manifold equipped with an almost G–structure.

Example A.1 Taking $G = \mathrm{U}(n) \subset \mathrm{SO}(2n)$ yields the usual notion of an almost complex manifold. Almost-contact manifolds correspond to $G = \mathrm{U}(n) \subset \mathrm{SO}(2n+1)$.

If the n–dimensional sphere S^n admits an almost G–structure, a result of Kahn [43, Theorem 2] implies that for any two n–dimensional almost G–manifolds M and N, there exists an almost G–structure on $M \# N$ which is compatible with the given ones on M and N in the complement of the disks used to form the connected sum. In general, this structure is not unique, so the connected sum $M \# N$ is not well-defined as an almost G–manifold. However, we will show in Section A.1 that a choice of almost G–structure β on S^n induces a canonical almost G–structure on the connected sum of any two almost G–manifolds. Hence, any such β gives rise to a connected sum operation $(M, N) \mapsto M \#_\beta N$ for almost G–manifolds. Moreover, the set of almost G–structures on S^n forms a group under this operation (with identity β). In Section A.2, we will show that this group acts on the set of almost G–structures of any n–dimensional almost G–manifold.

A.1 Connected sums of almost G–manifolds

Let S^n be the unit sphere in $\mathbb{R}^{n+1}$, equipped with its standard orientation as the boundary of the unit disk D^{n+1}. We will write its points as pairs $(x, z) \in \mathbb{R}^n \times \mathbb{R}$. Define

$$D_- = \bigl\{(x,z) \in S^n \mid z < \tfrac{1}{2}\bigr\}, \quad D_+ = \bigl\{(x,z) \in S^n \mid z > -\tfrac{1}{2}\bigr\}, \quad A = D_- \cap D_+, \quad C_\pm = D_\pm \setminus A.$$

Note that D_- and D_+ are open disks, C_- and C_+ are closed disks, A is an open annulus, and $S^n = D_- \cup D_+ = C_- \sqcup A \sqcup C_+$.

Let M and N be smooth connected oriented n–dimensional manifolds. Choose orientation-preserving embeddings $i_+ \colon \overline{D}_+ \to M$ and $i_- \colon \overline{D}_- \to N$. We define the connected sum $M \# N = M \#_{i_+, i_-} N$ by

$$M \# N = \bigl(M \setminus i_+(C_+) \sqcup N \setminus i_-(C_-)\bigr)/{\sim},$$

where $i_+(x) \sim i_-(x)$ for every $x \in A$.

[9] This assumption is used in the proof of Proposition A.6.

We will now explain how to construct a classifying map for the tangent bundle of $M \# N$. The following elementary fact from topology will be useful.

Proposition A.2 *Let $i: A \to X$ be a cofibration. Assume A is contractible. Then for any connected space Y and continuous maps $F: X \to Y$ and $f: A \to Y$, there exists a map $F': X \to Y$ homotopic to F such that $F' \circ i = f$.*

Let $\tau_S: S^n \to B\mathrm{SO}(n)$ be a classifying map for TS^n. Let τ_M and τ_N be classifying maps for TM and TN such that $\tau_M \circ i_+ = \tau_S|_{\overline{D}_+}$ and $\tau_N \circ i_- = \tau_S|_{\overline{D}_-}$ (such maps always exist by Proposition A.2). Define $\tau_{M\#N}$ to be the unique map $M \# N \to B\mathrm{SO}(n)$ which coincides with τ_M on $M \setminus i_+(C_+)$ and with τ_N on $N \setminus i_-(C_-)$.

Proposition A.3 *The map $\tau_{M\#N}$ is a classifying map for $T(M \# N)$.*

We start with an easy topological lemma.

Lemma A.4 *Let E be an oriented vector bundle over a manifold M^n and let $i: D^n \to M^n$ be an embedding. Then any automorphism of i^*E can be extended to an automorphism of E.*

Proof Let ϕ be an automorphism of i^*E. Since D^n is contractible, we can trivialize i^*E and think of ϕ as a map $D^n \to \mathrm{GL}^+(n)$. Clearly $\phi|_{\partial D^n}$ is nullhomotopic, and since $\mathrm{GL}^+(n)$ is connected, we can extend ϕ to a map $\tilde{\phi}: D^n_2 \to \mathrm{GL}^+(n)$ which is constant with value $\mathrm{Id} \in \mathrm{GL}^+(n)$ near ∂D^n_2. Using a tubular neighborhood of $i(\partial D^n) \subset M$, we can also extend i to an embedding $\tilde{i}: D^n_2 \to M^n$. Then $\tilde{\phi}$ gives us an automorphism of $\tilde{i}^*E$ which is equal to the identity over a neighborhood of $\partial D^n_2 \subset D^n_2$ and hence extends trivially to an automorphism of E. □

Proof of Proposition A.3 Let $\tilde{\gamma}_n \to B\mathrm{SO}(n)$ be the universal bundle over $B\mathrm{SO}(n)$. We want to show that the tangent bundle $T(M \# N)$ of the connected sum is isomorphic to $\tau^*_{M\#N}\tilde{\gamma}_n$.

$T(M \# N)$ is obtained by gluing $T(M \setminus i_+(C_+))$ and $T(N \setminus i_-(C_-))$ along the maps $di_+: TA \to T(M \setminus i_+(C_+))$ and $di_-: TA \to T(N \setminus i_-(C_-))$. Because of our assumption that $\tau_M \circ i_+ = \tau_S|_{\overline{D}_+}$ and $\tau_N \circ i_- = \tau_S|_{\overline{D}_-}$, we have that $\tau^*_{M\#N}\tilde{\gamma}_n$ is obtained by gluing $(\tau_M|_{M\setminus i_+(C_+)})^*\tilde{\gamma}_n$ and $(\tau_N|_{N\setminus i_-(C_-)})^*\tilde{\gamma}_n$ along bundle maps $(\tau_S|_A)^*\tilde{\gamma}_n \to (\tau_M|_{M\setminus i_+(C_+)})^*\tilde{\gamma}_n$ and $(\tau_S|_A)^*\tilde{\gamma}_n \to (\tau_N|_{N\setminus i_-(C_-)})^*\tilde{\gamma}_n$ covering $i_+: A \to M \setminus i_+(C_+)$ and $i_-: A \to N \setminus i_-(C_-)$, respectively. Hence, in order to show that $T(M \# N)$ is isomorphic to $\tau^*_{M\#N}\tilde{\gamma}_n$, it suffices to construct a commutative diagram

$$\begin{array}{ccc} T(M \setminus i_+(C_+)) & \dashrightarrow & (\tau_M|_{M\setminus i_+(C_+)})^*\tilde{\gamma}_n \\ {\scriptstyle di_+}\uparrow & & \uparrow \\ TA & \dashrightarrow & (\tau_S|_A)^*\tilde{\gamma}_n \\ {\scriptstyle di_-}\downarrow & & \downarrow \\ T(N \setminus i_-(C_-)) & \dashrightarrow & (\tau_N|_{N\setminus i_-(C_-)})^*\tilde{\gamma}_n \end{array}$$

where the horizontal arrows are bundle isomorphisms.

Start by fixing an isomorphism $\phi\colon TS^n \to \tau_S^* \widetilde{\gamma}_n$, and let the middle arrow of the diagram be the restriction of ϕ to TA. To get the top and bottom arrows, it suffices to find bundle isomorphisms completing the following commutative squares:

$$\begin{array}{ccc} TM & \dashrightarrow & \tau_M^* \widetilde{\gamma}_n \\ {\scriptstyle di_+}\uparrow & & \uparrow \\ T\overline{D}_+ & \xrightarrow{\phi} & (\tau_S|_{\overline{D}_+})^* \widetilde{\gamma}_n \end{array} \qquad \begin{array}{ccc} T\overline{D}_- & \xrightarrow{\phi} & (\tau_S|_{\overline{D}_-})^* \widetilde{\gamma}_n \\ {\scriptstyle di_-}\downarrow & & \downarrow \\ TN & \dashrightarrow & \tau_N^* \widetilde{\gamma}_n \end{array}$$

This is possible by Lemma A.4. □

We are now ready to define the connected sum of two almost G–manifolds.

Definition A.5 Suppose that S^n admits an almost G–structure, and fix a choice β of one such structure. Let β_M and β_N be almost G–structures on M and N respectively. We define an almost G–structure $\beta_M \#_\beta \beta_N$ on $M \# N$ as follows.

Pick maps $\widetilde{\tau}_S\colon S^n \to BG$, $\widetilde{\tau}_M\colon M \to BG$ and $\widetilde{\tau}_N\colon N \to BG$ representing β, β_M and β_N, respectively. By Proposition A.2, we can assume that $\widetilde{\tau}_M \circ i_+ = \widetilde{\tau}_S|_{\overline{D}_+}$ and $\widetilde{\tau}_N \circ i_- = \widetilde{\tau}_S|_{\overline{D}_-}$. Hence, there is a unique map

$$\widetilde{\tau}_{M\#N} = \widetilde{\tau}_M \#_{\widetilde{\tau}_S} \widetilde{\tau}_N\colon M \# N \to BG$$

which coincides with $\widetilde{\tau}_M$ on $M \setminus i_+(C_+)$ and with $\widetilde{\tau}_N$ on $N \setminus i_-(C_-)$. By Proposition A.3, the composition

$$M \# N \xrightarrow{\widetilde{\tau}_{M\#N}} BG \longrightarrow B\mathrm{SO}(n)$$

is a classifying map for $T(M \# N)$. Hence, we can (and do) define $\beta_M \#_\beta \beta_N$ to be the homotopy class of $\widetilde{\tau}_{M\#N}$.

Proposition A.6 *The almost G–structure $\beta_M \#_\beta \beta_N$ is well-defined, ie independent of the choice of $\widetilde{\tau}_S$, $\widetilde{\tau}_M$ and $\widetilde{\tau}_N$.*

Proof Let $\widetilde{\tau}_S^j$, $\widetilde{\tau}_M^j$ and $\widetilde{\tau}_N^j$ represent β, β_M and β_N, respectively, where $j \in \{0,1\}$. As in Definition A.5, we assume that $\widetilde{\tau}_M^j \circ i_+ = \widetilde{\tau}_S^j|_{\overline{D}_+}$ and $\widetilde{\tau}_N^j \circ i_- = \widetilde{\tau}_S^j|_{\overline{D}_-}$.

Fix a homotopy $\widetilde{\tau}_S^t$ between $\widetilde{\tau}_S^0$ and $\widetilde{\tau}_S^1$. We will show that there exist homotopies $\widetilde{\tau}_M^t$ and $\widetilde{\tau}_N^t$ such that $\widetilde{\tau}_M^t \circ i_+ = \widetilde{\tau}_S^t|_{\overline{D}_+}$ and $\widetilde{\tau}_N^t \circ i_- = \widetilde{\tau}_S^t|_{\overline{D}_-}$. This implies that $\widetilde{\tau}_{M\#N}^0$ is homotopic to $\widetilde{\tau}_{M\#N}^1$ and hence that $\beta_M \#_\beta \beta_N$ is well-defined.

Pick an arbitrary homotopy $h\colon M \times I \to BG$ between $\widetilde{\tau}_M^0$ and $\widetilde{\tau}_M^1$ and define a map

$$g\colon \overline{D}_+ \times \partial I^2 \to BG$$

by $g(x,t,0) = h(i_+(x),t)$, $g(x,0,s) = \widetilde{\tau}^0_S(x)$, $g(x,1,s) = \widetilde{\tau}^1_S(x)$ and $g(x,t,1) = \widetilde{\tau}^t_S(x)$. We can extend g to a map $\widehat{g}\colon \overline{D}_+ \times I^2 \to BG$ since the obstruction to doing so lies in

$$H^2(\overline{D}_+ \times I^2, \overline{D}_+ \times \partial I^2; \pi_1(BG)) \cong \pi_1(BG) \cong \pi_0(G),$$

which is trivial by our assumption that G is connected.

Let

$$f\colon \bigl(M \times (I \times \{0\} \cup \{0\} \times I \cup \{1\} \times I)\bigr) \cup \bigl(i_+(\overline{D}_+) \times I^2\bigr) \to BG$$

be defined by

- $f(x,t,0) = h(x,t)$, $f(x,0,s) = \widetilde{\tau}^0_M(x)$ and $f(x,1,s) = \widetilde{\tau}^1_M(x)$ for $x \in M$,
- $f(x,t,s) = \widehat{g}(i_+^{-1}(x),t,s)$ for $x \in i_+(\overline{D}_+)$.

Since $i_+\colon \overline{D}_+ \to M$ is a cofibration, the domain of f is a retract of $M \times I^2$. We can therefore extend f to a map $\widehat{f}\colon M \times I^2 \to BG$. Restricting $\widehat{f}$ to $M \times I \times \{1\}$ then provides us with a homotopy $\widetilde{\tau}^t_M$ such that $\widetilde{\tau}^t_M \circ i_+ = \widetilde{\tau}^t_S|_{\overline{D}_+}$.

The same argument gives us a homotopy $\widetilde{\tau}^t_N$ such that $\widetilde{\tau}^t_N \circ i_- = \widetilde{\tau}^t_S|_{\overline{D}_-}$, so this completes the proof. □

Definition A.7 If $M = (M, \beta_M)$ and $N = (N, \beta_N)$ are almost G–manifolds, their connected sum (with respect to β) is the almost G–manifold $M \#_\beta N := (M \# N, \beta_M \#_\beta \beta_N)$.

As usual, there is an ambiguity in the notation $M \#_\beta N$ since the construction of the connected sum involves a choice of embeddings $i_+\colon \overline{D}_+ \to M$, $i_-\colon \overline{D}_- \to N$. However, the result is independent of these choices up to the appropriate notion of equivalence, as one would expect.

Definition A.8 A *diffeomorphism of almost G–manifolds* $f\colon (M, \beta_M) \to (N, \beta_N)$ consists of a smooth diffeomorphism $f\colon M \to N$ such that $f^*\beta_N = \beta_M$.

Proposition A.9 *The connected sum $M \#_\beta N$ is well-defined up to diffeomorphism of almost G–manifolds. More precisely, given any orientation-preserving embeddings $i_+, j_+\colon \overline{D}_+ \to M$ and $i_-, j_-\colon \overline{D}_- \to N$, there exists an orientation-preserving diffeomorphism $\phi\colon M \#_{i_+,i_-} N \to M \#_{j_+,j_-} N$ such that*

$$\phi^*(\beta_M \#_{j_+,j_-,\beta} \beta_N) = \beta_M \#_{i_+,i_-,\beta} \beta_N$$

for any almost G–structures β_M and β_N on M and N.

Proof This follows from the isotopy extension theorem as in the smooth case. □

Remark A.10 (connected sums of contact manifolds) Suppose that (M_1, α_1) and (M_2, α_2) are contact manifolds. Then one can form the connected sum $(M_1 \# M_2, \alpha_1 \# \alpha_2)$, which is also a contact manifold. The connected sum is obtained by choosing Darboux balls in M_1, M_2 and connecting them by a "neck". This operation can also be understood as a contact surgery along a 0–sphere. We refer to [8, Section 6.2] and [46, Section 3] for more details.

Let $\beta \in \mathfrak{alm}_{U(n)}(S^{2n-1})$ be the almost-contact structure induced by the standard contact structure on the sphere. Then the operation of connected sum (with respect to β) of almost $U(n-1)$–manifolds defined in Definition A.7, and the operation of connected sum of contact manifolds described above, commute with the forgetful map from contact manifolds to almost-contact manifolds. This can be shown as in the proof of Proposition A.3, replacing $B\mathrm{SO}(n)$ with $BU(n-1)$.

The main properties of the connected sum in the smooth case have analogs for almost G–manifolds:

Proposition A.11 *Let M, N and P be connected almost G–manifolds of dimension n, and let β and β' be almost G–structures on S^n. Then*:

(1) $M \#_\beta (S^n, \beta) \cong M$.

(2) $(M \#_\beta N) \#_{\beta'} P \cong M \#_\beta (N \#_{\beta'} P)$.

Proof If one takes $i_-\colon \overline{D}_- \to S^n$ to be the inclusion map, then the connected sum $M \# S^n$ is canonically identified with M as a smooth manifold. If one further takes $\tilde{\tau}_N = \tilde{\tau}_S$ in Definition A.5, then this identification is compatible with the almost G–structures on $M \# S^n$ and M. This proves that $M \#_\beta (S^n, \beta) \cong M$.

To prove that $(M \#_\beta N) \#_\beta P \cong M \#_\beta (N \#_\beta P)$, choose embeddings $i_+\colon \overline{D}_+ \to M$, $i_-\colon \overline{D}_- \to N$, $j_+\colon \overline{D}_+ \to N$ and $j_-\colon \overline{D}_- \to P$. If we assume that i_- and j_+ have disjoint images, then i_- induces an embedding $\overline{D}_- \to N \#_{j_+, j_-} P$, j_+ induces an embedding $\overline{D}_+ \to M \#_{i_+, i_-} N$, and there is a canonical identification of smooth manifolds

$$(M \#_{i_+,i_-} N) \#_{j_+,j_-} P \cong M \#_{i_+,i_-} (N \#_{j_+,j_-} P).$$

Moreover, this identification is compatible with the almost G–structures in the sense that for any choice of maps $\tilde{\tau}_S$, $\tilde{\tau}'_S$, $\tilde{\tau}_M$, $\tilde{\tau}_N$ and $\tilde{\tau}_P$, the following diagram commutes:

$$\begin{array}{ccc} (M \#_{i_+,i_-} N) \#_{j_+,j_-} P & \xrightarrow{(\tilde{\tau}_M \#_{\tilde{\tau}_S} \tilde{\tau}_N) \#_{\tilde{\tau}'_S} \tilde{\tau}_P} & BG \\ \downarrow \cong & & \\ M \#_{i_+,i_-} (N \#_{j_+,j_-} P) & \xrightarrow{\tilde{\tau}_M \#_{\tilde{\tau}_S} (\tilde{\tau}_N \#_{\tilde{\tau}'_S} \tilde{\tau}_P)} & BG \end{array}$$

□

A.2 The group of almost G–structures on the sphere

We will denote the set of almost G–structures on a manifold M by $\mathfrak{alm}_G(M)$. More generally, if $A \subset M$ is a closed subset and β_0 is an almost G–structure on some open neighborhood of A, then $\mathfrak{alm}_G(M, A; \beta_0)$ will denote the set of almost G–structures on M which agree with β_0 near A.

In this section, we will show that $\#_\beta$ is a group operation on $\mathfrak{alm}_G(S^n)$, with β as identity element. The resulting group will be denoted by $\mathfrak{alm}_G^\beta(S^n)$. We will then show that $\mathfrak{alm}_G^\beta(S^n)$ acts on $\mathfrak{alm}_G(M)$, and more generally on $\mathfrak{alm}_G(M, A; \beta_0)$ if $M \setminus A$ is connected.

Proposition A.12 *Given any $\beta_1 \in \mathfrak{alm}_G(S^n)$, there exists a $\beta_2 \in \mathfrak{alm}_G(S^n)$ such that $\beta_1 \#_\beta \beta_2 = \beta$.*

Proof Recall the decomposition $S^n = C_- \cup A \cup C_+$ introduced at the beginning of Section A.1. We will use the notation $\langle \tau^-, \tau^A, \tau^+ \rangle$ to denote the unique (assuming it exists) map $S^n \to BG$ which coincides with the given maps $\tau^-\colon C_- \to BG$, $\tau^A\colon \bar{A} \to BG$ and $\tau^+\colon C_+ \to BG$ on C_-, $\bar{A}$ and C_+, respectively.

Let $\tilde{\tau}_S = \langle \tau_S^-, \tau_S^A, \tau_S^+ \rangle$ be a representative for β. Given β_1 and β_2 in $\mathfrak{alm}_G(S^n)$, we can choose representatives of the form $\langle \tau_1^-, \tau_S^A, \tau_S^+ \rangle$ and $\langle \tau_S^-, \tau_S^A, \tau_2^+ \rangle$ by Proposition A.2. Then $\beta_1 \#_\beta \beta_2$ is represented by $\langle \tau_1^-, \tau_S^A, \tau_2^+ \rangle$. Hence, all we need to show is that for any $\tau_1^-\colon C_- \to BG$, there exists $\tau_2^+\colon C_+ \to BG$ such that $\langle \tau_1^-, \tau_S^A, \tau_2^+ \rangle$ is homotopic to $\tilde{\tau}_S$. This again follows from Proposition A.2. □

Corollary A.13 *We have that $(\mathfrak{alm}_G(S^n), \#_\beta)$ is a group with identity β.*

Proof This follows from Propositions A.11 and A.12. □

Remark A.14 The group $(\mathfrak{alm}_G(S^n), \#_\beta)$ is independent of β up to isomorphism. Indeed, given any $x, y, \beta, \beta' \in \mathfrak{alm}_G(S^n)$, it follows from Proposition A.11 that

$$(x \#_\beta \beta') \#_{\beta'} (y \#_\beta \beta') = (x \#_\beta (\beta' \#_{\beta'} y)) \#_\beta \beta' = (x \#_\beta y) \#_\beta \beta',$$

which implies that the map

$$(\mathfrak{alm}_G(S^n), \#_\beta) \to (\mathfrak{alm}_G(S^n), \#_{\beta'}), \quad x \mapsto x \#_\beta \beta',$$

is a group isomorphism.

Given orientation-preserving embeddings $i_+\colon \bar{D}_+ \to M$ and $i_-\colon \bar{D}_- \to S^n$, the results of Section A.1 give us a well-defined map

$$\mathfrak{alm}_G(M) \times \mathfrak{alm}_G^\beta(S^n) \to \mathfrak{alm}_G(M \#_{i_+,i_-} S^n).$$

For the remainder of this section, we will take i_- to be the inclusion map $\bar{D}_- \hookrightarrow S^n$. Then $M \#_{i_+,i_-} S^n$ is canonically identified with M (regardless of what i_+ is) and we get a map

$$\mathfrak{alm}_G(M) \times \mathfrak{alm}_G^\beta(S^n) \to \mathfrak{alm}_G(M). \tag{A-1}$$

By Proposition A.11, this is a group action. Note that the diffeomorphism $\phi\colon M \to M$ appearing in the statement of Proposition A.9 (applied to $N = S^n$) can be chosen to be isotopic to the identity, which implies that the map (A-1) is independent of i_+.

If we assume that the image of the embedding $i_+\colon \overline{D}_+ \to M$ is disjoint from A, then it follows directly from Definition A.5 that the subset $\mathfrak{alm}_G(M, A; \beta_0) \subset \mathfrak{alm}_G(M)$ is invariant under the map (A-1). If $M \setminus A$ is connected, then the resulting action on $\mathfrak{alm}_G(M, A; \beta_0)$ doesn't depend on the choice of i_+.

References

[1] **R Abraham**, *Bumpy metrics*, from "Global analysis" (S-s Chern, S Smale, editors), Proc. Sympos. Pure Math. XIV, Amer. Math. Soc., Providence, RI (1970) 1–3 MR Zbl

[2] **P Albers**, **B Bramham**, **C Wendl**, *On nonseparating contact hypersurfaces in symplectic 4–manifolds*, Algebr. Geom. Topol. 10 (2010) 697–737 MR Zbl

[3] **M R R Alves**, **A Pirnapasov**, *Reeb orbits that force topological entropy*, Ergodic Theory Dynam. Systems 42 (2022) 3025–3068 MR Zbl

[4] **E Bao**, **K Honda**, *Definition of cylindrical contact homology in dimension three*, J. Topol. 11 (2018) 1002–1053 MR Zbl

[5] **M S Borman**, **Y Eliashberg**, **E Murphy**, *Existence and classification of overtwisted contact structures in all dimensions*, Acta Math. 215 (2015) 281–361 MR Zbl

[6] **F Bourgeois**, *A Morse–Bott approach to contact homology*, PhD thesis, Stanford University (2002) Zbl Available at `https://www.proquest.com/docview/305591502`

[7] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR Zbl

[8] **F Bourgeois**, **O van Koert**, *Contact homology of left-handed stabilizations and plumbing of open books*, Commun. Contemp. Math. 12 (2010) 223–263 MR Zbl

[9] **J Bowden**, *Two-dimensional foliations on four-manifolds*, PhD thesis, Ludwig Maximilian University of Munich (2010) Available at `https://edoc.ub.uni-muenchen.de/12551/1/Bowden_Jonathan.pdf`

[10] **R Casals**, **J B Etnyre**, *Non-simplicity of isocontact embeddings in all higher dimensions*, Geom. Funct. Anal. 30 (2020) 1–33 MR Zbl

[11] **R Casals**, **E Murphy**, *Legendrian fronts for affine varieties*, Duke Math. J. 168 (2019) 225–323 MR Zbl

[12] **R Casals**, **E Murphy**, **F Presas**, *Geometric criteria for overtwistedness*, J. Amer. Math. Soc. 32 (2019) 563–604 MR Zbl

[13] **R Casals**, **D M Pancholi**, **F Presas**, *The Legendrian Whitney trick*, Geom. Topol. 25 (2021) 3229–3256 MR Zbl

[14] **B Chantraine**, *Lagrangian concordance of Legendrian knots*, Algebr. Geom. Topol. 10 (2010) 63–85 MR Zbl

[15] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Floer theory for Lagrangian cobordisms*, J. Differential Geom. 114 (2020) 393–465 MR Zbl

[16] **K Cieliebak**, **Y Eliashberg**, *From Stein to Weinstein and back: symplectic geometry of affine complex manifolds*, American Mathematical Society Colloquium Publications 59, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[17] **A Connes**, *Noncommutative differential geometry*, Inst. Hautes Études Sci. Publ. Math. 62 (1985) 257–360 MR Zbl

[18] **M Dörner**, **H Geiges**, **K Zehmisch**, *Open books and the Weinstein conjecture*, Q. J. Math. 65 (2014) 869–885 MR Zbl

[19] **T Ekholm**, *Rational SFT, linearized Legendrian contact homology, and Lagrangian Floer cohomology*, from "Perspectives in analysis, geometry, and topology" (I Itenberg, B Jöricke, M Passare, editors), Progr. Math. 296, Springer (2012) 109–145 MR Zbl

[20] **T Ekholm**, *Holomorphic curves for Legendrian surgery*, preprint (2019) arXiv 1906.07228

[21] **T Ekholm**, **J Etnyre**, **L Ng**, **M Sullivan**, *Filtrations on the knot contact homology of transverse knots*, Math. Ann. 355 (2013) 1561–1591 MR Zbl

[22] **T Ekholm**, **J B Etnyre**, **L Ng**, **M G Sullivan**, *Knot contact homology*, Geom. Topol. 17 (2013) 975–1112 MR Zbl

[23] **T Ekholm**, **K Honda**, **T Kálmán**, *Legendrian knots and exact Lagrangian cobordisms*, J. Eur. Math. Soc. 18 (2016) 2627–2689 MR Zbl

[24] **Y Eliashberg**, **A Givental**, **H Hofer**, *Introduction to symplectic field theory*, from "Visions in mathematics: GAFA 2000 special volume, II" (N Alon, J Bourgain, A Connes, M Gromov, V Milman, editors), Birkhäuser, Boston, MA (2000) 560–673 MR Zbl

[25] **Y Eliashberg**, **N Mishachev**, *Introduction to the h–principle*, Graduate Studies in Mathematics 48, Amer. Math. Soc., Providence, RI (2002) MR Zbl

[26] **Y Eliashberg**, **E Murphy**, *Lagrangian caps*, Geom. Funct. Anal. 23 (2013) 1483–1514 MR Zbl

[27] **J B Etnyre**, **L L Ng**, *Legendrian contact homology in* $\mathbb{R}^3$, from "Surveys in differential geometry 2020: surveys in 3–manifold topology and geometry" (I Agol, D Gabai, editors), Surv. Differ. Geom. 25, International, Boston, MA ([2022] ©2022) 103–161 MR Zbl

[28] **J B Etnyre**, **D S Vela-Vick**, *Torsion and open book decompositions*, Int. Math. Res. Not. 2010 (2010) 4385–4398 MR Zbl

[29] **H Geiges**, *An introduction to contact topology*, Cambridge Studies in Advanced Mathematics 109, Cambridge Univ. Press (2008) MR Zbl

[30] **M Gerstenhaber**, **C W Wilkerson**, *On the deformation of rings and algebras, V: Deformation of differential graded algebras*, from "Higher homotopy structures in topology and mathematical physics" (J McCleary, editor), Contemp. Math. 227, Amer. Math. Soc., Providence, RI (1999) 89–101 MR Zbl

[31] **E Giroux**, *Géométrie de contact: de la dimension trois vers les dimensions supérieures*, from "Proceedings of the International Congress of Mathematicians, II" (T Li, editor), Higher Ed., Beijing (2002) 405–414 MR Zbl

[32] **P Griffiths**, **J Morgan**, *Rational homotopy theory and differential forms*, 2nd edition, Progr. Math. 16, Springer (2013) MR Zbl

[33] **J Gutt**, *The Conley–Zehnder index for a path of symplectic matrices*, preprint (2012) arXiv 1201.3728v2

[34] **V Hinich**, *Homological algebra of homotopy algebras*, Comm. Algebra 25 (1997) 3291–3323 MR Zbl

[35] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR Zbl

[36] **H Hofer**, **K Wysocki**, **E Zehnder**, *A general Fredholm theory, I: A splicing-based differential geometry*, J. Eur. Math. Soc. 9 (2007) 841–876 MR Zbl

[37] **H Hofer**, **K Wysocki**, **E Zehnder**, *A general Fredholm theory, II: Implicit function theorems*, Geom. Funct. Anal. 19 (2009) 206–293 MR Zbl

[38] **H Hofer**, **K Wysocki**, **E Zehnder**, *A general Fredholm theory, III: Fredholm functors and polyfolds*, Geom. Topol. 13 (2009) 2279–2387 MR Zbl

[39] **K Honda**, **Y Huang**, *Convex hypersurface theory in contact topology*, preprint (2019) arXiv 1907.06025

[40] **U Hryniewicz**, **A Momin**, **P A S Salomão**, *A Poincaré–Birkhoff theorem for tight Reeb flows on S^3*, Invent. Math. 199 (2015) 333–422 MR Zbl

[41] **M Hutchings**, *Mean action and the Calabi invariant*, J. Mod. Dyn. 10 (2016) 511–539 MR Zbl

[42] **M Hutchings**, **J Nelson**, *Cylindrical contact homology for dynamically convex contact forms in three dimensions*, J. Symplectic Geom. 14 (2016) 983–1012 MR Zbl

[43] **P J Kahn**, *Obstructions to extending almost X–structures*, Illinois J. Math. 13 (1969) 336–357 MR Zbl

[44] **W Klingenberg**, *Lectures on closed geodesics*, Grundl. Math. Wissen. 230, Springer (1978) MR Zbl

[45] **M Klukas**, *Open books and exact symplectic cobordisms*, Internat. J. Math. 29 (2018) 1850026, 19 MR Zbl

[46] **O van Koert**, *Lecture notes on stabilization of contact open books*, Münster J. Math. 10 (2017) 425–455 MR Zbl

[47] **A Lazarev**, **M Markl**, *Disconnected rational homotopy theory*, Adv. Math. 283 (2015) 303–361 MR Zbl

[48] **O Lazarev**, *Maximal contact and symplectic structures*, J. Topol. 13 (2020) 1058–1083 MR Zbl

[49] **W Li**, *Lagrangian cobordism functor in microlocal sheaf theory, I*, J. Topol. 16 (2023) 1113–1166 MR Zbl

[50] **J-L Loday**, *Cyclic homology*, 2nd edition, Grundl. Math. Wissen. 301, Springer (1998) MR Zbl

[51] **J-L Loday**, **B Vallette**, *Algebraic operads*, Grundl. Math. Wissen. 346, Springer (2012) MR Zbl

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

[53] **A Momin**, *Contact homology of orbit complements and implied existence*, J. Mod. Dyn. 5 (2011) 409–472 Zbl

[54] **A Moreno**, **R Siefring**, *Holomorphic curves in the presence of holomorphic hypersurface foliations*, preprint (2019) arXiv 1902.02700v1

[55] **E Murphy**, *Loose Legendrian embeddings in high dimensional contact manifolds*, preprint (2019) arXiv 1201.2245v5

[56] **L Ng**, *Framed knot contact homology*, Duke Math. J. 141 (2008) 365–406 MR Zbl

[57] **L Ng**, *A topological introduction to knot 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) 485–530 MR Zbl

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

[59] **D M Pancholi**, **S Pandit**, *Iso-contact embeddings of manifolds in codimension* 2, J. Symplectic Geom. 20 (2022) 471–497 MR Zbl

[60] **J Pardon**, *An algebraic approach to virtual fundamental cycles on moduli spaces of pseudo-holomorphic curves*, Geom. Topol. 20 (2016) 779–1034 MR Zbl

[61] **J Pardon**, *Contact homology and virtual fundamental cycles*, J. Amer. Math. Soc. 32 (2019) 825–919 MR Zbl

[62] **A Ranicki**, *High-dimensional knot theory: algebraic surgery in codimension* 2, Springer (1998) MR Zbl

[63] **J M Sabloff**, **L Traynor**, *Obstructions to Lagrangian cobordisms between Legendrians via generating families*, Algebr. Geom. Topol. 13 (2013) 2733–2797 MR Zbl

[64] **R Siefring**, *Intersection theory of punctured pseudoholomorphic curves*, Geom. Topol. 15 (2011) 2351–2457 MR Zbl

[65] **I Ustilovsky**, *Contact homology and contact structures on* S^{4m+1}, PhD thesis, Stanford University (1999) Available at `https://www.proquest.com/docview/304526632`

[66] **B Vallette**, *Homotopy theory of homotopy algebras*, Ann. Inst. Fourier (Grenoble) 70 (2020) 683–738 MR Zbl

[67] **C Wendl**, *A hierarchy of local symplectic filling obstructions for contact* 3*–manifolds*, Duke Math. J. 162 (2013) 2197–2283 MR Zbl

[68] **C Wendl**, *Lectures on symplectic field theory*, preprint (2016) arXiv 1612.01009

[69] **C Wendl**, *Lectures on contact* 3*–manifolds, holomorphic curves and intersection theory*, Cambridge Tracts in Mathematics 220, Cambridge Univ. Press (2020) MR Zbl

[70] **Z Zhou**, *Infinite not contact isotopic embeddings in* $(S^{2n-1}, \xi_{\mathrm{std}})$ *for* $n \geqslant 4$, Bull. Lond. Math. Soc. 55 (2023) 149–155 MR Zbl

Department of Mathematics, Harvard University
Cambridge, MA, United States

Department of Mathematics, Stanford University
Stanford, CA, United States

`lcote@math.harvard.edu`, `fsimon@stanford.edu`

Proposed: Leonid Polterovich Received: 19 December 2020
Seconded: Ciprian Manolescu, Paul Seidel Revised: 23 January 2022

$\mathcal{G}eometry$ & $\mathcal{T}opology$ **28**:1 (2024) 127–160
DOI: 10.2140/gt.2024.28.127
Published: 27 February 2024

The desingularization of the theta divisor of a cubic threefold as a moduli space

AREND BAYER
SJOERD VIKTOR BEENTJES
SOHEYLA FEYZBAKHSH
GEORG HEIN
DILETTA MARTINELLI
FATEMEH REZAEE
BENJAMIN SCHMIDT

We show that the moduli space $\overline{M}_X(v)$ of Gieseker stable sheaves on a smooth cubic threefold X with Chern character $v = \left(3, -H, -\frac{1}{2}H^2, \frac{1}{6}H^3\right)$ is smooth and of dimension four. Moreover, the Abel–Jacobi map to the intermediate Jacobian of X maps it birationally onto the theta divisor Θ, contracting only a copy of $X \subset \overline{M}_X(v)$ to the singular point $0 \in \Theta$.

We use this result to give a new proof of a categorical version of the Torelli theorem for cubic threefolds, which says that X can be recovered from its Kuznetsov component $\mathrm{Ku}(X) \subset \mathrm{D^b}(X)$. Similarly, this leads to a new proof of the description of the singularity of the theta divisor, and thus of the classical Torelli theorem for cubic threefolds, ie that X can be recovered from its intermediate Jacobian.

14D20; 14F08, 14J30, 14J45

1 Introduction

Moduli spaces of sheaves provide examples of algebraic varieties with an interesting and rich geometry and they have been widely studied in the past few decades. In particular, there are many strong results regarding moduli spaces on surfaces, while the situation on threefolds is less understood. We refer to Huybrechts and Lehn [23] for a more detailed account of the theory, which has been revolutionized by the introduction of stability conditions on triangulated categories by Bridgeland [12].

Perhaps the main player of the seminal paper by Clemens and Griffiths [14] on the geometry of cubic threefolds is the theta divisor Θ of its intermediate Jacobian $J(X)$. Various authors have studied parametrizations of the theta divisor by moduli spaces of sheaves; see Artebani, Kloosterman and Pacini [3], Beauville [9] and Iliev [24].

In this paper, we find a new, and in a sense most efficient, parametrization of this type: a smooth four-dimensional moduli space of stable sheaves isomorphic to the desingularization of the theta divisor.

Let $X \subset \mathbb{P}^4$ be a smooth cubic threefold over $\mathbb{C}$ and H the hyperplane section. Let $\overline{M}_X(v)$ be the moduli space of Gieseker-semistable sheaves on X with Chern character $v := \left(3, -H, -\frac{1}{2}H^2, \frac{1}{6}H^3\right)$.

Theorem 7.1 *The moduli space $\overline{M}_X(v)$ is smooth and irreducible of dimension 4. More precisely, it is the blowup of Θ in its unique singular point. The exceptional divisor is isomorphic to the cubic threefold X itself, and parametrizes non-locally-free sheaves in $\overline{M}_X(v)$.*

Moduli space in the Kuznetsov component

The original motivation for our analysis of the moduli space $\overline{M}_X(v)$ comes from the study of moduli spaces of stable objects in a full triangulated subcategory $\mathrm{Ku}(X) \subset \mathrm{D}^{\mathrm{b}}(X)$ called the *Kuznetsov component*. It is defined through the semiorthogonal decomposition

$$\mathrm{D}^{\mathrm{b}}(X) = \langle \mathrm{Ku}(X), \mathcal{O}_X, \mathcal{O}_X(H) \rangle.$$

See Kuznetsov [25] for details on the decomposition and on the Kuznetsov component.

Stability conditions on $\mathrm{Ku}(X)$ have been constructed in Bernardara, Macrì, Mehrotra and Stellari [11] and Bayer, Lahoz, Macrì and Stellari [5]. These stability conditions are Serre-invariant, which roughly means that stability of an object is preserved by the action of the Serre functor of $\mathrm{Ku}(X)$; see Section 8 for the precise definition. This property allows us to study stability of objects irrespective of the specific construction of stability conditions.

The class v in Theorem 7.1 is chosen as the class of the projection K_P of a skyscraper sheaf $\mathcal{O}_P$ for a point $P \in X$, which is defined by the short exact sequence

$$0 \to K_P \to \mathcal{O}^{\oplus 4} \to I_P(1) \to 0.$$

These are the non-locally-free torsion-free slope-stable sheaves appearing in Theorem 7.1, and we show that they are also stable as objects of $\mathrm{Ku}(X)$ with respect to any Serre-invariant stability condition. Hence, the moduli space $M_\sigma(v)$ of σ-stable objects in $\mathrm{Ku}(X)$ of Chern character v contains X, yet its expected dimension is four. This was our first clue that this moduli space is of interest. Indeed, our next result says that the moduli spaces $M_\sigma(v)$ and $\overline{M}_X(v)$ agree entirely.

Theorem 1.1 (Theorem 8.7 and Proposition 8.10) *Let σ be an arbitrary Serre-invariant stability condition on $\mathrm{Ku}(X)$. Then the moduli space $M_\sigma(v)$ is isomorphic to the moduli space $\overline{M}_X(v)$.*

To summarize, we project the structure sheaf of a point into the Kuznetsov component and take its moduli space. It obviously contains X but is bigger. It is the resolution of the theta divisor, with X as the exceptional divisor. Thus, we recover X from $\mathrm{Ku}(X)$ or from the intermediate Jacobian, ie we obtain new proofs of both the categorical and classical Torelli theorem for cubic threefolds:

Theorem 1.2 (Corollary 7.6 and Theorem 8.1) *Let X_1 and X_2 be smooth cubic threefolds. The following are equivalent:*

(i) *X_1 and X_2 are isomorphic.*

(ii) *$\mathrm{Ku}(X_1)$ and $\mathrm{Ku}(X_2)$ are equivalent as triangulated categories.*

(iii) *$J(X_1)$ and $J(X_2)$ are isomorphic as principally polarized abelian varieties.*

Proof ideas

The proof of Theorem 7.1 relies on two classical ingredients. Firstly, we use the fact that any irreducible theta divisor is normal, due to Ein and Lazarsfeld [16]. Secondly, we use a characterization of the theta divisor of the intermediate Jacobian in terms of twisted cubics; see Proposition 2.2. This was proved by Beauville in [9], but it can also be deduced from the description of Θ as differences of lines in Clemens and Griffiths [14]; see Remark 2.3.

The strategy to prove Theorem 7.1 is to vary the notion of stability and reach a detailed description of the objects that belong to the moduli space $\overline{M}_X(v)$ through wall-crossing. Since X has Picard rank one, Gieseker stability cannot be varied. This is where the derived category comes into play in the form of tilt-stability introduced in Bridgeland [13] for K3 surfaces, and then further generalized to other surfaces and threefolds in Arcara and Bertram [2] and Bayer, Macrì and Toda [7]. In fact, we give a complete description of the wall and chamber structure; see Section 6. Once a set-theoretic description of $\overline{M}_X(v)$ has been reached, we use standard deformation theory arguments in Corollary 6.9 to deduce that it is smooth and of dimension four.

To prove Theorem 8.7, we first show the claim for the specific stability condition constructed in Bayer, Lahoz, Macrì and Stellari [5] which are Serre-invariant by Pertusi and Yang [35]. We then prove in a completely separate argument that our moduli space is independent of the choice of Serre-invariant stability conditions σ. The essential ingredient in this last argument is the weak Mukai lemma from [35].

Related work

In the recent paper [1], Altavilla, Petković and Rota studied moduli spaces of some torsion sheaves in the Kuznetsov components of Fano threefolds with Picard rank one and index two. In the case of cubic threefolds they study $M_\sigma([S^2(K_P)])$ (S is the Serre functor on $\mathrm{Ku}(X)$), but do not obtain our detailed geometric description. A key difference is that in their case the moduli space in the Kuznetsov component is different from the moduli space of Gieseker-semistable sheaves.

Classical Torelli is the implication (iii) $\Longrightarrow$ (i) in Theorem 1.2, which was first proved in Clemens and Griffiths [14]. The implication (ii) $\Longrightarrow$ (iii) was first established in Bernardara, Macrì, Mehrotra and Stellari [11, Theorem 1.1], where it was shown that the Fano variety of lines $F(X)$ can be recovered from $\mathrm{Ku}(X)$ as a moduli space of stable objects. Thus, one obtains the intermediate Jacobian $J(X)$ as the Albanese variety of $F(X)$. A more recent argument for (ii) $\Longrightarrow$ (iii) can be deduced from Perry's categorical construction of intermediate Jacobians [34, Section 5.3], when the equivalence is given by a Fourier–Mukai kernel on $X_1 \times X_2$. Instead, our paper gives a very direct geometric argument for (ii) $\Longrightarrow$ (i), as well as a variant of the proof of classical Torelli via the description of the singularity of theta divisor implied by Theorem 7.1.

Since this article originally appeared on the arXiv, Feyzbakhsh and Pertusi [17] and Zhang [40] proved uniqueness of Serre-invariant stability conditions on $\mathrm{Ku}(X)$. Proposition 8.10 in the last section could now be obtained as an immediate corollary.

Acknowledgements

This work originated from a working group at the workshop *Semiorthogonal decompositions, stability conditions and sheaves of categories* in Toulouse in May 2018, a meeting of the CIMI network "Semiorthogonal decompositions and stability condition". We are deeply grateful to the organizers of the workshop for the nice and productive atmosphere. In particular, we would like to thank Aaron Bertram, Naoki Koseki, Marin Petković, Maxim Smirnov and Benjamin Sung for very useful discussions we had in Toulouse at the beginning of the project. We would also like to express our gratitude to Marcello Bernardara, Emanuele Macrì, Alex Perry, Laura Pertusi and Richard Thomas for many useful comments on the subject of this paper. In addition, we are very grateful to the referee for many detailed comments, making the exposition more accurate and readable.

Bayer, Beentjes, Feyzbakhsh, Martinelli and Rezaee were in part supported by ERC starting grant 337039. Bayer and Rezaee were in part supported by ERC consolidator grant 819864 WallCrossAG. Rezaee was partially supported by EPSRC grant EP/T015896/1 and Maxwell Institute Research Fellowship EP/W522648/1. Feyzbakhsh was partially supported by EPSRC postdoctoral fellowship EP/T018658/1. This material is also partially based upon work supported by the NSF under grant DMS-1440140 while Bayer, Martinelli and Rezaee were in residence at the MSRI in Berkeley, California, during the Spring 2019 semester.

2 Cubic threefolds and intermediate Jacobians

Let $X \subset \mathbb{P}^4$ be a smooth cubic threefold. In their celebrated article [14], Clemens and Griffiths introduced the *intermediate Jacobian* of X. It is the complex torus defined as

$$J(X) := H^{2,1}(X)^\vee / H_3(X,\mathbb{Z}) = H^1(\Omega_X^2)^\vee / H_3(X,\mathbb{Z}).$$

It turns out that $J(X)$ is a principally polarized abelian variety of dimension five.

Let $\{Z_b\}_{b\in\mathcal{B}}$ be a family of 1–cycles over a variety $\mathcal{B}$. The choice of a basepoint $b_0 \in \mathcal{B}$ leads to an Abel–Jacobi map $\Psi_{\mathcal{B}}\colon \mathcal{B} \to J(X)$ as follows. For any $b \in \mathcal{B}$ the cycle $Z_b - Z_{b_0}$ has degree 0, ie it is homologically trivial, and can be written as the boundary $\partial\Gamma$ for a 3–chain Γ. The integral $\int_\Gamma$ is an element in $H^{1,2}(X)^\vee$ whose class in $J(X)$ is the image of the Abel–Jacobi map. By [19, Theorem 2.20] the map $\Psi_{\mathcal{B}}$ is algebraic along the smooth locus of $\mathcal{B}$.

If $Z_b = C$ is a smooth curve, then the induced morphism on tangent spaces has been described by Welters; see [39, Section 2]. Recall that the tangent space of the Hilbert scheme at C is naturally given by $H^0(\mathcal{N}_{C/X})$, where $\mathcal{N}_{C/X}$ is the normal bundle. The tangent space of $J(X)$ at any point is given by $H^{1,2}(X)^\vee = H^1(\Omega_X^2)^\vee$. By definition, the *infinitesimal Abel–Jacobi map*

$$\psi_C\colon H^0(\mathcal{N}_{C/X}) \to H^1(\Omega_X^2)^\vee$$

is the map of tangent spaces induced by $\Psi_{\mathcal{B}}$. We get a dual morphism

$$\psi_C^\vee\colon H^1(\Omega_X^2) \to H^0(\mathcal{N}_{C/X})^\vee.$$

Lemma 2.1 *The following diagram is commutative and has exact rows and columns*:

$$\begin{array}{ccccc}
 & & 0 & & \\
 & & \downarrow & & \\
 & & H^0(\mathcal{I}_C(H)) & & \\
 & & \downarrow & & \\
 & & H^0(\mathcal{O}_X(H)) & \xrightarrow{\cong} & H^1(\Omega_X^2) \\
 & & \downarrow & & \downarrow{\scriptstyle \psi_C^\vee} \\
H^0(\mathcal{N}_{C/\mathbb{P}^4}(-2H)) & \longrightarrow & H^0(\mathcal{O}_C(H)) & \longrightarrow & H^0(\mathcal{N}_{C/X})^\vee
\end{array}$$

Proof This is mostly [39, Lemma 2.8] and the preceding construction of the morphisms. The map $H^0(\mathcal{O}_X(H)) \to H^1(\Omega_X^2)$ is the connecting morphism in a long exact sequence

$$H^0(\Omega_{\mathbb{P}^4}^3 \otimes \mathcal{O}_X(3H)) \to H^0(\mathcal{O}_X(H)) \to H^1(\Omega_X^2) \to H^1(\Omega_{\mathbb{P}^4}^3 \otimes \mathcal{O}_X(3H)).$$

The wedge product induces a perfect pairing $\Omega_{\mathbb{P}^4}^3 \otimes \Omega_{\mathbb{P}^4} \to \mathcal{O}_{\mathbb{P}^4}(-5)$. Therefore, $\Omega_{\mathbb{P}^4}^3 = T_{\mathbb{P}^4}(-5)$. For $i = 0, 1$ we have

$$H^i(T_{\mathbb{P}^4} \otimes \mathcal{O}_X(-2H)) = 0. \qquad \square$$

Recall that the Lefschetz hyperplane theorem says that the hyperplane section $H \in \mathrm{Pic}(X)$ generates the Picard group. One can use twisted cubics to characterize the theta divisor of $J(X)$. A proof of the following result can be found in [9, Proposition 5.2]. Let $\mathcal{T}$ be the open locus of smooth twisted cubics in the Hilbert scheme of X, and let $\overline{\mathcal{T}}$ be its closure.

Proposition 2.2 *The Abel–Jacobi map $\varphi\colon \overline{\mathcal{T}} \to J(X)$ with basepoint of class H^2 is algebraic. Its image is a theta divisor $\Theta \subset J(X)$ and its generic fiber is isomorphic to $\mathbb{P}^2$.*

Remark 2.3 Proposition 2.2 can be deduced from the description of Θ as differences of lines as well. We give a rough sketch of the argument here.

Let F be the Fano variety of lines on X. According to [14] the morphism $F \times F \to J(X)$ that maps $(L, L') \mapsto [L]-[L']$ is generically a 6-to-1 cover of Θ.

Since a twisted cubic $C \subset X$ lies in a unique cubic surface $Y \subset X$, the morphism $\mathcal{T} \to J(X)$ factors via the moduli space $\mathcal{F}$ of pairs (D, Y), where Y is a cubic surface and D is the divisor class of a twisted cubic. The generic fiber of the morphism $\mathcal{T} \to \mathcal{F}$ is given by $\mathbb{P}(H^0(\mathcal{O}_Y(D)) = \mathbb{P}^2$. Indeed, $\mathcal{O}_Y(D)$ is the pullback of $\mathcal{O}_{\mathbb{P}^2}(1)$ if Y is written as the blowup of six general points in $\mathbb{P}^2$.

If D is the class of a twisted cubic on a smooth cubic surface, then $D - H^2$ can be written as the difference of two lines on a cubic surface. Therefore, the Abel–Jacobi morphism maps onto Θ. Moreover, there are precisely six ways to write $D - H^2$ as the difference of two lines. Together with the fact that $F \times F \to J(X)$ is generically a 6-to-1 cover of Θ, we get that $\mathcal{F} \to \Theta$ has degree 1.

Lemma 2.4 *Let $\mathbb{P}^1 \cong C \subset X \subset \mathbb{P}^4$ be a twisted cubic. Then*

$$\mathcal{N}_{C/\mathbb{P}^4} = \mathcal{O}_{\mathbb{P}^1}(5)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^1}(3), \quad h^0(\mathcal{N}_{C/X}) = 6 \quad \text{and} \quad h^1(\mathcal{N}_{C/X}) = 0.$$

In particular, the Hilbert scheme $\mathcal{T}$ is smooth of dimension six.

Proof We have a short exact sequence

$$0 \to \mathcal{N}_{C/\mathbb{P}^3} = \mathcal{O}_{\mathbb{P}^1}(5)^{\oplus 2} \to \mathcal{N}_{C/\mathbb{P}^4} \to \mathcal{N}_{\mathbb{P}^3/\mathbb{P}^4} \otimes \mathcal{O}_C = \mathcal{O}_{\mathbb{P}^1}(3) \to 0.$$

Since $\mathrm{Ext}^1(\mathcal{O}_{\mathbb{P}^1}(3), \mathcal{O}_{\mathbb{P}^1}(5)) = 0$, we get $\mathcal{N}_{C/\mathbb{P}^4} = \mathcal{O}_{\mathbb{P}^1}(5)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^1}(3)$. Next, we have a short exact sequence

$$0 \to \mathcal{N}_{C/X} \to \mathcal{N}_{C/\mathbb{P}^4} = \mathcal{O}_{\mathbb{P}^1}(5)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^1}(3) \to \mathcal{N}_{X/\mathbb{P}^4} \otimes \mathcal{O}_C = \mathcal{O}_{\mathbb{P}^1}(9) \to 0.$$

Thus, $\mathcal{N}_{C/X}$ has degree 4 and can only be $\mathcal{O}_{\mathbb{P}^1}(m) \oplus \mathcal{O}_{\mathbb{P}^1}(4-m)$ for some $-1 \le m \le 5$. The claim about the cohomology of $\mathcal{N}_{C/X}$ holds for each of them. □

Lemma 2.5 *Along the locus of smooth curves $\mathcal{T} \subset \overline{\mathcal{T}}$, the Abel–Jacobi morphism φ has differential of rank four.*

Proof Let $C \subset X$ be a smooth twisted cubic. Clearly, restriction maps $H^0(\mathcal{O}_X(H)) \cong \mathbb{C}^5$ surjectively onto $H^0(\mathcal{O}_C(H)) \cong \mathbb{C}^4$. By Lemma 2.4, we have $h^0(N_{C/\mathbb{P}^4}(-2H)) = h^0(\mathcal{O}_{\mathbb{P}^1}(-1)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^1}(-3)) = 0$.

By Lemma 2.1, we get a commutative diagram

$$\begin{array}{ccc} \mathbb{C}^5 & \xrightarrow{\cong} & H^1(\Omega_X^2) \\ \downarrow\!\!\!\downarrow & & \downarrow \psi_C^\vee \\ \mathbb{C}^4 & \hookrightarrow & H^0(\mathcal{N}_{C/X})^\vee \end{array}$$

Therefore, $\psi_C^\vee$ has rank four. □

The singularities of the theta divisor were computed in [33, page 348]. Another proof was given in [8, Main Theorem and Proposition 2]. We will not need this full description and instead rely only on normality.

Theorem 2.6 [16, Theorem 1] *Any irreducible theta divisor of an abelian variety is normal.*

Lemma 2.7 *Up to numerical equivalence, the Todd class of X is $\mathrm{td}(X) = \left(1, H, \frac{2}{3}H^2, \frac{1}{3}H^3\right)$. In particular, for any $E \in \mathrm{D^b}(X)$,*

$$\chi(E) = \mathrm{ch}_3(E) + H \cdot \mathrm{ch}_2(E) + \tfrac{2}{3}H^2 \cdot \mathrm{ch}_1(E) + \tfrac{1}{3}H^3 \cdot \mathrm{ch}_0(E).$$

Proof By Kodaira vanishing $H^i(\mathcal{O}_X) = 0$ for $i \neq 0$, and therefore, $\chi(\mathcal{O}_X) = 1$. By the Hirzebruch–Riemann–Roch Theorem we get $\mathrm{td}_3(X) = \chi(\mathcal{O}_X) = \frac{1}{3}H^3$. Similarly, Kodaira vanishing implies $H^i(\mathcal{O}_X(-H)) = 0$ for $i \neq 0$. Again by Hirzebruch–Riemann–Roch,

$$0 = \chi(\mathcal{O}_X(-H)) = -\tfrac{1}{6}H^3 + H \cdot \tfrac{1}{2}H^2 - \mathrm{td}_2(X) \cdot H + \tfrac{1}{3}H^3.$$

Since X has Picard rank one, this is only possible if $\mathrm{td}_2(X) = \frac{2}{3}H^2$. □

Lemma 2.8 *The numerical Chow ring $\mathrm{CH}_n^*(X)$ has a basis given by 1, H, $\frac{1}{3}H^2$ and $\frac{1}{3}H^3$. In particular, if $E \in \mathrm{D^b}(X)$, then $\mathrm{ch}_2(E) \in \frac{1}{6}H^2 \cdot \mathbb{Z}$, and $\mathrm{ch}_3(E) \in \frac{1}{6}H^3 \cdot \mathbb{Z}$.*

Proof Since $\mathrm{Pic}(X)$ is generated by H, the group $\mathrm{CH}_\mathrm{n}^2(X)$ is generated by a rational multiple of H^2. A general hyperplane section of X is a smooth cubic surface, which contains lines. The class of such a line is $\frac{1}{3}H^2$. Since $H^3 = 3$, the class has to be indivisible. Since $\frac{1}{3}H^3$ is the class of a point, the group $\mathrm{CH}_\mathrm{n}^3(X)$ must be generated by it.

The claim about second Chern characters follows directly from $\mathrm{ch}_2(E) = \frac{1}{2}c_1^2(E) - c_2(E)$. The claim about $\mathrm{ch}_3(E)$ follows from Lemma 2.7 and the fact that $\chi(E) \in \mathbb{Z}$. □

3 Divisors on hyperplane sections

We need to understand the singularities that can occur on hyperplane sections of X.

Proposition 3.1 *Any cubic hyperplane section $Y = V \cap X \subset \mathbb{P}^4$ is normal and integral.*

Proof Since hypersurfaces satisfy condition S2, by Serre's condition [10, Section 031S], it is enough to show that Y has isolated singularities. Assume for contradiction that Y contains a curve C of singular points. Let F and x be the defining equations of X and V, respectively. Then $\partial F/\partial x$ is a homogeneous degree 2 polynomial and hence vanishes somewhere along C. At such a point, all partial derivatives of F vanish, hence it is a singular point of X, a contradiction. □

In order to deal with singular hyperplane sections, we need to recall the relation between Weil divisors and rank-one reflexive sheaves on integral normal varieties. This is very similar to the standard relation between line bundles and Cartier divisors. We refer to [10, Tag 0EBK] or [36] for proofs of the following facts. They can also be found in [22] in more generality.

Let Y be a normal integral projective variety. By $\mathrm{Cl}(Y)$ we denote the *group of Weil divisors modulo rational equivalence*. For two rank-one reflexive sheaves $L_1, L_2 \in \mathrm{Coh}(Y)$ we can define a new rank-one reflexive sheaf by $(L_1 \otimes L_2)^{\vee\vee}$. This defines a group law for rank-one reflexive sheaves on Y, where inverses are given by $L \mapsto L^\vee$. For any effective prime divisor D one can define a rank-one reflexive sheaf $\mathcal{O}_Y(D) := \mathcal{I}_D^\vee$. This can be linearly extended to any divisor.

Proposition 3.2 (i) *The group of isomorphism classes of rank-one reflexive sheaves is isomorphic to* $\mathrm{Cl}(Y)$ *under the homomorphism* $D \mapsto \mathcal{O}_Y(D)$.

(ii) *To every nonzero section* $s \in H^0(L)$ *of a rank-one reflexive sheaf* L, *one can associate an effective divisor* D *on* Y.

(iii) *For any effective Weil divisor* D *on* Y, *there is a section* $s \in H^0(\mathcal{O}_Y(D))$ *such that the associated divisor is given by* D.

(iv) *Two sections* $s_1, s_2 \in H^0(L)$ *define the same divisor if they satisfy* $s_1 = \lambda s_2$ *for some* $\lambda \in \mathbb{C}^*$.

4 Notions of stability

In this section, we recall a number of notions of stability for sheaves. Let X be a smooth projective threefold, and let H be an ample divisor on X.

Definition 4.1 [32; 38] (i) For any $E \in \mathrm{Coh}(X)$, the *Mumford–Takemoto slope* is defined as

$$\mu(E) := \begin{cases} \dfrac{H^2 \cdot \mathrm{ch}_1(E)}{H^3 \cdot \mathrm{ch}_0(E)} & \text{for } \mathrm{ch}_0(E) \neq 0, \\ +\infty & \text{for } \mathrm{ch}_0(E) = 0. \end{cases}$$

(ii) A sheaf $E \in \mathrm{Coh}(X)$ is *slope-(semi)stable* if for any nontrivial proper subsheaf $F \hookrightarrow E$ the inequality $\mu(F) < (\leq) \mu(E/F)$ holds.

From the definition it follows immediately that if $\operatorname{Pic}(X) = \mathbb{Z} \cdot H$ and E is slope-semistable with $\gcd\bigl(\operatorname{ch}_0(E), H^2 \operatorname{ch}_1(E)/H^3\bigr) = 1$, then E is slope-stable.

While slope-stability suffices to construct moduli spaces of vector bundles on curves, a refinement is necessary in higher dimensions.

Definition 4.2 We define a preorder on the polynomial ring $\mathbb{R}[m]$ as follows.

- (i) For all nonzero $f \in \mathbb{R}[m]$, we have $f \prec 0$.
- (ii) If $\deg(f) > \deg(g)$ for nonzero $f, g \in \mathbb{R}[m]$, then $f \prec g$.
- (iii) Let $\deg(f) = \deg(g)$ for nonzero $f, g \in \mathbb{R}[m]$, and let a_f and a_g be the leading coefficients of f and g, respectively. Then $f \preceq g$ if and only if $f(m)/a_f \le g(m)/a_g$ for all $m \gg 0$.
- (iv) If $f, g \in \mathbb{R}[m]$ with $f \preceq g$ and $g \preceq f$, we write $f \asymp g$.

For any $E \in \operatorname{Coh}(X)$, we denote its *Hilbert polynomial* and the terms $\alpha_i(E)$ by

$$P(E, m) := \chi(E(mH)) = \sum_{i=0}^{3} \alpha_i(E) m^i.$$

Moreover, let $P_2(E, m) = \sum_{i=1}^{3} \alpha_i(E) m^i$.

Definition 4.3 (i) The sheaf E is Gieseker-(semi)stable if for all nontrivial proper subsheaves $F \subset E$, the inequality $P(F, m) \prec (\preceq) P(E, m)$ holds.

- (ii) The sheaf E is 2–Gieseker-(semi)stable if for all nontrivial proper subsheaves $F \subset E$, the inequality $P_2(F, m) \prec (\preceq) P_2(E/F, m)$ holds.

Note that for 2–Gieseker-semistability we could have equivalently asked $P_2(F, m) \preceq P_2(E, m)$, but for 2–Gieseker-stability, $P_2(F, m) \prec P_2(E, m)$ is a stronger condition that is almost never fulfilled for all such subsheaves. These notions imply each other as follows:

$$\begin{array}{ccccc} \text{slope-stable} & \Longrightarrow & \text{2–Gieseker-stable} & \Longrightarrow & \text{Gieseker-stable} \\ & & & & \Downarrow \\ \text{slope-semistable} & \Longleftarrow & \text{2–Gieseker-semistable} & \Longleftarrow & \text{Gieseker-semistable} \end{array}$$

The intermediate notion of 2–Gieseker stability is not classical and will just appear in the technical parts of our arguments.

Due to [18; 30; 31; 37] there exists a projective moduli space $\overline{M}_X(v)$ parametrizing S–equivalence classes of Gieseker-semistable sheaves with Chern character v. Here two semistable sheaves are called *S–equivalent* if they have the same stable factors, up to order and isomorphism, in their Jordan–Hölder filtrations:

Proposition 4.4 [23, Proposition 1.5.2] *Any Gieseker-semistable sheaf $E \in \mathrm{Coh}(X)$ has a filtration*

$$0 = E_0 \hookrightarrow E_1 \hookrightarrow \cdots \hookrightarrow E_n = E$$

such that the factors $A_i := E_i/E_{i-1}$ are Gieseker-stable with $P(A_i, m) \asymp P(E, m)$ for $i = 1, \dots, n$. The sheaf

$$\bigoplus_{i=1}^{n} A_i$$

is uniquely determined (up to isomorphism) by E.

Moreover, any sheaf E has a *Harder–Narasimhan filtration* into Gieseker-semistable factors.

Proposition 4.5 [23, Theorem 1.3.4] *Let $E \in \mathrm{Coh}(X)$. There is a unique filtration*

$$0 = E_0 \hookrightarrow E_1 \hookrightarrow \cdots \hookrightarrow E_n = E$$

such that the factors $A_i := E_i/E_{i-1}$ are Gieseker-semistable with

$$P(A_1, m) \succ P(A_2, m) \succ \cdots \succ P(A_n, m).$$

Based on Bridgeland stability on surfaces, the notion of tilt stability was introduced in [7]. It is not quite a Bridgeland stability condition, but it turns out to suffice for our purposes. The basic idea is to change the category in which subobjects are taken when defining stability. This is done via the theory of tilting introduced in [20]. As before, let X be a smooth projective threefold with an ample divisor H.

Definition 4.6 For any $\beta \in \mathbb{R}$, we define two full additive subcategories of $\mathrm{Coh}(X)$:

$$\mathcal{F}_\beta(X) := \{E \in \mathrm{Coh}(X) : \text{any slope-semistable factor } F \text{ of } E \text{ satisfies } \mu(F) \le \beta\},$$
$$\mathcal{T}_\beta(X) := \{E \in \mathrm{Coh}(X) : \text{any slope-semistable factor } F \text{ of } E \text{ satisfies } \mu(F) > \beta\}.$$

The category

$$\mathrm{Coh}^\beta(X) := \langle \mathcal{T}_\beta(X), \mathcal{F}_\beta(X)[1] \rangle$$

is the full additive subcategory of those $E \in \mathrm{D^b}(X)$ for which $\mathcal{H}^0(E) \in \mathcal{T}_\beta(X)$, $\mathcal{H}^{-1}(E) \in \mathcal{F}_\beta(X)$ and $\mathcal{H}^i(E) = 0$ for all $i \neq -1, 0$.

Note that $\mathrm{Hom}(T, F) = 0$ for all $T \in \mathcal{T}_\beta(X)$ and $F \in \mathcal{F}_\beta(X)$, by semistability. It is well known that the category $\mathrm{Coh}^\beta(X)$ is abelian. A sequence of morphisms

$$0 \to A \to B \to C \to 0$$

in $\mathrm{Coh}^\beta(X)$ is a short exact sequence if and only if the induced sequence

$$A \to B \to C \to A[1]$$

is a distinguished triangle in $\mathrm{D^b}(X)$.

To simplify notation, we define for any $E \in \mathrm{D}^{\mathrm{b}}(X)$ its *twisted Chern character* $\mathrm{ch}^{\beta}(E) := \mathrm{ch}(E) \cdot e^{-\beta H}$. Note that when $\beta \in \mathbb{Z}$, this is nothing but $\mathrm{ch}(E \otimes \mathcal{O}_X(-\beta H))$.

Definition 4.7 For $\alpha > 0$, $\beta \in \mathbb{R}$ and $E \in \mathrm{Coh}^{\beta}(X)$, we define a slope function

$$\nu_{\alpha,\beta}(E) := \frac{H \cdot \mathrm{ch}_2^{\beta}(E) - \frac{1}{2}\alpha^2 H^3 \cdot \mathrm{ch}_0^{\beta}(E)}{H^2 \cdot \mathrm{ch}_1^{\beta}(E)},$$

where again division by zero needs to be interpreted as $+\infty$. Analogously to slope-stability, an object $E \in \mathrm{Coh}^{\beta}(X)$ is called $\nu_{\alpha,\beta}$*–(semi)stable* if for all nontrivial proper subobjects $F \hookrightarrow E$ in $\mathrm{Coh}^{\beta}(X)$ the inequality $\nu_{\alpha,\beta}(F) < (\le)\ \nu_{\alpha,\beta}(E/F)$ holds.

If it is clear from context, we will sometimes abuse notation and write *tilt-(semi)stable* instead of $\nu_{\alpha,\beta}$–(semi)stable. Note that by definition, any $E \in \mathrm{Coh}^{\beta}(X)$ satisfies $H^2 \cdot \mathrm{ch}_1^{\beta}(E) \ge 0$. Therefore, this function plays the same role in $\mathrm{Coh}^{\beta}(X)$ as the rank does in $\mathrm{Coh}(X)$.

As previously, Harder–Narasimhan filtrations exist. However, note that a version of Jordan–Hölder filtrations exists, but the stable factors are not unique up to order.

The notion of 2–Gieseker stability occurs as a limit of tilt stability as follows.

Proposition 4.8 [13, Proposition 14.2] *Let $E \in \mathrm{D}^{\mathrm{b}}(X)$ and $\beta < \mu(E)$. Then $E \in \mathrm{Coh}^{\beta}(X)$ and E is $\nu_{\alpha,\beta}$–(semi)stable for $\alpha \gg 0$ if and only if $E \in \mathrm{Coh}(X)$ and E is 2–Gieseker-(semi)stable.*

The statement in [13] is for K3 surfaces, but the same proof works in our setting. If $\beta > \mu(E)$ the situation is slightly more complicated. The following proposition is a combination of [6, Lemma 2.7] and [27, Proposition 3.1].

Proposition 4.9 *Take a $\nu_{\alpha,\beta}$–semistable object $E \in \mathrm{Coh}^{\beta}(X)$. If $\beta \ne \mu(E)$, then $\mathcal{H}^{-1}(E)$ is a reflexive sheaf, and if $\beta \ge \mu(E)$ and $\alpha \gg 0$, then $\mathcal{H}^{-1}(E)$ is a torsion-free slope-semistable sheaf and $\mathcal{H}^0(E)$ is supported in dimension less than or equal to one.*

Semistable sheaves satisfy the Bogomolov inequality; see [23, Theorem 3.4.1]. A version for tilt stability was proved in [7, Corollary 7.3.2].

Theorem 4.10 (Bogomolov inequality) *Let $E \in \mathrm{Coh}^{\beta}(X)$ be $\nu_{\alpha,\beta}$–semistable. Then*

$$\Delta_H(E) := (H^2 \cdot \mathrm{ch}_1(E))^2 - 2(H^3 \cdot \mathrm{ch}_0(E))(H \cdot \mathrm{ch}_2(E)) \ge 0.$$

Most applications of tilt stability come from varying (α, β) and determining what that means for the stability of a given set of objects. We visualize the parameter space of tilt stability, $(\alpha, \beta) \in \mathbb{R}^2$ with $\alpha > 0$, as the upper half-plane via $i\alpha + \beta$. For a given class $v \in K_0(X)$, it turns out that there is a locally finite wall and chamber structure such that stability only changes as we cross a wall. These walls are

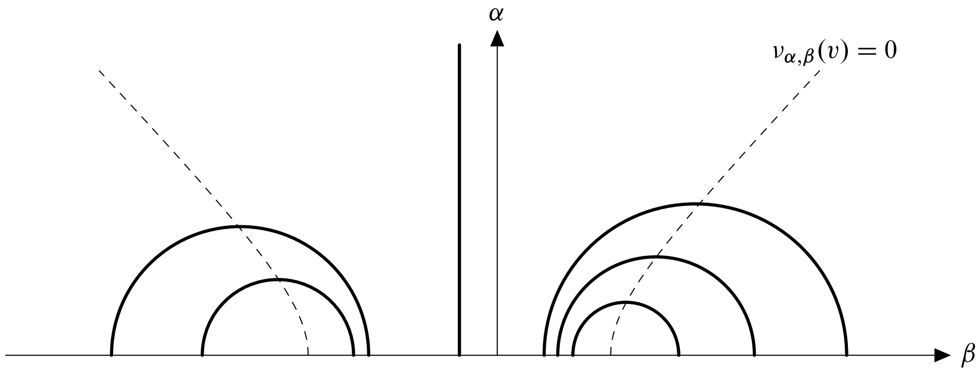


Figure 1: Walls are nested semicircles or a unique vertical wall (Theorem 4.12(ii)).

either semicircles with center on the β–axis or vertical lines; see Figures 1 and 2. In the following, we recall what this means formally.

For $v \in K_0(X)$ we write $\mathrm{ch}(v)$, $\mu(v)$, $\nu_{\alpha,\beta}(v)$ and $\Delta(v)$ to mean the appropriate versions where E is replaced by v.

Definition 4.11 For $v, w \in K_0(X)$, we define

$$W(v,w) := \{(\alpha,\beta) \in \mathbb{R}_{>0} \times \mathbb{R} : \nu_{\alpha,\beta}(v) = \nu_{\alpha,\beta}(w)\}.$$

The set $W(v,w)$ is a *numerical wall* if $W(v,w) \neq \varnothing$ and $W(v,w) \neq \mathbb{R}_{>0} \times \mathbb{R}$, ie if it is a proper nontrivial subset of the upper half-plane.

Numerical walls in tilt stability have a rather simple structure, as shown in [28]:

Theorem 4.12 (nested wall theorem) *Let $v \in K_0(X)$ with $\Delta(v) \geq 0$.*

(i) *A numerical wall for v is either a semicircle centered along the β–axis, or a vertical line parallel to the α–axis in the upper half-plane.*

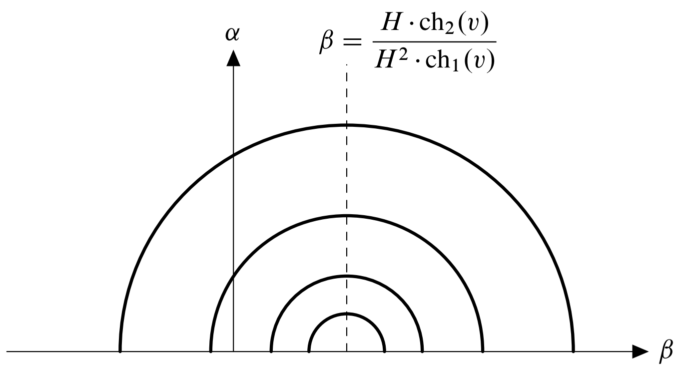


Figure 2: Walls are nested semicircles (Theorem 4.12(iii)).

(ii) *If* $\mathrm{ch}_0(v) \neq 0$, *then there is a unique numerical vertical wall for* v *given by* $\beta = \mu(v)$. *The remaining numerical walls for* v *are split into two sets of nested semicircles, whose apexes lie on the hyperbola* $\nu_{\alpha,\beta}(v) = 0$. *In particular, no two distinct walls intersect.*

(iii) *If* $\mathrm{ch}_0(v) = 0$ *and* $H^2 \cdot \mathrm{ch}_1(v) \neq 0$, *then every numerical wall for* v *is a semicircle, whose apex lies on the ray* $\beta = (H \cdot \mathrm{ch}_2(v))/(H^2 \cdot \mathrm{ch}_1(v))$.

The following is a well-known consequence of the fact that walls do not intersect.

Corollary 4.13 *Let*

$$0 \to F \to E \to G \to 0$$

be a short exact sequence of $\nu_{\alpha,\beta}$*–semistable objects in* $\mathrm{Coh}^{\beta_0}(X)$ *for some* $(\alpha_0, \beta_0) \in W(F, E)$. *Then this is a short exact sequence of* $\nu_{\alpha,\beta}$*–semistable objects in* $\mathrm{Coh}^{\beta}(X)$ *for any* $(\alpha, \beta) \in W(E, F)$.

Definition 4.14 Let $v \in K_0(X)$. A numerical wall W for v is called an *actual wall* for v if there is a short exact sequence

$$0 \to F \to E \to G \to 0$$

of $\nu_{\alpha,\beta}$–semistable objects in $\mathrm{Coh}^{\beta}(X)$ for one $(\alpha, \beta) \in W(F, E)$ such that $W = W(F, E)$ and $\mathrm{ch}(E) = v$.

The above corollary implies that this is a short exact sequence in $\mathrm{Coh}^{\beta}(X)$ for all $(\alpha, \beta) \in W(F, E)$. Determining walls is the key technique in this paper. It will allow us to classify sheaves with certain Chern characters in terms of short exact sequences; see Theorem 6.1. Note that the condition $W(F, E) \neq \mathbb{R}_{>0} \times \mathbb{R}$ implies $\nu_{\alpha,\beta}(F) > \nu_{\alpha,\beta}(E)$ on one side of such a wall. We say that the short exact sequence

$$0 \to F \to E \to G \to 0,$$

or sometimes the wall $W(F, E)$, *destabilizes* E.

Proposition 4.15 [6, Appendix A] *If an actual wall is induced by a short exact sequence of tilt-semistable objects* $0 \to F \to E \to G \to 0$, *then*

$$\Delta_H(F) + \Delta_H(G) \leq \Delta_H(E),$$

and equality can only occur if either F *or* G *is a sheaf supported in dimension zero.*

It turns out that walls of large radius can only be induced by subobjects of small rank. The following precise statement is close to [15, Proposition 8.3]. A proof of this version can be found in [29, Lemma 2.4] for the case of nonnegative ranks. The case of nonpositive ranks has the exact same proof, with reversed signs.

Proposition 4.16 *Assume that an object E is destabilized by a semicircular wall induced by a subobject $F \hookrightarrow E$ or quotient $E \twoheadrightarrow F$ with $\mathrm{ch}_0(F) > \mathrm{ch}_0(E) \geq 0$ or $\mathrm{ch}_0(F) < \mathrm{ch}_0(E) \leq 0$. Then the radius ρ of $W(F,E)$ satisfies*

$$\rho^2 \leq \frac{\Delta_H(E)}{4(H^3 \cdot \mathrm{ch}_0(F))(H^3 \cdot \mathrm{ch}_0(F) - H^3 \cdot \mathrm{ch}_0(E))}.$$

Tilt stability interacts nicely with the derived dual $\mathbb{D}(\,\cdot\,) := \mathbf{R}\mathrm{Hom}(\,\cdot\,, \mathcal{O}_X)[1]$.

Proposition 4.17 [7, Proposition 5.1.3] *Suppose that $E \in \mathrm{Coh}^{\beta}(X)$ is a $\nu_{\alpha,\beta}$–semistable object with $\nu_{\alpha,\beta}(E) \neq \infty$. Then there is a $\nu_{\alpha,-\beta}$–semistable object $\widetilde{E} \in \mathrm{Coh}^{-\beta}(X)$, a torsion sheaf T supported in dimension zero, and a distinguished triangle*

$$\widetilde{E} \to \mathbb{D}(E) \to T[-1] \to \widetilde{E}[1].$$

The following proposition seems to be well known to experts, but we could find no proof in the literature.

Proposition 4.18 *Let $E \in \mathrm{Coh}(X)$ be torsion-free. Then $E[1]$ is tilt-stable along the vertical wall $\beta = \mu(E)$ if and only if E is slope-stable and reflexive. In particular, slope-stable reflexive sheaves do not get destabilized along the vertical wall.*

Proof If E is slope-unstable, then $E \notin \mathrm{Coh}^{\mu(E)}(X)$. Assume that E is strictly slope-semistable. Then there is a short exact sequence of slope-semistable sheaves

$$0 \to F \to E \to G \to 0$$

such that $\mu(F) = \mu(G)$. Taking a shift by one, this becomes a short exact sequence in $\mathrm{Coh}^{\mu(E)}(X)$ with $\nu_{\alpha,\mu(E)}(F[1]) = \nu_{\alpha,\mu(E)}(G[1])$.

Assume that E is not reflexive, but slope-stable. Then we have a short exact sequence in $\mathrm{Coh}^{\mu(E)}(X)$ given by

$$0 \to T \to E[1] \to E^{\vee\vee}[1] \to 0,$$

where T is a nontrivial sheaf supported in dimension less than or equal to one. However, this sequence makes $E[1]$ strictly tilt-semistable along $\beta = \mu(E)$.

Assume conversely that E is a slope-semistable reflexive sheaf. Then it is an object in $\mathrm{Coh}^{\mu(E)}(X)$ of maximal phase, and in particular tilt-semistable. If it is strictly semistable, then it admits a short exact sequence

$$0 \to F \to E[1] \to G[1] \to 0,$$

where F, $G[1]$, $\mathcal{H}^{-1}(F)[1]$ and $\mathcal{H}^0(F)$ are also of maximal phase. In particular, $\mathcal{H}^{-1}(F)$ and G are torsion-free and slope-semistable of slope $\mu(E)$, and $\mathcal{H}^0(F)$ has support of dimension at most one.

Consider the long exact sequence

$$0 \to \mathcal{H}^{-1}(F) \to E \to G \to \mathcal{H}^0(F) \to 0.$$

Since we assume that E is strictly stable, this is a contradiction unless $\mathcal{H}^{-1}(F)=0$. Taking duals we get an exact sequence

$$0 \to G^\vee \to E^\vee \to \mathcal{E}xt^1(F,\mathcal{O}_X).$$

Since F is supported in dimension less than or equal to one, this implies $\mathcal{E}xt^1(F,\mathcal{O}_X)=0$ and $G^\vee \cong E^\vee$. Hence, $E \subsetneq G = G^{\vee\vee} = E^{\vee\vee}$, a contradiction to E being reflexive. □

From now on, we assume $X \subset \mathbb{P}^4$ is a smooth cubic threefold. In the later sections, we need the following result of [26, Proposition 3.2], which improves the Bogomolov inequality in the case of a Fano threefold of Picard rank one. Be aware that our notation differs from Li's.

Theorem 4.19 *Let E be a tilt-stable with $\mathrm{ch}_0(E) \neq 0$ for some $\alpha > 0$, $\beta \in \mathbb{R}$. If $-\frac{1}{2} \le \mu_H(E) \le \frac{1}{2}$, then*

$$\frac{H \cdot \mathrm{ch}_2(E)}{H^3 \cdot \mathrm{ch}_0(E)} \le 0.$$

In the case of cubic threefolds, direct sums of line bundles can be detected among semistable sheaves or objects by their Chern characters, as follows.

Proposition 4.20 (i) *If E is slope-semistable, or $\nu_{\alpha,\beta}$–semistable for some $\alpha > 0$ and $\beta < 0$, with $\mathrm{ch}(E) = (r,0,0,eH^3)$ where $r > 0$, then $e \le 0$. If, additionally, $e = 0$, then $E \cong \mathcal{O}_X^{\oplus r}$.*

(ii) *If E is $\nu_{\alpha,\beta}$–semistable for some $\alpha > 0$ and $\beta > 0$, with $\mathrm{ch}(E) = (-r,0,0,eH^3)$ where $r > 0$, then $e = 0$ and $E \cong \mathcal{O}_X^{\oplus r}[1]$.*

Proof In either case, Proposition 4.15 and $\Delta(E) = 0$ imply that E has no semicircular walls.

We first claim that the only slope-stable reflexive sheaf of class $(r,0,0,eH^3)$ is $\mathcal{O}_X$. Assume otherwise. By Proposition 4.18, such an E is also stable at the vertical wall $\beta = 0$, and thus, it is $\nu_{\alpha,\beta}$–stable for all $\alpha > 0$ and $\beta \in \mathbb{R}$. Since $\nu_{0,\beta}(E) = -\frac{1}{2}\beta > -\frac{1}{2}\beta - 1 = \nu_{0,\beta}(\mathcal{O}_X(-2H)[1])$ and both objects are stable for $\alpha \ll 1$ and $\beta \in (-2,0)$, we have $\mathrm{Ext}^2(\mathcal{O}_X, E) = \mathrm{Hom}(E, \mathcal{O}_X(-2H)[1]) = 0$. Similarly, from $\nu_{\alpha,\beta}$–stability for $\alpha \ll 1$ and $\beta \in (0,2)$ we obtain $\mathrm{Ext}^2(E,\mathcal{O}_X) = \mathrm{Hom}(\mathcal{O}_X(2H), E[1]) = 0$. However, at least one of $\chi(\mathcal{O}_X, E) = r + 3e$ or $\chi(E, \mathcal{O}_X) = r - 3e$ is positive, and so E admits a morphism from $\mathcal{O}_X$ or a morphism to $\mathcal{O}_X$. As both are reflexive and slope-stable of slope 0, this shows $E \cong \mathcal{O}_X$.

Now consider an object E as in case (i). Then $E[1]$ is $\nu_{\alpha,0}$–semistable. By Proposition 4.18, its Jordan–Hölder factors are either of the form $F[1]$ for a slope-stable reflexive sheaf F with $\mathrm{ch}(F) = (r_F, 0, d_F H^2, e_F H^3)$, or a torsion sheaf supported in dimension ≤ 1. In fact, Proposition 4.15 shows $d_F = 0$ in the former case, and thus, $F = \mathcal{O}_X$ by the previous case, and that the torsion sheaves are supported in dimension zero. As $-3e$ is the total length of the torsion sheaves, we get $e \le 0$. If $e = 0$, all factors are isomorphic to $\mathcal{O}_X[1]$ and the claim follows from $\mathrm{Ext}^1(\mathcal{O}_X, \mathcal{O}_X) = 0$.

In case (ii), we again consider a Jordan–Hölder filtration with respect to $\nu_{\alpha,0}$–stability. Let $E_i \hookrightarrow E_{i+1}$ be the first filtration step where the quotient E_{i+1}/E_i is a zero-dimensional torsion sheaf T, should one exist. Then $E_i = \mathcal{O}_X[1]^{\oplus k}$ for some $k > 0$. Since $\mathrm{Ext}^1(T, \mathcal{O}_X[1]) = H^1(T)^\vee = 0$, we have $E_{i+1} = E_i \oplus T$, and so T is a subobject of E. This contradicts stability of E for $\beta > 0$. Thus, $E = \mathcal{O}_X[1]^{\oplus r}$, as claimed. □

5 Construction of sheaves

In this section, we introduce the sheaves that make up our moduli space $\overline{M}_X(v)$. It turns out that all of them are at least reflexive, and the generic one is a vector bundle. From now on $X \subset \mathbb{P}^4$ is an arbitrary smooth cubic threefold.

Let $Y \subset X$ be an arbitrary hyperplane section, D be an effective Weil divisor on Y, and $V \subset H^0(\mathcal{O}_Y(D))$ be a nontrivial subspace. Then we define $\mathcal{E}_{D,V} \in \mathrm{D^b}(X)$ to be the cone of the induced morphism $\mathcal{O}_X \otimes V \to \mathcal{O}_Y(D)$. Moreover, let $E_{D,V} := \mathcal{H}^{-1}(\mathcal{E}_{D,V})$. Hence, we have a long exact sequence

$$0 \to E_{D,V} \to \mathcal{O}_X \otimes V \to \mathcal{O}_Y(D) \to \mathcal{H}^0(\mathcal{E}_{D,V}) \to 0.$$

If $V = H^0(\mathcal{O}_Y(D))$, we will drop V, and just write $\mathcal{E}_D$ and E_D.

Lemma 5.1 *The sheaf $E_{D,V}$ is slope-stable and reflexive. If, additionally, $\mathcal{H}^0(\mathcal{E}_{D,V}) = 0$, then $E_{D,V}$ is a vector bundle.*

Proof The quotient $(\mathcal{O}_X \otimes V)/E_{D,V}$ embeds into $\mathcal{O}_Y(D)$. Since Y is integral by Proposition 3.1, the sheaf $(\mathcal{O}_X \otimes V)/E_{D,V}$ must be supported on Y. Therefore, $\mathrm{ch}_{\leq 1}(E_{D,V}) = (\dim V, -H)$ is primitive and it is enough to show that $E_{D,V}$ is slope-semistable. If not, let $F \subset E_{D,V}$ be the slope-semistable subsheaf in the Harder–Narasimhan filtration of $E_{D,V}$. Then $\mu(F) > \mu(E_{D,V})$ and the quotient $E_{D,V}/F$ is torsion-free. Since F is also a subsheaf of $\mathcal{O}_X \otimes V$, we must have $\mu(F) = 0$. Let $\mathrm{ch}(F) = (r, 0, dH^2, eH^3)$. The quotient $(\mathcal{O}_X \otimes V)/F$ satisfies $\mathrm{ch}((\mathcal{O}_X \otimes V)/F) = (\dim V - r, 0, -dH^2, -eH^3)$. By the snake lemma this quotient is either torsion-free or has a torsion subsheaf purely supported on Y. However, if it is not torsion-free, then its torsion-free quotient would destabilize $\mathcal{O}_X \otimes V$, a contradiction. As a torsion-free quotient of $\mathcal{O}_X \otimes V$ with slope zero, $(\mathcal{O}_X \otimes V)/F$ has to be slope-semistable as well.

The classical Bogomolov inequalities $\Delta_H(F) \geq 0$ and $\Delta_H((\mathcal{O}_X \otimes V)/F) \geq 0$ imply $d = 0$. Applying Proposition 4.20 to both F and $(\mathcal{O}_X \otimes V)/F$ implies $e = 0$, and finally, $F = \mathcal{O}_X^{\oplus r}$. However, by construction, $E_{D,V}$ has no global sections, a contradiction.

To see that $E_{D,V}$ is reflexive it suffices to show that $\mathcal{E}xt^q(E_{D,V}, \mathcal{O}_X) = 0$ for $q \geq 2$ and $\mathcal{E}xt^1(E_{D,V}, \mathcal{O}_X)$ is supported in dimension zero. If additionally $\mathcal{E}xt^1(E_{D,V}, \mathcal{O}_X) = 0$, then $E_{D,V}$ is a vector bundle.

Clearly, $\mathcal{E}xt^q(\mathcal{O}_X \otimes V, \mathcal{O}_X) = 0$ for $q \neq 0$. Because $\mathcal{O}_Y(D)$ is a rank-one reflexive sheaf on the codimension one subvariety Y, the quotient $(\mathcal{O}_X \otimes V)/E_{D,V} \subset \mathcal{O}_Y(D)$ is purely supported on Y. We can use [23, Proposition 1.1.10] to see that $\mathcal{E}xt^q((\mathcal{O}_X \otimes V)/E_{D,V}, \mathcal{O}_X) = 0$ for all $q \neq 1, 2$, and $\mathcal{E}xt^2((\mathcal{O}_X \otimes V)/E_{D,V}, \mathcal{O}_X)$ is supported in dimension zero. The long exact sequence obtained from dualizing the short exact sequence

$$0 \to E_{D,V} \to \mathcal{O}_X \otimes V \to (\mathcal{O}_X \otimes V)/E_{D,V} \to 0 \tag{1}$$

implies the required vanishings.

If additionally $\mathcal{H}^0(\mathcal{E}_{D,V}) = 0$, then $(\mathcal{O}_X \otimes V)/E_{D,V} = \mathcal{O}_Y(D)$ is a reflexive sheaf on the codimension-one subvariety Y, and we can use [23, Proposition 1.1.10] again to see that $\mathcal{E}xt^2(\mathcal{O}_Y(D), \mathcal{O}_X) = 0$. The same long exact sequence as above now implies $\mathcal{E}xt^1(E_{D,V}, \mathcal{O}_X) = 0$. □

Note that we will use this lemma for the case $\operatorname{ch}(\mathcal{O}_Y(D)) = \left(0, H, \frac{1}{2}H^2, -\frac{1}{6}H^3\right)$. It will turn out that in this case $h^0(\mathcal{O}_Y(D)) = 3$ for any such D (see Theorem 6.1) and we will choose $V = H^0(\mathcal{O}_Y(D))$. Moreover, we will show that in that case $\mathcal{H}^0(\mathcal{E}_D) = 0$, ie $\mathcal{O}_Y(D)$ is globally generated; see Theorem 6.1. A straightforward computation shows that in this example $\operatorname{ch}(E_D) = \left(3, -H, -\frac{1}{2}H^2, \frac{1}{6}H^3\right)$.

Corollary 5.2 *Let $P \in X$. Then $h^0(\mathcal{I}_P(H)) = 4$ and the sheaf K_P defined through the exact sequence*

$$0 \to K_P \to \mathcal{O}_X^{\oplus 4} \to \mathcal{I}_P(H) \to 0 \tag{2}$$

satisfies $\operatorname{ch}(K_P) = \left(3, -H, -\frac{1}{2}H^2, \frac{1}{6}H^3\right)$. Moreover, K_P is reflexive and slope-stable, and locally free except at P.

Proof By choosing an embedding $K_P \hookrightarrow \mathcal{O}_X^{\oplus 3}$ we get a short exact sequence

$$0 \to K_P \to \mathcal{O}_X^{\oplus 3} \to \mathcal{I}_{P/Y}(H) \to 0$$

for some hyperplane section Y. The statement then follows from Lemma 5.1 by choosing $D = H$ and $V = H^0(\mathcal{I}_{P/Y}(H)) \subset H^0(\mathcal{O}_Y(H))$.

From the defining short exact sequence (2) one immediately sees that K_P is locally free away from P (as it is the kernel of a surjective map of vector bundles), and not locally free at P (as $\operatorname{Ext}^2(\mathcal{O}_P, K_P) = \operatorname{Ext}^1(\mathcal{O}_P, I_P(H)) \neq 0$). □

6 Variation of stability

In this section, we investigate semistable sheaves with Chern character

$$v := \left(3, -H, -\tfrac{1}{2}H^2, \tfrac{1}{6}H^3\right).$$

The main goal is to use wall-crossing to prove the following theorem, which gives a set-theoretic description of the moduli space $\overline{M}_X(v)$.

Theorem 6.1 (i) *Suppose that D is a Weil divisor on a (possibly singular) hyperplane section Y with $\mathrm{ch}(\mathcal{O}_Y(D)) = \left(0, H, \frac{1}{2}H^2, -\frac{1}{6}H^3\right)$. Then $\mathcal{O}_Y(D)$ is globally generated, and $h^0(\mathcal{O}_Y(D)) = 3$. In particular, there exists a smooth twisted cubic C in Y of class D.*

(ii) *A sheaf E with Chern character v is Gieseker-semistable if and only if it is either equal to the reflexive sheaf K_P for a point $P \in X$ as in (2), or the vector bundle E_D for a Weil divisor D on a hyperplane section $Y \subset X$ as in (1) with $\mathrm{ch}(\mathcal{O}_Y(D)) = \left(0, H, \frac{1}{2}H^2, -\frac{1}{6}H^3\right)$.*

Note that since $\mathrm{ch}_1(E) = -H$, any Gieseker-semistable sheaf of class v is slope-stable. The argument will essentially boil down to a detailed analysis of the numerical wall W defined by

$$\alpha^2 + \left(\beta - \tfrac{1}{2}\right)^2 = \tfrac{1}{4}. \tag{3}$$

At this wall, the short exact sequences (2) and (1) become destabilizing short exact sequences in $\mathrm{Coh}^\beta(X)$ in the form

$$0 \to \mathcal{O}_Y(D) \to E_D[1] \to \mathcal{O}_X[1]^{\oplus 3} \to 0 \quad \text{and} \quad 0 \to I_P(H) \to K_P[1] \to \mathcal{O}_X[1]^{\oplus 4} \to 0.$$

Moreover, we can show that every object gets destabilized, and the destabilizing short exact sequence must be of one of these types; see Lemma 6.8.

6.1 Classification of some torsion sheaves

In this section, we prove the following proposition.

Proposition 6.2 *The wall W of equation (3) is the unique actual wall in tilt stability for objects G with Chern character $\mathrm{ch}(G) = \left(0, H, \frac{1}{2}H^2, -\frac{1}{6}H^3\right)$.*

(i) *Above W the moduli space of tilt-semistable objects is the moduli space of Gieseker-semistable sheaves, and contains precisely the following two types of sheaves G:*

   (a) *$G = \mathcal{I}_{P/Y}(H)$ for $Y \in |H|$ and $P \in Y$, and*

   (b) *$G = \mathcal{O}_Y(D)$, where D is a Weil divisor on some $Y \in |H|$.*

(ii) *Below W the moduli space of tilt-semistable objects contains precisely the following two types of objects G:*

   (a) *the unique nontrivial extensions*

$$0 \to \mathcal{O}_X[1] \to G_P \to \mathcal{I}_P(H) \to 0 \tag{4}$$

   for points $P \in X$, and

   (b) *$G = \mathcal{O}_Y(D)$, where D is a Weil divisor on some $Y \in |H|$.*

We start by dealing with slightly more general objects without fixing ch_3.

Lemma 6.3 *The wall W of equation* (3) *is the unique actual wall in tilt stability for objects G with Chern character* $\mathrm{ch}_{\le 2}(G) = \left(0, H, \frac{1}{2}H^2\right)$. *If G is strictly semistable along W, then any Jordan–Hölder filtration of G is given by either*

$$0 \to \mathcal{I}_Z(H) \to G \to \mathcal{O}_X[1] \to 0 \quad \text{or} \quad 0 \to \mathcal{O}_X[1] \to G \to \mathcal{I}_Z(H) \to 0,$$

where $Z \subset X$ is a zero-dimensional subscheme of length $\frac{1}{6}H^3 - \mathrm{ch}_3(G)$.

Proof All walls for $\left(0, H, \frac{1}{2}H^2\right)$ intersect the vertical ray $\beta = \frac{1}{2}$. If G is strictly semistable along some numerical wall intersecting $\beta = \frac{1}{2}$, then there is a short exact sequence in $\mathrm{Coh}^{1/2}(X)$ of tilt-semistable objects

$$0 \to A \to G \to B \to 0$$

with equal tilt-slope. Let $\mathrm{ch}_{\le 2}(A) = (r, cH, dH^2)$. By definition of $\mathrm{Coh}^{1/2}(X)$ and the fact that neither A nor B can have infinite tilt-slope, we get

$$0 < H^2 \cdot \mathrm{ch}_1^{1/2}(A) = H^3\left(c - \tfrac{1}{2}r\right) < H^2 \cdot \mathrm{ch}_1^{1/2}(G) = H^3.$$

Therefore, $c = \frac{1}{2}r + \frac{1}{2}$, and in particular, r is odd. We will deal with the case $r < 0$. If $r > 0$, then B has negative rank and one simply has to exchange the roles of A and B in the following argument.

For $\left(\alpha, \frac{1}{2}\right) \in W(A, G)$ we have

$$-\alpha^2 r + 2d - \tfrac{1}{4}r - \tfrac{1}{2} = \nu_{\alpha,1/2}(A) = \nu_{\alpha,1/2}(G) = 0.$$

Since $\alpha^2 > 0$, this implies $d < \frac{1}{8}r + \frac{1}{4}$. The fact

$$0 \le \frac{\Delta_H(A)}{(H^3)^2} = -2dr + \tfrac{1}{4}r^2 + \tfrac{1}{2}r + \tfrac{1}{4}$$

implies $d \ge \frac{1}{8}r + (1/8r) + \frac{1}{4}$. Since $d \in \frac{1}{6}\mathbb{Z}$, these restrictions on d are only possible for $r \in \{-1, -3\}$. If $r = -3$, then $\mathrm{ch}_{\le 2}(A) = \left(-3, -H, -\frac{1}{6}H^2\right)$. This case is immediately ruled out by Theorem 4.19. If $r = -1$, then $\mathrm{ch}_{\le 2}(A) = (-1, 0, 0)$, and by Proposition 4.20, we know $A = \mathcal{O}_X[1]$. Then $\mathrm{ch}(B) = \left(1, H, \frac{1}{2}H^2, \mathrm{ch}_3(G)\right)$. By Proposition 4.15, there is no semicircular wall for B, and by Proposition 4.8, the object B has to be a 2–Gieseker-stable sheaf. Since $\mathrm{ch}(B(-H)) = \left(1, 0, 0, \mathrm{ch}_3(G) - \frac{1}{6}H^3\right)$, the remaining statement follows by applying Proposition 4.20 to $B(-H)$. □

The next step is to gain further control over the third Chern character.

Lemma 6.4 *Let G be a $\nu_{\alpha,\beta}$–semistable object with* $\mathrm{ch}_{\le 2}(G) = \left(0, H, \frac{1}{2}H^2\right)$. *Then* $\mathrm{ch}_3(G) \le \frac{1}{6}H^3$. *If* $\mathrm{ch}_3(G) = \frac{1}{6}H^3$ *and (α, β) is above W, then $G \cong \mathcal{O}_Y(H)$ for some $Y \in |H|$.*

Proof We may assume $\mathrm{ch}_3(G) \ge \frac{1}{6}H^3$. By Lemma 6.3, the only possible wall is given by W. Therefore, G has to be tilt-semistable along W. Since W lies below the numerical wall $W(G, \mathcal{O}_X(-H)[1])$, we get $\mathrm{ext}^2(\mathcal{O}_X(H), G) = \mathrm{hom}(G, \mathcal{O}_X(-H)[1]) = 0$. Thus,

$$\mathrm{hom}(\mathcal{O}_X(H), G) \ge \chi(\mathcal{O}_X(H), G) = \mathrm{ch}_3(G) + \tfrac{1}{6}H^3 > 0.$$

Therefore, W is a wall for G and by Lemma 6.3, the destabilizing sequence is

$$0 \to \mathcal{O}_X(H) \to G \to \mathcal{O}_X[1] \to 0.$$

This implies $G = \mathcal{O}_Y(H)$ for some $Y \in |H|$ and $\mathrm{ch}_3(G) = \frac{1}{6}H^3$. □

Proof of Proposition 6.2 Assume that G is strictly tilt-semistable along W. Then Lemma 6.3 splits our problem into two cases.

Firstly, assume that G fits into a nonsplitting short exact sequence

$$0 \to \mathcal{I}_P(H) \to G \to \mathcal{O}_X[1] \to 0$$

for a point $P \in X$. Then clearly $G = \mathcal{I}_{P/Y}(H)$ for some $Y \in |H|$. This object is tilt-stable above W, and tilt-unstable below W by precisely this sequence.

Secondly, assume that G fits into a nonsplitting short exact sequence

$$0 \to \mathcal{O}_X[1] \to G \to \mathcal{I}_P(H) \to 0 \tag{5}$$

for some $P \in X$. By Serre duality, $\mathrm{Ext}^1(\mathcal{I}_P(H), \mathcal{O}_X[1]) = h^1(\mathcal{I}_P(-H)) = 1$ and hence, there is a unique G for each $P \in X$. Clearly, this object is tilt-unstable above W. Assume it is also tilt-unstable below W. Then there is a short exact sequence $0 \to A \to G \to B \to 0$ destabilizing G below the wall. However, G is strictly semistable at W, and by Lemma 6.3, this implies $B = \mathcal{O}_X[1]$. However, that means the short exact sequence (5) splits, a contradiction.

Lastly, assume that G is $\nu_{\alpha,\beta}$–stable for all (α, β). By Proposition 4.17, $\mathbb{D}(G)$ lies in a distinguished triangle

$$\widetilde{G} \to \mathbb{D}(G) \to T[-1] \to \widetilde{G}[1], \tag{6}$$

where T is a torsion sheaf supported in dimension zero and $\widetilde{G} \in \mathrm{Coh}^{-\beta}(X)$ is $\nu_{\alpha,-\beta}$–semistable. If $\mathrm{ch}_3(T) = t$, then $\mathrm{ch}(\widetilde{G}) = \big(0, H, -\frac{1}{2}H^2, -\frac{1}{6}H^3 + t\big)$. Thus, $\widetilde{G}$ is a pure sheaf supported on a hyperplane section $Y \in |H|$. We can compute

$$\mathrm{ch}(\widetilde{G} \otimes \mathcal{O}_X(H)) = \big(0, H, \tfrac{1}{2}H^2, -\tfrac{1}{6}H^3 + t\big).$$

Thus, Lemma 6.4 gives $t = 0$ or $t = 1$, and if $t = 1$, then $\widetilde{G} \otimes \mathcal{O}_X(H) \cong \mathcal{O}_Y(H)$, ie $\widetilde{G} \cong \mathcal{O}_Y(H)$. Hence there is a nontrivial morphism $\mathcal{O}_X \to \widetilde{G}$. Since $\hom(\mathcal{O}_X, T[-i]) = 0$ for $i > 0$, The triangle (6) shows that there is a nontrivial morphism $\mathcal{O}_X \to \mathbb{D}(G)$. Dualizing this morphism leads to a nontrivial morphism $G \to \mathcal{O}_X[1]$. However, this is in contradiction to the assumption that G is stable along W.

If $t = 0$, then $\mathbb{D}(G) = \widetilde{G}$ is a sheaf, so $\mathcal{E}xt^q(G, \mathcal{O}_X) = 0$ for $q > 1$. Thus, [23, Proposition 1.1.10] implies that G is reflexive and supported on a hyperplane section $Y \in |H|$. This means $G = \mathcal{O}_Y(D)$ for some Weil divisor D on Y. □

6.2 Set-theoretic description of the moduli space

We now prepare the proof of Theorem 6.1.

Lemma 6.5 *There are no walls along $\beta = -1$ for tilt-semistable objects E with Chern character* $\mathrm{ch}_{\leq 2}(E) = \left(3, -H, -\frac{1}{2}H^2\right)$.

Proof Assume there is such a wall induced by a short exact sequence

$$0 \to A \to E \to B \to 0$$

with $\mathrm{ch}^{-1}_{\leq 2}(A) = (r, xH, yH^2)$. Then

$$0 < H \cdot \mathrm{ch}^{-1}_1(A) = xH^3 < H \cdot \mathrm{ch}^{-1}_1(E) = 2H^3$$

implies $x = 1$. By exchanging the roles of A and B if necessary, we may assume $r \geq 2$.

Using $\Delta_H(A) \geq 0$ we get $y \leq 1/2r$. A straightforward computation shows that there exists $\alpha > 0$ with $\nu_{\alpha,-1}(A) = \nu_{\alpha,-1}(E)$ if and only if $y > 0$. Since $y \in \frac{1}{6}\mathbb{Z}$, this is only possible if $y = \frac{1}{6}$ and $r \in \{2, 3\}$. Both cases $\mathrm{ch}^{-1}_{\leq 2}(A) = \left(3, H, \frac{1}{6}H^2\right)$ and $\mathrm{ch}^{-1}_{\leq 2}(A) = \left(2, H, \frac{1}{6}H^2\right)$ are directly ruled out by Theorem 4.19. □

Proposition 6.6 *Take a slope-stable sheaf E of Chern character $(3, -H, \mathrm{ch}_2, \mathrm{ch}_3)$. Then $H \cdot \mathrm{ch}_2 \leq -\frac{1}{2}H^3$, and if $\mathrm{ch}_2 \cdot H = -\frac{1}{2}H^3$, then $\mathrm{ch}_3 \leq \frac{1}{6}H^3$. In particular, this implies that any slope-stable sheaf of Chern character v is a reflexive sheaf.*

Proof Since E is slope-stable, the classical Bogomolov inequality gives

$$\Delta_H(E) = (H^3)^2 - 2(3H^3)(H \cdot \mathrm{ch}_2(E)) \geq 0,$$

which implies $H \cdot \mathrm{ch}_2(E) \leq \frac{1}{6}H^3$. The case $H \cdot \mathrm{ch}_2(E) = \frac{1}{6}H^3$ is immediately ruled out by Theorem 4.19. Since $c_2(E) = \frac{1}{2}H^2 - \mathrm{ch}_2(E)$ has to be an integral class, we are left to rule out $H \cdot \mathrm{ch}_2(E) = -\frac{1}{6}H^3$. Assume $H \cdot \mathrm{ch}_2(E) = -\frac{1}{6}H^3$. We may assume that E is a reflexive sheaf. If not, we replace it by the double dual $E^{\vee\vee}$, which satisfies $H \cdot \mathrm{ch}_2(E) \leq H \cdot \mathrm{ch}_2(E^{\vee\vee})$. By the first part of the argument $H \cdot \mathrm{ch}_2(E^{\vee\vee}) = -\frac{1}{6}H^3$ holds as well.

We first show that $\mathrm{ext}^2(E, E) = 0$. Since $H^3 \cdot \mathrm{ch}^{-1/2}_1(E) = \frac{1}{2}H$, any destabilizing subobject $F \subset E$ along $\beta = -\frac{1}{2}$ must satisfy $H^3 \cdot \mathrm{ch}^{-1/2}_1(F) = \frac{1}{2}H$ or $H^3 \cdot \mathrm{ch}^{-1/2}_1(F) = 0$. Thus, either F or the quotient E/F have infinite tilt-slope, a contradiction. This means E is $\nu_{\alpha,-1/2}$–stable for all $\alpha > 0$.

By Proposition 4.18, the object $E[1]$ is tilt-stable for $\beta = 0$ and $\alpha \gg 0$. Since $H^3 \cdot \mathrm{ch}_1(E[1]) = H^3$, the same type of argument as above shows that there cannot be any wall along $\beta = 0$. Hence, $E(-2H)[1]$ is $\nu_{\alpha,\beta}$–stable for $\beta = -2$ and any $\alpha > 0$.

A straightforward computation shows that $W(E, E(-2H)[1])$ intersects both the vertical lines $\beta = -2$ and $\beta = -\frac{1}{2}$. Therefore, E and $E(-2H)[1]$ are tilt-stable for any $(\alpha, \beta) \in W(E, E(-2H)[1])$ and have the same phase, and thus, $\mathrm{ext}^2(E, E) = \hom(E, E(-2H)[1]) = 0$. Since E is stable, we know $\hom(E, E) = 1$ and hence, $3 = \chi(E, E) = 1 - \mathrm{ext}^1(E, E) - \mathrm{ext}^3(E, E) \leq 1$, a contradiction.

Now assume $H \cdot \mathrm{ch}_2 = -\frac{1}{2}H^3$. We know $E \in \mathrm{Coh}^{\beta}(X)$ is $\nu_{\alpha,\beta}$–stable for $\alpha \gg 0$ and $\beta < -\frac{1}{3}$. By Lemma 6.5, we have that E is $\nu_{\alpha,-1}$–stable for any $\alpha > 0$. One can easily compute

$$\nu_{0,-1}(\mathcal{O}_X(-2H)[1]) < \nu_{0,-1}(E),$$

which implies $h^2(E) = \hom(E, \mathcal{O}_X(-2H)[1]) = 0$. Moreover, since $\mu(E) = -\frac{1}{3} < \mu(\mathcal{O}_X)$, we get $\hom(\mathcal{O}_X, E) = 0$. Therefore, $\chi(E) = \mathrm{ch}_3(E) - \frac{1}{6}H^3 \leq 0$, as claimed.

Lastly, assume that a slope-stable sheaf E of Chern character v is not reflexive. We have a short exact sequence

$$0 \to E \to E^{\vee\vee} \to T \to 0.$$

Since $E^{\vee\vee}$ is also slope-stable, and both $H \cdot \mathrm{ch}_2(E)$ and $H \cdot \mathrm{ch}_3(E)$ are maximal, one gets $\mathrm{ch}(E) = \mathrm{ch}(E^{\vee\vee})$. This is only possible if $T = 0$. □

To prove Theorem 6.1, we start in the large volume limit.

Lemma 6.7 *Take $\beta > -\frac{1}{3}$. An object $\widetilde{E} \in \mathrm{Coh}^{\beta}(X)$ of Chern character $-v$ is $\nu_{\alpha,\beta}$–semistable for $\alpha \gg 0$ if and only if $\widetilde{E} \cong E[1]$ for a slope-stable reflexive sheaf E.*

Proof Take a $\nu_{\alpha,\beta}$–semistable object $\widetilde{E}$ of class $-v$. Proposition 4.9 implies that $\mathcal{H}^{-1}(\widetilde{E})$ is a slope-stable reflexive sheaf and $\mathcal{H}^0(\widetilde{E})$ is a torsion sheaf supported in dimension ≤ 1. Therefore,

$$\mathrm{ch}(\mathcal{H}^{-1}(\widetilde{E})) = \big(3, -H, -\tfrac{1}{2}H^2 + \mathrm{ch}_2(\mathcal{H}^0(\widetilde{E})), \tfrac{1}{6}H^3 + \mathrm{ch}_3(\mathcal{H}^0(\widetilde{E}))\big).$$

By Proposition 6.6, this is only possible if $\mathrm{ch}_2(\mathcal{H}^0(\widetilde{E})) = \mathrm{ch}_3(\mathcal{H}^0(\widetilde{E})) = 0$, ie $\mathcal{H}^0(\widetilde{E}) = 0$.

Conversely, any slope-stable reflexive sheaf E of class v is $\nu_{\alpha,\beta}$–stable for $\alpha \gg 0$ and $\beta < \mu(E) = -\frac{1}{3}$. Proposition 4.18 implies that $E[1]$ is $\nu_{\alpha,\beta}$–stable for $\alpha \gg 0$ and $\beta > \mu(E) = -\frac{1}{3}$. □

Next, we move down from the large volume limit and investigate walls for objects of class $-v$. Note that all walls to the right of the vertical wall must intersect $\beta = -\frac{1}{3}$.

Lemma 6.8 *The wall W of equation* (3) *is the unique actual wall for objects with Chern character $-v$ to the right of the vertical wall. There are no tilt-semistable objects below W. Any tilt-semistable $\widetilde{E}$ with Chern character $-v$ fits into one of the following two cases*:

(i) *$\widetilde{E}$ fits into a short exact sequence*

$$0 \to \mathcal{O}_Y(D) \to \widetilde{E} \to \mathcal{O}_X^{\oplus 3}[1] \to 0,$$

where D is a Weil divisor on hyperplane section $Y \in |H|$.

(ii) *$\widetilde{E}$ fits into a short exact sequence*

$$0 \to \mathcal{I}_P(H) \to \widetilde{E} \to \mathcal{O}_X^{\oplus 4}[1] \to 0,$$

where $P \in X$.

Proof Let $\widetilde{E}$ be a tilt-semistable object with Chern character $-v$. Let W' be a wall strictly above W induced by a short exact sequence $0 \to F \to \widetilde{E} \to G \to 0$. Then the wall W' contains points $(\alpha, 0)$ with $\alpha > 0$. In particular, $0 < H \cdot \mathrm{ch}_1(F) < H \cdot \mathrm{ch}_1(\widetilde{E}) = H^3$, a contradiction.

Since the wall $W(\mathcal{O}_X(2H), \widetilde{E})$ is larger than W, we get $\hom(\widetilde{E}, \mathcal{O}_X[3]) = \hom(\mathcal{O}_X(2H), \widetilde{E}) = 0$ and

$$\hom(\widetilde{E}, \mathcal{O}_X[1]) = \hom(\widetilde{E}, \mathcal{O}_X) + \mathrm{ext}^2(\widetilde{E}, \mathcal{O}_X) - \chi(\widetilde{E}, \mathcal{O}_X) \geq -\chi(\widetilde{E}, \mathcal{O}_X) = 3.$$

Clearly, any morphism $\widetilde{E} \to \mathcal{O}_X[1]$ destabilizes $\widetilde{E}$ below W.

Let $r := \hom(\widetilde{E}, \mathcal{O}_X[1]) \geq 3$. We get a short exact sequence of tilt-semistable objects along W given by

$$0 \to G \to \widetilde{E} \to \mathcal{O}_X^{\oplus r}[1] \to 0.$$

If $r \geq 4$, then Proposition 4.16 says

$$\frac{1}{4} \leq \frac{1}{r(r-3)},$$

ie $r \leq 4$. For $r = 4$, we get $\mathrm{ch}(G(-H)) = \left(1, 0, 0, -\frac{1}{3}H^3\right)$ and so $G = \mathcal{I}_P(H)$ for some $P \in X$.

If $r = 3$, then $\mathrm{ch}(G) = \left(0, H, \frac{1}{2}H^2, -\frac{1}{6}H^3\right)$. Assume G is not of the form $\mathcal{O}_Y(D)$ for some Weil divisor D on a hyperplane section $Y \in |H|$. Then Proposition 6.2 implies that G has to be strictly semistable along our wall W. Since $\widetilde{E}$ is tilt-semistable above the wall, we know $\mathrm{Hom}(\mathcal{O}_X[1], E) = 0$. Therefore, Lemma 6.3 shows that there is a short exact sequence

$$0 \to \mathcal{I}_P(H) \to G \to \mathcal{O}_X[1] \to 0$$

for a point $P \in X$. But then there is an inclusion $\mathcal{I}_P(H) \hookrightarrow \widetilde{E}$ and we are in the second case. □

Proof of Theorem 6.1 Let D be a Weil divisor on a hyperplane section $Y \in |H|$ with $\mathrm{ch}(\mathcal{O}_Y(D)) = \left(0, H, \frac{1}{2}H^2, -\frac{1}{6}H^3\right)$. By Proposition 6.2, the sheaf $\mathcal{O}_Y(D)$ is tilt-stable for all $\alpha > 0$ and $\beta \in \mathbb{R}$. A straightforward computation shows that the numerical wall $W(\mathcal{O}_Y(D), \mathcal{O}_X(-2H)[1])$ is nonempty, and therefore, $h^2(\mathcal{O}_Y(D)) = \hom(\mathcal{O}_Y(D), \mathcal{O}_X(-2H)[1]) = 0$. We conclude

$$h^0(\mathcal{O}_Y(D)) = \chi(\mathcal{O}_Y(D)) + h^1(\mathcal{O}_Y(D)) + h^3(\mathcal{O}_Y(D)) \geq \chi(\mathcal{O}_Y(D)) = 3.$$

We pick a three-dimensional subspace $V \subset h^0(\mathcal{O}_Y(D))$ to get an object $\mathcal{E}_{D,V} \in \mathrm{D^b}(X)$ as in Section 5. By Lemma 5.1, the sheaf $E_{D,V} = \mathcal{H}^{-1}(\mathcal{E}_{D,V})$ is slope-stable and reflexive. If $\mathcal{H}^0(\mathcal{E}_{D,V}) \neq 0$, then $E_{D,V}$ has a Chern character in contradiction to Proposition 6.6. This shows that $\mathcal{O}_Y(D)$ is globally generated.

Since $E_{D,V}$ is slope-stable, we know $h^0(E_{D,V}) = 0$ and $h^3(E_{D,V}) = \hom(E_{D,V}, \mathcal{O}_X(-2H)) = 0$. Moreover, as in the proof of Proposition 6.6 we get $h^2(E_{D,V}) = 0$. This implies $h^1(E_{D,V}) = -\chi(E_{D,V}) = 0$.

The long exact sequence obtained from taking sheaf cohomology of

$$0 \to E_{D,V} \to \mathcal{O}_X \otimes V \to \mathcal{O}_Y(D) \to 0$$

implies $H^i(\mathcal{O}_Y(D)) = 0$ for $i > 0$ and $h^0(\mathcal{O}_Y(D)) = 3$. Therefore, $V = H^0(\mathcal{O}_Y(D))$ and for each D there is a unique slope-stable sheaf $E_D = E_{D,V}$.

Let $U \subset Y$ be the smooth locus of Y. By Proposition 3.1, we know that Y is normal, and therefore, $Y \backslash U$ has dimension zero. In particular, a general section of $\mathcal{O}_Y(D)$ leads to a curve completely contained in U. Since we work in characteristic 0, we can use a version of Bertini's theorem [21, Corollary III.10.9, Remark III.10.9.1, Remark III.10.9.2] on the open subset U to see that a general section cuts out a smooth curve C. By adjunction,

$$\begin{aligned}\mathrm{ch}(K_C) &= \mathrm{ch}(\mathcal{O}_Y(-H+D)_{|D}) = \mathrm{ch}(\mathcal{O}_Y(-H+D)) - \mathrm{ch}(\mathcal{O}_Y(-H)) \\ &= \left(0, H, -\tfrac{1}{2}H^2, -\tfrac{1}{6}H^3\right) - \left(0, H, -\tfrac{3}{2}H^2, \tfrac{7}{6}H^3\right) = \left(0, 0, H^2, -\tfrac{4}{3}H^3\right),\end{aligned}$$

which shows that C is of degree 3 with $\chi(K_C) = -1$, ie a twisted cubic. This completes the proof of part (i).

For part (ii), we already showed in Corollary 5.2 that K_P is slope-stable for any $P \in X$. Conversely, if E is slope-stable, we can immediately conclude by Lemma 6.8. □

As a consequence we can already infer that our moduli space $\overline{M}_X(v)$ is smooth.

Corollary 6.9 *Every Gieseker-semistable sheaf E with $\mathrm{ch}(E) = \left(3, -H, -\tfrac{1}{2}H^2, \tfrac{1}{6}H^3\right)$ satisfies*

$$\mathrm{Ext}^i(E,E) = \begin{cases} \mathbb{C} & \text{if } i = 0, \\ \mathbb{C}^4 & \text{if } i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, the moduli space $\overline{M}_X(v)$ is smooth and 4–dimensional.

Proof Since $(3, -H)$ is primitive, we know that E is slope-stable. Therefore, $\hom(E,E) = 1$. Moreover, we must have $\mathrm{Ext}^3(E,E) = \mathrm{Hom}(E, E(-2H))^\vee = 0$. By Lemma 6.5, the sheaf E is $\nu_{\alpha,-1}$–stable for any $\alpha > 0$. Proposition 6.6 shows that $E(-2H)$ is reflexive, so its shift $E(-2H)[1]$ lies in the heart $\mathrm{Coh}^{\beta=-1}(X)$ and it is $\nu_{\alpha,-1}$–stable for any $\alpha > 0$ by Lemma 6.8. Since

$$\nu_{0,-1}(E) = 0 > -\tfrac{1}{2} = \nu_{0,-1}(E(-2H)[1]),$$

we get $\mathrm{Ext}^2(E,E) = \mathrm{Hom}(E, E(-2H)[1]) = 0$. We can conclude that

$$\mathrm{ext}^1(E,E) = \hom(E,E) - \chi(E,E) = 4.$$

□

7 Proof of the main theorem

Recall that $\overline{M}_X(v)$ is the moduli space of Gieseker-semistable sheaves with Chern character

$$v := \left(3, -H, -\tfrac{1}{2}H^2, \tfrac{1}{6}H^3\right)$$

and $M_X(v) \subset \overline{M}_X(v)$ is the open locus of Gieseker-semistable vector bundles. The aim of this section is to prove the following theorem.

Theorem 7.1 *The moduli space $\overline{M}_X(v)$ is smooth and irreducible of dimension 4. Moreover, there is an Abel–Jacobi morphism $\Psi\colon \overline{M}_X(v) \to J(X)$ sending $E \mapsto \widetilde{c}_2(E) - H^2$, whose image is a theta divisor Θ in the intermediate Jacobian $J(X)$. The theta divisor has a unique singular point, and $\overline{M}_X(v)$ is the blowup of Θ in this point. The exceptional divisor is isomorphic to the cubic threefold X itself.*

We have already shown that $\overline{M_X(v)}$ is smooth of dimension 4 in Corollary 6.9. By Proposition 2.2, the image of $\varphi\colon \overline{\mathcal{T}} \to J(X)$ is $\Theta \subset J(X)$, where $\mathcal{T}$ is the open locus of smooth twisted cubics in the Hilbert scheme of X, and $\overline{\mathcal{T}}$ is its closure. By Theorem 2.6, we know that Θ is normal.

Proposition 7.2 *There is a surjective map $\varphi'\colon \mathcal{T} \to M_X(v)$ that sends a twisted cubic C to the vector bundle E_C. The map $\varphi|_{\mathcal{T}}\colon \mathcal{T} \to J(X)$ factors through φ':*

$$\begin{array}{ccc} & \mathcal{T} & \\ {}^{\varphi'}\swarrow & & \searrow^{\varphi|_{\mathcal{T}}} \\ M_X(v) & \xrightarrow{\ \Psi|_{M_X(v)}\ } & J(X) \end{array}$$

Therefore, the image of $\Psi\colon \overline{M}_X(v) \to J(X)$ is $\Theta \subset J(X)$.

Proof Let C be a twisted cubic in X. Then it lies in a unique hyperplane section Y. There is a short exact sequence

$$0 \to \mathcal{O}_Y \to \mathcal{O}_Y(C) \to T \to 0,$$

where T is a sheaf supported on C with rank one. Therefore, $\widetilde{\mathrm{ch}}_{\le 2}(\mathcal{O}_Y(C)) = \bigl(0, H, C - \frac{1}{2}H^2\bigr)$ and we get $\widetilde{\mathrm{ch}}_{\le 2}(E_C) = \bigl(3, -H, \frac{1}{2}H^2 - C\bigr)$. It follows that $\widetilde{c}_2(E_C) = C$. Thus, the composition $\Psi|_{M_X(v)} \circ \varphi'\colon \mathcal{T} \to M_X(v) \to J(X)$ is the Abel–Jacobi map $\varphi\colon \overline{\mathcal{T}} \to J(X)$ restricted to $\mathcal{T}$. Surjectivity of φ' is a direct consequence of Theorem 6.1. □

Lemma 7.3 *The morphism $i\colon X \to \overline{M}_X(v)$ that maps $P \mapsto K_P$ is an embedding with normal bundle $\mathcal{O}_X(-H)$.*

Proof We interpret X as the moduli spaces of twisted ideal sheaves $\mathcal{I}_P(H)$ for all $P \in X$. By definition of K_P, we have a canonical short exact sequence

$$0 \to K_P \to \mathcal{O}_X^{\oplus 4} \to \mathcal{I}_P(H) \to 0. \tag{7}$$

The appropriate version in families, considered below, induces the morphism i. It is injective, as P is the unique point where K_P is not locally free by Corollary 5.2.

Applying $\mathrm{Hom}(\cdot, K_P)$ to (7), we get an isomorphism $\mathrm{Ext}^1(K_P, K_P) \cong \mathrm{Ext}^2(\mathcal{I}_P(H), K_P)$. Next, we apply the functor $\mathrm{Hom}(\mathcal{I}_P(H), \cdot)$ to (7) to show that the induced morphism on tangent spaces $\mathrm{Ext}^1(\mathcal{I}_P(H), \mathcal{I}_P(H)) \hookrightarrow \mathrm{Ext}^2(\mathcal{I}_P(H), K_P) = \mathrm{Ext}^1(K_P, K_P)$ is an embedding. Since both X and $\overline{M}_X(v)$ are smooth, the morphism is an embedding.

To determine the normal bundle, we need a relative version of the previous arguments to determine the cokernel of this embedding as a line bundle on X. The universal family inducing i is given by the sheaf $\mathcal{K}$ on $X \times X$ fitting into the short exact sequence

$$0 \to \mathcal{K} \to p^*\Omega_{\mathbb{P}^4}|_X(H) \to \mathcal{I}_\Delta(0, H) \to 0,$$

where $p\colon X \times X \to X$ is the projection to the first factor. The pullback of the tangent bundle via i is $i^*T_{\overline{M}_X(v)} = \mathcal{H}^1(p_*\,\mathcal{H}om(\mathcal{K}, \mathcal{K}))$. Since $p_*\,\mathcal{H}om(p^*\Omega_{\mathbb{P}^4}|_X(H), \mathcal{K}) = 0$, we have an isomorphism

$$\mathcal{H}^1(p_*\,\mathcal{H}om(\mathcal{K}, \mathcal{K})) = \mathcal{H}^2(p_*\,\mathcal{H}om(\mathcal{I}_\Delta(0, H), \mathcal{K}).$$

The differential d_i of i fits into the four-term long exact sequence

$$\begin{aligned} 0 \to T_X = \mathcal{H}^1\big(p_*\,\mathcal{H}om(\mathcal{I}_\Delta(0, H), \mathcal{I}_\Delta(0, H))\big) &\xrightarrow{d_i} \mathcal{H}^2\big(p_*\,\mathcal{H}om(\mathcal{I}_\Delta(0, H), \mathcal{K})\big) \\ &\to \mathcal{H}^2\big(p_*\,\mathcal{H}om(\mathcal{I}_\Delta(0, H), p^*\Omega_{\mathbb{P}^4}|_X(H))\big) \to \mathcal{H}^2\big(p_*\,\mathrm{Hom}(\mathcal{I}_\Delta(0, H), \mathcal{I}_\Delta(0, H))\big) \to 0. \end{aligned}$$

Using Grothendieck duality and the projection formula, the third term becomes

$$\Omega_{\mathbb{P}^4}|_X(H) \otimes \mathcal{H}^1(p_*\mathcal{I}_\Delta(0, -H))^\vee = \Omega_{\mathbb{P}^4}|_X(H) \otimes \mathcal{H}^0(p_*\mathcal{O}_\Delta(0, -H))^\vee = \Omega_{\mathbb{P}^4}|_X(2H).$$

A similar computation using the short exact sequence $I_\Delta \hookrightarrow \mathcal{O}_X \boxtimes \mathcal{O}_X \twoheadrightarrow \mathcal{O}_\Delta$ gives

$$\mathcal{H}^2(p_*\,\mathrm{Hom}(\mathcal{I}_\Delta, \mathcal{I}_\Delta) = \Omega_X(2H)$$

for the fourth term. Thus, the cokernel of d_i is isomorphic to $\mathcal{N}^\vee_{X/\mathbb{P}^4}(2H) = \mathcal{O}_X(-H)$, as claimed. □

Lemma 7.4 *The morphism Ψ induces an isomorphism $M_X(v) \to \Theta \setminus \{0\}$. Moreover, Ψ contracts the irreducible divisor $\overline{M}_X(v) \setminus M_X(v)$ to the zero point. In particular, Θ is smooth away from* 0.

Proof By Lemma 5.1 and Corollary 5.2, the locus $\overline{M}_X(v) \setminus M_X(v)$ coincides with vector bundles E_C associated to a twisted cubic C. By Lemma 2.5, the map $\varphi|_T$ has full rank four on tangent spaces. Thus, the commutative diagram in Proposition 7.2 implies that $\Psi|_{M_X(v)}$ has full rank four on tangent spaces. Since $M_X(v)$ is smooth of dimension four, $\Psi|_{M_X(v)}$ must be injective on tangent spaces. In particular, the morphism $\Psi|_{M_X(v)}$ must have finite fibers. Since $\varphi|_T$ has generically connected fibers by Proposition 2.2, the same holds for $\Psi|_{M_X(v)}$. Since Θ is normal, Zariski's main theorem implies that $\Psi|_{M_X(v)}$ is an open embedding. Since Θ is singular at the origin, we must have $\Psi(M_X(v)) \subset \Theta - \{0\}$.

By definition, $\tilde{c}_2(K_P) = H^2$ and we get $\Psi(K_P) = 0$. Thus $\Psi^{-1}(0) = \overline{M}_X(v) \setminus M_X(v)$, and the image of $M_X(v)$ is indeed $\Theta \setminus \{0\}$ by Proposition 2.2. □

We can finish the proof of Theorem 7.1 with the following lemma.

Lemma 7.5 *The formal neighborhood of* $0 \in \Theta$ *is isomorphic to the vertex of the affine cone over* $X \subset \mathbb{P}^4$*. Moreover, we have an isomorphism* $\overline{M}_X(v) = \mathrm{Bl}_0(\Theta)$*. Thus,* X *is the union of all rational curves on* $\overline{M}_X(v)$*, and the unique divisor contracted by any morphism to a complex abelian variety.*

Proof The first two claims are scheme-theoretic enhancements of the set-theoretic statements in the previous lemma, which hold for any contraction of a divisor with ample conormal bundle to a point. We will only sketch the arguments.

Since the normal bundle of $X \subset \overline{M}_X(v)$ is antiample, by Artin's contractibility criterion [4, Corollary 6.12] there is a contraction $\Psi' \colon \overline{M}_X(v) \to N$ to an algebraic space N of finite type over $\mathbb{C}$ that is an isomorphism away from X, and contracts X to a point $0 \in N$. Moreover, by Artin's construction in [4, Theorem 6.2], the formal neighborhood of $0 \in N$ is given by the affinization of the formal neighborhood of $X \subset \overline{M}_X(v)$. More precisely, if $\mathcal{I}$ is the ideal of X, then it is given by

$$\operatorname{Spec} \varprojlim_n H^0(X, \mathcal{O}_{\overline{M}_X(v)}/\mathcal{I}^{n+1}) = \operatorname{Spec} \varprojlim_n \bigoplus_{0 \le k \le n} H^0(X, \mathcal{O}_X(k)),$$

ie the completion of the vertex of the affine cone over X. Since the image of every infinitesimal neighborhood of X under Ψ is affine, it factors via its affinization. Taking the limit, we see that Ψ factors via Ψ' both in the formal neighborhood of X, and in its complement. Hence (eg by [4, Theorem 3.1]) we get an induced morphism $j \colon N \to \Theta$ factoring Ψ. As j is bijective on points and has normal target, it is an isomorphism.

For the last claim, note that X is uniruled, hence the union U of all rational curves in $\overline{M}_X(v)$ contains X. If there was any other rational curve C not contained in X, then $\Psi \colon C \to \Theta$ is a nonconstant map from a rational to an abelian variety, a contradiction. □

Corollary 7.6 *If* X_1 *and* X_2 *are smooth projective threefolds with* $J(X_1) = J(X_2)$ *as principally polarized abelian varieties, then* $X_1 = X_2$*.*

Proof As in the classical argument, this is an immediate consequence of the description of the singularity of the theta divisor in Lemma 7.5. □

8 Kuznetsov component

The bounded derived category of a cubic threefold X admits a semiorthogonal decomposition

$$\mathrm{D}^{\mathrm{b}}(X) = \langle \mathrm{Ku}(X), \mathcal{O}_X, \mathcal{O}_X(1) \rangle,$$

whose nontrivial part $\mathrm{Ku}(X)$ is called the Kuznetsov component. The goal of this section is to give a new proof of the following theorem.

Theorem 8.1 *Let X_1 and X_2 be smooth cubic threefolds. Then $\mathrm{Ku}(X_1)$ and $\mathrm{Ku}(X_2)$ are equivalent as triangulated categories if and only if X_1 and X_2 are isomorphic.*

Let S be the Serre functor of $\mathrm{Ku}(X)$. By [25, Lemmas 4.1 and 4.2], for any object $F \in \mathrm{Ku}(X)$, we have

$$S(F) = L_{\mathcal{O}_X}(F \otimes \mathcal{O}_X(H))[1], \tag{8}$$

where $L_{\mathcal{O}_X}$ is the left mutation functor with respect to $\mathcal{O}_X$. By [11, Proposition 2.7], the numerical Grothendieck group $\mathcal{N}(\mathrm{Ku}(X))$ is a two-dimensional lattice

$$\mathcal{N}(\mathrm{Ku}(X)) \cong \mathbb{Z}^2 \cong \mathbb{Z}[\mathcal{I}_\ell] \oplus \mathbb{Z}[S(\mathcal{I}_\ell)],$$

where $\mathcal{I}_\ell$ is the ideal sheaf of a line ℓ in X. With respect to this basis, the Euler characteristic $\chi(-,-)$ on $\mathcal{N}(\mathrm{Ku}(X))$ has the form

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

For any line ℓ in X, we know $\mathrm{ch}(\mathcal{I}_\ell) = \left(1, 0, -\frac{1}{3}H^2, 0\right)$. The Chern character of our second basis vector of $\mathrm{ch}(\mathrm{Ku}(X))$, and the action of the Serre functor S on our chosen basis are given as follows.

Lemma 8.2 *We have $\mathrm{ch}(S(I_\ell)) = \left(2, -H, -\frac{1}{6}H^2, \frac{1}{6}H^3\right)$ and $\mathrm{ch}(S^2(I_\ell)) = \left(1, -H, \frac{1}{6}H^2, \frac{1}{6}H^3\right)$. Thus, the class $[S^2(\mathcal{I}_\ell)]$ in $\mathcal{N}(\mathrm{Ku}(X))$ is equal to $[S(\mathcal{I}_\ell)] - [\mathcal{I}_\ell]$.*

Proof By (8) we have $[S(E)] = -[E(H)] + \chi(E(H))[\mathcal{O}_X]$ for $E \in \mathrm{Ku}(X)$. Hence $\mathrm{ch}(I_\ell(H)) = \left(1, H, \frac{1}{6}H^2, -\frac{1}{6}H^3\right)$ and $\chi(I_\ell(H)) = 3$ imply the formula for $\mathrm{ch}(S(I_\ell))$. The formula for $\mathrm{ch}(S^2(I_\ell))$ follows from the last claim, which in turn follows from the Euler characteristic form above with

$$\chi(I_\ell, S^2(I_\ell)) = \chi(S^2(I_\ell), S(I_\ell)) = \chi(S(I_\ell), I_\ell) = 0 = \chi([I_\ell], [S(\mathcal{I}_\ell)] - [\mathcal{I}_\ell]),$$
$$\chi(S(I_\ell), S^2(I_\ell)) = \chi(I_\ell, S(I_\ell)) = -1 = \chi([S(I_\ell)], [S(\mathcal{I}_\ell)] - [\mathcal{I}_\ell]). \qquad \square$$

For a point $P \in X$, the sheaf K_P, which is defined through the sequence (2), lies in the Kuznetsov component $\mathrm{Ku}(X)$.

Lemma 8.3 *Let $[A]$ be a class in $\mathcal{N}(\mathrm{Ku}(X))$ such that $\chi([A],[A]) = -3$. Then, up to a sign, $[A]$ is either $[K_P] = [\mathcal{I}_\ell] + [S(\mathcal{I}_\ell)]$, or $[S(K_P)] = -[\mathcal{I}_\ell] + 2[S(\mathcal{I}_\ell)]$, or $[S^2(K_p)] = -2[\mathcal{I}_\ell] + [S(\mathcal{I}_\ell)]$.*

Let $\sigma^0_{\alpha,-1/2} = (\mathrm{Coh}^0_{\alpha,-1/2}(X), Z^0_{\alpha,-1/2})$ be the weak stability condition on $\mathrm{D^b}(X)$ constructed in [5, Proposition 2.14]. Here $\mathrm{Coh}^0_{\alpha,-1/2}(X)$ is the usual double tilt and

$$Z^0_{\alpha,-1/2}(E) = H^2 \cdot \mathrm{ch}_1^{-1/2}(E) + i\left(H \cdot \mathrm{ch}_2^{-1/2}(E) - \tfrac{1}{2}\alpha^2 H^3 \cdot \mathrm{ch}_0(E)\right). \tag{9}$$

As proven in [5, Theorem 6.8], for $0 < \alpha \ll 1$ it induces the stability condition $\sigma(\alpha) = (\mathcal{A}(\alpha), Z(\alpha))$ on $\mathrm{Ku}(X)$, where

$$\mathcal{A}(\alpha) := \mathrm{Coh}^0_{\alpha,-1/2}(X) \cap \mathrm{Ku}(X) \quad \text{and} \quad Z(\alpha) := Z^0_{\alpha,-1/2}|_{\mathrm{Ku}(X)}.$$

Lemma 8.4 *There is an embedding $M_X(v) \hookrightarrow M_{\sigma(\alpha)}([\mathcal{I}_\ell]+[S(\mathcal{I}_\ell)])$ from the moduli space $M_X(v)$ for $v = \mathrm{ch}(\mathcal{I}_\ell)+\mathrm{ch}(S(\mathcal{I}_\ell)) = \left(3, -H, -\frac{1}{2}H^2, \frac{1}{6}H^3\right)$ to $M_{\sigma(\alpha)}([\mathcal{I}_\ell]+[S(\mathcal{I}_\ell)])$, which parametrizes $\sigma(\alpha)$–semistable objects in $\mathrm{Ku}(X)$ of class $[\mathcal{I}_\ell]+[S(\mathcal{I}_\ell)] \in \mathcal{N}(\mathrm{Ku}(X))$.*

Proof According to Lemma 6.5 there is no wall for objects of Chern character v to the left of the vertical wall. Thus, E is $\nu_{\alpha,-1/2}$–stable for any $\alpha > 0$. Since $\sigma^0_{\alpha,-1/2}$ is just a rotation of $\nu_{\alpha,-1/2}$, we obtain that E is $\sigma^0_{\alpha,-1/2}$–stable. By Theorem 6.1(ii), the sheaf $E \in \mathrm{Ku}(X)$ lies in the Kuznetsov component. Thus, E is $\sigma(\alpha)$–stable. Note that the object E could be destabilized by objects with $Z^0_{\alpha,-1/2} = 0$ after rotation. But we know that these are all sheaves supported in dimension zero and would not be in $\mathrm{Ku}(X)$ and therefore, E is stable after restriction to $\mathrm{Ku}(X)$. □

Corollary 5.6 of [35] implies that the stability condition $\sigma(\alpha)$ is S–invariant, ie $S \cdot \sigma(\alpha) = \sigma(\alpha)\cdot \widetilde{g}$ for $\widetilde{g} \in \widetilde{\mathrm{GL}}^+(2,\mathbb{R})$. Thus, there is an isomorphism

$$S\colon M_{\sigma(\alpha)}(2[I_\ell]-[S(\mathcal{I}_\ell)]) \to M_{\sigma(\alpha)}([\mathcal{I}_\ell]+[S(\mathcal{I}_\ell)]), \quad E \mapsto S(E). \tag{10}$$

The following proposition is a slight strengthening of [1, Theorem 1.2], which describes all elements of the moduli space. The idea of the proof is the same as [1, Lemma 2.2].

Proposition 8.5 *Any $\sigma(\alpha)$–semistable object in $\mathrm{Ku}(X)$ of class $2[\mathcal{I}_\ell]-[S(\mathcal{I}_\ell)]$ is of the form $G[2k]$ for $k \in \mathbb{Z}$, where G is either equal to $G_P(-H)$ described in (4) for a point $P \in X$, or $\mathcal{O}_Y(D-H)$, where D is a Weil divisor on some $Y \in |H|$.*

Proof Lemma 8.2 implies $\mathrm{ch}(G) = \left(0, H, -\frac{1}{2}H^2, -\frac{1}{6}H^3\right)$. Since G is $\sigma(\alpha)$–semistable, its shift $G[2k]$ lies in the heart $\mathcal{A}(\alpha)$ for some $k \in \mathbb{Z}$. We know its image under the stability function $Z(\alpha)$ is equal to $-H^3$, so it has maximum phase in the heart $\mathcal{A}(\alpha)$, which immediately implies $G[2k]$ is $\sigma^0_{\alpha,-1/2}$–semistable. We claim that $G[2k]$ has no subobject $Q \in \mathrm{Coh}^0_{\alpha,-1/2}$ with $Z^0_{\alpha,-1/2}(Q) = 0$, so it is $\nu_{\alpha,-1/2}$–semistable. Assume for a contradiction that there is such a subobject Q. By the definition of $\mathrm{Coh}^0_{\alpha,-1/2}(X)$, it is a sheaf supported in dimension zero. Thus, $\hom(\mathcal{O}_X, Q) \neq 0$. Since $\mathcal{O}_X \in \mathrm{Coh}^0_{\alpha,-1/2}(X)$, we have $\hom(\mathcal{O}_X, (G[2k]/Q)[-1]) = 0$. Therefore, $\hom(\mathcal{O}_X, G[2k]) \neq 0$, which is not possible because $G[2k] \in \mathrm{Ku}(X)$. Finally, since $G[2k]$ is $\nu_{\alpha,-1/2}$–semistable for $0 < \alpha \ll 1$, the claim follows by Proposition 6.2(ii). □

Remark 8.6 Since the class $2[\mathcal{I}_\ell]-[S(\mathcal{I}_\ell)]$ is primitive in $\mathcal{N}(\mathrm{Ku}(X))$, any $\sigma(\alpha)$–semistable object of this class is $\sigma(\alpha)$–stable if we choose α sufficiently small.

We now describe the image of the semistable objects $G \in M_{\sigma(\alpha)}(2[\mathcal{I}_\ell]-[S(\mathcal{I}_\ell)])$ under the Serre functor S. If $G = G_P(-H)$, then by (4), we know there is a distinguished triangle

$$\mathcal{O}_X[1] \to G_P \to \mathcal{I}_P(H) \to \mathcal{O}_X[2],$$

which gives $L_{\mathcal{O}_X}(G_P) = L_{\mathcal{O}_X}(\mathcal{I}_P(H)) = K_P[1]$, so

$$S(G_P) = K_P[2]. \tag{11}$$

If $G = \mathcal{O}_Y(D-H)$, then $G(H) = \mathcal{O}_Y(D)$ is of class $\left(0, H, \frac{1}{2}H^2, -\frac{1}{6}H^3\right)$, and lies in a distinguished triangle

$$\mathcal{O}_X^{\oplus 3} \to \mathcal{O}_Y(D) \to E_D[1] \to \mathcal{O}_X^{\oplus 3}[1].$$

Thus,

$$S(G) = L_{\mathcal{O}_X}(\mathcal{O}_Y(D))[1] = L_{\mathcal{O}_X}(E_D[1])[1] = E_D[2]. \tag{12}$$

Combining (11) and (12) with Lemma 8.4 implies the next result.

Theorem 8.7 *The moduli space $M_{\sigma(\alpha)}([\mathcal{I}_\ell] + [S(\mathcal{I}_\ell)])$ is isomorphic to the moduli space $\overline{M}_X(v)$ parametrizing Gieseker-stable sheaves of class v.*

The next step is to show that we can replace $\sigma(\alpha)$ by any S–invariant stability condition on $\mathrm{Ku}(X)$.

Lemma 8.8 [35, Lemmas 5.8 and 5.10] *Let σ be an S–invariant stability condition on $\mathrm{Ku}(X)$ and $F \in \mathrm{Ku}(X)$ be σ–semistable of phase $\varphi(F)$. Then*

(i) $\varphi(F) < \varphi(S(F)) < \varphi(F) + 2$,

(ii) $\dim \mathrm{Ext}^1(F, F) \geq 2$.

For cubic threefolds, we also have a weak version of the Mukai lemma for K3 surfaces.

Lemma 8.9 (weak Mukai lemma [35, Lemma 5.11]) *Let σ be an S–invariant stability condition. Let $A \to E \to B$ be a triangle in $\mathrm{Ku}(X)$ such that $\hom(A, B) = 0$ and the σ–semistable factors of A have phase greater than or equal to the phase of the σ–semistable factors of B. Then*

$$\dim_{\mathbb{C}} \mathrm{Ext}^1(A, A) + \dim_{\mathbb{C}} \mathrm{Ext}^1(B, B) \leq \dim_{\mathbb{C}} \mathrm{Ext}^1(E, E).$$

Proposition 8.10 *Let σ_1 and σ_2 be two S–invariant stability conditions on $\mathrm{Ku}(X)$. An object $E \in \mathrm{Ku}(X)$ of class $[\mathcal{I}_\ell] + [S(\mathcal{I}_\ell)]$ is σ_1–stable if and only if it is σ_2–stale.*

Proof By [35, Proposition 4.6], $\mathcal{I}_\ell$ and $S(\mathcal{I}_\ell)$ are σ–stable with respect to any S–invariant stability condition. Thus, Lemma 8.8 implies that

$$\varphi_\sigma(\mathcal{I}_\ell) < \varphi_\sigma(S(\mathcal{I}_\ell)) < \varphi_\sigma(\mathcal{I}_\ell) + 2. \tag{13}$$

Take a σ_1–stable object $E \in \mathrm{Ku}(X)$ of class $[\mathcal{I}_\ell] + [S(\mathcal{I}_\ell)]$. Since σ_1 is S–invariant, Lemma 8.8 gives

$$\varphi_{\sigma_1}(E) < \varphi_{\sigma_1}(S(E)) < \varphi_{\sigma_1}(E) + 2.$$

Thus, for $i < 0$ or $i \geq 2$, we get

$$\hom(E, E[i]) = \hom(E[i], S(E)) = 0.$$

Since E is σ_1–stable, we get $\hom(E,E)=1$, which gives

$$\hom(E,E[1]) = -\chi(E,E)+1 = 4.$$

Suppose now for a contradiction that E is σ_2–unstable. There is a distinguished triangle of destabilizing objects $F_1 \to E \to F_2 \to F_1[1]$ with respect to σ_2. We may assume F_1 is σ_2–semistable. Thus, Lemma 8.8 implies that

$$\hom(F_1, F_1[1]) \geq 2. \tag{14}$$

Since the phase of F_1 is bigger than the phase of σ_2–semistable factors of F_2, we have

$$\hom(F_1, F_2) = 0. \tag{15}$$

Thus, the weak Mukai lemma (Lemma 8.9) implies

$$\hom(F_1, F_1[1]) + \hom(F_2, F_2[1]) \leq \hom(E, E[1]) = 4.$$

By (14), we get $\hom(F_2, F_2[1]) \leq 2$. If $\hom(F_2, F_2[1]) = 0$ or 1, then all its σ_2–semistable factors would satisfy the same property by the weak Mukai lemma (Lemma 8.9), which is not possible by Lemma 8.8. Therefore,

$$\hom(F_1, F_1[2]) = \hom(F_2, F_2[1]) = 2,$$

and [35, Lemma 5.12] implies that F_1 and F_2 are σ_2–stable. This gives $\chi(F_i, F_i) = -1$ for $i = 1, 2$, so $[F_i]$ is either $\pm[\mathcal{I}_\ell]$, or $\pm[S(\mathcal{I}_\ell)]$, or $\pm([S(\mathcal{I}_\ell)] - [\mathcal{I}_\ell])$. Since there are only 2 stable factors and the object E is of class $[\mathcal{I}_\ell] + [S(\mathcal{I}_\ell)]$, the destabilizing objects must be of class $[\mathcal{I}_\ell]$ and $[S(\mathcal{I}_\ell)]$. Thus, [35, Proposition 4.6] implies that the destabilizing objects are $\mathcal{I}_\ell[2k]$ and $S(\mathcal{I}_{\ell'})[2k']$ for two lines ℓ, ℓ' and integers $k, k' \in \mathbb{Z}$.

Let $F_1 = \mathcal{I}_\ell[2k]$ and $F_2 = S(\mathcal{I}_{\ell'})[2k']$. Since E is σ_1–stable, we have $\varphi_{\sigma_1}(F_1) < \varphi_{\sigma_1}(F_2)$, thus (13) gives $k \leq k'$. But F_1 and F_2 are the destabilizing objects with respect to σ_2, hence $\varphi_{\sigma_2}(F_1) > \varphi_{\sigma_2}(F_2)$ and (13) gives $k'+1 \leq k$, which is not possible. By a similar argument, we reach a contradiction if $F_1 = S(\mathcal{I}_{\ell'})[2k']$ and $F_2 = \mathcal{I}_\ell[2k]$. Finally, note that E cannot be strictly σ_2–semistable because the phases of $\mathcal{I}_\ell[2k]$ and $S(\mathcal{I}_\ell)[2k']$ cannot be equal, by (13). □

Proof of Theorem 8.1 As a cubic threefold has free Picard group of rank one, the first implication is obvious. As for the second implication, assume there is an exact equivalence $\Phi\colon \mathrm{Ku}(X_1) \to \mathrm{Ku}(X_2)$. Lemma 8.3 implies that, up to composing with a power of the Serre functor of $\mathrm{Ku}(X_1)$ and shift functor, we may assume $[\Phi_*(K_P)] = [K_{P'}]$ for points P and P' in X_1 and X_2, respectively. Take an S–invariant stability condition σ on $\mathrm{Ku}(X_1)$. Theorem 8.7 and Proposition 8.10 imply that

$$M_{X_1}(v) \cong M_\sigma(\mathrm{Ku}(X_1), [K_P]) \cong M_{\varphi\cdot\sigma}(\mathrm{Ku}(X_2), [K_{P'}]). \tag{16}$$

Since the Serre functor commutes with autoequivalences, $\varphi \cdot \sigma$ is an S–invariant stability condition on $\mathrm{Ku}(X_2)$. Thus, Theorem 8.7 gives

$$M_{\varphi\cdot\sigma}(\mathrm{Ku}(X_2), [K_{P'}]) \cong M_{X_2}(v).$$

Combining this with (16) gives $M_{X_1}(v) \cong M_{X_2}(v)$. By Lemma 7.5, we know X_1 and X_2 are the unique exceptional divisors of $M_{X_1}(v)$ and $M_{X_2}(v)$ which get contracted by any map to a complex abelian variety. Thus, $X_1 \cong X_2$. □

List of symbols

X	smooth cubic threefold in $\mathbb{P}^4$ over $\mathbb{C}$
H	the ample generator of $\mathrm{Pic}(X)$
Y	a hyperplane section of X
$\mathrm{D^b}(X)$	bounded derived category of coherent sheaves on X
$\mathrm{Ku}(X)$	the Kuznetsov component inside $\mathrm{D^b}(X)$
$\mathrm{CH}^*(X)$	the Chow ring of X
$\mathrm{CH}^*_{\mathrm{n}}(X)$	the numerical Chow ring of X, obtained as $\mathrm{CH}^*(X)$ modulo numerical equivalence
$\mathcal{H}^i(E)$	the i^{th} cohomology sheaf of a complex $E \in \mathrm{D^b}(X)$
$H^i(E)$	the i^{th} sheaf cohomology group of a complex $E \in \mathrm{D^b}(X)$
$\mathrm{ch}(E)$	total Chern character of an object $E \in \mathrm{D^b}(X)$ up to numerical equivalence
$c(E)$	total Chern class of an object $E \in \mathrm{D^b}(X)$ up to numerical equivalence
$\widetilde{\mathrm{ch}}(E)$	total Chern character of an object $E \in \mathrm{D^b}(X)$ up to rational equivalence
$\tilde{c}(E)$	total Chern class of an object $E \in \mathrm{D^b}(X)$ up to rational equivalence
$\mathrm{ch}_{\leq l}(E)$	$(\mathrm{ch}_0(E), \dots, \mathrm{ch}_l(E))$
$\widetilde{\mathrm{ch}}_{\leq l}(E)$	$(\widetilde{\mathrm{ch}}_0(E), \dots, \widetilde{\mathrm{ch}}_l(E))$

References

[1] **M Altavilla**, **M Petković**, **F Rota**, *Moduli spaces on the Kuznetsov component of Fano threefolds of index 2*, Épijournal Géom. Algébrique 6 (2022) art. id. 13 MR Zbl

[2] **D Arcara**, **A Bertram**, *Bridgeland-stable moduli spaces for K–trivial surfaces*, J. Eur. Math. Soc. 15 (2013) 1–38 MR Zbl

[3] **M Artebani**, **R Kloosterman**, **M Pacini**, *A new model for the theta divisor of the cubic threefold*, Matematiche (Catania) 58 (2003) 201–236 MR Zbl

[4] **M Artin**, *Algebraization of formal moduli, II: Existence of modifications*, Ann. of Math. 91 (1970) 88–135 MR Zbl

[5] **A Bayer**, **M Lahoz**, **E Macrì**, **P Stellari**, *Stability conditions on Kuznetsov components*, Ann. Sci. Éc. Norm. Supér. 56 (2023) 517–570 MR Zbl

[6] **A Bayer**, **E Macrì**, **P Stellari**, *The space of stability conditions on abelian threefolds, and on some Calabi–Yau threefolds*, Invent. Math. 206 (2016) 869–933 MR Zbl

[7] **A Bayer**, **E Macrì**, **Y Toda**, *Bridgeland stability conditions on threefolds, I: Bogomolov–Gieseker type inequalities*, J. Algebraic Geom. 23 (2014) 117–163 MR Zbl

[8] **A Beauville**, *Les singularités du diviseur Θ de la Jacobienne intermédiaire de l'hypersurface cubique dans* $\mathbb{P}^4$, from "Algebraic threefolds" (A Conte, editor), Lecture Notes in Math. 947, Springer (1982) 190–208 MR Zbl

[9] **A Beauville**, *Vector bundles on the cubic threefold*, from "Symposium in honor of C H Clemens" (A Bertram, J A Carlson, H Kley, editors), Contemp. Math. 312, Amer. Math. Soc., Providence, RI (2002) 71–86 MR Zbl

[10] **P Belmans**, **A J de Jong**, et al., *The Stacks project*, electronic reference (2005–) Available at `http://stacks.math.columbia.edu`

[11] **M Bernardara**, **E Macrì**, **S Mehrotra**, **P Stellari**, *A categorical invariant for cubic threefolds*, Adv. Math. 229 (2012) 770–803 MR Zbl

[12] **T Bridgeland**, *Stability conditions on triangulated categories*, Ann. of Math. 166 (2007) 317–345 MR Zbl

[13] **T Bridgeland**, *Stability conditions on $K3$ surfaces*, Duke Math. J. 141 (2008) 241–291 MR Zbl

[14] **C H Clemens**, **P A Griffiths**, *The intermediate Jacobian of the cubic threefold*, Ann. of Math. 95 (1972) 281–356 MR Zbl

[15] **I Coskun**, **J Huizenga**, *The ample cone of moduli spaces of sheaves on the plane*, Algebr. Geom. 3 (2016) 106–136 MR Zbl

[16] **L Ein**, **R Lazarsfeld**, *Singularities of theta divisors and the birational geometry of irregular varieties*, J. Amer. Math. Soc. 10 (1997) 243–258 MR Zbl

[17] **S Feyzbakhsh**, **L Pertusi**, *Serre-invariant stability conditions and Ulrich bundles on cubic threefolds*, Épijournal Géom. Algébrique 7 (2023) art. id. 1 MR Zbl

[18] **D Gieseker**, *On the moduli of vector bundles on an algebraic surface*, Ann. of Math. 106 (1977) 45–60 MR Zbl

[19] **P A Griffiths**, *Periods of integrals on algebraic manifolds, II: Local study of the period mapping*, Amer. J. Math. 90 (1968) 805–865 MR Zbl

[20] **D Happel**, **I Reiten**, **S O Smalø**, *Tilting in abelian categories and quasitilted algebras*, Mem. Amer. Math. Soc. 575, Amer. Math. Soc., Providence, RI (1996) MR Zbl

[21] **R Hartshorne**, *Algebraic geometry*, Graduate Texts in Math. 52, Springer (1977) MR Zbl

[22] **R Hartshorne**, *Generalized divisors on Gorenstein schemes*, K–Theory 8 (1994) 287–339 MR Zbl

[23] **D Huybrechts**, **M Lehn**, *The geometry of moduli spaces of sheaves*, 2nd edition, Cambridge Univ. Press (2010) MR Zbl

[24] **A Iliev**, *Minimal sections of conic bundles*, Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. 2 (1999) 401–428 MR Zbl

[25] **A G Kuznetsov**, *Derived category of a cubic threefold and the variety V_{14}*, Tr. Mat. Inst. Steklova 246 (2004) 183–207 MR In Russian: translated in Proc. Steklov Inst. Math. 246 (2004) 171–194

[26] **C Li**, *Stability conditions on Fano threefolds of Picard number 1*, J. Eur. Math. Soc. 21 (2019) 709–726 MR Zbl

[27] **J Lo**, **Y More**, *Some examples of tilt-stable objects on threefolds*, Comm. Algebra 44 (2016) 1280–1301 MR Zbl

[28] **A Maciocia**, *Computing the walls associated to Bridgeland stability conditions on projective surfaces*, Asian J. Math. 18 (2014) 263–279 MR Zbl

[29] **E Macrì**, **B Schmidt**, *Derived categories and the genus of space curves*, Algebr. Geom. 7 (2020) 153–191 MR Zbl

[30] **M Maruyama**, *Moduli of stable sheaves. I*, J. Math. Kyoto Univ. 17 (1977) 91–126 MR Zbl

[31] **M Maruyama**, *Moduli of stable sheaves. II*, J. Math. Kyoto Univ. 18 (1978) 557–614 MR Zbl

[32] **D Mumford**, *Projective invariants of projective structures and applications*, from "Proceedings of the International Congress of Mathematicians" (V Stenström, editor), Institut Mittag-Leffler, Djursholm (1963) 526–530 MR Zbl

[33] **D Mumford**, *Prym varieties, I*, from "Contributions to analysis: a collection of papers dedicated to Lipman Bers" (L V Ahlfors, I Kra, B Maskit, L Nirenberg, editors), Academic (1974) 325–350 MR Zbl

[34] **A Perry**, *The integral Hodge conjecture for two-dimensional Calabi–Yau categories*, Compos. Math. 158 (2022) 287–333 MR Zbl

[35] **L Pertusi**, **S Yang**, *Some remarks on Fano threefolds of index two and stability conditions*, Int. Math. Res. Not. 2022 (2022) 12940–12983 MR Zbl

[36] **K Schwede**, *Generalized divisors and reflexive sheaves*, lecture notes (2007) Available at `https://www.math.utah.edu/~schwede/Notes/GeneralizedDivisors.pdf`

[37] **C T Simpson**, *Moduli of representations of the fundamental group of a smooth projective variety, I*, Inst. Hautes Études Sci. Publ. Math. 79 (1994) 47–129 MR Zbl

[38] **F Takemoto**, *Stable vector bundles on algebraic surfaces*, Nagoya Math. J. 47 (1972) 29–48 MR Zbl

[39] **G E Welters**, *Abel–Jacobi isogenies for certain types of Fano threefolds*, Mathematical Centre Tracts 141, Mathematisch Centrum, Amsterdam (1981) MR Zbl

[40] **S Zhang**, *Bridgeland moduli spaces for Gushel–Mukai threefolds and Kuznetsov's Fano threefold conjecture*, preprint (2021) arXiv 2012.12193v2

AB, SVB: *School of Mathematics and Maxwell Institute, University of Edinburgh*
Edinburgh, United Kingdom

SF: *Department of Mathematics, Imperial College London*
London, United Kingdom

GH: *Fakultät für Mathematik, Universität Duisburg-Essen*
Essen, Germany

DM: *Korteweg-de Vries Institute for Mathematics, University of Amsterdam*
Amsterdam, Netherlands

FR: *Centre for Mathematical Sciences, University of Cambridge*
Cambridge, United Kingdom

FR: *ETH Zürich*
Zürich, Switzerland

BS: *Institut für Algebraische Geometrie, Gottfried Wilhelm Leibniz Universität Hannover*
Hannover, Germany

arend.bayer@ed.ac.uk, sjoerd.beentjes@ed.ac.uk, s.feyzbakhsh@imperial.ac.uk, georg.hein@uni-due.de, d.martinelli@uva.nl, fr414@cam.ac.uk, bschmidt@math.uni-hannover.de

Proposed: Mark Gross
Seconded: Richard P Thomas, Jim Bryan

Received: 21 January 2021
Revised: 19 February 2022

Geometry & Topology **28**:1 (2024) 161–266
DOI: 10.2140/gt.2024.28.161
Published: 27 February 2024

Coarse-median preserving automorphisms

ELIA FIORAVANTI

This paper has three main goals.

First, we study fixed subgroups of automorphisms of right-angled Artin and Coxeter groups. If φ is an untwisted automorphism of a RAAG, or an arbitrary automorphism of a RACG, we prove that $\mathrm{Fix}\,\varphi$ is finitely generated and undistorted. Up to replacing φ with a power, we show that $\mathrm{Fix}\,\varphi$ is quasiconvex with respect to the standard word metric. This implies that $\mathrm{Fix}\,\varphi$ is a virtual retract and a special group in the sense of Haglund and Wise.

By contrast, there exist "twisted" automorphisms of RAAGs for which $\mathrm{Fix}\,\varphi$ is undistorted but not of type F (hence not special), of type F but distorted, or even infinitely generated.

Secondly, we introduce the notion of "coarse-median preserving" automorphism of a coarse median group, which plays a key role in the above results. We show that automorphisms of RAAGs are coarse-median preserving if and only if they are untwisted. On the other hand, all automorphisms of Gromov-hyperbolic groups and right-angled Coxeter groups are coarse-median preserving. These facts also yield new or more elementary proofs of Nielsen realisation for RAAGs and RACGs.

Finally, we show that, for every special group G (in the sense of Haglund and Wise), every infinite-order, coarse-median preserving outer automorphism of G can be realised as a homothety of a finite-rank median space X equipped with a "moderate" isometric G–action. This generalises the classical result, due to Paulin, that every infinite-order outer automorphism of a hyperbolic group H projectively stabilises a small H–tree.

20F65, 20F67; 20E36, 20F28, 20F34, 20F36, 20F55

1 Introduction

This paper is inspired by the following, at first sight unrelated, questions.

Question 1 Given a finitely generated group G and $\varphi \in \operatorname{Aut} G$, what is the structure of the subgroup of fixed points $\operatorname{Fix} \varphi \leq G$?

Question 2 Given a finitely generated group G and $\varphi \in \operatorname{Aut} G$, when can we realise φ as a homothety of a nonpositively curved metric space X equipped with a "nice" G–action by isometries?

Our motivation comes from the theory of automorphisms of free groups. When $G = F_n$, a complete answer to Question 1 was first conjectured by Peter Scott in 1978, and later proved — after work by Dyer and Scott [46], Jaco and Shalen [72], Gersten [56; 57], Culler [36], Goldstein and Turner [58], Cooper [33] and Cohen and Lustig [32], among others — by Bestvina and Handel [12]:

> *For every* $\varphi \in \operatorname{Aut} F_n$*, the fixed subgroup* $\operatorname{Fix} \varphi \leq F_n$ *is generated by at most* n *elements.*

In particular, $\operatorname{Fix} \varphi$ is finitely generated, free, and quasiconvex in F_n.

Bestvina and Handel's proof is based on the extension of several ideas of Nielsen–Thurston theory from surfaces to graphs. Specifically, every homotopy equivalence between finite graphs is homotopic to a (*relative*) *train track map* [12; 11]. This result is also a key ingredient in providing the following answer to Question 2, by Gaboriau, Jaeger, Levitt and Lustig [53]:

> *For every* $\varphi \in \operatorname{Aut} F_n$*, there exists an action by homotheties* $F_n \rtimes_\varphi \mathbb{Z} \curvearrowright T$*, where* T *is an* $\mathbb{R}$*–tree and the restriction* $F_n \curvearrowright T$ *is isometric, minimal, and has trivial arc-stabilisers.*

If φ is exponentially growing, then $F_n \curvearrowright T$ has dense orbits and $\operatorname{Fix} \varphi$ is elliptic.

We are interested in Question 2 because of its connections to Question 1. Indeed, if one admits the existence of an F_n–tree as above, it is possible to give more elementary proofs of the Scott conjecture, which are completely independent of the complicated machinery of train tracks and instead rely on an "index theory" for F_n–trees; see Gaboriau, Levitt and Lustig [55] and Gaboriau and Levitt [54].

More generally, a satisfactory answer to Question 2 was obtained by Paulin [88] for all *Gromov-hyperbolic groups* G. If $\phi \in \operatorname{Out} G$ has infinite order, then it can be similarly realised as a homothety of a *small* G–tree, ie an $\mathbb{R}$–tree with a minimal isometric G–action such that no G–stabiliser of an arc contains a copy of the free group F_2.

Paulin's proof is abstract in nature, but his result can be pictured quite concretely in the case when $G = \pi_1(S)$ for a closed surface S: Thurston [97] showed that ϕ is induced by a homeomorphism of S that preserves a projective measured singular foliation on S; the $\mathbb{R}$–tree T can then be constructed by lifting this singular foliation to the universal cover $\widetilde{S}$ and considering its leaf space.

It is natural to wonder if the above discussion is specific to hyperbolic groups. This might be suggested by the fact that automorphism groups of one-ended hyperbolic groups can essentially be understood in terms of mapping class groups of finite-type surfaces (see Levitt [78] and Sela [93]), for which Nielsen–Thurston theory is available.

In recent years, the study of outer automorphisms of groups other than $\pi_1(S)$ and F_n has gained significant traction. The groups $\mathrm{Out}\,\mathcal{A}_\Gamma$ — where $\mathcal{A}_\Gamma$ is a right-angled Artin group (RAAG) — are particularly appealing in this context, as they can exhibit a variety of interesting behaviours ranging between the extremal cases of $\mathrm{Out}\,F_n$ and $\mathrm{Out}\,\mathbb{Z}^n = \mathrm{GL}_n\mathbb{Z}$.

One may look at the large body of work on $\mathrm{Out}\,F_n$ hoping to extract a blueprint that will direct the study of the groups $\mathrm{Out}\,\mathcal{A}_\Gamma$. This has proved a successful approach in some cases, remarkably with the definition of analogues of Outer Space (see Bregman, Charney and Vogtmann [20] and Charney, Stambaugh and Vogtmann [25]) and its consequences for the study of homological properties. However, there are limits to such analogies: in practice, techniques that are tailored to general RAAGs and based on induction on the complexity of the graph Γ seem to provide the most effective approach to many problems; see for instance Charney and Vogtmann [27; 28], Day and Wade [43], Day, Sale and Wade [42] and Guirardel and Sale [61].

Our aim is to investigate Questions 1 and 2 when G is a RAAG or, more generally, a cocompactly cubulated group. These are just two of the many basic questions that have been fully solved for $\mathrm{Out}\,F_n$, but have so far remained out of the limelight for the groups $\mathrm{Out}\,\mathcal{A}_\Gamma$.

One quickly realises that it is necessary to impose some restrictions on $\varphi \in \mathrm{Aut}\,\mathcal{A}_\Gamma$ if the two questions are to be fruitfully addressed. To begin with, it is not hard to construct automorphisms of $F_2 \times \mathbb{Z}$ whose fixed subgroup is infinitely generated (Example 4.13), which would prevent us from relying on the tools of geometric group theory in relation to Question 1. In addition, when $G = \mathbb{Z}^n$, it should heuristically always be possible to equivariantly collapse the space X in Question 2 to a copy of $\mathbb{R}$, which forces $\varphi \in \mathrm{GL}_n\mathbb{Z}$ to have a positive eigenvalue.

We choose to consider the subgroup of *untwisted automorphisms* $U(\mathcal{A}_\Gamma) \le \mathrm{Aut}\,\mathcal{A}_\Gamma$, which was introduced by Day in [41] (with the name of "long-range automorphisms") and further studied by Charney, Stambaugh and Vogtmann [25] and Hensel and Kielak [69]. This can be defined as the subgroup generated by a certain subset of the Laurence–Servatius generators for $\mathrm{Aut}\,\mathcal{A}_\Gamma$ (see Laurence [75] and Servatius [94]), excluding generators that "resemble" too closely elements of $\mathrm{GL}_n\mathbb{Z}$.

The subgroup $U(\mathcal{A}_\Gamma) \le \mathrm{Aut}\,\mathcal{A}_\Gamma$ displays stronger similarities to $\mathrm{Aut}\,F_n$ and often makes up a large portion of the entire group $\mathrm{Aut}\,\mathcal{A}_\Gamma$. For instance, $U(F_n) = \mathrm{Aut}\,F_n$ and $U(\mathcal{A}_\Gamma)$ always contains the kernel of the homomorphism $\mathrm{Aut}\,\mathcal{A}_\Gamma \to \mathrm{GL}_n\mathbb{Z}$ induced by the $(\mathrm{Aut}\,\mathcal{A}_\Gamma)$–action on the abelianisation of $\mathcal{A}_\Gamma$.

Our first result is a novel, *coarse geometric* characterisation of untwisted automorphisms. This will play a fundamental role in addressing both Questions 1 and 2 in the rest of the paper.

Recall that every right-angled Artin group $\mathcal{A}_\Gamma$ is equipped with a median operator $\mu\colon \mathcal{A}_\Gamma^3 \to \mathcal{A}_\Gamma$ coming from the fact that $\mathcal{A}_\Gamma$ is naturally identified with the 0–skeleton of a CAT(0) cube complex (the universal cover of its Salvetti complex); see Chepoi [31]. Thus, one can consider those automorphisms of $\mathcal{A}_\Gamma$ with respect to which μ is coarsely equivariant.

More generally, it makes sense to study such automorphisms for any *coarse median group* (G,μ). This remarkably broad class of groups was introduced by Bowditch in [15] and contains all Gromov-hyperbolic groups, as well as all groups admitting a geometric action on a CAT(0) cube complex, and all *hierarchically hyperbolic groups* in the sense of Behrstock, Hagen and Sisto [6, Definition 1.21].

Definition An automorphism φ of a coarse median group (G,μ) is *coarse-median preserving*[1] (*CMP*) if there exists a constant $C \geq 0$ such that

$$\varphi(\mu(g_1,g_2,g_3)) \approx_C \mu(\varphi(g_1),\varphi(g_2),\varphi(g_3)) \quad \text{for all } g_1,g_2,g_3 \in G,$$

where $x \approx_C y$ means $d(x,y) \leq C$ with respect to some fixed word metric d on G.

It is easy to see that CMP automorphisms form a subgroup of $\operatorname{Aut} G$ containing all inner automorphisms.[2] Thus, it makes sense to speak of CMP *outer* automorphisms, as this property does not depend on the specific lift to $\operatorname{Aut} G$.

It turns out that, in the setting of right-angled Artin groups, CMP automorphisms coincide with untwisted automorphisms, perhaps explaining the closer analogy between $U(\mathcal{A}_\Gamma)$ and $\operatorname{Aut} F_n$. In particular, every element of $\operatorname{Aut} F_n$ is CMP, while only a finite subgroup of $\operatorname{Aut} \mathbb{Z}^n$ is CMP.

More precisely, we have the following. We endow right-angled Artin/Coxeter groups with the coarse median structure induced by the action on the universal cover of the Salvetti/Davis complex.

Proposition A (1) *All automorphisms of hyperbolic groups are CMP.*

(2) *All automorphisms of right-angled Coxeter groups are CMP.*

(3) *Automorphisms of right-angled Artin groups are CMP if and only if they are untwisted.*

Part (1) is due to the fact that hyperbolic groups admit a unique coarse median structure, which was shown in [83]; see Example 2.28 below. That CMP automorphisms of RAAGs are untwisted can be easily deduced from the proof, due to Laurence [75], that elementary automorphisms generate the automorphism group. We prove the rest of Proposition A in Section 3.4.

[1] This terminology is motivated in Section 2.6; see Remark 2.25.

[2] Here it is important that our definition of *coarse median group* (Definition 2.24) is slightly stronger than Bowditch's original definition [15], in that we require μ to be coarsely G–equivariant. The difference between the two notions is analogous to the distinction between *hierarchically hyperbolic groups* and groups that are just a *hierarchically hyperbolic space*.

Our first result on Question 1 applies to all CMP automorphisms of *cocompactly cubulated* groups, ie those groups that admit a proper cocompact action on a CAT(0) cube complex.

We remark that, in addition to Proposition A, examples of CMP automorphisms of cubulated groups are provided by [52, Theorem E], which characterises when a generalised Dehn twist preserves the coarse median structure induced by the cubulation.

Theorem B *Let G be a cocompactly cubulated group, with the induced coarse median structure. If $\varphi \in \operatorname{Aut} G$ is coarse-median preserving, then:*

(1) $\operatorname{Fix}\varphi$ *is finitely generated and undistorted in G.*

(2) $\operatorname{Fix}\varphi$ *is itself cocompactly cubulated.*

Both parts of this result fail badly for "twisted" automorphisms of right-angled Artin groups. For every finite graph Γ, there exist automorphisms $\psi \in \operatorname{Aut}(\mathcal{A}_\Gamma \times \mathbb{Z})$ with $\operatorname{Fix}\psi = BB_\Gamma \times \mathbb{Z}$, where $BB_\Gamma \leq \mathcal{A}_\Gamma$ denotes the Bestvina–Brady subgroup [8]; see Example 4.13. When finitely generated, BB_Γ is quadratically distorted in $\mathcal{A}_\Gamma$ as soon as $\mathcal{A}_\Gamma$ is directly irreducible and noncyclic; see Tran [98]. Even when $\operatorname{Fix}\psi$ is finitely generated and undistorted, one can ensure that $\operatorname{Fix}\psi$ not be of type F, which implies that $\operatorname{Fix}\psi$ is not cocompactly cubulated. These examples can be easily extended to RAAGs that do not split as products.

We emphasise that the cubulation of $\operatorname{Fix}\varphi$ provided by Theorem B does not arise from a *convex* subcomplex of the cubulation of G in general, but just from a *median subalgebra* of it; see Section 2.2 for a definition. In fact, the subgroup $\operatorname{Fix}\varphi$ need not be *quasiconvex* in G, as can be observed for the automorphism $\varphi \in \operatorname{Aut}\mathbb{Z}^2$ that swaps the standard generators, where $\operatorname{Fix}\varphi$ is the diagonal subgroup of $\mathbb{Z}^2$.

Nevertheless, in many situations, $\operatorname{Fix}\varphi$ does turn out to be quasiconvex in the ambient group. We prove this fact in the context of right-angled Artin and Coxeter groups, where it has the remarkable consequence that $\operatorname{Fix}\varphi$ is a retract of a finite-index subgroup of the ambient group; see Haglund and Wise [68, Section 6].

Theorem C *Consider the right-angled Artin group $\mathcal{A}_\Gamma$ or the right-angled Coxeter group $\mathcal{W}_\Gamma$. There are finite-index subgroups $U_0(\mathcal{A}_\Gamma) \leq U(\mathcal{A}_\Gamma)$ and $\operatorname{Aut}_0 \mathcal{W}_\Gamma \leq \operatorname{Aut}\mathcal{W}_\Gamma$ such that, for any automorphism φ lying in either of these subgroups:*

(1) $\operatorname{Fix}\varphi$ *is quasiconvex in $\mathcal{A}_\Gamma$ or $\mathcal{W}_\Gamma$ with respect to their standard word metric, ie geodesics in their standard Cayley graph with endpoints in* $\operatorname{Fix}\varphi$ *stay uniformly close to* $\operatorname{Fix}\varphi$.

(2) *In particular,* $\operatorname{Fix}\varphi$ *is a virtual retract and it is a special group in the Haglund–Wise sense.*

For the experts, the finite-index subgroups in Theorem C are generated by the elementary automorphisms known as inversions, folds and partial conjugations; see Section 3.4 and Remark 3.27. Quasiconvexity of $\operatorname{Fix}\varphi$ can alternatively be characterised saying that $\operatorname{Fix}\varphi$ acts properly and cocompactly on a convex subcomplex of the universal cover of the Salvetti/Davis complex, or, again, in coarse median terms; see Definition 2.30, Remark 2.31 and Lemma 3.2.

In light of Theorem C, it is only natural to wonder what isomorphism types of special groups can arise as $\mathrm{Fix}\,\varphi$, and whether their complexity can be bounded in any way in terms of the ambient group, in the spirit of Scott's conjecture. We only provide a very partial result on these questions (Corollary E), leaving a more detailed treatment for later work. The main proof ingredient, which we believe is of independent interest, is the following construction of $U_0(\mathcal{A}_\Gamma)$–invariant Bass–Serre trees for most right-angled Artin groups.

Proposition D *Let $\mathcal{A}_\Gamma$ be directly irreducible, freely irreducible and noncyclic. Then there exists an amalgamated product splitting $\mathcal{A}_\Gamma = \mathcal{A}_+ *_{\mathcal{A}_0} \mathcal{A}_-$, with $\mathcal{A}_\pm$ and $\mathcal{A}_0$ parabolic subgroups of $\mathcal{A}_\Gamma$, such that the corresponding Bass–Serre tree $\mathcal{A}_\Gamma \curvearrowright T$ is $U_0(\mathcal{A}_\Gamma)$–invariant. That is: for every $\varphi \in U_0(\mathcal{A}_\Gamma)$, there exists an isometry $f : T \to T$ satisfying $f \circ g = \varphi(g) \circ f$ for all $g \in \mathcal{A}_\Gamma$.*

Corollary E *Consider a right-angled Artin group $\mathcal{A}_\Gamma$ and $\varphi \in U_0(\mathcal{A}_\Gamma)$.*

(1) *If $\mathcal{A}_\Gamma$ splits as a direct product $\mathcal{A}_1 \times \mathcal{A}_2$, then $\varphi(\mathcal{A}_i) = \mathcal{A}_i$ and $\mathrm{Fix}\,\varphi = \mathrm{Fix}\,\varphi|_{\mathcal{A}_1} \times \mathrm{Fix}\,\varphi|_{\mathcal{A}_2}$.*

(2) *If $\mathcal{A}_\Gamma$ is directly irreducible, then the subgroup $\mathrm{Fix}\,\varphi \le \mathcal{A}_\Gamma$ splits as a (possibly trivial) finite graph of groups with vertex and edge groups of the form $\mathrm{Fix}\,\varphi|_P$, for proper parabolic subgroups $P \le \mathcal{A}_\Gamma$ with $\varphi(P) = P$ and $\varphi|_P \in U_0(P)$.*

The same two results hold for right-angled Coxeter groups $\mathcal{W}_\Gamma$ and automorphisms $\varphi \in \mathrm{Aut}_0\,\mathcal{W}_\Gamma$.

We now turn to Question 2, which is the second main focus of the paper. Recall that Paulin [88] showed that, for every Gromov-hyperbolic group G, every infinite-order element of $\mathrm{Out}\,G$ can be realised as a homothety of a small, isometric G–tree.

Our main result on Question 2, generalises Paulin's theorem to CMP automorphisms of *special groups G*, in the Haglund–Wise sense [68; 90]. This is a broad class of groups including right-angled Artin groups, finite-index subgroups of right-angled Coxeter groups, as well as free and surface groups and a number of other hyperbolic examples.

Note that *small* G–actions on $\mathbb{R}$–trees are not the right notion to consider in this context. Indeed, if a special group G has a small action on an $\mathbb{R}$–tree T, then every arc stabiliser is free abelian and the work of Rips and Bestvina–Feighn implies that G splits over an abelian subgroup; see Bestvina and Feighn [10, Theorem 9.5]. However, there exist special groups that admit an infinite-order CMP outer automorphism, but do not split over any abelian subgroup (eg the RAAG $\mathcal{A}_\Gamma$ with Γ as in Figure 1, by Groves and Hull [59]).

In fact, due to the lack of hyperbolicity, it is reasonable to expect that $\mathbb{R}$–trees will need to be replaced by higher-dimensional analogues.

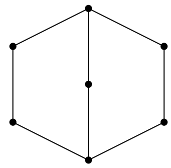


Figure 1

The correct setting seems to be provided by the simultaneous generalisation of $\mathbb{R}$–trees and CAT(0) cube complexes known as *median spaces*. These are those metric spaces (X,d) such that, for all $x_1,x_2,x_3\in X$, there exists a unique point $m(x_1,x_2,x_3)$ (known as their *median*) satisfying

$$d(x_i,x_j)=d(x_i,m(x_1,x_2,x_3))+d(m(x_1,x_2,x_3),x_j) \quad \text{for all } 1\le i<j\le 3.$$

A connected median space X is said to have *rank* $\le r$ if all its locally compact subsets have topological dimension $\le r$. Rank-1 connected median spaces are precisely $\mathbb{R}$–trees.

The following is our main result on Question 2 (a more general statement for infinite abelian subgroups of $\operatorname{Out} G$ is Theorem 7.25). Note that, although higher-rank median spaces are never nonpositively curved, they always admit a canonical, bi-Lipschitz equivalent CAT(0) metric;[3] see Bowditch [17].

Theorem F *Let G be the fundamental group of a compact special cube complex. Suppose G has trivial centre. Let $\phi\in\operatorname{Out} G$ be infinite-order and coarse-median preserving. Then*:

(1) *There is a geodesic, finite-rank median space X and an action by homotheties $G\rtimes_\phi\mathbb{Z}\curvearrowright X$.*
(2) *The restriction $G\curvearrowright X$ is isometric, minimal, with unbounded orbits, and "moderate".*
(3) *If $\varphi\in\operatorname{Aut} G$ represents ϕ, then the subgroup $\operatorname{Fix}\varphi\le G$ fixes a point of X.*
(4) *If ϕ and ϕ^{-1} are subexponentially growing, then the action $G\rtimes_\phi\mathbb{Z}\curvearrowright X$ is isometric.*

As for actions on $\mathbb{R}$–trees, we say that $G\curvearrowright X$ is *minimal* if X does not contain any proper, G–invariant convex subsets. We propose the notion of "moderate" action on a median space as a higher-rank generalisation of the notion of small action on an $\mathbb{R}$–tree.

Definition (moderate actions) Let G be a group and X be a median space.

(1) A *k–cube* in X is a median subalgebra $C\subseteq X$ isomorphic to the product $\{0,1\}^k$.
(2) An isometric action $G\curvearrowright X$ is *moderate* if, for every $k\ge 1$ and every k–cube $C\subseteq X$, the subgroup of G fixing C pointwise contains a copy of $\mathbb{Z}^k$ in its centraliser.

Any 2–element subset of X is a 1–cube. Thus, if G is hyperbolic and $G\curvearrowright X$ is moderate, the intersection of any two point-stabilisers must be virtually cyclic. In particular, if G is torsionfree hyperbolic and T is an $\mathbb{R}$–tree, then the action $G\curvearrowright T$ is moderate if and only if it is small. We remark that, when G is hyperbolic, the space X provided by Theorem F is indeed an $\mathbb{R}$–tree.

[3] The reader should keep in mind the case of $\mathbb{R}^n$, where the ℓ^1 metric is median and the Euclidean metric is CAT(0).

We would like to emphasise that Theorem F does not provide any *lower* bounds to the rank of the median space X. In particular, we still do not have an answer to the following:

Question 3 (1) Can we always take the median space X in Theorem F to be an $\mathbb{R}$–tree?

(2) If G is a directly and freely irreducible RAAG, can we even take X to be a simplicial tree?

We have seen that, when $\mathcal{A}_\Gamma$ is directly and freely irreducible, Proposition D yields a $U_0(\mathcal{A}_\Gamma)$–invariant simplicial $\mathcal{A}_\Gamma$–tree. However, it remains unclear if such a simplicial tree can always be taken to be moderate and, more importantly, if it can be constructed so that $\operatorname{Fix}\varphi$ is elliptic.

We conclude this overview by highlighting two more results. These fall outside the main purpose of this text, but they are almost immediate consequences of the techniques used in this paper and we find them of independent interest. We prove them at the end of Section 4.2.

Recall that the property of being cocompactly cubulated does not, in general, pass to finite-index *overgroups*. Many examples of this are provided by crystallographic groups (see Hagen [62]): for instance, the $(3,3,3)$ triangle group has $\mathbb{Z}^2$ as a finite-index subgroup, but it is not itself cocompactly cubulated.

The following is a criterion for cubulating finite-index overgroups. Its proof is loosely inspired by the idea of Guirardel cores (see Guirardel [60] and Hagen and Wilton [65]), but it requires none of the technical machinery. Instead, it is a simple consequence of Proposition 4.1 (or the earlier result of Bowditch [18, Proposition 4.1]).

Corollary G *Let G be a group with a cocompactly cubulated finite-index subgroup H. Suppose that the coarse median structure on G induced by the cubulation of H is G–invariant (it is automatically H–invariant). Then G is cocompactly cubulated.*

Along with Proposition A, the previous corollary implies the following version of Nielsen realisation for automorphisms of right-angled Artin and Coxeter groups.

Corollary H (Nielsen realisation for RA$*$Gs) *Consider one of the following two settings:*

(1) *A centreless right-angled Artin group $G=\mathcal{A}_\Gamma$ and a finite subgroup $F\leq \operatorname{Out}\mathcal{A}_\Gamma$ contained in the projection to outer automorphisms of the untwisted subgroup $U(\mathcal{A}_\Gamma)\leq \operatorname{Aut}\mathcal{A}_\Gamma$.*

(2) *A centreless right-angled Coxeter group $G=\mathcal{W}_\Gamma$ and any finite subgroup $F\leq \operatorname{Out}\mathcal{W}_\Gamma$.*

In either case, F can be realised as a group of automorphisms of a compact, nonpositively curved, cube (orbi)complex Q with $G=\pi_1 Q$.

Part (2) is new, while part (1) is originally due to Hensel and Kielak [69]. When $F\leq U_0(\mathcal{A}_\Gamma)$, they constructed Q quite explicitly via a glueing construction, ensuring that $\dim Q=\dim \mathcal{X}_\Gamma$. By comparison, our approach does not offer much control on dimension (except $\dim Q\leq \#F\cdot\dim\mathcal{X}_\Gamma$), but it provides a much more elementary proof of the existence of some Q.

We expect our complex Q to be special, but this would require additional arguments in the proof (the only delicate point being lack of interosculations). We also think it should be possible to "trim" Q into having the optimal dimension $\dim \mathcal{X}_\Gamma$ by relying on the "panel collapse" procedure of Hagen and Touikan [64] (or small variations thereof), but the details seem too technical to be discussed here.

1.1 On the proof of Theorems B and C

The two theorems are proved in Section 4 under the aliases of Theorem 4.10 and Corollaries 4.34 and 4.35.

Regarding Theorem B, the starting observation is that $\operatorname{Fix}\varphi$ is an *approximate median subalgebra* of the group G; see Definition 2.33 and Lemma 2.35. Fixing a proper cocompact action on a CAT(0) cube complex $G \curvearrowright \mathcal{Z}$, the proof then takes place in three steps.

(1) If a subgroup $H \leq G$ is an approximate median subalgebra, H is finitely generated (Proposition 4.11). We prove this by relying on a straightforward adaptation of an argument due to Paulin [86] in the context of hyperbolic groups. Paulin's argument is itself a generalisation of Cooper's proof [33] in the case when the group G is free (a result originally due to Gersten [57] from the early 80s).

(2) Approximate median subalgebras of CAT(0) cube complexes are always at finite Hausdorff distance from actual median subalgebras (Proposition 4.1 or [18, Proposition 4.1]).

(3) Applying the previous step to H–orbits in $\mathcal{Z}$, we obtain an H–invariant median subalgebra $M \subseteq \mathcal{Z}^{(0)}$ such that $H \curvearrowright M$ is cofinite. Along with the fact that H is finitely generated, this yields a cocompact cubulation that quasi-isometrically embeds into $\mathcal{Z}$ (Lemma 4.12), though not necessarily as a convex subcomplex.

A similar strategy gives a new proof of W Neumann's result [82] that fixed subgroups of automorphisms of hyperbolic groups are quasiconvex; see also Minasyan and Osin [81]. Indeed, recall that, although not all hyperbolic groups are cocompactly cubulated, they are all coarse median, and all their automorphisms φ are CMP by Proposition A. It is easy to see that all coarsely connected, approximate median subalgebras of hyperbolic spaces are quasiconvex. As above, this implies that $\operatorname{Fix}\varphi$ is quasiconvex.

When dealing with nonhyperbolic groups, quasiconvexity is significantly harder to ensure and the proof of Theorem C requires additional work. Namely, assuming that $\varphi \in U_0(\mathcal{A}_\Gamma)$ or $\varphi \in \operatorname{Aut}_0 \mathcal{W}_\Gamma$, we need to show that $(\operatorname{Fix}\varphi)$–orbits in the Salvetti complex $\mathcal{X}_\Gamma$ or Davis complex $\mathcal{Y}_\Gamma$ are quasiconvex (in the coarse median sense; see Definition 2.30, Remark 2.31 and Lemma 3.2).

The proof of this is based on a quasiconvexity criterion for median subalgebras of CAT(0) cube complexes (Proposition 4.25). The most important ingredients are the fact that $\mathcal{X}_\Gamma$ and $\mathcal{Y}_\Gamma$ do not contain "infinite staircases" (Section 4.3), and certain properties that distinguish elements of $U_0(\mathcal{A}_\Gamma)$ and $\operatorname{Aut}_0 \mathcal{W}_\Gamma$ from more general CMP automorphisms in $U(\mathcal{A}_\Gamma)$ and $\operatorname{Aut} \mathcal{W}_\Gamma$ (Lemmas 4.30 and 4.32).

We conclude by mentioning that other important tools for the study of undistortion and quasiconvexity of subgroups of cubulated groups were recently developed by Beeker and Lazarovich in [2] and [3, Theorem 1.2(2)], and Dani and Levcovitz [38, Theorem A], based on extensions of the classical machinery of *Stallings folds* [95; 96] from graphs to higher-dimensional cube complexes. These techniques play no role in our arguments, but it is possible that they can be used to give alternative proofs of certain special cases of Theorems B and C.

1.2 On the proof of Theorem F

Keeping the case of Out F_n in mind, as described eg in Gaboriau, Jaeger, Levitt and Lustig [53, Section 2], there are two main obstacles to overcome:

(a) No good analogue of (*relative*) *train track maps* is available to represent homotopy equivalences between nonpositively curved cube complexes.

(b) It is not known if (isometric) actions on finite-rank median spaces are completely determined by their length function. There are results of this type for actions on $\mathbb{R}$–trees (see Culler and Morgan [37]) and cube complexes (see Beyrer and Fioravanti [13; 14]), but their extension to a general median setting would require some significantly new ideas.

The proof of Theorem F is made up of two main steps, which we now describe. In this sketch, we restrict our attention to the construction of the homothetic action $G \rtimes_\phi \mathbb{Z} \curvearrowright X$ (parts (1) and (2) of the theorem). Parts (3) and (4) follow, respectively, from parts (1) and (2) of Remark 7.27.

Let G be a special group, let $\mathcal{Z}$ be a CAT(0) cube complex, and let $\rho\colon G \to \operatorname{Aut}\mathcal{Z}$ be the homomorphism corresponding to a proper, cocompact, cospecial action $G \curvearrowright \mathcal{Z}$. Equip G with the coarse median structure arising from $\mathcal{Z}$. Let $\varphi \in \operatorname{Aut} G$ be a coarse-median preserving automorphism projecting to an infinite-order element of $\operatorname{Out} G$.

Step 1 *There exist a finite-rank median space X, an isometric action $G \curvearrowright X$ with unbounded orbits, and a homeomorphism $H\colon X \to X$ satisfying $H \circ g = \varphi(g) \circ H$ for all $g \in G$.*

In order to prove this, we consider the sequence of homomorphisms $\rho_n := \rho \circ \varphi^n$ and the sequence of G–actions on cube complexes $G \curvearrowright \mathcal{Z}_n$ that they induce. We then fix a nonprincipal ultrafilter ω, choose basepoints $p_n \in \mathcal{Z}_n$ and scaling factors $\epsilon_n > 0$, and consider the ultralimit

$$(X, p) := \lim_{\omega}(\epsilon_n \mathcal{Z}_n, p_n).$$

This is easily seen to be a finite-rank median space and, for a suitable choice of p_n and λ_n, the actions $G \curvearrowright \mathcal{Z}_n$ converge to an isometric action $G \curvearrowright X$ with unbounded orbits.

So far this is just a classical Bestvina–Paulin construction; see Bestvina [7] and Paulin [85]. The actual subtleties lie in the definition of the map $H\colon X \to X$. By the Milnor–Schwarz lemma, there exists a quasi-isometry $h\colon \mathcal{Z} \to \mathcal{Z}$ satisfying $h \circ g = \varphi(g) \circ h$ for all $g \in G$. We would like to define H as the ultralimit of the corresponding sequence of quasi-isometries $\mathcal{Z}_n \to \mathcal{Z}_n$, but this might displace the basepoint $p \in X$ by an infinite amount.

In order to rule out this eventuality, we rely on an argument similar to the one used in Paulin [88] for hyperbolic groups. On closer inspection, Paulin's argument only requires the following property, which is satisfied by nonelementary hyperbolic groups.

Definition Let G be a infinite group with a (fixed) Cayley graph $(\mathcal{G}, d)$. We say that G is *uniformly nonelementary* (*UNE*) if there exists a constant $c > 0$ with the following property. For every finite generating set $S \subseteq G$ and for all $x, y \in \mathcal{G}$, we have

$$d(x, y) \le c \cdot \max_{s \in S}[d(x, sx) + d(y, sy)].$$

The important part of this definition is that the constant c *does not* depend on the generating set S. Note that the UNE property is independent of the specific choice of $\mathcal{G}$; cf Definition 2.36.

Our main contribution to Step 1 is the proof of the following fact (Corollary 7.23), which is potentially of independent interest.

Theorem I *Let G be the fundamental group of a compact special cube complex. If G has trivial centre, then G is uniformly nonelementary.*

Now, let $m\colon X^3 \to X$ denote the median operator of the median space X. The fact that $\varphi \in \operatorname{Aut} G$ is coarse-median preserving easily implies that the homeomorphism $H : X \to X$ arising from the above construction satisfies $H(m(x, y, z)) = m(H(x), H(y), H(z))$ for all $x, y, z \in X$. However, H need not be a homothety at this stage.

Step 2 *There exists a G–invariant (pseudo)metric $\eta\colon X \times X \to [0, +\infty)$ such that (X, η) is a median space with the same median operator m, and H is a homothety with respect to η.*

Since $H : X \to X$ preserves the median operator m, there is an action of H on the space of all G–invariant median pseudometrics on X that induce m. More precisely, we show that H gives a homeomorphism of a certain space of (projectivised) median pseudometrics on X, and that the latter is a compact *absolute retract* (*AR*). The existence of the required pseudometric η then follows from the Lefschetz fixed point theorem for homeomorphisms of compact ANRs. This is discussed mainly in Sections 6.2 and 7.4; see especially Corollaries 6.23 and 7.24.

Once the pseudometric η is obtained, we can pass to the quotient metric space to obtain a genuine median space.

1.3 Further questions

We would like to highlight four questions raised by our results.

As mentioned earlier, every hyperbolic group admits a unique *coarse median structure* (Definition 2.22). At the opposite end of the spectrum, any RAAG for which $U(\mathcal{A}_\Gamma)$ has infinite index in $\operatorname{Aut}\mathcal{A}_\Gamma$ will admit infinitely many $\mathcal{A}_\Gamma$–invariant coarse median structures.

Right-angled Coxeter groups $\mathcal{W}_\Gamma$ seem to place themselves in between these two extremal situations: they can admit infinitely many distinct coarse median structures — eg because every RAAG is a finite-index subgroup of a RACG; see Davis and Januszkiewicz [40] — but it is not clear which of these structures are $\mathcal{W}_\Gamma$–invariant. For instance, Proposition A(2) implies that all Coxeter generating sets of $\mathcal{W}_\Gamma$ give rise to the same coarse median structure (which fails for Artin generating sets of $\mathcal{A}_\Gamma$).

Question 4 Does each RACG $\mathcal{W}_\Gamma$ have only *finitely many* $\mathcal{W}_\Gamma$–invariant coarse median structures?

As an example of why one might expect this kind of rigidity, we suggest looking at the difference between the RAAG $\mathbb{Z}^n$ and the RACG $(D_\infty)^n$, where D_∞ is the infinite dihedral group. The space of $\mathbb{Z}^n$–invariant coarse median structures on $\mathbb{Z}^n$ (equivalently, on $\mathbb{R}^n$) is uncountable, simply because it is endowed with a natural $GL_n\mathbb{R}$–action and we can consider the orbit of the standard structure. However, of the structures in this orbit, only finitely many are $(D_\infty)^n$–invariant.

The second question naturally arises from Theorem C and was already mentioned above:

Question 5 Consider $\varphi \in U_0(\mathcal{A}_\Gamma)$ or $\varphi \in \operatorname{Aut}_0 \mathcal{W}_\Gamma$.

(1) What isomorphism types of special groups can arise as $\operatorname{Fix}\varphi$ for some choice of φ and Γ? When $\varphi \in U_0(\mathcal{A}_\Gamma)$, is $\operatorname{Fix}\varphi$ itself a right-angled Artin group?

(2) Can we bound the "complexity" of $\operatorname{Fix}\varphi$ in terms of $\#\Gamma^{(0)}$, in the spirit of Scott's conjecture?

Regarding part (1) of Question 5, note that every RAAG can arise as the fixed subgroup of some element of $U_0(\mathcal{A}_\Gamma)$, simply because we can always take $\varphi = \mathrm{id}$. One can easily construct more elaborate examples using this observation as a starting point.

One can also wonder about fixed subgroups of automorphisms of general coarse median groups G. By Lemma 2.35, this reduces to understanding subgroups that are *approximate median subalgebras* (Definition 2.33). We study these subgroups when G is cocompactly cubulated (Theorem 4.10), but some of our arguments should work more generally (especially the proof of Proposition 4.11).

Question 6 Let (G, μ) be a finite-rank coarse median group. Let a subgroup $H \leq G$ be an approximate median subalgebra.

(1) Is H finitely generated?

(2) Is H undistorted? Which properties of G does H retain?

For instance, when G is hierarchically hyperbolic, I do not know if H must be finitely generated. However, assuming that it is, the second part of the question has a positive answer: H is undistorted and hierarchically hyperbolic. This is evident from Bowditch's axioms (B1)–(B10) for (weak) hierarchically hyperbolic spaces [18, Section 7] and the coarse median characterisation of hierarchy paths [18, Theorem 1.1].

We emphasise that our definition of *coarse median group* (Definition 2.24) is slightly stronger than Bowditch's original definition [15], in that we require μ to be coarsely G–equivariant.

Our last question regards UNE groups. It is clear that UNE groups have finite centre, and it is not hard to show that nonelementary hyperbolic groups are UNE. All other examples of UNE groups that we are aware of are provided by Theorem I.

Are there other interesting examples or nonexamples of UNE groups? Given the proof of Theorem I, a positive answer to the following seems likely:

Question 7 Are hierarchically hyperbolic groups with finite centre UNE?

Outline of the paper

Section 2 mostly contains background material on median algebras, cube complexes and coarse median groups. An exception is Section 2.4, which reviews some of the results of [51]. The latter will be helpful, mostly in Sections 6 and 7, for some of the more technical arguments in the proof of Theorem F.

In Section 3, we consider cocompactly cubulated groups G and study a notion of *convex-cocompactness* for subgroups of G, which is a special instance of quasiconvexity in coarse median spaces (Definition 2.30). Section 3.2 studies cyclic, convex-cocompact subgroups of RAAGs (whose generators we call *label-irreducible*). Section 3.4 contains the proof of Proposition A.

Section 4 is concerned with fixed subgroups of CMP automorphisms. First, Sections 4.1 and 4.2 are devoted to the proof of Theorem B. Then Section 4.3 studies staircases in cube complexes, allowing us to formulate a quasiconvexity criterion for median subalgebras in Section 4.4. Finally, Section 4.5 restricts to Salvetti and Davis complexes, proving Theorem C.

Section 5 is completely independent from the subsequent part of the paper and can be safely skipped. It only contains the proof of Proposition D and Corollary E.

Finally, Sections 6 and 7 are the most technical parts of the paper and they contain the bulk of the proof of Theorem F. In Section 6, we consider group actions on finite-rank median algebras and develop a criterion for the existence of a (projectively) invariant metric (as required for Step 2 of the proof sketch for Theorem F). In Section 7, we study ultralimits of actions on Salvetti complexes, in order to obtain the properties needed to apply the results of Section 6. Theorems F and I are proved in Section 7.4.

Acknowledgements I would like to thank Mark Hagen for noticing that a result similar to Proposition 4.1 already appeared in [18], Camille Horbez for mentioning that ARs have good fixed-point properties, Robert Kropholler for pointing me to [98] and [62], Alessandro Sisto for discussions related to Definition 2.36, and Ric Wade for suggesting a major simplification to my original argument for Proposition A in the case of RAAGs. I thank Jason Behrstock, Corey Bregman, Ruth Charney, Ashot Minasyan and Karen Vogtmann for other interesting conversations, and the referee for many useful comments.

I am grateful to Ursula Hamenstädt and the Max Planck Institute for Mathematics in Bonn for their hospitality and financial support while part of this work was being completed.

2 Preliminaries

2.1 Frequent notation and identities

Throughout the paper, all groups will be equipped with the discrete topology. Thus, we will refer to *properly discontinuous* actions on topological spaces simply as *proper* actions.

If G is a group and $F \subseteq G$ is a subset, we denote by $\langle F \rangle$ the subgroup of G generated by F. We denote by $Z_G(F)$ the *centraliser* of the subset F, ie the subgroup of elements of G commuting with all elements of F.

If (X, d) is a metric space, $A \subseteq X$ is a subset, and $R \geq 0$ is a real number, we denote by $\mathcal{N}_R(A)$ the closed R–neighbourhood of A. If $x, y \in X$, we write $x \approx_R y$ with the meaning of $d(x, y) \leq R$.

Consider a group action on a set $G \curvearrowright X$. If η is a G–invariant pseudometric on X, we write, for every $x \in X$, $g \in G$, and $F \subseteq G$,

$$\ell(g, \eta) = \inf_{x \in X} \eta(x, gx), \quad \tau_F^{\eta}(x) = \max_{f \in F} \eta(x, fx), \quad \overline{\tau}_F^{\eta} = \inf_{x \in X} \tau_F^{\eta}(x).$$

When X is a metric space and we do not name its metric explicitly, we also write $\ell(g, X)$, τ_F^X and $\overline{\tau}_F^X$. If X is equipped with several G–actions originating from homomorphisms $\rho_n \colon G \to \operatorname{Isom} X$, we will write $\ell(g, \rho_n)$, $\tau_F^{\rho_n}$, $\overline{\tau}_F^{\rho_n}$ in order to avoid confusion.

If $S \subseteq G$ is a finite generating set, we denote by $|\cdot|_S$ and $\|\cdot\|_S$ the associated *word length* and *conjugacy length*, respectively:

$$|g|_S = \inf\{k \mid g = s_1 \cdot \dots \cdot s_k,\ s_i \in S^{\pm}\} \quad \text{and} \quad \|g\|_S = \inf_{h \in G} |hgh^{-1}|_S.$$

The following useful identities will be repeatedly used in this text. We consider a G–action on a set X, a G–invariant pseudometric η, a point $x \in X$, and finite generating sets $S, S_1, S_2 \subseteq G$. We have

$$\eta(x, gx) \leq |g|_S \cdot \tau_S^{\eta}(x), \quad \ell(g, \eta) \leq \|g\|_S \cdot \overline{\tau}_S^{\eta}, \quad \tau_{S_1}^{\eta}(x) \leq |S_1|_{S_2} \cdot \tau_{S_2}^{\eta}(x),$$

where we have defined $|S_1|_{S_2} := \max_{s \in S_1} |s|_{S_2}$.

2.2 Median algebras

In this and the next section, we only fix notation and prove a few simple facts that do not appear elsewhere in the literature. For a comprehensive introduction to median algebras and median spaces, the reader can consult [29, Sections (2)–(4)], [15, Sections (4)–(6)] and [50, Section 2].

A median algebra is a pair (M, m), where M is a set and $m\colon M^3 \to M$ is a map satisfying, for all $a, b, c, x \in M$,

$$m(a,a,b) = a, \quad m(a,b,c) = m(b,c,a) = m(b,a,c), \quad m(m(a,x,b),x,c) = m(a,x,m(b,x,c)).$$

The third identity, usually known as the 4–*point condition*, is sometimes replaced by a different identity involving 5 points (for instance, in [89; 29; 15; 50]). The equivalence of the two conditions [74; 1] is quite nontrivial, but not required in the rest of the paper.

A map $\phi\colon M \to N$ between median algebras is a *median morphism* if, for all $x, y, z \in M$, we have $\phi(m(x,y,z)) = m(\phi(x), \phi(y), \phi(z))$. We denote by $\operatorname{Aut} M$ the group of median automorphisms of M. Throughout the paper, all group actions on median algebras will be by (median) automorphisms, unless stated otherwise.

A subset $S \subseteq M$ is a *median subalgebra* if $m(S \times S \times S) \subseteq S$. A subset $C \subseteq M$ is *convex* if $m(C \times C \times M) \subseteq C$. Helly's lemma states that any finite family of pairwise-intersecting convex subsets of M has nonempty intersection [89, Theorem 2.2]. We say that C is *gate-convex* if it admits a *gate-projection*, ie a map $\pi_C\colon M \to C$ with the property that $m(z, \pi_C(z), x) = \pi_C(z)$ for all $x \in C$ and $z \in M$. Gate-convex subsets are convex, and convex subsets are median subalgebras. Each gate-convex subset admits a unique gate-projection, and gate-projections are median morphisms.

The *interval* $I(x,y)$ between points $x, y \in M$ is defined as the set $\{z \in M \mid m(x,y,z) = z\}$. Note that $I(x,y)$ is gate-convex with projection given by the map $z \mapsto m(x,y,z)$. Intervals can be used to give an alternative description of convexity: a subset $C \subseteq M$ is convex if and only if $I(x,y) \subseteq C$ for all $x, y \in C$.

A *halfspace* is a subset $\mathfrak{h} \subseteq M$ such that both $\mathfrak{h}$ and $\mathfrak{h}^* := M \setminus \mathfrak{h}$ are convex and nonempty. A *wall* is a set of the form $\mathfrak{w} = \{\mathfrak{h}, \mathfrak{h}^*\}$, where $\mathfrak{h}$ and $\mathfrak{h}^*$ are halfspaces. We say that $\mathfrak{w}$ is the wall *bounding* $\mathfrak{h}$, and that $\mathfrak{h}$ and $\mathfrak{h}^*$ are the halfspaces *associated* to $\mathfrak{w}$.

Two halfspaces $\mathfrak{h}_1$ and $\mathfrak{h}_2$ are *transverse* if all four intersections $\mathfrak{h}_1 \cap \mathfrak{h}_2$, $\mathfrak{h}_1^* \cap \mathfrak{h}_2$, $\mathfrak{h}_1 \cap \mathfrak{h}_2^*$ and $\mathfrak{h}_1^* \cap \mathfrak{h}_2^*$ are nonempty. If $\mathfrak{w}_1$ and $\mathfrak{w}_2$ are the walls bounding $\mathfrak{h}_1$ and $\mathfrak{h}_2$, we also say that $\mathfrak{w}_1$ is *transverse* to $\mathfrak{w}_2$ and $\mathfrak{h}_2$. If $\mathcal{U}$ and $\mathcal{V}$ are sets of walls or halfspaces, we say that $\mathcal{U}$ and $\mathcal{V}$ are *transverse* if every element of $\mathcal{U}$ is transverse to every element of $\mathcal{V}$. If $\mathcal{H}$ is a set of halfspaces, we write $\mathcal{H}^* := \{\mathfrak{h}^* \mid \mathfrak{h} \in \mathcal{H}\}$.

We denote by $\mathscr{W}(M)$ and $\mathscr{H}(M)$, respectively, the set of all walls and all halfspaces of M. Given subsets $A, B \subseteq M$, we write

$$\mathscr{H}(A|B) = \{\mathfrak{h} \in \mathscr{H}(M) \mid A \subseteq \mathfrak{h}^*, B \subseteq \mathfrak{h}\}, \quad \mathscr{W}(A|B) = \{\mathfrak{w} \in \mathscr{W}(M) \mid \mathfrak{w} \cap \mathscr{H}(A|B) \neq \varnothing\}.$$

If $\mathfrak{w}_1$ and $\mathfrak{w}_2$ are walls bounding disjoint halfspaces $\mathfrak{h}_1$ and $\mathfrak{h}_2$, we set

$$\mathscr{W}(\mathfrak{w}_1|\mathfrak{w}_2) := \mathscr{W}(\mathfrak{h}_1|\mathfrak{h}_2) \setminus \{\mathfrak{w}_1, \mathfrak{w}_2\}.$$

If $A, B \subseteq M$ are nonempty, then $\mathscr{H}(A|B)$ admits minimal elements under inclusion. This follows from Zorn's lemma since, for every totally ordered subset $\mathscr{C} \subseteq \mathscr{H}(A|B)$, the intersection of all halfspaces in $\mathscr{C}$ is again a halfspace in $\mathscr{H}(A|B)$. Note that any two minimal elements $\mathfrak{h}_1, \mathfrak{h}_2 \in \mathscr{H}(A|B)$ are transverse, since $\mathfrak{h}_1 \cap \mathfrak{h}_2$ and $\mathfrak{h}_1^* \cap \mathfrak{h}_2^*$ are nonempty and there is no inclusion relation between $\mathfrak{h}_1$ and $\mathfrak{h}_2$.

If $\mathfrak{w} \in \mathscr{W}(A|B)$, we say that the wall $\mathfrak{w}$ *separates* A and B. Any two disjoint convex subsets of M are separated by at least one wall [89, Theorem 2.8]; in particular, distinct points of M are always separated by a wall.

Given a subset $A \subseteq M$, we also introduce

$$\mathscr{H}_A(M) := \{\mathfrak{h} \in \mathscr{H}(M) \mid \mathfrak{h} \cap A \neq \varnothing, \mathfrak{h}^* \cap A \neq \varnothing\}, \quad \mathscr{W}_A(M) := \{\mathfrak{w} \in \mathscr{W}(M) \mid \mathfrak{w} \subseteq \mathscr{H}_A(M)\}.$$

Equivalently, a wall $\mathfrak{w}$ lies in $\mathscr{W}_A(M)$ if and only if it separates two points of A.

Remark 2.1 If $\mathcal{U} \subseteq \mathscr{H}(M)$ and $\mathcal{V} \subseteq \mathscr{H}(N)$ are subsets, we say that a map $\phi\colon \mathcal{U} \to \mathcal{V}$ is a *morphism of pocsets* if, for all $\mathfrak{h}, \mathfrak{k} \in \mathcal{U}$ with $\mathfrak{h} \subseteq \mathfrak{k}$, we have $\phi(\mathfrak{h}) \subseteq \phi(\mathfrak{k})$ and $\phi(\mathfrak{h}^*) = \phi(\mathfrak{h})^*$.

Every median morphism $\phi\colon M \to N$ induces a morphism of pocsets $\phi^*\colon \mathscr{H}_{\phi(M)}(N) \to \mathscr{H}(M)$ defined by $\phi^*(\mathfrak{h}) = \phi^{-1}(\mathfrak{h})$. When $\phi\colon M \to N$ is surjective, we obtain a map $\phi^*\colon \mathscr{H}(N) \to \mathscr{H}(M)$ that is injective and preserves transversality.

Remark 2.2 (1) If $S \subseteq M$ is a subalgebra, we have a map $\mathrm{res}_C\colon \mathscr{H}_C(M) \to \mathscr{H}(C)$ given by $\mathrm{res}_C(\mathfrak{h}) = \mathfrak{h} \cap C$. This is a morphism of pocsets and, by [15, Lemma 6.5], it is a surjection.

(2) If $C \subseteq M$ is convex, then the map res_C is also injective and it preserves transversality. In particular, the sets $\mathscr{H}(C)$ and $\mathscr{H}_C(M)$ are naturally identified in this case.
Indeed, if $\mathfrak{h}, \mathfrak{k} \in \mathscr{H}_C(M)$ are intersecting halfspaces, Helly's lemma guarantees that $\mathfrak{h} \cap C$ and $\mathfrak{k} \cap C$ intersect too. Moreover, we have $\mathfrak{h} = \mathfrak{k}$ if and only if $\mathfrak{h} \cap \mathfrak{k}^*$ and $\mathfrak{h}^* \cap \mathfrak{k}$ are empty.

(3) If C is gate-convex with projection π_C, then $\mathrm{res}_C \circ \pi_C^* = \mathrm{id}_{\mathscr{H}(C)}$ and $\pi_C^* \circ \mathrm{res}_C = \mathrm{id}_{\mathscr{H}_C(M)}$.

If $C_1, C_2 \subseteq M$ are gate-convex subsets with gate-projections π_1, π_2, then $\mathscr{H}(x|C_i) = \mathscr{H}(x|\pi_i(x))$ for all $x \in M$. We say that $x_1 \in C_1$ and $x_2 \in C_2$ are a *pair of gates* if $\pi_2(x_1) = x_2$ and $\pi_1(x_2) = x_1$. Pairs of gates always exist and satisfy $\mathscr{H}(x_1|x_2) = \mathscr{H}(C_1|C_2)$.

The *standard k–cube* is the finite set $\{0,1\}^k$ equipped with the median operator m determined by a majority vote on each coordinate. A subset $S \subseteq M$ is a *k–cube* if it is a median subalgebra isomorphic to the standard k–cube. In particular, any subset of M with cardinality 2 is a 1–cube.

Remark 2.3 An important example of median algebra is provided by the 0–skeleton of any CAT(0) cube complex X; see [31]. The vertex set of any k–cell of X is a k–cube in the above sense, but the converse does not hold. For instance, in the standard tiling of $\mathbb{R}^n$, every set of the form $\{a_1, b_1\} \times \cdots \times \{a_n, b_n\}$ with $a_i < b_i$ is a k–cube according to the above notion. To avoid confusion, when dealing with cube complexes we will refer to k–cubes in $X^{(0)}$ as *generalised k–cubes*.

The *rank* of M, denoted by $\operatorname{rk} M$, is the largest cardinality of a set of pairwise-transverse walls of M. Equivalently, $\operatorname{rk} M$ is the supremum of the integers k such that M contains a k–cube (assuming $\operatorname{rk} M$ is at most countable); see [15, Proposition 6.2]. We will be exclusively interested in median algebras of finite rank.

We will need the following criterion, which summarises Lemmas 2.9 and 2.11 in [51]. If $\mathcal{H} \subseteq \mathscr{H}(M)$, we denote by $\bigcap \mathcal{H} \subseteq M$ the intersection of all halfspaces in $\mathcal{H}$.

Lemma 2.4 *Let M be a finite-rank median algebra. Partially order $\mathscr{H}(M)$ by inclusion.*

(1) *Let $\mathcal{H} \subseteq \mathscr{H}(M)$ be a set of pairwise intersecting halfspaces. Suppose that every chain in $\mathcal{H}$ admits a lower bound in $\mathcal{H}$. Then $\bigcap \mathcal{H}$ is a nonempty convex subset of M.*

(2) *A convex subset $C \subseteq M$ is gate-convex if and only if there does not exist a chain $\mathscr{C} \subseteq \mathscr{H}_C(M)$ such that $\bigcap \mathscr{C}$ is nonempty and disjoint from C.*

If $A \subseteq M$ is a subset, we denote by $\langle A \rangle$ the median subalgebra generated by A, ie the smallest subalgebra of M containing A. We also denote by $\operatorname{Hull} A$ the smallest convex subset of M that contains A; this coincides with the intersection of all halfspaces of M that contain A.

The sets $\langle A \rangle$ and $\operatorname{Hull} A$ are best understood in terms of the following operators:

$$\mathcal{M}(A) = \mathcal{M}^1(A) := m(A \times A \times A), \qquad \mathcal{M}^{n+1}(A) := \mathcal{M}(\mathcal{M}^n(A)),$$
$$\mathcal{J}(A) = \mathcal{J}^1(A) := m(A \times A \times M) = \bigcup_{x,y \in A} I(x, y), \qquad \mathcal{J}^{n+1}(A) := \mathcal{J}(\mathcal{J}^n(A)).$$

It is clear that $\operatorname{Hull} A = \bigcup_{n \geq 1} \mathcal{J}^n(A)$ and $\langle A \rangle = \bigcup_{n \geq 1} \mathcal{M}^n(A)$.

Remark 2.5 When $\operatorname{rk} M = r$ is finite, [15, Lemma 6.4] shows that already $\mathcal{J}^r(A) = \operatorname{Hull} A$. A similar result holds for $\langle A \rangle$ and the operator $\mathcal{M}$ (see Proposition 4.2 below), but its proof will require considerable work.

If M_1 and M_2 are median algebras, we denote by $M_1 \times M_2$ their *product*. This is the median algebra with underlying set $M_1 \times M_2$ and the only median operator for which both coordinate projections are median morphisms.

The set $\mathscr{W}(M_1 \times M_2)$ is naturally partitioned into two transverse subsets $\mathscr{W}_1$ and $\mathscr{W}_2$. A wall lies in $\mathscr{W}_1$ if and only if it separates two points in one (equivalently, every) fibre $M_1 \times \{*\}$; halfspaces associated to walls in $\mathscr{W}_1$ are unions of fibres $\{*\} \times M_2$. The set $\mathscr{W}_2$ is defined similarly, swapping the roles played by the two indices. Since all fibres are gate-convex in $M_1 \times M_2$, Remark 2.2 gives natural identifications between $\mathscr{W}_i$ and $\mathscr{W}(M_i)$.

In finite rank, product splittings can be completely characterised in terms of walls. The following is [51, Lemma 2.12]; also see [23, Lemma 2.5] in the special case of cube complexes.

Lemma 2.6 *For a finite-rank median algebra M, the following are equivalent*:

(1) *M splits as a product of median algebras $M_1 \times M_2$, where neither M_i is a singleton.*

(2) *There exists a partition $\mathscr{W}(M) = \mathscr{W}_1 \sqcup \mathscr{W}_2$, where the $\mathscr{W}_i$ are nonempty and transverse.*

When this happens, the set $\mathscr{W}_i$ is identified with $\mathscr{W}(M_i)$ as described above.

2.3 Compatible metrics on median algebras

A metric space (X, d) is a *median space* if, for all $x_1, x_2, x_3 \in X$, there exists a unique point $m(x_1, x_2, x_3)$ in X such that

$$d(x_i, x_j) = d(x_i, m(x_1, x_2, x_3)) + d(m(x_1, x_2, x_3), x_j)$$

for all $1 \leq i < j \leq 3$. In this case, the map $m\colon X^3 \to X$ gives a median algebra (X, m).

Remark 2.7 (rank of median spaces) We define the *rank* of X as the rank of the underlying median algebra (X, m). If X is a connected median space, then this notion of rank coincides with the supremum of the topological dimensions of the locally compact subsets of X. The latter is the definition of rank that we used in the introduction. One inequality follows from Theorem 2.2 and Lemma 7.6 in [15], while the other from [17, Proposition 5.6].

For the purposes of this paper, it is convenient to think of median spaces in terms of the following notion. Let M be a median algebra.

Definition 2.8 A pseudometric $\eta\colon M \times M \to [0, +\infty)$ is *compatible* if, for every $x, y, z \in M$,

$$\eta(x, y) = \eta(x, m(x, y, z)) + \eta(m(x, y, z) + y).$$

Thus, we can equivalently define median spaces as pairs (M, d), where M is a median algebra and d is a compatible metric on M.

We write $\mathcal{D}(M)$ and $\mathcal{PD}(M)$, respectively, for the sets of all compatible metrics and all compatible pseudometrics on M. In the presence of a group action $G \curvearrowright M$, we write $\mathcal{D}^G(M)$ and $\mathcal{PD}^G(M)$ for the subsets of G–invariant (pseudo)metrics (or just $\mathcal{D}^g(M)$ and $\mathcal{PD}^g(M)$ if $G = \langle g \rangle$).

To avoid confusion, we will normally denote compatible *metrics* by the letter δ, and general compatible *pseudometrics* by the letter η.

Consider a gate-convex subset $C \subseteq M$ and its gate-projection $\pi_C\colon M \to C$. For every pseudometric $\eta \in \mathcal{PD}(M)$, the maps $\pi_C\colon M \to C$ and $m\colon M^3 \to M$ are 1–Lipschitz, in the sense that

$$\eta(\pi_C(x), \pi_C(y)) \le \eta(x, y), \quad \eta(m(x, y, z), m(x', y', z')) \le \eta(x, x') + \eta(y, y') + \eta(z, z').$$

This can be proved as in Lemma 2.13 and Corollary 2.15 of [29]. In addition, gate-projections are nearest-point projections, in the sense that $\eta(x, \pi_C(x)) = \eta(x, C)$ for all $x \in M$.

If $\delta \in \mathcal{D}(M)$ and (M, δ) is complete, then a subset $C \subseteq M$ is gate-convex if and only if it is convex and closed in the topology induced by δ; see [29, Lemma 2.13].

If M is the 0–skeleton of a CAT(0) cube complex X, then a natural compatible metric on M is given by the restriction of the *combinatorial metric* on X: this is just the intrinsic path metric of the 1–skeleton of X. All cube complexes in this paper will be implicitly endowed with their combinatorial metric, rather than the CAT(0) metric. All geodesics will be assumed to be *combinatorial geodesics*.

Remark 2.9 A *halfspace-interval* is a set of the form $\mathscr{H}(x|y) \subseteq \mathscr{H}(M)$ for $x, y \in M$. Let $\mathscr{B}(M) \subseteq 2^{\mathscr{H}(M)}$ denote the σ–algebra generated by halfspace-intervals. We say that a subset $\mathcal{H} \subseteq \mathscr{H}(M)$ is $\mathscr{B}$–measurable if it lies in $\mathscr{B}(M)$.

Every $\eta \in \mathcal{PD}(M)$ induces a measure ν_η on $\mathscr{B}(M)$ such that $\nu_\eta(\mathscr{H}(x|y)) = \eta(x, y)$ for all $x, y \in M$; see eg [29, Theorem 5.1]. If $\eta \in \mathcal{PD}^G(M)$, then ν_η is G–invariant.

Lemma 2.10 *Let (X, d) be a median space. Let $A \subseteq X$ be a subset such that $\mathcal{J}(A) \subseteq \mathcal{N}_R(A)$ for some $R \ge 0$. Then, for every $D \ge 0$, we have*

$$\mathcal{J}(\mathcal{N}_D(A)) \subseteq \mathcal{N}_{2D+R}(A).$$

In addition, if $\operatorname{rk} X = r$, *we have* $\operatorname{Hull} A \subseteq \mathcal{N}_{2^r R}(A)$.

Proof If $z \in \mathcal{J}(\mathcal{N}_D(A))$, there exist $x, y \in \mathcal{N}_D(A)$ and $z \in I(x, y)$. Consider points $x', y' \in A$ with $d(x, x'), d(y, y') \le D$. Set $z' = m(x', y', z)$. Since $z' \in \mathcal{J}(A)$, we have $d(z', A) \le R$. Furthermore,

$$d(z, z') = d(m(x, y, z), m(x', y', z)) \le d(x, x') + d(y, y') \le 2D.$$

In conclusion, $d(z, A) \le d(z, z') + d(z', A) \le 2D + R$, as required.

Proceeding by induction, it is straightforward to obtain $\mathcal{J}^i(A) \subseteq \mathcal{N}_{(2^i-1)R}(A)$ for every $i \ge 0$. If $\operatorname{rk} X = r$, we have $\operatorname{Hull} A = \mathcal{J}^r(A)$ by Remark 2.5, hence $\operatorname{Hull} A \subseteq \mathcal{N}_{(2^r-1)R}(A) \subseteq \mathcal{N}_{2^r R}(A)$. □

2.4 Convex cores in median algebras

In this subsection, we collect a few facts proved in [51] extending the notion of "essential core" [23, Section 3] from actions on cube complexes to general actions on finite-rank median algebras (even with no invariant metric or topology). These results will only play a role in the proofs of Theorems F and I (especially in Sections 6 and 7). The reader only interested in the other results mentioned in the introduction can safely read this subsection with CAT(0) cube complexes in mind, just to familiarise themselves with our notation.

Let M be a median algebra of finite rank r.

Definition 2.11 We say that $g \in \operatorname{Aut} M$ acts

(1′) *nontransversely* if there does not exist a wall $\mathfrak{w} \in \mathscr{W}(X)$ such that $\mathfrak{w}$ and $g\mathfrak{w}$ are transverse;

(2′) *stably without inversions* if there do not exist $n \in \mathbb{Z}$ and $\mathfrak{h} \in \mathscr{H}(X)$ with $g^n\mathfrak{h} = \mathfrak{h}^*$.

An action $G \curvearrowright M$ by automorphisms is

(1) *nontransverse* if every $g \in G$ acts nontransversely;

(2) *without wall inversions* if every $g \in G$ acts stably without inversions;

(3) *essential* if, for every $\mathfrak{h} \in \mathscr{H}(M)$, there exists $g \in G$ with $g\mathfrak{h} \subsetneq \mathfrak{h}$.

Remark 2.12 If there exists $\delta \in \mathcal{D}^G(M)$ such that (M, δ) is connected, then $G \curvearrowright M$ is without wall inversions. This follows from [50, Proposition B] when (M, δ) is complete, and from [51, Remark 4.3] in general.

Keeping the notation of [51], each action $G \curvearrowright M$ determines sets of halfspaces

$$\mathcal{H}_1(G) := \{\mathfrak{h} \in \mathscr{H}(M) \mid \exists g \in G \text{ such that } g\mathfrak{h} \subsetneq \mathfrak{h}\},$$
$$\overline{\mathcal{H}}_{1/2}(G) := \{\mathfrak{h} \in \mathscr{H}(M) \setminus \mathcal{H}_1(G) \mid \exists g \in G \text{ such that } g\mathfrak{h}^* \cap \mathfrak{h}^* = \varnothing \text{ and } g\mathfrak{h} \neq \mathfrak{h}^*\},$$
$$\overline{\mathcal{H}}_0(G) := \{\mathfrak{h} \in \mathscr{H}(M) \mid \forall g \in G \text{ either } g\mathfrak{h} \in \{\mathfrak{h}, \mathfrak{h}^*\} \text{ or } g\mathfrak{h} \text{ and } \mathfrak{h} \text{ are transverse}\}.$$

As observed in [51, Section 3.1], we have a G–invariant partition

$$\mathscr{H}(M) = \overline{\mathcal{H}}_0(G) \sqcup \mathcal{H}_1(G) \sqcup \overline{\mathcal{H}}_{1/2}(G) \sqcup \overline{\mathcal{H}}_{1/2}(G)^*.$$

We write $\mathcal{W}_1(G)$ and $\mathcal{W}_0(G)$ for the sets of walls bounding the halfspaces in $\mathcal{H}_1(G)$ and $\overline{\mathcal{H}}_0(G)$.

Definition 2.13 The *reduced core* $\overline{\mathcal{C}}(G)$ is the intersection of all halfspaces lying in $\overline{\mathcal{H}}_{1/2}(G)$.

We adopt the convention that $\overline{\mathcal{C}}(G) = M$ when $\overline{\mathcal{H}}_{1/2}(G)$ is empty. We will write $\overline{\mathcal{C}}(G, M)$ (and $\mathcal{H}_\bullet(G, M)$, $\mathcal{W}_\bullet(G, M)$) if it is necessary to specify the ambient median algebra. We just write $\overline{\mathcal{C}}(g)$ (and $\mathcal{H}_\bullet(g)$, $\mathcal{W}_\bullet(g)$) if $G = \langle g \rangle$.

Theorem 2.14 [51] *Let G be finitely generated and let $G \curvearrowright M$ be without wall inversions.*

(1) *The reduced core $\overline{\mathcal{C}}(G)$ is nonempty, G–invariant and convex.*

Suppose in addition that $\mathcal{D}^G(M) \neq \varnothing$.

(2) *There is a G–fixed point in M if and only if $\mathcal{H}_1(G) = \varnothing$.*

(3) *The sets $\mathcal{W}_1(G)$ and $\mathcal{W}_0(G)$ are transverse and $\mathscr{W}_{\overline{\mathcal{C}}(G)}(M) = \mathcal{W}_0(G) \sqcup \mathcal{W}_1(G)$.*

(4) *The resulting partition of $\mathscr{W}(\overline{\mathcal{C}}(G))$ gives a product splitting $\overline{\mathcal{C}}(G) = \overline{\mathcal{C}}_0(G) \times \overline{\mathcal{C}}_1(G)$. The normaliser of the image of G in* Aut M *leaves $\overline{\mathcal{C}}(G)$ invariant, preserving the two factors. The action $G \curvearrowright \overline{\mathcal{C}}_1(G)$ is essential, while $G \curvearrowright \overline{\mathcal{C}}_0(G)$ fixes a point.*

Proof We just refer the reader to the relevant statements in [51]. Part (1) follows from Theorem 3.17(2). The two implications in part (2) are obtained from Proposition 3.23(2) and Lemma 4.5(1), respectively. Part (3) is a consequence of Lemma 4.5 and Lemma 3.22(2). Finally, part (4) follows from Remark 3.16 and the previous parts. □

Remark 2.15 If G acts on a CAT(0) cube complex X and $M = X^{(0)}$, then the action $G \curvearrowright \overline{\mathcal{C}}_1(G)$ in Theorem 2.14(4) is easily identified as the *G–essential core* of Caprace and Sageev; cf [23, Section 3.3]. In particular, note that Theorem 2.14 strengthens [23, Proposition 3.5], showing that the G–essential core always embeds G–equivariantly as a convex subcomplex of X.

Theorem 2.16 *If $g \in$* Aut M *acts nontransversely and stably without inversions, then*

(1) *the reduced core $\overline{\mathcal{C}}(g)$ is gate-convex, and*

(2) *for every $x \in M$ and every $\eta \in \mathcal{PD}^g(M)$, we have $\eta(x, gx) = \ell(g, \eta) + 2\eta(x, \overline{\mathcal{C}}(g))$.*

Proof Part (1) is [51, Proposition 3.36] and part (2) is [51, Proposition 4.9(3)]. □

Note that $\overline{\mathcal{C}}(G)$ is not gate-convex in general, even when $G \curvearrowright M$ is an isometric action of a finitely generated free group on a complete $\mathbb{R}$–tree. See [51, Example 3.37].

Remark 2.17 Part (2) of Theorem 2.16 implies that, if $\delta \in \mathcal{D}^g(M)$ and (M, δ) is a geodesic space, then g is *semisimple*: either g fixes a point of M or g translates along a $\langle g \rangle$–invariant geodesic.

The next two remarks will only be needed in Section 7.

Remark 2.18 Let $g \in$ Aut M act nontransversely and stably without inversions, with $\mathcal{D}^g(M) \neq \varnothing$.

(1) Each $\mathfrak{h} \in \mathcal{H}_1(g)$ satisfies $\bigcap_{n \in \mathbb{Z}} g^n \mathfrak{h} = \varnothing$; see [51, Lemma 4.5(1)].

(2) A halfspace $\mathfrak{h}$ lies in $\mathfrak{h} \in \bar{\mathcal{H}}_0(g)$ if and only if $g\mathfrak{h} = \mathfrak{h}$, and it lies in $\mathcal{H}_1(g)$ if and only if either $g\mathfrak{h} \subsetneq \mathfrak{h}$ or $g\mathfrak{h} \supsetneq \mathfrak{h}$. This follows from Remarks 3.33 and 3.34 in [51], after observing that $\mathcal{H}_1(g) \subseteq \mathscr{H}_{\bar{C}(g)}(M)$ (eg by part (1) of this remark).

(3) Let $N \subseteq M$ be a $\langle g \rangle$–invariant median subalgebra. By Remark 2.2, intersecting the halfspaces of M with N, we obtain a surjective restriction map $\mathrm{res}_N : \mathscr{H}_N(M) \to \mathscr{H}(N)$. Parts (1) and (2) show that:

- If $\mathfrak{h} \in \bar{\mathcal{H}}_0(g, M) \cap \mathscr{H}_N(M)$, then $g \cdot \mathrm{res}_N(\mathfrak{h}) = \mathrm{res}_N(\mathfrak{h})$ and $\mathrm{res}_N(\mathfrak{h}) \in \bar{\mathcal{H}}_0(g, N)$.
- If $\mathfrak{h} \in \bar{\mathcal{H}}_{1/2}(g, M) \cap \mathscr{H}_N(M)$, then either $\mathrm{res}_N(\mathfrak{h}) \in \bar{\mathcal{H}}_{1/2}(g, N)$ or $g \cdot \mathrm{res}_N(\mathfrak{h}) = \mathrm{res}_N(\mathfrak{h})^*$.
- We have $\mathcal{H}_1(g, M) \subseteq \mathscr{H}_N(M)$ and $\mathrm{res}_N(\mathcal{H}_1(g, M)) = \mathcal{H}_1(g, N)$.

Remark 2.19 Let $g \in \mathrm{Aut}\, M$ act nontransversely and stably without inversions. Let ν_η be the measure introduced in Remark 2.9. Part (2) of Theorem 2.16 shows that $\ell(g, \eta) = \nu_\eta(\mathscr{H}(x|gx))$ for any $x \in \bar{C}(g)$. In view of parts (1) and (2) of Remark 2.18, the set $\mathscr{H}(x|gx) \sqcup \mathscr{H}(gx|x)$ is a $\mathscr{B}$–measurable fundamental domain for the action $\langle g \rangle \curvearrowright \mathcal{H}_1(g)$. It follows that, for *any* fundamental domain $\Omega \in \mathscr{B}(M)$ for the action $\langle g \rangle \curvearrowright \mathcal{H}_1(g)$, we have $\ell(g, \eta) = \frac{1}{2}\nu_\eta(\Omega)$.

2.5 Two constructions involving cube complexes

2.5.1 Restriction quotients Restriction quotients of CAT(0) cube complexes were originally introduced in [23, page 860]. Our interest is due to the fact that the Salvetti blowups and collapses from [25] are a particular instance of this construction, which can actually be phrased purely in median-algebra terms. This is mainly needed in the proof of Proposition A(3) in Section 3.4, though it will also be useful in Sections 3.1 and 7.3.

A map $f : X \to Y$ between cube complexes is said to be *cubical* if, on every cube $c \subseteq X$, it factors as a projection of c onto one of its faces, followed by an isomorphism onto a cube of Y.

Let X be a CAT(0) cube complex. The *carrier* of a hyperplane $\mathfrak{w} \in \mathscr{W}(X)$ is the smallest convex subcomplex of X that contains all edges crossing $\mathfrak{w}$. It naturally splits as a product $C \times [0, 1]$, where $C \times \{0\}$ and $C \times \{1\}$ are convex subcomplexes of X on the two sides of $\mathfrak{w}$.

Given a hyperplane $\mathfrak{w} \in \mathscr{W}(X)$, we can construct a new CAT(0) cube complex Y by *collapsing* $\mathfrak{w}$: we remove from X the interior of the carrier $C \times (0, 1)$ and we identify the isomorphic subcomplexes $C \times \{0\}$ and $C \times \{1\}$. The natural collapse map $X \to Y$ is a cubical map.

Now, consider a set of hyperplanes $\mathcal{U} \subseteq \mathscr{W}(X)$. The *restriction quotient* of X determined by $\mathcal{U}$ is the CAT(0) cube complex $X(\mathcal{U})$ obtained by collapsing all hyperplanes in $\mathscr{W}(X) \setminus \mathcal{U}$ (which usually involves infinitely many collapses). It has one vertex for every connected component of the complement in X of the union of the hyperplanes in $\mathcal{U}$, with two vertices joined by an edge exactly when the corresponding components are separated by a single element of $\mathcal{U}$. Let $\pi_{\mathcal{U}}\colon X \to X(\mathcal{U})$ be the natural collapse, which is again a cubical map.

If $G \curvearrowright X$ is an action and the subset $\mathcal{U} \subseteq \mathscr{W}(X)$ is G–invariant, then the restriction quotient $X(\mathcal{U})$ is also equipped with a natural G–action and the collapse map $\pi_{\mathcal{U}}$ is G–equivariant.

Proposition 2.20 *Consider* CAT(0) *cube complexes X, Y and a surjective cubical map $\pi : X \to Y$. Then the following are equivalent*:

(1) *There exists a subset $\mathcal{U} \subseteq \mathscr{W}(X)$ and an isomorphism $Y \cong X(\mathcal{U})$ with respect to which π corresponds to the natural collapse $\pi_{\mathcal{U}}\colon X \to X(\mathcal{U})$.*

(2) *For every vertex $v \in Y$, the preimage $\pi^{-1}(v)$ is a convex subcomplex of X.*

(3) *The restriction $\pi : X^{(0)} \to Y^{(0)}$ is a median morphism.*

If X and Y are equipped with G–actions and π is G–equivariant, then the set $\mathcal{U}$ is G–invariant.

Proof The equivalence of (1) and (2) was shown in [71, Theorem 4.4]. Fibres of median morphisms between median algebras are always convex, so (3) implies (2). Finally, (1) $\Longrightarrow$ (3) can be shown by observing that single hyperplane-collapses are median morphisms. □

2.5.2 Roller boundaries In two proofs (Proposition 4.11 and, briefly, Lemma 3.13), we will need the notion of Roller boundary of a CAT(0) cube complex X, denoted by ∂X. We list here the (well-known) properties that we will use.

The 0–skeleton of any CAT(0) cube complex X has a natural structure of median algebra; see for instance [31, Theorem 6.1] and [89, Theorem 10.3]. The ℓ^1 metric on X, denoted by d, is a compatible metric in the sense of Definition 2.8. Thus, the pair $(X^{(0)}, d)$ is a median space. The notions of "halfspace" and "wall" coincide with the usual notion of halfspace and hyperplane in CAT(0) cube complexes. Thus, we write $\mathscr{W}(X)$ and $\mathscr{H}(X)$ with the meaning of $\mathscr{W}(X^{(0)})$ and $\mathscr{H}(X^{(0)})$.

We can embed $X^{(0)} \hookrightarrow 2^{\mathscr{H}(X)}$ by mapping each vertex v to the subset $\sigma_v \subseteq \mathscr{H}(X)$ of halfspaces that contain it. This is a median morphism if we endow $2^{\mathscr{H}(X)}$ with the structure of median algebra given by

$$m(\sigma_1, \sigma_2, \sigma_3) = (\sigma_1 \cap \sigma_2) \cup (\sigma_2 \cap \sigma_3) \cup (\sigma_3 \cap \sigma_1).$$

The space $2^{\mathscr{H}(X)}$ is compact with the product topology, and we can consider the closure $\overline{X}$ of $X^{(0)}$ inside it. We define the *Roller boundary* ∂X as the set $\overline{X} \setminus X^{(0)}$.

For us, the only important facts will be:

(1) The subset $\overline{X} = X \sqcup \partial X \subseteq 2^{\mathscr{H}(X)}$ is a median subalgebra and $X^{(0)}$ is convex in $\overline{X}$.

(2) The median $m\colon \overline{X}^3 \to \overline{X}$ is continuous with respect to the topology that $\overline{X}$ inherits from $2^{\mathscr{H}(X)}$. With this topology, $\overline{X}$ is compact and totally disconnected. If X is locally finite, the subset $X^{(0)} \subseteq \overline{X}$ is discrete.

(3) If $\mathfrak{h} \in \mathscr{H}(X)$, its closure $\overline{\mathfrak{h}}$ inside $\overline{X}$ is gate-convex. In fact, $\overline{\mathfrak{h}}$ and $\overline{\mathfrak{h}^*}$ are complementary halfspaces of the median algebra $\overline{X}$. The gate-projection $\pi_{\mathfrak{h}}\colon \overline{X} \to \overline{\mathfrak{h}}$ takes $X^{(0)}$ to $\mathfrak{h}$.

(4) Two halfspaces $\mathfrak{h}, \mathfrak{k} \in \mathscr{H}(X)$ are said to be *strongly separated* if $\mathfrak{h} \cap \mathfrak{k} = \varnothing$ and no halfspace of X is transverse to both $\mathfrak{h}$ and $\mathfrak{k}$; see [4]. If $\mathfrak{h}$ and $\mathfrak{k}$ are strongly separated, then the gate-projection $\pi_{\mathfrak{h}}\colon \overline{X} \to \overline{\mathfrak{h}}$ maps $\overline{\mathfrak{k}}$ to a single point.

The reader can consult [47, Sections 2.3–2.4] and [50, Theorem 4.14] for more details on facts (1)–(3). Fact (4) follows, for example, from Corollary 2.22 and Lemma 2.23 in [49].

2.6 Coarse median structures

Coarse median spaces were introduced by Bowditch in [15]. We present the following equivalent definition from [83].

Definition 2.21 Let X be a metric space. A *coarse median* on X is a map $\mu\colon X^3 \to X$ for which there exists a constant $C \geq 0$ such that, for all $a, b, c, x \in X$, we have

(1) $\mu(a,a,b) = a$ and $\mu(a,b,c) = \mu(b,c,a) = \mu(b,a,c)$,

(2) $\mu(\mu(a,x,b),x,c) \approx_C \mu(a,x,\mu(b,x,c))$,

(3) $d(\mu(a,b,c),\mu(x,b,c)) \leq C d(a,x) + C$.

Note that part (2) of the definition is an approximate version of the 4–point condition, from our definition of median algebras at the beginning of Section 2.2.

There is an appropriate notion of *rank* also for coarse median spaces. Since this notion will play no significant role in our paper (except when we briefly mention it at the end of Section 7.1), we simply refer the reader to [15; 83; 84] for more details.

The following notion of *coarse median structure* is different from the one in [84, Definition 2.8], but it is hard to imagine this being cause for confusion.

Definition 2.22 Two coarse medians $\mu_1, \mu_2\colon X^3 \to X$ are *at bounded distance* if there exists a constant $C \geq 0$ such that $\mu_1(x,y,z) \approx_C \mu_2(x,y,z)$ for all $x, y, z \in X$. A *coarse median structure* on X is an equivalence class $[\mu]$ of coarse medians pairwise at bounded distance. A *coarse median space* is a pair $(X, [\mu])$ where X is a metric space and $[\mu]$ is a coarse median structure on it.

Remark 2.23 Let $f : X \to Y$ be a quasi-isometry with a coarse inverse denoted by $f^{-1}\colon Y \to X$. If $\mu\colon X^3 \to X$ is a coarse median on X, then

$$(f_*\mu)(x,y,z) := f\bigl(\mu(f^{-1}(x), f^{-1}(y), f^{-1}(z))\bigr)$$

is a coarse median on Y. If $[\mu_1] = [\mu_2]$, then $[f_*\mu_1] = [f_*\mu_2]$.

If $\mathrm{QI}(X)$ is the group of quasi-isometries $X \to X$ up to bounded distance (as defined for example in [44, Definition 8.22]), the above defines a natural left action of $\mathrm{QI}(X)$ on the set of coarse median structures on X.

Definition 2.24 A *coarse median* group is a pair $(G, [\mu])$ where G is a finitely generated group equipped with a word metric and $[\mu]$ is a G–invariant coarse median structure on G.

The requirement that $[\mu]$ be G–invariant can be equivalently stated as follows: for each $g \in G$, there exists a constant $C(g) \geq 0$ such that $g\mu(g_1, g_2, g_3) \approx_{C(g)} \mu(gg_1, gg_2, gg_3)$ for all $g_1, g_2, g_3 \in G$.

Note that Definition 2.24 is stronger than Bowditch's original definition from [15], which did not ask for $[\mu]$ to be G–invariant. Definition 2.24 is better suited to our needs in this paper, but it is not QI–invariant or even commensurability-invariant (unlike Bowditch's).

These two definitions of coarse median group parallel the notions of HHS and HHG from [5; 6]. Namely, every hierarchically hyperbolic group is a coarse median group in the sense of Definition 2.24, while any group that admits a structure of hierarchically hyperbolic space is coarse median in the sense of Bowditch [19] (we will simply refer to these as "groups with a coarse median structure").

Remark 2.25 If G is finitely generated, any group automorphism $\varphi\colon G \to G$ is bi-Lipschitz with respect to any word metric on G. The resulting homomorphism $\mathrm{Aut}\, G \to \mathrm{QI}(G)$ defines an $(\mathrm{Aut}\, G)$–action on the set of coarse median structures on G that takes G–invariant structures to G–invariant structures. If $(G, [\mu])$ is a coarse median group, then every inner automorphism of G fixes $[\mu]$, and we obtain an action of $\mathrm{Out}\, G$ on the $(\mathrm{Aut}\, G)$–orbit of $[\mu]$.

Definition 2.26 Let $(G, [\mu])$ be a coarse median group. We say that $\phi \in \mathrm{Out}\, G$ (or $\varphi \in \mathrm{Aut}\, G$) is *coarse-median preserving* if it fixes $[\mu]$. We denote by $\mathrm{Out}(G, [\mu]) \leq \mathrm{Out}\, G$ and $\mathrm{Aut}(G, [\mu]) \leq \mathrm{Aut}\, G$ the subgroups of coarse-median preserving automorphisms.

Thus $\varphi \in \mathrm{Aut}\, G$ is coarse-median preserving exactly when, fixing a word metric on G, there exists a constant $C \geq 0$ such that, for all $g_i \in G$,

$$\varphi(\mu(g_1, g_2, g_3)) \approx_C \mu(\varphi(g_1), \varphi(g_2), \varphi(g_3)).$$

Remark 2.27 Let $G \curvearrowright X$ be a proper cocompact action on a CAT(0) cube complex. Any orbit map $o\colon G \to X$ is a quasi-isometry that can be used to pull back the median operator $m\colon X^3 \to X$ to a coarse median structure $[\mu_X] := o_*^{-1}[m]$ on G. It is straightforward to check that $[\mu_X]$ is independent of all choices involved (though the notation is slightly improper, as $[\mu_X]$ does depend on the specific G–action on X). We refer to $[\mu_X]$ as the *coarse median structure induced by* $G \curvearrowright X$.

Let us write gx for the action of $g \in G$ on $x \in X$ according to $G \curvearrowright X$. Then, every $\varphi \in \operatorname{Aut} G$ gives rise to a twisted G–action on X, which we denote by $G \curvearrowright X^{\varphi}$, and is defined as $g \cdot x = \varphi^{-1}(g)x$. Note that $\varphi_*[\mu_X] = [\mu_{X^{\varphi}}]$ and thus $\varphi \operatorname{Out}(G, [\mu_X])\varphi^{-1} = \operatorname{Out}(G, [\mu_{X^{\varphi}}])$.

Each of the structures $[\mu_{X^{\varphi}}]$ is G–invariant. In particular, $(G, [\mu_X])$ is a coarse median group.

Example 2.28 Every geodesic Gromov-hyperbolic space X is equipped with a natural coarse median structure $[\mu]$ represented by the operators μ that map each triple (x, y, z) to an approximate incentre for a geodesic triangle with vertices x, y, z; cf [15, Section 3]. In fact, by [83, Theorem 4.2], this is the only coarse median structure that X can be endowed with. It follows that $[\mu]$ is preserved by every quasi-isometry of X.

In particular, all automorphisms of Gromov-hyperbolic groups are coarse-median preserving. Alternatively, it is not hard to prove this last fact directly, relying on the Morse lemma and the observation that group automorphisms are quasi-isometries with respect to any word metric.

Example 2.29 Equipping $\mathbb{Z}^n$ with the median operator μ associated to its ℓ^1 metric, we obtain a coarse median group $(\mathbb{Z}^n, [\mu])$. An automorphism $\varphi \in \operatorname{Aut} \mathbb{Z}^n = \mathrm{GL}_n\mathbb{Z}$ is coarse-median preserving if and only if it lies in the signed permutation group $O(n, \mathbb{Z}) \leq \mathrm{GL}_n\mathbb{Z}$, ie if it can be realised as an automorphism of the standard tiling of $\mathbb{R}^n$ by unit cubes. This will follow from Proposition A(3) once we prove it in Section 3.4 (though it also is easily shown by hand).

We end this subsection with the definitions of *quasiconvex* subsets and *approximate median subalgebras*, which will play an important role in Sections 3 and 4.

Definition 2.30 Let $(X, [\mu])$ be a coarse median space. A subset $A \subseteq X$ is *quasiconvex* if there exists $R \geq 0$ such that $\mu(A \times A \times X) \subseteq \mathcal{N}_R(A)$.

This notion is clearly independent of the chosen representative μ of the structure $[\mu]$. Moreover, by Definition 2.21(3), if subsets A and B have finite Hausdorff distance, then A is quasiconvex if and only if B is.

By Example 2.28, Definition 2.30 extends the usual notion of quasiconvexity in hyperbolic spaces. The next remark shows that this is also the notion of quasiconvexity appearing in the statement of Theorem C. We will discuss in Section 3.1 other equivalent notions of quasiconvexity in (nonhyperbolic) cube complexes.

Remark 2.31 Let G be a right-angled Artin/Coxeter group. Let $G \curvearrowright X$ be the action on the universal cover of the Salvetti/Davis complex and let $[\mu_X]$ be the induced coarse median structure on G, as in Remark 2.27. Recall that, for a subset $A \subseteq X^{(0)}$, the set $\mathcal{J}(A) = \mu_X(A \times A \times X)$ is the union of all geodesics joining points of A.

Since the standard Cayley graph of G is precisely the 1–skeleton of X, a subgroup $H \leq G$ is quasiconvex as defined in the statement of Theorem C if and only if we have $\mathcal{J}(H \cdot x) \subseteq \mathcal{N}_R(H \cdot x)$ for some $x \in X$ and $R \geq 0$. This is clearly equivalent to quasiconvexity of H with respect to the coarse median structure $[\mu_X]$.

Remark 2.32 If X is a finite-rank median space, then a subset $A \subseteq X$ is quasiconvex if and only if $d_{\mathrm{Haus}}(A, \mathrm{Hull}\, A) < +\infty$. This follows from Lemma 2.10.

A similar, weaker notion is that of *approximate median subalgebra*.

Definition 2.33 Let $(X, [\mu])$ be a coarse median space. A subset $A \subseteq X$ is an *approximate median subalgebra* if there exists $R \geq 0$ such that $\mu(A \times A \times A) \subseteq \mathcal{N}_R(A)$.

Again, the definition only depends on the structure $[\mu]$ and passes on to all subsets of X at finite Hausdorff distance from A. An analogue of Remark 2.32 also holds, but it is more complicated and will be discussed in Section 4.1.

If φ is a coarse-median preserving automorphism of a coarse median group $(G, [\mu])$, the fixed subgroup $\mathrm{Fix}\, \varphi \leq G$ is in general not quasiconvex (for instance, consider the automorphism of $\mathbb{Z}^2$ that swaps the standard generators). However, it is always an approximate median subalgebra, as the next two lemmas show. This will be important in the proof of Theorem B.

Lemma 2.34 *Let G be a finitely generated group and let d be a word metric on G. For every $\varphi \in \mathrm{Aut}\, G$, there exist functions $\zeta_1, \zeta_2 \colon \mathbb{N} \to \mathbb{R}_{>0}$, with ζ_1 linear, such that, for every $g \in G$,*

$$\zeta_1\big(d(g, \varphi(g))\big) \leq d(g, \mathrm{Fix}\, \varphi) \leq \zeta_2\big(d(g, \varphi(g))\big).$$

Proof For the first inequality, note that $\varphi \colon G \to G$ is C–bi-Lipschitz with respect to d, for some constant $C \geq 0$. If $g' \in \mathrm{Fix}\, \varphi$ is an element closest to g, we have

$$d(g, \varphi(g)) \leq d(g, g') + d(\varphi(g'), \varphi(g)) \leq (1 + C) \cdot d(g, g') = (1 + C) \cdot d(g, \mathrm{Fix}\, \varphi).$$

Thus, we can take $\zeta_1(t) := t/(1 + C)$.

Regarding the second inequality, suppose for the sake of contradiction that there does not exist a function ζ_2 so that it is satisfied. Then, there exist elements $g_n \in G$ with $d(g_n, \mathrm{Fix}\, \varphi) \to +\infty$, but $d(g_n, \varphi(g_n)) \leq D$ for some $D \geq 0$. Passing to a subsequence, we can assume that $\varphi(g_n) = g_n x$ for some $x \in G$ and all n. Thus $g_n g_m^{-1} \in \mathrm{Fix}\, \varphi$, hence $d(g_n, \mathrm{Fix}\, \varphi) = d(g_m, \mathrm{Fix}\, \varphi)$ for all $n, m \geq 0$, contradicting the fact that the distances $d(g_n, \mathrm{Fix}\, \varphi)$ diverge. □

Lemma 2.35 *Let $(G, [\mu])$ be a coarse median group. If $\varphi \in \mathrm{Aut}(G, [\mu])$, then $\mathrm{Fix}\, \varphi \leq G$ is an approximate median subalgebra.*

Proof Since $\varphi \in \mathrm{Aut}(G, [\mu])$, there is a constant C such that

$$\varphi(\mu(x, y, z)) \approx_C \mu(\varphi(x), \varphi(y), \varphi(z)) \quad \text{for all } x, y, z \in G.$$

Thus, if $x, y, z \in \mathrm{Fix}\,\varphi$, we have $\varphi(\mu(x, y, z)) \approx_C \mu(x, y, z)$. Lemma 2.34 gives a constant C' such that $d(\mu(x, y, z), \mathrm{Fix}\,\varphi) \le C'$ for all $x, y, z \in \mathrm{Fix}\,\varphi$, as required. □

2.7 UNE actions and groups

The following (seemingly novel) notion will play an important role in the proof of Theorem F, especially in Sections 6.2, 7.1 and 7.4.

Definition 2.36 Let G be a finitely generated group and let (X, d) be a (pseudo)metric space.

(1) An isometric action $G \curvearrowright X$ is *uniformly nonelementary* (*UNE*) if there exists a constant $c > 0$ with the following property. For every finite generating set $S \subseteq G$ and for all $x, y \in X$,

$$d(x, y) \le c \cdot [\tau_S^d(x) + \tau_S^d(y)].$$

We say that $G \curvearrowright X$ is *c–uniformly nonelementary* (*c–UNE*) when we need to specify c.

(2) An infinite group G is *UNE* if it admits a UNE, proper, cocompact action on a geodesic metric space.

The previous definition differs slightly from the one given in the introduction, but it is easily seen to be equivalent.

Remark 2.37 If G is infinite and an action $G \curvearrowright X$ is proper and cocompact, then there exists $\epsilon > 0$ such that, for every generating set $S \subseteq G$ and every $x \in X$, we have $\tau_S^d(x) \ge \epsilon$.

Along with the Milnor–Schwarz lemma, this can be used to show that a group is UNE if and only if *every* proper, cocompact action on a geodesic space is UNE. Equivalently, if the action of G on its locally finite Cayley graphs is UNE.

Example 2.38 (1) Nonelementary hyperbolic groups are UNE (for instance, this is implicitly shown in the last two paragraphs of the proof of [88, Lemme 3.1]).

(2) Fundamental groups of compact special cube complexes with finite centre are UNE. We will obtain this in Corollary 7.23.

(3) UNE groups have finite centre.

3 Cubical convex-cocompactness

This section is devoted to *convex-cocompact* subgroups of cocompactly cubulated groups (Definition 3.1). First, in Section 3.1, we discuss the relationship between convex-cocompactness and coarse median quasiconvexity. Then, Section 3.2 discusses basic properties of *cyclic*, convex-cocompact subgroups of RAAGs. Finally, Proposition A is proved in Section 3.4.

The reader who is not interested in the proofs of Theorems F and I can safely skip Section 3.3, which is devoted to some of the finer properties of convex-cocompact subgroups of RAAGs and is more technical. Its results will only be needed in Section 7.

3.1 Cubical convex-cocompactness in general

Let $G \curvearrowright X$ be a proper cocompact action on a CAT(0) cube complex. In particular, X is finite-dimensional and locally finite.

Definition 3.1 A subgroup $H \leq G$ is *convex-cocompact* in $G \curvearrowright X$ if there exists an H–invariant, convex subcomplex $C \subseteq X$ that is acted upon cocompactly by H.

Despite the similarity in terminology, we emphasise that the above is much weaker than the notion of "boundary convex-cocompactness" due to Cordes and Durham [34]. For instance, all convex-cocompact subgroups of RAAGs are free if we consider the Cordes–Durham notion [73], whereas every special group is a convex-cocompact subgroup of some RAAG acting on its Salvetti complex according to Definition 3.1; see [68].

Let $[\mu_X]$ be the coarse median structure on G induced by $G \curvearrowright X$ as in Remark 2.27. Recall that quasiconvex subsets of coarse median spaces were introduced in Definition 2.30. For the notion of H–essential core, see Remark 2.15 or [23, Section 3.3].

The following is just a restating of some well-known facts. The equivalence of the first two parts is due to Haglund; see [66, Theorem H] and [91].

Lemma 3.2 *The following are equivalent for a subgroup $H \leq G$:*

(1) *H is convex-cocompact in $G \curvearrowright X$.*
(2) *H is quasiconvex in $(G, [\mu_X])$.*
(3) *H is finitely generated and acts cocompactly on the H–essential core of $H \curvearrowright X$.*

Proof Let us begin with the equivalence of (1) and (2). Picking a vertex $v \in X$, condition (2) holds if and only if there exists a constant R' such that $m(H \cdot v, H \cdot v, G \cdot v) \subseteq \mathcal{N}_{R'}(H \cdot v)$. Since G acts cocompactly and m is 1–Lipschitz in each component, this is equivalent to the existence of R'' with

$$\mathcal{J}(H \cdot v) = m(H \cdot v, H \cdot v, X) \subseteq \mathcal{N}_{R''}(H \cdot v).$$

It is clear that this holds when (1) is satisfied, so (1) $\Longrightarrow$ (2).

Conversely, if (2) holds, then $H \cdot v$ is quasiconvex in X and Remark 2.32 implies that $\mathrm{Hull}(H \cdot v)$ is at finite Hausdorff distance from $H \cdot v$. Since X is locally finite, this means that H acts cocompactly on $\mathrm{Hull}(H \cdot v)$, hence H is convex-cocompact.

We now show the equivalence of (1) and (3). First, if $C \subseteq X$ is convex and H–invariant, the H–essential core of $H \curvearrowright X$ is a restriction quotient of C (as defined in Section 2.5). Thus, if H acts cocompactly on C, it also acts cocompactly on the H–essential core. Moreover, the action $H \curvearrowright C$ is proper and cocompact, which implies that H is finitely generated. This proves (1) $\Longrightarrow$ (3).

Conversely, let X' be the cubical subdivision. Since H is finitely generated and $H \curvearrowright X'$ has no inversions, the essential core of $H \curvearrowright X'$ embeds H–equivariantly as a convex subcomplex of X'; see Remark 2.15. This shows that (3) $\Longrightarrow$ (1). □

Recalling that automorphisms of G are bi-Lipschitz with respect to word metrics on G, the equivalence of (1) and (2) in Lemma 3.2 has the following straightforward consequence.

Corollary 3.3 *If $\varphi \in \mathrm{Aut}(G, [\mu_X])$, then a subgroup $H \leq G$ is convex-cocompact in $G \curvearrowright X$ if and only if $\varphi(H)$ is.*

Example 3.4 If G is Gromov-hyperbolic, then a subgroup $H \leq G$ is convex-cocompact in $G \curvearrowright X$ if and only if H is quasiconvex in G (again since (1) $\Longleftrightarrow$ (2) in Lemma 3.2). In particular, the notion of convex-cocompactness is independent of the chosen cubulation of G in this case. A quick look at the standard cubulation of $\mathbb{Z}^2$ immediately shows that the latter does not hold in general.

3.2 Label-irreducible elements in RAAGs

This subsection studies convex-cocompact *cyclic* subgroups of right-angled Artin groups. Let Γ be a finite simplicial graph. Let $\mathcal{A} = \mathcal{A}_\Gamma$ be a RAAG and $\mathcal{X} = \mathcal{X}_\Gamma$ the universal cover of its Salvetti complex. Set $r = \dim \mathcal{X}$.

The Cayley graph of $\mathcal{A}$ corresponding to the standard generating set $\Gamma^{(0)}$ is naturally identified with the 1–skeleton of the CAT(0) cube complex $\mathcal{X}$. Thus, every edge of $\mathcal{X}$ is labelled by a vertex of Γ. Observing that edges crossing the same hyperplane have the same label, we obtain a map $\gamma \colon \mathscr{W}(\mathcal{X}) \to \Gamma^{(0)}$.

We can apply the discussion in Section 2.4 to the standard action $\mathcal{A} \curvearrowright \mathcal{X}$ (or, to be precise, the action on the 0–skeleton of $\mathcal{X}$). Every element of $\mathcal{A}$ acts nontransversely and stably without inversions. For every $g \in \mathcal{A} \setminus \{1\}$, the reduced core $\overline{\mathcal{C}}(g)$ is the union of all axes of g.

A hyperplane of $\mathcal{X}$ lies in $\mathcal{W}_1(g)$ if and only if it is crossed by one (equivalently, all) axis of g. Hyperplanes lie in $\mathcal{W}_0(g)$ when they are preserved by g; equivalently, when they are transverse to all elements of $\mathcal{W}_1(g)$, or, again, when they separate two axes of g.

The factor $\bar{\mathcal{C}}_1(g)$ is $\langle g\rangle$–equivariantly isomorphic to the convex hull in $\mathcal{X}$ of any axis of g. The factor $\bar{\mathcal{C}}_0(g)$ is fixed pointwise by g and it is isomorphic to $\mathcal{X}_\Lambda$, where $\Lambda \subseteq \Gamma$ is the maximal subgraph all of whose vertices are joined by an edge to all vertices in $\gamma(\mathcal{W}_1(g))$.

For a simplicial graph Δ, we denote by Δ^o the *opposite* of Δ. This the graph that has the same vertex set as Δ and an edge between two vertices exactly when they are not connected by an edge in Δ.

Definition 3.5 Consider $g \in \mathcal{A} \setminus \{1\}$.

(1) We define $\Gamma(g) := \gamma(\mathcal{W}_1(g)) \subseteq \Gamma^{(0)}$. These are precisely the standard generators of $\mathcal{A}$ that appear in the cyclically reduced words representing elements conjugate to g.

(2) We say that g is *label-irreducible* if the full subgraph of Γ spanned by $\Gamma(g)$ does not split as a nontrivial join (ie its opposite graph is connected). Equivalently, g is contracting [26] within a parabolic subgroup of $\mathcal{A}$.

Two label-irreducible elements $g, h \in \mathcal{A}$ are *independent* if $\langle g, h\rangle \not\simeq \mathbb{Z}$. If g, h are independent and commute, then $\langle g, h\rangle \simeq \mathbb{Z}^2$. We will also use the following result of Servatius; see eg [94, Proposition III.1].

Lemma 3.6 *If $g, h \in \mathcal{A}$ are commuting, independent, label-irreducible elements, then every vertex of $\Gamma(g)$ is joined to every vertex of $\Gamma(h)$ by an edge of Γ.*

To each element $g \in \mathcal{A}$, we can associate a canonical collection of label-irreducible elements $g_1, \dots, g_k$, called the *label-irreducible components* of g, as shown in the next result.

Lemma 3.7 (label-irreducible components) *For every element $g \in \mathcal{A}$, the following hold.*

(1) *We can write $g = g_1 \cdot \dots \cdot g_k$ for pairwise-commuting, pairwise-independent label-irreducibles $g_i \in \mathcal{A}$. In addition, $0 \le k \le r$ and the g_i are unique up to permutation.*

(2) *The sets $\mathcal{W}_1(g_i)$ are transverse to each other and $\mathcal{W}_1(g_i) \subseteq \mathcal{W}_0(g_j)$ for $i \neq j$. In addition,*

$$\mathcal{W}_1(g) = \mathcal{W}_1(g_1) \sqcup \dots \sqcup \mathcal{W}_1(g_k), \quad \ell(g, \mathcal{X}) = \ell(g_1, \mathcal{X}) + \dots + \ell(g_k, \mathcal{X}),$$
$$\bar{\mathcal{C}}_1(g) \simeq \bar{\mathcal{C}}_1(g_1) \times \dots \times \bar{\mathcal{C}}_1(g_k), \qquad \bar{\mathcal{C}}(g) = \bar{\mathcal{C}}(g_1) \cap \dots \cap \bar{\mathcal{C}}(g_k).$$

(3) *Centralisers satisfy $Z_{\mathcal{A}}(g) = Z_{\mathcal{A}}(g_1) \cap \dots \cap Z_{\mathcal{A}}(g_k)$. Moreover, $Z_{\mathcal{A}}(g)$ splits as the direct product of a parabolic subgroup of $\mathcal{A}$ and a copy of $\mathbb{Z}^k$ freely generated by roots of $g_1, \dots, g_k$.*

Proof Since label-irreducibility is invariant under taking conjugates, we assume throughout the proof that g is cyclically reduced. If g is the identity, we can simply take $k = 0$ and the entire lemma holds trivially. Suppose instead that $g \neq 1$.

We begin with part (1). Let $\Lambda_1, \dots, \Lambda_k$ be the connected components of the subgraph of Γ^o spanned by $\Gamma(g)$. In Γ, every vertex of Λ_i is joined by an edge to every vertex of Λ_j with $j \neq i$. Thus, permuting the letters in a word representing g, we can write $g = g_1 \cdot \dots \cdot g_k$, where each g_i is cyclically reduced and $\Gamma(g_i) = \Lambda_i^{(0)}$. The elements g_i commute pairwise and, since each Λ_i is connected, they are all label-irreducible. It is clear that g_i and g_j are independent for $i \neq j$.

Uniqueness of the g_i up to permutations follows from the fact that, by Lemma 3.6, $\Gamma(g_1), \dots, \Gamma(g_k)$ must coincide with the vertex sets of $\Lambda_1, \dots, \Lambda_k$ in any such decomposition of g. Furthermore, choosing a vertex from each Λ_i, we obtain a k–clique in Γ, so $k \leq r$. This proves part (1).

We now prove part (2). Since g_i is cyclically reduced, there exists a (combinatorial) axis $\alpha_i \subseteq \mathcal{X}_{\Gamma(g_i)}$ through the identity. Note that the product $\mathcal{X}_{\Gamma(g_1)} \times \dots \times \mathcal{X}_{\Gamma(g_k)} \subseteq \mathcal{X}$ is preserved by all g_i and that each g_i leaves invariant every hyperplane of $\mathcal{X}_{\Gamma(g_j)}$ for all $j \neq i$. Thus, the sets $\mathcal{W}_1(g_i)$ are transverse to each other and $\mathcal{W}_1(g_i) \subseteq \mathcal{W}_0(g_j)$ for $i \neq j$. The equality $\mathcal{W}_1(g) = \mathcal{W}_1(g_1) \sqcup \dots \sqcup \mathcal{W}_1(g_k)$ now follows by observing that $\alpha_1 \times \dots \times \alpha_k$ contains an axis of g. The product splitting of $\overline{\mathcal{C}}_1(g)$ can be deduced from the transverse partition of $\mathcal{W}_1(g)$ using Lemma 2.6.

If Ω_i is a fundamental domain for the $\langle g_i \rangle$–action on $\mathcal{W}_1(g_i)$, the previous paragraph implies that $\Omega_1 \sqcup \dots \sqcup \Omega_k$ is a fundamental domain for the $\langle g \rangle$–action on $\mathcal{W}_1(g)$. Taking cardinalities, this shows that $\ell(g, \mathcal{X}) = \ell(g_1, \mathcal{X}) + \dots + \ell(g_k, \mathcal{X})$. Finally, the characterisation of $\overline{\mathcal{C}}(g)$ can be deduced from the fact that this is the set of points of $\mathcal{X}$ that realise the translation length $\ell(g, \mathcal{X})$.

We conclude with part (3). The inclusion $Z_{\mathcal{A}}(g_1) \cap \dots \cap Z_{\mathcal{A}}(g_k) \leq Z_{\mathcal{A}}(g)$ is clear. Conversely, if $h \in \mathcal{A}$ commutes with g, uniqueness in part (1) implies that the elements hg_ih^{-1} coincide with the g_i up to permutation. Since $\Gamma(hg_ih^{-1}) = \Gamma(g_i)$, it follows that $hg_ih^{-1} = g_i$ for each i. Hence $h \in Z_{\mathcal{A}}(g_1) \cap \dots \cap Z_{\mathcal{A}}(g_k)$, as required. The last statement is Servatius' centraliser theorem from [94, Section III]. □

Remark 3.8 For every $H \leq \mathcal{A}$, there exists a finite subset $F \subseteq H$ such that $Z_{\mathcal{A}}(H) = Z_{\mathcal{A}}(F)$.

Indeed, we have observed in Lemma 3.7(3) that the centraliser of every element of $\mathcal{A}$ splits as a product of a free abelian group and a parabolic subgroup of $\mathcal{A}$. It follows that every descending chain of centralisers of subsets of $\mathcal{A}$ eventually stabilises, since this is true of chains of parabolics.

We conclude this subsection by showing that label-irreducibles are precisely those elements $g \in \mathcal{A}$ such that the subgroup $\langle g \rangle$ is convex-cocompact in $\mathcal{A} \curvearrowright \mathcal{X}$. After a couple of preliminary results, this is shown below in Lemma 3.11.

Lemma 3.9 *Every connected full subgraph $\Lambda \subseteq \Gamma^o$ has diameter $\leq 2r - 1$.*

Proof Suppose towards a contradiction that there exist vertices $x, y \in \Lambda$ and a shortest path $\sigma \subseteq \Lambda$ joining them, with σ made up of $2r$ edges. Let σ_i be the i^{th} vertex of Γ^o met by σ, with $\sigma_0 = x$ and $\sigma_{2r} = y$. Since σ is shortest and Λ is full, no two of the $r+1$ vertices $\sigma_0, \sigma_2, \dots, \sigma_{2r}$ are joined by an edge of Γ^o. Thus, they form an $(r+1)$–clique in Γ, contradicting the fact that $r = \dim \mathcal{X}$. □

Lemma 3.10 *Let $g \in \mathcal{A}$ be label-irreducible. Then, for every $\mathfrak{u} \in \mathcal{W}_1(g)$, there exists a point $x \in \bar{\mathcal{C}}(g)$ such that $\mathscr{W}(x|gx) \subseteq \mathscr{W}(\mathfrak{u}|g^{4r-2}\mathfrak{u})$. In particular, $\gamma(\mathscr{W}(\mathfrak{u}|g^{4r-2}\mathfrak{u})) = \Gamma(g)$.*

Proof Pick a point y on an axis of g so that $\mathfrak{u} \in \mathscr{W}(y|gy)$. Set $x = g^{2r-1}y$ and consider a hyperplane $\mathfrak{w} \in \mathscr{W}(x|gx)$. Since g is label-irreducible, the full subgraph of Γ^o spanned by $\Gamma(g)$ is connected. By Lemma 3.9, there exists a sequence $\sigma_0 = \gamma(\mathfrak{u}), \sigma_1, \dots, \sigma_k = \gamma(\mathfrak{w})$ of vertices in $\Gamma(g)$ such that $k \le 2r-1$ and consecutive σ_i are not joined by an edge of Γ. Set $\sigma_j = \sigma_k$ for $k < j \le 2r-1$.

For $0 \le i \le 2r-1$, pick a hyperplane $\mathfrak{w}_i \in \mathscr{W}(g^i y|g^{i+1}y)$ with $\gamma(\mathfrak{w}_i) = \sigma_i$, making sure that $\mathfrak{w}_0 = \mathfrak{u}$ and $\mathfrak{w}_{2r-1} = \mathfrak{w}$. Since σ_i and σ_{i+1} are not joined by an edge, the hyperplanes $\mathfrak{w}_i$ and $\mathfrak{w}_{i+1}$ are not transverse. Since these hyperplanes are all crossed by an axis of g, we conclude that each $\mathfrak{w}_i$ separates the $\mathfrak{w}_j$ with $j < i$ from those with $j > i$. In particular, $\mathfrak{u}$ and $\mathfrak{w}$ are not transverse.

The same argument shows that $\mathfrak{w}$ and $g^{4r-2}\mathfrak{u}$ are not transverse, hence $\mathfrak{w} \in \mathscr{W}(\mathfrak{u}|g^{4r-2}\mathfrak{u})$. Since $\mathfrak{w} \in \mathscr{W}(x|gx)$ was arbitrary, we have shown that $\mathscr{W}(x|gx) \subseteq \mathscr{W}(\mathfrak{u}|g^{4r-2}\mathfrak{u})$. □

Lemma 3.11 (1) *If g is label-irreducible and $\alpha \subseteq \mathcal{X}$ is an axis, then $d_{\mathrm{Haus}}(\alpha, \mathrm{Hull}\,\alpha) \le (8r-4)\ell(g, \mathcal{X})$.*

(2) *An element $g \in \mathcal{A} \setminus \{1\}$ is label-irreducible if and only if $\langle g \rangle$ is convex-cocompact in $\mathcal{A} \curvearrowright \mathcal{X}$.*

Proof Assuming part (1), we first prove part (2). Using the third characterisation of convex-cocompactness in Lemma 3.2 and the fact that $\bar{\mathcal{C}}_1(g)$ is equivariantly isomorphic to $\mathrm{Hull}\,\alpha$, part (1) shows that label-irreducible elements are convex-cocompact. Conversely, if g is not label-irreducible, the nontrivial splitting of $\bar{\mathcal{C}}_1(g)$ provided by Lemma 3.7(2) implies that $\langle g \rangle$ cannot act cocompactly on $\bar{\mathcal{C}}_1(g)$.

Let us now prove part (1). Considering a point $p \in \mathrm{Hull}\,\alpha$, it is enough to obtain the inequality $d(p, \alpha) \le (8r-4)\ell(g, \mathcal{X})$.

Every element of $\mathscr{H}_{\mathrm{Hull}\,\alpha}(\mathcal{X})$ intersects α in a subray. Let $\mathcal{H}_+$ be the subset of halfspaces intersecting α in a positive subray (ie containing all points $g^n z$ with $n \ge 0$, for a suitable choice of $z \in \alpha$). Any two maximal halfspaces lying in $\mathcal{H}_+$ and not containing p are transverse. It follows that there are only finitely many such maximal halfspaces, which we denote by $\mathfrak{h}_1, \dots, \mathfrak{h}_k$.

A negative subray of α is contained in $\mathfrak{h}_1^* \cap \dots \cap \mathfrak{h}_k^*$, so we can pick a point $x \in \alpha \cap \mathfrak{h}_1^* \cap \dots \cap \mathfrak{h}_k^*$. In particular, x does not lie in any halfspaces of $\mathcal{H}_+$ that do not contain p; hence $\mathscr{H}(x|p) \subseteq \mathcal{H}_+$. Let $y \in \alpha$ be the point with $d(x, p) = d(x, y)$ and $\mathscr{H}(x|y) \subseteq \mathcal{H}_+$. Setting $m = m(x, p, y)$, we note that every $\mathfrak{j} \in \mathscr{H}(m|p)$ is transverse to every $\mathfrak{k} \in \mathscr{H}(m|y)$. Indeed, $m \in \mathfrak{j}^* \cap \mathfrak{k}^*$, $p \in \mathfrak{j} \cap \mathfrak{k}^*$ and $y \in \mathfrak{j}^* \cap \mathfrak{k}$, while $\mathfrak{j} \cap \mathfrak{k}$ is nonempty because $\mathfrak{j}$ and $\mathfrak{k}$ both lie in $\mathcal{H}_+$.

Now, suppose for the sake of contradiction that $d(p, y) > (8r-4)\ell(g, \mathcal{X})$. Since we chose y with $d(x, p) = d(x, y)$, we have $d(p, m) = d(m, y) > (4r-2)\ell(g, \mathcal{X})$. Now $\mathscr{W}(p|m) \subseteq \mathscr{W}_{\mathrm{Hull}\,\alpha}(\mathcal{X}) = \mathcal{W}_1(g)$, a set on which $\langle g^{4r-2} \rangle$ acts with exactly $(4r-2)\ell(g, \mathcal{X})$ orbits. Thus, there exists a hyperplane $\mathfrak{u} \in \mathscr{W}(p|m)$ such that $g^{4r-2}\mathfrak{u} \in \mathscr{W}(p|m)$. Lemma 3.10 implies that $\gamma(\mathscr{W}(p|m)) = \Gamma(g)$. Similarly, we obtain $\gamma(\mathscr{W}(m|y)) = \Gamma(g)$. This contradicts the fact that $\mathscr{W}(p|m)$ is transverse to $\mathscr{W}(m|y)$. □

3.3 More on convex-cocompactness in RAAGs

The results in this subsection will only be used in Section 7 and can be skipped by the reader uninterested in the proof of Theorems F and I.

First, we discuss additional properties of label-irreducible elements of RAAGs. Our aim is obtaining *uniform* control on the extent to which axes of distinct label-irreducibles can track each other. The main result here is Lemma 3.13, along with its direct consequence Corollary 3.14. Both results will be fundamental building blocks in the proof that centreless special groups are UNE.

Then, in the second part of the subsection, we study general convex-cocompact subgroups of RAAGs, proving only a couple of simple properties related to label-irreducible components.

3.3.1 Additional properties of label-irreducible elements We maintain the notation introduced at the beginning of Section 3.2. Recall that $r = \dim \mathcal{X}$.

Recall that the carrier of a hyperplane $\mathfrak{w} \in \mathscr{W}(\mathcal{X})$ is the smallest convex subcomplex of $\mathcal{X}$ that contains all edges crossing $\mathfrak{w}$. A hyperplane of $\mathcal{X}$ separates two points in the carrier of $\mathfrak{w}$ if and only if it is either equal or transverse to $\mathfrak{w}$. If two hyperplanes $\mathfrak{u}$ and $\mathfrak{w}$ have intersecting carriers, then they are transverse if and only if $\gamma(\mathfrak{u})$ and $\gamma(\mathfrak{w})$ are joined by an edge of Γ.

Lemma 3.12 *If $g, h \in \mathcal{A}$ and $\Gamma(g) \subseteq \gamma\big(\mathscr{W}_{\bar{C}(g)}(\mathcal{X}) \cap \mathscr{W}_{\bar{C}(h)}(\mathcal{X})\big)$, then $\bar{C}(g) \cap \bar{C}(h) \neq \varnothing$.*

Proof Suppose for the sake of contradiction that $\bar{C}(g)$ and $\bar{C}(h)$ are disjoint. Then there exists a hyperplane $\mathfrak{v}$ separating them, which we pick so that the carrier of $\mathfrak{v}$ intersects $\bar{C}(g)$. This guarantees that g admits an axis α that intersects the carrier of $\mathfrak{v}$.

Since $\mathfrak{v}$ separates $\bar{C}(g)$ and $\bar{C}(h)$, it is transverse to $\mathscr{W}_{\bar{C}(g)}(\mathcal{X}) \cap \mathscr{W}_{\bar{C}(h)}(\mathcal{X})$, so the vertex $\gamma(\mathfrak{v})$ is connected by an edge of Γ to all elements of $\Gamma(g)$. Observing that all hyperplanes crossed by α are labelled by elements of $\Gamma(g)$ and recalling that α intersects the carrier of $\mathfrak{v}$, we deduce that all hyperplanes crossed by α are transverse to $\mathfrak{v}$. In other words, $\mathfrak{v}$ is transverse to $\mathcal{W}_1(g)$, hence $\mathfrak{v} \in \mathcal{W}_0(g) \subseteq \mathscr{W}_{\bar{C}(g)}(\mathcal{X})$. This is the required contradiction. □

Lemma 3.13 *Let $g, h \in \mathcal{A}$ be label-irreducible. If there exist hyperplanes $\mathfrak{u}, \mathfrak{w} \in \mathscr{W}(\mathcal{X})$ such that $\{\mathfrak{u}, g^{4r}\mathfrak{u}, \mathfrak{w}, h^{4r}\mathfrak{w}\} \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(h)$, then $\langle g, h\rangle \simeq \mathbb{Z}$.*

Proof The proof will consist of three steps.

Step 1 *We can assume that $1 \in \mathcal{A} \cong \mathcal{X}^{(0)}$ lies in $\bar{C}(g) \cap \bar{C}(h)$, and that $\Gamma(g) = \Gamma(h) = \Gamma^{(0)}$.*

Since $\mathcal{W}_1(g) \cap \mathcal{W}_1(h)$ contains any hyperplane separating two of its elements, we have $\mathscr{W}(\mathfrak{u}|g^{4r-2}\mathfrak{u}) \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(h)$. Lemma 3.10 yields

$$\Gamma(g) = \gamma(\mathscr{W}(\mathfrak{u}|g^{4r-2}\mathfrak{u})) \subseteq \gamma(\mathcal{W}_1(g) \cap \mathcal{W}_1(h)) \subseteq \Gamma(h).$$

One the one hand, this allows us to apply Lemma 3.12 and deduce that $\bar{\mathcal{C}}(g)\cap\bar{\mathcal{C}}(h)\neq\varnothing$. On the other, this shows that $\Gamma(g)\subseteq\Gamma(h)$ and the inclusion $\Gamma(h)\subseteq\Gamma(g)$ is obtained similarly, so $\Gamma(g)=\Gamma(h)$.

Conjugating g and h by any $x\in\bar{\mathcal{C}}(g)\cap\bar{\mathcal{C}}(h)$, we can assume that $1\in\bar{\mathcal{C}}(g)\cap\bar{\mathcal{C}}(h)$. Equivalently, g and h are cyclically reduced, so they lie in the parabolic subgroup $\mathcal{A}_{\Gamma(g)}=\mathcal{A}_{\Gamma(h)}\leq\mathcal{A}$. Replacing $\mathcal{A}$ with $\mathcal{A}_{\Gamma(g)}$ does not alter the properties in the statement of the lemma, so we can assume that $\Gamma(g)=\Gamma(h)=\Gamma^{(0)}$.

Step 2 *Assume without loss of generality that $\ell(g,\mathcal{X})\leq\ell(h,\mathcal{X})$. Possibly replacing g and h with their inverses and conjugating them, there exists a geodesic $\sigma\subseteq\mathcal{X}$ from 1 to g such that*

- *the union $\rho:=\bigcup_{i\geq 0}g^i\sigma$ is a ray and contains h and h^2 (viewing $1,g,h,h^2$ as vertices of $\mathcal{X}$), and*
- *if $\tau\subseteq\rho$ is the arc joining 1 to h, then $h\cdot\tau$ is the arc of ρ joining h to h^2.*

Let $\mathfrak{k}\in\mathscr{H}(\mathcal{X})$ be a halfspace bounded by $h^{4r-2}\mathfrak{w}\in\mathcal{W}_1(g)\cap\mathcal{W}_1(h)$. Possibly replacing g and/or h with their inverses, we have $g\mathfrak{k}\subsetneq\mathfrak{k}$ and $h\mathfrak{k}\subsetneq\mathfrak{k}$. Since $\Gamma^{(0)}=\Gamma(h)$, Lemma 3.10 shows that $\mathfrak{w}$ and $h^{4r-2}\mathfrak{w}$ are strongly separated in $\mathcal{X}$.

The subray contained in $\mathfrak{k}^*$ of any (combinatorial) axis of g defines a point ξ in the Roller boundary $\partial\mathcal{X}$ such that $g\xi=\xi$ and $\xi\in h^{-4r+2}\mathfrak{k}^*$ (recall that this halfspace is bounded by $\mathfrak{w}$). Similarly, there exists $\eta\in\partial\mathcal{X}$ with $h\eta=\eta$ and $\eta\in h^{-4r+2}\mathfrak{k}^*$. Since the halfspaces $h^{-4r+2}\mathfrak{k}^*$ and $\mathfrak{k}$ are strongly separated, the gate-projections of ξ and η to $\mathfrak{k}$ coincide and they are a vertex $x\in\bar{\mathcal{C}}(g)\cap\bar{\mathcal{C}}(h)$. Conjugating g and h by x, we can assume that $x=1$.

Label $\mathfrak{k}_1\supsetneq\mathfrak{k}_2\supsetneq\cdots\supsetneq\mathfrak{k}_m$ the elements of $\mathscr{H}(1|h^2)$ bounded by hyperplanes with label $\gamma(\mathfrak{w})$. Set $\mathfrak{k}_0:=\mathfrak{k}$ and observe that $\mathfrak{k}_m=h^2\mathfrak{k}$, which is bounded by $h^{4r}\mathfrak{w}\in\mathcal{W}_1(g)\cap\mathcal{W}_1(h)$. In conclusion,

$$\xi,\eta\notin h^{-4r+2}\mathfrak{k}\supsetneq\mathfrak{k}=\mathfrak{k}_0\supsetneq\mathfrak{k}_1\supsetneq\cdots\supsetneq\mathfrak{k}_m=h^2\mathfrak{k}.$$

Note that the hyperplanes bounding the $\mathfrak{k}_i$ all lie in $\mathcal{W}_1(g)\cap\mathcal{W}_1(h)$. Since $1\in\bar{\mathcal{C}}(g)\cap\bar{\mathcal{C}}(h)$, there exist an axis of h and an axis of g each crossing all hyperplanes bounding the $\mathfrak{k}_i$. Hence there exist $1\leq t\leq s$ such that $g\mathfrak{k}_j=\mathfrak{k}_{j+t}$ for all $0\leq j\leq m-t$, and $h\mathfrak{k}_i=\mathfrak{k}_{i+s}$ for all $0\leq i\leq m-s$.

Let x_i be the gate-projection of $x=1$ to $\mathfrak{k}_i$. Note that this is also the gate-projection to $\mathfrak{k}_i$ of ξ and η. Since $g\xi=\xi$ and $h\eta=\eta$, we must have $gx_j=x_{j+t}$ and $hx_i=x_{i+s}$ for all $1\leq j\leq m-t$ and $1\leq i\leq m-s$. In particular, since $x_0=1$, we have $h=x_s$, $h^2=x_{2s}=x_m$ and $g=x_t$.

Observe that each x_i is also the gate-projection to $\mathfrak{k}_i$ of each x_j with $j<i$. Thus, we can construct a (combinatorial) geodesic σ from 1 to g by concatenating arbitrary geodesics σ_j from x_j to x_{j+1} for $0\leq j<t$. The union $\rho=\bigcup_{i\geq 0}g^i\sigma$ is a ray since $1\in\bar{\mathcal{C}}(g)$. Let $k,l\geq 1$ be the integers with $0\leq s-kt<t$ and $0\leq 2s-lt<t$. Since σ contains the points $g^{-k}h=x_{s-kt}$ and $g^{-l}h^2=x_{2s-lt}$, it is clear that h and h^2 lie on the ray ρ.

Finally, note that we can choose the geodesics σ_j so that the following compatibility condition is satisfied: *whenever there exist $f\in\mathcal{A}$ and $0\leq i,j<t$ with $fx_i=x_j$ and $fx_{i+1}=x_{j+1}$, we have $f\sigma_i=\sigma_j$.* This

is possible because the action $\mathcal{A} \curvearrowright \mathcal{X}$ is free and so the element f is uniquely determined by i and j (when it exists). Now, given $0 \le j < s$, the arc of the ray ρ joining x_{s+j} to x_{s+j+1} is precisely $g^{a_j}\sigma_{b_j}$, where $s+j = a_j t + b_j$ and $0 \le b_j < t$. The element $g^{-a_j}h$ maps x_j and x_{j+1} to x_{b_j} and x_{b_j+1}, so it takes σ_j to σ_{b_j} by our construction. Thus $h\sigma_j = g^{a_j}\sigma_{b_j}$ is contained in ρ for every $0 \le j < s$. This proves the second condition in the statement of Step 2.

Step 3 *We have $\langle g, h\rangle \simeq \mathbb{Z}$.*

Let $S \cong \Gamma^{(0)}$ be the standard generating set of $\mathcal{A}$. Let $F(S)$ be the free group freely generated by S, and let $\pi : F(S) \to \mathcal{A}$ be the surjective homomorphism that takes each generator of $F(S)$ to the corresponding standard generator of $\mathcal{A}$. Let $w_g \in F(S)$ be the word spelled by the labels of the edges met moving from 1 to g along the geodesic σ. Let $w_h \in F(S)$ be the word spelled moving from 1 to h along the ray $\rho = \bigcup_{i\ge 0} g^i\sigma$. It is clear that $\pi(w_g) = g$ and $\pi(w_h) = h$.

From Step 2, we have $w_h = w_g^p a$ for some $p \ge 1$ and an initial subword a of w_g, and $w_h^2 = w_g^{p+1}ab$ for some word b such that $w_g^{p+1}ab$ is reduced in $F(S)$. It follows that $w_g^p a w_g^p a = w_g^{p+1}ab$ in $F(S)$, where both sides of the equality are reduced words. Looking at the first $(p+1)|w_g| + |a|$ letters on the left, we deduce that $aw_g = w_g a$. Hence $\langle w_g, w_h\rangle = \langle w_g, a\rangle$ is a cyclic subgroup of $F(S)$. We conclude that $\langle g, h\rangle = \pi\left(\langle w_g, w_h\rangle\right) \simeq \mathbb{Z}$. □

Corollary 3.14 *Consider two elements $g, h \in \mathcal{A}$. Suppose that g is label-irreducible. Assume in addition that* ***one*** *of the following conditions is satisfied*:

(1) *There exists $\mathfrak{w} \in \mathcal{W}_1(g)$ such that h preserves $\mathfrak{w}$ and $g^{4r}\mathfrak{w}$.*

(2) *There exist hyperplanes $\mathfrak{u}, \mathfrak{w} \in \mathscr{W}(\mathcal{X})$ with $\{\mathfrak{u}, \mathfrak{w}, h^{4r}\mathfrak{u}, g^{4r}\mathfrak{w}\} \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(h)$.*

Then g and h commute in $\mathcal{A}$.

Proof Assume first that there exists $\mathfrak{w} \in \mathcal{W}_1(g)$ such that $\mathfrak{w}$ and $g^{4r}\mathfrak{w}$ are preserved by h. Then $\{\mathfrak{w}, g^{4r}\mathfrak{w}\} = \{\mathfrak{w}, (hgh^{-1})^{4r}\mathfrak{w}\} \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(hgh^{-1})$. Since g and hgh^{-1} are label-irreducible, Lemma 3.13 implies that $\langle g, hgh^{-1}\rangle \simeq \mathbb{Z}$. Observing that $\ell(g, \mathcal{X}) = \ell(hgh^{-1}, \mathcal{X})$, we deduce that hgh^{-1} must coincide with either g or g^{-1}. The second option cannot occur in a right-angled Artin group, hence $hgh^{-1} = g$, as required.

Suppose now that there exist hyperplanes $\mathfrak{u}, \mathfrak{w}$ with $\{\mathfrak{u}, \mathfrak{w}, h^{4r}\mathfrak{u}, g^{4r}\mathfrak{w}\} \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(h)$. In light of Lemma 3.7(2), there exist (possibly equal) irreducible components h_1, h_2 of h, such that $\{\mathfrak{u}, g^{4r}\mathfrak{u}\} \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(h_1)$ and $\{\mathfrak{w}, h^{4r}\mathfrak{w}\} = \{\mathfrak{w}, h_2^{4r}\mathfrak{w}\} \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(h_2)$.

Since g is label-irreducible and $\gamma(\mathscr{W}(\mathfrak{u}|g^{4r}\mathfrak{u})) = \Gamma(g)$ by Lemma 3.10, no element of $\mathcal{W}_1(g)$ can be transverse to both $\mathfrak{u}$ and $g^{4r}\mathfrak{u}$. Hence $h_1 = h_2$, otherwise $\mathcal{W}_1(h_1)$ and $\mathcal{W}_1(h_2)$ would be transverse. Thus $\{\mathfrak{u}, g^{4r}\mathfrak{u}, \mathfrak{w}, h_2^{4r}\mathfrak{w}\} \subseteq \mathcal{W}_1(g) \cap \mathcal{W}_1(h_2)$ and Lemma 3.13 yields $\langle g, h_2\rangle \simeq \mathbb{Z}$. Now, a power of g coincides with a power of h_2, hence it commutes with h. It follows that g and h commute. □

We conclude with the following lemma, which is actually independent from the notion of label-irreducibility and from the discussion in the rest of this subsection, albeit in a similar spirit.

Lemma 3.15 *Consider elements $g_1, g_2 \in \mathcal{A}$ and vertices $x_1, x_2 \in \mathcal{X}$ such that the two sets $\mathscr{W}(x_i|g_i x_i)$ are transverse. Then g_1 and g_2 commute and we have $\mathcal{W}_1(g_i) \subseteq \mathcal{W}_0(g_j)$ for $i \neq j$.*

Proof We begin with the following observation:

Claim *For $\mathfrak{w} \in \mathscr{W}(\mathcal{X})$ and $x, y \in \mathcal{X}$, the hyperplane $\mathfrak{w}$ is transverse to $\mathscr{W}(x|y)$ if and only if every vertex in the set $\gamma(\mathscr{W}(x|y))$ is joined by an edge of Γ to every vertex in the set $\{\gamma(\mathfrak{w})\} \cup \gamma(\mathscr{W}(x|\mathfrak{w}))$.*

Proof Suppose first that $\mathfrak{w}$ is transverse to $\mathscr{W}(x|y)$. Then, since every hyperplane in $\mathscr{W}(\mathfrak{w}|x)$ is disjoint from $\mathfrak{w}$, we have $\mathscr{W}(\mathfrak{w}|x) \subseteq \mathscr{W}(\mathfrak{w}|x, y)$. Since $\mathscr{W}(x|y)$ is transverse to $\mathscr{W}(\mathfrak{w}|x, y)$, it is also transverse to $\{\mathfrak{w}\} \cup \mathscr{W}(\mathfrak{w}|x)$. It follows that every vertex in $\gamma(\mathscr{W}(x|y))$ is joined by an edge to every vertex in $\{\gamma(\mathfrak{w})\} \cup \gamma(\mathscr{W}(x|\mathfrak{w}))$, as required.

Suppose now instead that $\mathfrak{w}$ is disjoint from a hyperplane $\mathfrak{u} \in \mathscr{W}(x|y)$. Choosing $\mathfrak{u}$ so that it is closest to $\mathfrak{w}$, we can assume that no hyperplane of $\mathscr{W}(x|y)$ separates $\mathfrak{w}$ and $\mathfrak{u}$. If the carriers of $\mathfrak{w}$ and $\mathfrak{u}$ intersect, then $\gamma(\mathfrak{w})$ and $\gamma(\mathfrak{u})$ cannot be joined by an edge, as $\mathfrak{u}$ and $\mathfrak{w}$ are disjoint. Otherwise, there exists a hyperplane $\mathfrak{v} \in \mathscr{W}(\mathfrak{u}|\mathfrak{w})$ such that its carrier intersects the carrier of $\mathfrak{u}$; in particular, $\gamma(\mathfrak{u})$ and $\gamma(\mathfrak{v})$ are not joined by an edge. Since $\mathfrak{v}$ does not separate x and y, we must have $\mathfrak{v} \in \mathscr{W}(\mathfrak{w}|x, y)$, hence $\gamma(\mathfrak{v}) \in \gamma(\mathscr{W}(x|\mathfrak{w}))$, as required. ◁

Consider for a moment $g \in \mathcal{A}$, $x \in \mathcal{X}$ and $n \geq 1$. Since $\mathscr{W}(x|g^n x)$ is contained in the union $\mathscr{W}(x|gx) \cup \cdots \cup \mathscr{W}(g^{n-1}x|g^n x)$, we have $\gamma(\mathscr{W}(x|g^n x)) \subseteq \gamma(\mathscr{W}(x|gx))$. Thus, the claim implies that a hyperplane $\mathfrak{w}$ is transverse to $\mathscr{W}(x|gx)$ if and only if it is transverse to $\bigcup_{n\in\mathbb{Z}} \mathscr{W}(x|g^n x)$.

Now, consider the situation in the statement of the lemma. If x_i' is the gate-projection of x_i to $\overline{\mathcal{C}}(g_i)$, we have $\mathscr{W}(x_i'|g_i x_i') \subseteq \mathscr{W}(x_i|g_i x_i)$ and $\mathcal{W}_1(g_i) = \bigcup_{n\in\mathbb{Z}} \mathscr{W}(x_i'|g_i^n x_i')$. It follows that the sets $\mathcal{W}_1(g_1)$ and $\mathcal{W}_1(g_2)$ are transverse, or, equivalently, $\mathcal{W}_1(g_i) \subseteq \mathcal{W}_0(g_j)$ for $i \neq j$. This implies that g_1 and g_2 commute (for instance, by decomposing g_i into label-irreducible components as in Lemma 3.7 and applying Corollary 3.14). □

3.3.2 Convex-cocompact subgroups of RAAGs Again, we keep the notation from Section 3.2. We will simply say that a subgroup $G \leq \mathcal{A}$ is *convex-cocompact* when G is convex-cocompact for the action $\mathcal{A} \curvearrowright \mathcal{X}$ (in the sense of Definition 3.1).

Lemma 3.16 *Let $G \leq \mathcal{A}$ be convex-cocompact. If $g \in G$ and $g = a_1 \cdot \cdots \cdot a_k$ is its decomposition into label-irreducible components $a_i \in \mathcal{A}$, then there exists $m \geq 1$ such that all a_i^m lie in G.*

Proof Let $A \leq G$ be a free abelian subgroup containing a power of g, such that no finite-index subgroup of A is contained in a free abelian subgroup of G of higher rank. Since G is convex-cocompact, Theorem 3.6 in [100] shows that there exists a convex, A–invariant, A–cocompact subcomplex $Y \subseteq \mathcal{X}$ that splits as a product $L_1 \times \cdots \times L_p$, where $A \simeq \mathbb{Z}^p$ and each L_i is a quasiline. Replacing each L_i with a subcomplex, we can assume that all quasilines are A–essential.

Note that Y must contain an axis of g in $\mathcal{X}$, hence its convex hull, which is isomorphic to:

$$\bar{\mathcal{C}}_1(g) = \bar{\mathcal{C}}_1(a_1) \times \cdots \times \bar{\mathcal{C}}_1(a_k).$$

Since each a_i is label-irreducible, Lemma 3.11 shows that $\bar{\mathcal{C}}_1(a_i)$ is an irreducible quasiline. Up to permuting the factors of Y, we can thus assume that $L_i \simeq \bar{\mathcal{C}}_1(a_i)$ for $1 \leq i \leq k$, where $k \leq p$.

Since the L_i are locally finite, none of the groups $\operatorname{Aut} L_i$ contains subgroups isomorphic to $\mathbb{Z}^2$. It follows that every projection of $\mathbb{Z}^p \simeq A \leq \prod_i \operatorname{Aut} L_i$ to a product of $(p-1)$ factors must have nontrivial kernel. Equivalently, there exist elements $h_i \in A$ such that h_i acts loxodromically on L_i, and fixes pointwise each L_j with $j \neq i$. For each $1 \leq i \leq k$, the elements h_i and a_i stabilise a common copy of $L_i \simeq \bar{\mathcal{C}}_1(a_i)$ inside Y, and act freely and cocompactly on it. It follows that h_i and a_i are commensurable, hence a power of a_i lies in $A \leq G$. This concludes the proof. □

The exponent m in Lemma 3.16 can be chosen independently of $g \in G$ due to the following.

Remark 3.17 Suppose that $G \leq \mathcal{A}$ is convex-cocompact and, more precisely, that there exists a G–invariant, convex subcomplex $Y \subseteq \mathcal{X}$ such that the action $G \curvearrowright Y^{(0)}$ has q orbits. Then, for every $g \in \mathcal{A}$ such that $\langle g \rangle \cap G \neq \{1\}$, there exists $1 \leq k \leq q$ such that $g^k \in G$.

Indeed, consider $N \geq 1$ such that $g^N \in G$. Since Y is G–invariant and acted upon without inversions, it contains an axis α for g^N; see [67]. Every axis of a power of g is, in fact, also an axis of g (this property is specific to the action $\mathcal{A} \curvearrowright \mathcal{X}$). Thus, picking any $x \in \alpha$, we have $g^i x \in Y$ for all $i \in \mathbb{Z}$. Hence there exist $0 \leq i < j \leq q$ such that $g^i x$ and $g^j x$ are in the same G–orbit. Since $\mathcal{A}$ acts freely on $\mathcal{X}$, we have $g^{j-i} \in G$ and $0 < j - i \leq q$.

3.4 CMP automorphisms of right-angled groups

This subsection is devoted to the proof of Proposition A. Automorphisms of hyperbolic groups were already discussed in Example 2.28, so we are only concerned with right-angled Artin/Coxeter groups.

Let Γ be a finite simplicial graph. Let $\mathcal{A} = \mathcal{A}_\Gamma$ and $\mathcal{W} = \mathcal{W}_\Gamma$ be, respectively, the right-angled Artin group and the right-angled Coxeter group defined by Γ.

We identify with $\Gamma^{(0)}$ the standard generating sets of $\mathcal{A}$ and $\mathcal{W}$. The standard Cayley graphs of $\mathcal{A}$ and $\mathcal{W}$ are 1–skeleta of CAT(0) cube complexes: the universal covers of the Salvetti and Davis complex, respectively. Thus, $\mathcal{A}$ and $\mathcal{W}$ are each endowed with a natural median operator μ_Γ.

Remark 3.18 We have $g \cdot \mu_\Gamma(x,y,z) = \mu_\Gamma(gx,gy,gz)$ for all elements g,x,y,z in $\mathcal{A}$ or $\mathcal{W}$. This implies that $(\mathcal{A},[\mu_\Gamma])$ and $(\mathcal{W},[\mu_\Gamma])$ are coarse median groups, in the sense of Definition 2.24.

Unlike hyperbolic groups, $\mathcal{A}$ and $\mathcal{W}$ can admit infinitely many different coarse median structures. For this reason, we will never omit the subscript in μ_Γ, in order to emphasise that this is the coarse median structure provided by the *standard* generating set of $\mathcal{A}$ or $\mathcal{W}$. Other Artin/Coxeter generating sets can a priori give different coarse median structures; it will be a consequence of Proposition A(2) that this does not actually happen in the Coxeter case.

It was shown by Laurence [75], Servatius [94] and Corredor and Gutierrez [35] that $\operatorname{Aut}\mathcal{A}$ and $\operatorname{Aut}\mathcal{W}$ are generated by finitely many *elementary automorphisms*. These take the same form in both cases.

- **Graph automorphisms** Every automorphism of the graph Γ gives a permutation of the standard generating sets that defines an automorphism of $\mathcal{A}$ and $\mathcal{W}$.
- **Inversions** ι_v for each $v \in \Gamma^{(0)}$. We have $\iota_v(v) = v^{-1}$ and $\iota_v(u) = u$ for all $u \in \Gamma^{(0)} \setminus \{v\}$.
- **Partial conjugations** $\kappa_{w,C}$ for $w \in \Gamma^{(0)}$ and a connected component C of $\Gamma \setminus \operatorname{st} w$. We have $\kappa_{w,C}(u) = w^{-1}uw$ if $u \in C^{(0)}$ and $\kappa_{w,C}(u) = u$ if $u \in \Gamma^{(0)} \setminus C$.
- **Transvections** $\tau_{v,w}$ for $v,w \in \Gamma^{(0)}$ with $\operatorname{lk} v \subseteq \operatorname{st} w$. They are defined by $\tau_{v,w}(v) = vw$ and $\tau_{v,w}(u) = u$ for all $u \in \Gamma^{(0)} \setminus \{v\}$.
  We refer to $\tau_{v,w}$ as a *fold* if v and w are not joined by an edge (equivalently, $\operatorname{lk} v \subseteq \operatorname{lk} w$), and as a *twist* if v and w are joined by an edge (equivalently, $\operatorname{st} v \subseteq \operatorname{st} w$).

Remark 3.19 Graph automorphisms and inversions can be realised as automorphisms of the Salvetti/Davis complex, so they preserve the operator μ_Γ (hence the coarse median structure $[\mu_\Gamma]$).

In the case of right-angled Artin groups, the following class of automorphisms was introduced by Day [41] and Charney, Stambaugh and Vogtmann [25].

Definition 3.20 An automorphism $\varphi \in \operatorname{Aut}\mathcal{A}$ is *untwisted* if it lies in the subgroup $U(\mathcal{A}) \leq \operatorname{Aut}\mathcal{A}$ generated by graph automorphisms, inversions, partial conjugations and folds.

We now proceed to prove parts (2) and (3) of Proposition A. We will treat separately the Coxeter and Artin cases, as the simplest arguments appear to be quite different in spirit. Still, in both situations, the following basic observation is important.

Remark 3.21 Let a group G act properly and cocompactly on two CAT(0) cube complexes X and Y. Let $[\mu_X]$ and $[\mu_Y]$ be the induced coarse median structures on G. If there exists an equivariant restriction-quotient map $\pi : X \to Y$, then $[\mu_X] = [\mu_Y]$. This is immediate from the third characterisation of restriction quotients in Proposition 2.20.

3.4.1 The Coxeter case Here our aim is to prove that all elements of $\operatorname{Aut}\mathcal{W}$ preserve the coarse median structure $[\mu_\Gamma]$. We will achieve this by showing that all elementary automorphisms of $\mathcal{W}$ restrict to graph automorphisms on certain finite-index Coxeter subgroups of $\mathcal{W}$. This guarantees that they are all coarse-median preserving.

Given a vertex $w \in \Gamma$, we denote by $\Delta(\Gamma, w)$ the *double* of $\Gamma \setminus \{w\}$ along $\operatorname{lk} w$. More precisely, $\Delta(\Gamma, w)$ is obtained from two disjoint copies of the graph $\Gamma \setminus \{w\}$ by identifying the two subgraphs corresponding to $\operatorname{lk} w$. We continue to denote by $\operatorname{lk} w$ the resulting subgraph of $\Delta(\Gamma, w)$, even though w does not appear in $\Delta(\Gamma, w)$ and so this is not the link of any vertex of $\Delta(\Gamma, w)$.

Let $\alpha_w : \mathcal{W} \to \mathbb{Z}/2\mathbb{Z}$ be the homomorphism that maps w to the generator of $\mathbb{Z}/2\mathbb{Z}$, and all other standard generators of $\mathcal{W}$ to the identity.

Lemma 3.22 *Consider a vertex $w \in \Gamma$. Then:*

(1) $\ker\alpha_w$ *is generated by* $\{x \mid x \in \Gamma \setminus \{w\}\} \sqcup \{w^{-1}yw \mid y \in \Gamma \setminus \operatorname{st} w\}$.

(2) *This is a Coxeter generating set giving an isomorphism between* $\ker\alpha_w$ *and* $\mathcal{W}_{\Delta(\Gamma,w)}$.

(3) *The coarse median structure* $[\mu_{\Delta(\Gamma,w)}]$ *induced on* $\ker\alpha_w$ *by this isomorphism with* $\mathcal{W}_{\Delta(\Gamma,w)}$ *coincides with the restriction of the coarse median structure* $[\mu_\Gamma]$ *on* $\mathcal{W}$.

Proof The first two parts are a straightforward application of the normal form for words in Coxeter groups; see eg [39, Chapter 3.4]. We instead focus on part (3).

Let $\mathcal{Y}_\Gamma$ and $\mathcal{Y}_{\Delta(\Gamma,w)}$ be the universal covers of the Davis complexes of $\mathcal{W}$ and $\mathcal{W}_{\Delta(\Gamma,w)}$. We aim to show that, under the above identification between $\ker\alpha_w$ and $\mathcal{W}_{\Delta(\Gamma,w)}$, the standard action $\mathcal{W}_{\Delta(\Gamma,w)} \curvearrowright \mathcal{Y}_{\Delta(\Gamma,w)}$ is a restriction quotient of the action $\ker\alpha_w \curvearrowright \mathcal{Y}_\Gamma$. This proves the lemma, since, by Remark 3.21, the two actions then induce the same coarse median structure on $\ker\alpha_w$.

First, if $\Omega \subseteq \mathcal{Y}_\Gamma$ is a fundamental domain for the $\mathcal{W}$–action, note that $\Omega \cup w\Omega$ is a fundamental domain for $\ker\alpha_w \curvearrowright \mathcal{Y}_\Gamma$. In addition, observe that $\ker\alpha_w$ contains the entire $\mathcal{W}$–stabiliser of a hyperplane $\mathfrak{u} \in \mathscr{W}(\mathcal{Y}_\Gamma)$ precisely when $\gamma(\mathfrak{u}) \notin \operatorname{st} w$. Thus, the orbit $\mathcal{W} \cdot \mathfrak{u}$ is made up of two $(\ker\alpha_w)$–orbits of hyperplanes when $\gamma(\mathfrak{u}) \notin \operatorname{st} w$, while it is a single $(\ker\alpha_w)$–orbit when $\gamma(\mathfrak{u}) \in \operatorname{st} w$. Combining these two observations, the reader should convince themselves that, starting with the action $\ker\alpha_w \curvearrowright \mathcal{Y}_\Gamma$ and collapsing the single orbit of hyperplanes $\mathfrak{u}$ with $\gamma(\mathfrak{u}) = w$, we obtain precisely the action $\mathcal{W}_{\Delta(\alpha,w)} \curvearrowright \mathcal{Y}_{\Delta(\alpha,w)}$, as required. □

Proposition 3.23 *All automorphisms of $\mathcal{W}$ preserve the coarse median structure* $[\mu_\Gamma]$.

Proof Recall that $\operatorname{Aut}\mathcal{W}$ is generated by graph automorphisms, partial conjugations and transvections, as defined above. We have already noticed in Remark 3.19 that graph automorphisms are coarse-median preserving. We make two additional observations.

- *Every partial conjugation* $\kappa_{w,C}$ *preserves the subgroup* $\ker \alpha_w \leq \mathcal{W}$*. The restriction of* $\kappa_{w,C}$ *to* $\ker \alpha_w$ *is a graph automorphism with respect to the identification* $\ker \alpha_w \simeq \mathcal{W}_{\Delta(\Gamma,w)}$ *constructed in Lemma 3.22.*

Indeed, the connected component $C \subseteq \Gamma \setminus \operatorname{st} w$ gets doubled to two connected components C', C'' of the graph $\Delta(\Gamma, w) \setminus \operatorname{lk} w$. These two subgraphs correspond to the sets of generators $C^{(0)}$ and $w^{-1}C^{(0)}w$ for $\ker \alpha_w$. The automorphism $\kappa_{w,C}$ swaps these two sets of generators, while fixing all generators of $\ker \alpha_w$ corresponding to vertices of $\Delta(\Gamma, w) \setminus (C' \cup C'')$. This is realised by an automorphism of the graph $\Delta(\Gamma, w)$.

- *Every transvection* $\tau_{v,w}$ *preserves* $\ker \alpha_v \leq \mathcal{W}$*. The restriction of* $\tau_{v,w}$ *to* $\ker \alpha_v$ *is a product of partial conjugations with respect to the identification* $\ker \alpha_v \simeq \mathcal{W}_{\Delta(\Gamma,v)}$*.*

Indeed, if $x \in \Gamma \setminus \{v\}$, we have $\tau_{v,w}(x) = x$ and $\tau_{v,w}(v^{-1}xv) = w^{-1}(v^{-1}xv)w$. Setting $A := \Gamma \setminus \operatorname{st} v \subseteq \Gamma$, we have $\Delta(\Gamma, v) = \operatorname{lk} v \sqcup A' \sqcup A''$, where the subgraphs A', A'' correspond, respectively, to the subsets $A^{(0)}, v^{-1}A^{(0)}v \subseteq \mathcal{W}$. Let $w' \in A'$ be the vertex originating from $w \in \Gamma$. The set A'' is a union of connected components of $\Delta(\Gamma, v) \setminus \operatorname{lk} v$. Since $\operatorname{lk} v \subseteq \operatorname{st} w$ in Γ, we have $\operatorname{lk} v \subseteq \operatorname{st} w'$ in $\Delta(\Gamma, v)$, hence A'' is also a union of connected components of $\Delta(\Gamma, v) \setminus \operatorname{st} w'$. The composition of the partial conjugations $\kappa_{w',C} \in \operatorname{Aut} \mathcal{W}_{\Delta(\Gamma,v)}$, as C ranges through these connected components, is precisely the restriction of $\tau_{v,w}$ to $\ker \alpha_v$.

Now, in view of Lemma 3.22 and Remark 3.19, partial conjugations and transvections each preserve the restriction of the coarse median structure $[\mu_\Gamma]$ to a finite-index subgroup of $\mathcal{W}$. Since finite-index subgroups are coarsely dense in $\mathcal{W}$, this implies that these automorphisms actually preserve $[\mu_\Gamma]$ itself, proving the proposition. □

3.4.2 The Artin case We begin by showing that *untwisted* automorphisms of $\mathcal{A}$ preserve the coarse median structure $[\mu_\Gamma]$. I present a proof that was suggested to me by Ric Wade, as it is much simpler than my original brute-force argument.

The main ingredient is the construction of Salvetti blowups from the work of Charney, Stambaugh and Vogtmann [25], which we record in the following lemma. Restriction quotients were discussed in Section 2.5.

Lemma 3.24 *Let* $\varphi \in U(\mathcal{A}_\Gamma)$ *be a fold or partial conjugation. Then there exists a proper cocompact action on a* CAT(0) *cube complex* $\mathcal{A}_\Gamma \curvearrowright Z$ *and two restriction-quotient maps* $\pi_1, \pi_2 \colon Z \to \mathcal{X}_\Gamma$ *such that, for all* $g \in \mathcal{A}_\Gamma$*, we have* $\pi_1 \circ g = g \circ \pi_1$ *and* $\pi_2 \circ g = \varphi(g) \circ \pi_2$*.*

Proof This holds more generally when φ is a Γ–*Whitehead automorphism*, as defined at the beginning of [25, Section 2.3]. Our statement is a straightforward rephrasing of [25, Lemma 3.2] in terms of universal covers. □

Corollary 3.25 *Automorphisms in* $U(\mathcal{A})$ *preserve the coarse median structure* $[\mu_\Gamma]$*.*

Proof Let $\varphi \in U(\mathcal{A}_\Gamma)$ be a fold or partial conjugation. Let the action $\mathcal{A}_\Gamma \curvearrowright Z$ and maps $\pi_1, \pi_2 \colon Z \to \mathcal{X}_\Gamma$ be as provided by Lemma 3.24, and let $[\mu_Z]$ be the coarse median structure on $\mathcal{A}_\Gamma$ induced by Z. Since π_1 is $\mathcal{A}_\Gamma$–equivariant, Remark 3.21 guarantees that $[\mu_\Gamma] = [\mu_Z]$.

On the other hand, π_2 becomes $\mathcal{A}_\Gamma$–equivariant if we endow $\mathcal{X}_\Gamma$ with the φ–twisted action: using the notation from Remark 2.27, this corresponds to replacing $\mathcal{X}_\Gamma$ with $\mathcal{X}_\Gamma^{\varphi^{-1}}$, which induces the coarse median structure $(\varphi^{-1})_*[\mu_\Gamma]$ on $\mathcal{A}_\Gamma$. Thus, another application of Remark 3.21 yields $(\varphi^{-1})_*[\mu_\Gamma] = [\mu_Z]$. We conclude that $\varphi_*[\mu_\Gamma] = [\mu_\Gamma]$.

This shows that all folds and partial conjugations preserve $[\mu_\Gamma]$. Graph automorphisms and inversions are also coarse-median preserving, by Remark 3.19. Since these four types of elementary automorphisms generate $U(\mathcal{A})$, this proves the corollary. □

In order to complete the proof of Proposition A(3), we are left to show that all coarse-median preserving automorphisms of $\mathcal{A}$ are untwisted. This can be easily deduced from the work of Laurence [75], as we now describe.

Proposition 3.26 *If $\varphi \in \operatorname{Aut} \mathcal{A}$ preserves the coarse median structure $[\mu_\Gamma]$, then $\varphi \in U(\mathcal{A})$.*

Proof In the terminology of [75, Section 2], an automorphism $\varphi \in \operatorname{Aut} \mathcal{A}$ is *conjugating* if it preserves the conjugacy class of each standard generator $v \in \Gamma$. More generally, φ is *simple* if, for every $v \in \Gamma$, the image $\varphi(v)$ is label-irreducible and $v \in \Gamma(\varphi(v))$; compare [75, Definition 5.3] and Definition 3.5 in our paper.

Consider a coarse-median preserving automorphism $\varphi = \varphi_0$. By [75, Corollary to Lemma 4.5], there exists a graph automorphism ψ_1 such that, setting $\varphi_1 := \psi_1\varphi$, we have $v \in \Gamma(\varphi_1(v))$ for every generator $v \in \Gamma$. Since graph automorphisms are coarse-median preserving, φ_1 is again coarse-median preserving. By Corollary 3.3 and Lemma 3.11(2), the element $\varphi_1(v)$ is label-irreducible for every $v \in \Gamma$. Thus, φ_1 is simple.

By the proofs of [75, Lemma 6.8] and [75, Corollary to Lemma 6.6], there exists a product of inversions, folds and partial conjugations ψ_2 such that $\varphi_2 := \varphi_1\psi_2$ is conjugating. Finally, by [75, Theorem 2.2], the automorphism φ_2 is a product of partial conjugations. This shows that $\varphi \in U(\mathcal{A})$, as required. □

3.4.3 Pure automorphisms We end this subsection by introducing the subgroups $U_0(\mathcal{A}) \leq U(\mathcal{A})$ and $\operatorname{Aut}_0 \mathcal{W} \leq \operatorname{Aut} \mathcal{W}$ generated by inversions, folds and partial conjugations (no graph automorphisms or twists, in both cases). These are the subgroups appearing in the statements of Theorem C and Proposition D, and we will study them further in Sections 4.5 and 5. For the time being, we limit ourselves to a few quick observations.

Remark 3.27 The subgroups $U_0(\mathcal{A}) \le U(\mathcal{A})$ and $\mathrm{Aut}_0\,\mathcal{W} \le \mathrm{Aut}\,\mathcal{W}$ have finite index. In the Coxeter case, see eg [92, Proposition 1.2]. In the Artin case, it suffices to observe that $U_0(\mathcal{A})$ is normalised by all graph automorphisms, and that the latter generate a finite subgroup of $U(\mathcal{A})$.

Remark 3.28 Although they do not appear in our chosen generating set for $U_0(\mathcal{A})$, graph automorphisms of $\mathcal{A}$ can still lie in $U_0(\mathcal{A})$. Indeed, confusing $\sigma \in \mathrm{Aut}\,\Gamma$ with the induced $\sigma \in \mathrm{Aut}\,\mathcal{A}$, we have $\sigma \in U_0(\mathcal{A})$ if and only if $\mathrm{lk}\,\sigma(v) = \mathrm{lk}\,v$ for every $v \in \Gamma$.

The "only if" part follows from Lemma 4.30. For the "if" part, it suffices to show that $\sigma \in U_0(\mathcal{A})$ whenever σ swaps two vertices of Γ with the same link and fixes the rest of Γ. In this case, σ is a product of 3 folds and 3 inversions, as described at the end of the proof of [43, Proposition 3.3].

Lemma 3.29 *If $\varphi(\mathcal{A}_\Delta) = \mathcal{A}_\Delta$ for a full subgraph $\Delta \subseteq \Gamma$ and $\varphi \in U_0(\mathcal{A})$, then $\varphi|_{\mathcal{A}_\Delta} \in U_0(\mathcal{A}_\Delta)$.*

Proof We begin with a general observation. As in the proof of Proposition 3.26, we can write $\varphi = \sigma\varphi_1$, where σ is a graph automorphism and φ_1 is a simple automorphism of $\mathcal{A}$. Moreover, simple automorphisms are products of inversions, folds and partial conjugations, so $\varphi_1 \in U_0(\mathcal{A})$. We conclude that $\sigma \in U_0(\mathcal{A})$, and Remark 3.28 shows that $\mathrm{lk}\,\sigma(v) = \mathrm{lk}\,v$ for every $v \in \Gamma$.

If $v \in \Delta$, then $v \in \Gamma(\varphi_1(v))$ because φ_1 is simple. Thus

$$\sigma(v) \in \sigma(\Gamma(\varphi_1(v))) = \Gamma(\sigma\varphi_1(v)) = \Gamma(\varphi(v)) \subseteq \Delta.$$

We deduce that $\sigma(\Delta) = \Delta$, and Remark 3.28 shows that $\sigma|_{\mathcal{A}_\Delta} \in U_0(\mathcal{A}_\Delta)$. Since σ and φ preserve $\mathcal{A}_\Delta$, so does φ_1, and it suffices to show that $\varphi_1|_{\mathcal{A}_\Delta} \in U_0(\mathcal{A}_\Delta)$.

It is clear that $\varphi_1|_{\mathcal{A}_\Delta}$ is a simple automorphism of $\mathcal{A}_\Delta$, so the fact that $\varphi_1|_{\mathcal{A}_\Delta} \in U_0(\mathcal{A}_\Delta)$ follows again from [75] as in the proof of Proposition 3.26. □

4 Fixed subgroups of CMP automorphisms

This section is devoted to fixed subgroups of coarse-median preserving automorphisms of cocompactly cubulated groups. Theorem B is proved in Sections 4.1 and 4.2, where we study the properties of those subgroups of cocompactly cubulated groups that are approximate median subalgebras; see Theorem 4.10. At the end of Section 4.2, we also prove Corollaries G and H.

Then in Sections 4.3 and 4.4, we develop a quasiconvexity criterion for approximate median subalgebras of cube complexes (Proposition 4.25). This is used to prove Theorem C in Section 4.5; see Corollaries 4.34 and 4.35.

The reader interested only in Theorems F and I can just read the proof that $\mathrm{Fix}\,\varphi$ is finitely generated (Proposition 4.11) and skip the rest of this section in its entirety.

4.1 Approximate median subalgebras

The goal of this subsection is to show that approximate median subalgebras (Definition 2.33) of median spaces stay close to actual subalgebras. This is an important ingredient in the proofs of Theorem B and Corollaries G and H, which will be discussed in the next subsection.

Shortly after the first draft of this paper appeared on arXiv, it was pointed out to me by Mark Hagen that a similar result appears in the work of Bowditch [18, Proposition 4.1], which I was not aware of. Although Propositions 4.1 and 4.2 below are more general and our proofs seem different, I want to emphasise that Bowditch's result would suffice for all applications in this paper.

Proposition 4.1 *If X is a finite-rank median space and $A \subseteq X$ is an approximate median subalgebra, then $d_{\mathrm{Haus}}(A, \langle A\rangle) < +\infty$.*

The only focus of this subsection will actually be the next result, which provides an analogue of Remark 2.5. From it, it is straightforward to deduce Proposition 4.1 proceeding as in Lemma 2.10, which we leave to the reader.

Proposition 4.2 *There exists a function $h\colon \mathbb{N} \to \mathbb{N}$ with the following property. If M is a median algebra of rank r and $A \subseteq M$ is a subset, then $\langle A\rangle = \mathcal{M}^{h(r)}(A)$.*

We now obtain a sequence of lemmas leading up to Proposition 4.8, which proves Proposition 4.2.

Let M be a median algebra. We denote by $\mathscr{M}(M)$ the collection of subsets of M of one of these three forms:

- $\mathfrak{h}$, where $\mathfrak{h}$ is a halfspace;
- $\mathfrak{h} \cup \mathfrak{k}$, where $\mathfrak{h}$ and $\mathfrak{k}$ are transverse halfspaces;
- $\mathfrak{h} \cup \mathfrak{k}$, where $\mathfrak{h}$ and $\mathfrak{k}$ are disjoint halfspaces.

Elements of $\mathscr{M}(M)$ are to median subalgebras what halfspaces of M are to convex subsets. More precisely, the following is a straightforward characterisation of the median subalgebra generated by a subset $A \subseteq M$; see for instance [99, II.4.25.7].

Lemma 4.3 *For every subset $A \subseteq M$, the median subalgebra $\langle A\rangle \subseteq M$ is the intersection of all elements of $\mathscr{M}(M)$ containing A.*

We will make repeated use of the following observation, without explicit mention:

Lemma 4.4 *Given points $a, b, c, d \in M$, the three sets $\mathscr{W}(a, b|c, d)$, $\mathscr{W}(a, c|b, d)$ and $\mathscr{W}(a, d|b, c)$ are transverse to each other.*

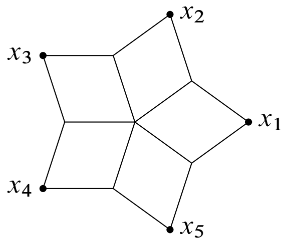


Figure 2: A pentagonal configuration in the 0–skeleton of a CAT(0) square complex.

It is also convenient to give a name to the situation in Figure 2.

Definition 4.5 An ordered 5–tuple $(x_1, x_2, x_3, x_4, x_5) \in M^5$ is a *pentagonal configuration* if the five sets $\mathscr{W}(x_{i-1}, x_i, x_{i+1}|x_{i+2}, x_{i+3})$ are all nonempty (indices are taken mod 5).

This requirement is invariant under cyclic permutations of the 5 points. Also note that, setting $\mathcal{W}_i := \mathscr{W}(x_{i-1}, x_i, x_{i+1}|x_{i+2}, x_{i+3})$, the sets $\mathcal{W}_i$ and $\mathcal{W}_{i+1}$ are transverse for all i mod 5.

Lemma 4.6 *Suppose that* $\operatorname{rk} M \leq 2$. *Consider* $x \in M$ *with* $x = m(m(a_1, a_2, a_3), b, c)$ *for points* $a_i, b, c \in M$. *Then one of the following happens*:

- *There exists* $1 \leq i \leq 3$ *such that* $x = m(a_i, b, c)$.
- *There exist* $1 \leq i < j \leq 3$ *such that either* $x = m(a_i, a_j, b)$ *or* $x = m(a_i, a_j, c)$.
- *We have* $x = m(a_1, a_2, a_3)$.
- *The points* a_1, a_2, a_3, b, c *can be ordered to form a pentagonal configuration.*

Proof Set $n = m(a_1, a_2, a_3)$. Consider the projections $\overline{a}_i = m(a_i, b, c)$ to the interval $I(b, c)$. Since gate-projections are median morphisms, we have $x = m(\overline{a}_1, \overline{a}_2, \overline{a}_3)$.

Claim 1 *If we are not in the* 1st *or* 3rd *case, we can assume that the four sets* $\mathscr{W}(x|\overline{a}_1)$, $\mathscr{W}(x|\overline{a}_2)$, $\mathscr{W}(x|\overline{a}_3)$ *and* $\mathscr{W}(a_1, a_2|b, c)$ *are all nonempty, and that* $\mathscr{W}(a_1, c|a_2, b) = \varnothing$.

Proof If one of the sets $\mathscr{W}(x|\overline{a}_i)$ is empty, then $x = \overline{a}_i$ and we are in the 1st case. On the other hand, if the sets $\mathscr{W}(a_i, a_j|b, c)$ are all empty for $i \neq j$, then we are in the 3rd case. Indeed, since $\mathscr{W}(n|b, c) \subseteq \bigcup_{i<j} \mathscr{W}(a_i, a_j|b, c)$, we have $n \in I(b, c)$, hence $x = m(n, b, c) = n = m(a_1, a_2, a_3)$.

Thus, up to permuting the a_i, we can assume that $\mathscr{W}(a_1, a_2|b, c) \neq \varnothing$. Since this is transverse to the transverse sets $\mathscr{W}(a_1, b|a_2, c)$ and $\mathscr{W}(a_1, c|a_2, b)$, one of the latter must be empty. Swapping b and c if necessary, we can assume that it is $\mathscr{W}(a_1, c|a_2, b)$. ◁

Claim 2 *If we are not in the* 4th *case either, we can further assume that* $\mathscr{W}(a_1, a_2, b|a_3, c) = \varnothing$.

Proof Note that the assumptions in Claim 1 are left unchanged if we simultaneously swap $b \leftrightarrow c$ and $a_1 \leftrightarrow a_2$. Thus, it suffices to show that we can suppose that at least one of the two sets $\mathscr{W}(b, a_1, a_2|c, a_3)$ and $\mathscr{W}(a_1, a_2, c|a_3, b)$ is empty.

In order to do so, we assume that $\mathscr{W}(b, a_1, a_2|c, a_3)$ and $\mathscr{W}(a_1, a_2, c|a_3, b)$ are both nonempty and show that (b, a_1, a_2, c, a_3) is a pentagonal configuration. This places us in the 4^{th} case.

Since $\mathscr{W}(a_1, c|a_2, b) = \varnothing$ and $x = m(\overline{a}_1, \overline{a}_2, \overline{a}_3)$, where $\overline{a}_i$ is the projection of a_i to $I(b, c)$, we have

$$\mathscr{W}(a_2, c, a_3|b, a_1) \supseteq \mathscr{W}(\overline{a}_2, \overline{a}_3|\overline{a}_1) = \mathscr{W}(x|\overline{a}_1) \neq \varnothing,$$
$$\mathscr{W}(a_3, b, a_1|a_2, c) \supseteq \mathscr{W}(\overline{a}_3, \overline{a}_1|\overline{a}_2) = \mathscr{W}(x|\overline{a}_2) \neq \varnothing.$$

Moreover, since $\mathscr{W}(a_1, a_3|b, c)$ is transverse to the nonempty transverse subsets $\mathscr{W}(b, a_1, a_2|c, a_3)$ and $\mathscr{W}(a_1, a_2, c|a_3, b)$, we have $\mathscr{W}(a_1, a_3|b, c) = \varnothing$. Hence $\mathscr{W}(c, a_3, b|a_1, a_2) = \mathscr{W}(c, b|a_1, a_2) \neq \varnothing$. $\lhd$

Claim 3 *Under these assumptions, we have* $\mathscr{W}(x|m(a_1, a_3, c)) = \mathscr{W}(b, c|a_1, a_3)$.

Proof By the properties of gate-projections, the set $\mathscr{W}(b|c)$ does not intersect any of the sets $\mathscr{W}(a_i|\overline{a}_i)$. Thus, since $x = m(\overline{a}_1, \overline{a}_2, \overline{a}_3)$, we must have

$$\begin{aligned}\mathscr{W}(x|m(a_1, a_3, c)) \cap \mathscr{W}(b|c) &= \mathscr{W}(m(a_1, a_2, a_3)|m(a_1, a_3, c)) \cap \mathscr{W}(b|c)\\ &= \mathscr{W}(a_1|a_3) \cap \mathscr{W}(a_2|c) \cap \mathscr{W}(b|c)\\ &= \mathscr{W}(a_1, a_2, b|a_3, c) \sqcup \mathscr{W}(a_2, a_3, b|a_1, c) = \varnothing,\end{aligned}$$

where we have used Claims 1 and 2 at the very end. Since $x \in I(b, c)$, we have $\mathscr{W}(x|b, c) = \varnothing$. Thus,

$$\mathscr{W}(x|m(a_1, a_3, c)) = \mathscr{W}(x, b, c|m(a_1, a_3, c)) = \mathscr{W}(b, c|a_1, a_3). \qquad \lhd$$

In order to conclude the proof of the lemma, suppose for the sake of contradiction that we are not in the 2^{nd} case, in addition to the assumptions of the claims. Then, Claim 3 implies

$$\varnothing \neq \mathscr{W}(x|m(a_1, a_3, c)) = \mathscr{W}(b, c|a_1, a_3).$$

On the other hand, since $\mathscr{W}(a_1, c|a_2, b)$ and $\mathscr{W}(a_1, a_2, b|a_3, c)$ are empty by Claims 1 and 2,

$$\varnothing \neq \mathscr{W}(x|\overline{a}_1) = \mathscr{W}(\overline{a}_2, \overline{a}_3|\overline{a}_1) = \mathscr{W}(c, a_2, a_3|b, a_1) \subseteq \mathscr{W}(c, a_3|b, a_1),$$
$$\varnothing \neq \mathscr{W}(x|\overline{a}_3) = \mathscr{W}(\overline{a}_1, \overline{a}_2|\overline{a}_3) = \mathscr{W}(a_1, a_2, c|a_3, b) \subseteq \mathscr{W}(a_1, c|a_3, b).$$

Since the three sets $\mathscr{W}(b, c|a_1, a_3)$, $\mathscr{W}(c, a_3|b, a_1)$, $\mathscr{W}(a_1, c|a_3, b)$ are pairwise transverse, this violates the assumption that $\operatorname{rk} M \leq 2$. This proves the lemma. $\square$

Corollary 4.7 *If T_1 and T_2 are rank-1 median algebras, then $\langle A \rangle = \mathcal{M}(A)$ for all $A \subseteq T_1 \times T_2$.*

Proof The product $T_1 \times T_2$ does not contain any pentagonal configurations. Otherwise, there would be walls $\mathfrak{w}_1, \mathfrak{w}_2, \mathfrak{w}_3, \mathfrak{w}_4, \mathfrak{w}_5$ with each $\mathfrak{w}_i$ transverse to $\mathfrak{w}_{i+1}$. If $\mathfrak{w}_1$ originates from the factor T_1, say, then $\mathfrak{w}_2$ must originate from T_2 and, continuing like this, $\mathfrak{w}_5$ again originates from T_1. Since $\mathfrak{w}_5$ and $\mathfrak{w}_1$ are transverse, this would contradict the fact that $\operatorname{rk} T_1 = 1$.

Thus, the 4th case of Lemma 4.6 never occurs, hence $\mathcal{M}^2(A) = \mathcal{M}(A)$ for all $A \subseteq T_1 \times T_2$. □

For the next result, let us consider the functions $f, g, h\colon \mathbb{N} \to \mathbb{N}$ given by

$$f(n) = 2^{2^n}, \quad g(n) = 1 + f\left(\tfrac{1}{2}n(n-1)\right), \quad h(n) = ng(n) + n.$$

Proposition 4.8 *Given a median algebra M and a subset $A \subseteq M$, the following hold*:

1. *If $\#A \le n$, then $\langle A \rangle = \mathcal{M}^{f(n)}(A)$.*
2. *If M can be embedded in a product of d rank-1 median algebras, then $\langle A \rangle = \mathcal{M}^{g(d)}(A)$.*
3. *If $\operatorname{rk} M \le r$, then $\langle A \rangle = \mathcal{M}^{h(r)}(A)$.*

Proof Part (1) is immediate from most constructions of the free median algebra on the set A; for instance, see [15, Lemma 4.2] and the subsequent paragraph.

Regarding part (2), let us fix an injective median morphism $M \hookrightarrow T_1 \times \cdots \times T_d$, where the T_i have rank 1. Let $\pi_{ij}\colon M \to T_i \times T_j$ denote the composition with the projection to $T_i \times T_j$. Given $x \in M$, Lemma 4.3 implies that $x \in \langle A \rangle$ if and only if $\pi_{ij}(x) \in \langle \pi_{ij}(A) \rangle$ for all $1 \le i < j \le d$.

Since each π_{ij} is a median morphism, Corollary 4.7 shows that

$$\langle \pi_{ij}(A) \rangle = \mathcal{M}(\pi_{ij}(A)) = \pi_{ij}(\mathcal{M}(A)).$$

Thus, given $x \in \langle A \rangle$, there exist points $m_{ij} \in \mathcal{M}(A)$ such that $\pi_{ij}(x) = \pi_{ij}(m_{ij})$. It follows that

$$x \in \langle \{m_{ij} \mid 1 \le i < j \le d\} \rangle.$$

Since there are at most $\frac{1}{2}d(d-1)$ distinct points m_{ij}, part (1) yields

$$\langle \{m_{ij} \mid 1 \le i < j \le d\} \rangle = \mathcal{M}^{g(d)-1}(\{m_{ij} \mid 1 \le i < j \le d\}) \subseteq \mathcal{M}^{g(d)-1}(\mathcal{M}(A)) = \mathcal{M}^{g(d)}(A).$$

Hence $\langle A \rangle \subseteq \mathcal{M}^{g(d)}(A)$.

Finally, let us prove part (3). Since $\operatorname{rk}\langle A \rangle \le \operatorname{rk} M$, we can safely assume that $M = \langle A \rangle$. Consider two points $a, b \in M$ and recall that the gate-projection $\pi_{ab}\colon M \to I(a,b)$ is given by $\pi_{ab}(x) = m(a,b,x)$. By Dilworth's lemma, the interval $I(a,b) \subseteq M$ can be embedded in a product of r rank-1 median algebras for all $a, b \in M$; cf [16, Proposition 1.4].

If $B \subseteq M$ is a subset with $\langle B \rangle = M$ and $a, b \in B$, then part (2) yields

$$I(a,b) = \pi_{ab}(M) = \pi_{ab}(\langle B \rangle) = \langle \pi_{ab}(B) \rangle = \mathcal{M}^{g(r)}(\pi_{ab}(B)) = \pi_{ab}(\mathcal{M}^{g(r)}(B)) \subseteq \mathcal{M}^{g(r)+1}(B).$$

It follows that $\mathcal{J}(B) \subseteq \mathcal{M}^{g(r)+1}(B)$ for every subset $B \subseteq M$ with $\langle B \rangle = M$. Observing that

$$\mathcal{J}^{k+1}(B) = \mathcal{J}(\mathcal{J}^k(B)) \subseteq \mathcal{M}^{g(r)+1}(\mathcal{J}^k(B)),$$

we inductively obtain $\mathcal{J}^m(B) \subseteq \mathcal{M}^{m(g(r)+1)}(B)$ for all $m \geq 1$. In particular, by Remark 2.5,

$$\langle A \rangle \subseteq \text{Hull}\, A = \mathcal{J}^r(A) \subseteq \mathcal{M}^{r(g(r)+1)}(A) = \mathcal{M}^{h(r)}(A).$$

This concludes the proof of the proposition. □

Remark 4.9 The bounds provided by Proposition 4.8 are highly nonsharp. For instance, if $\operatorname{rk} M \leq 2$, a slightly more careful use of Lemma 4.6 would show that $\langle A \rangle = \mathcal{M}^2(A)$ for every $A \subseteq M$, while the proposition only gives $\langle A \rangle = \mathcal{M}^{244}(A)$. For the purposes of this paper, we only care that such bounds exist and only depend on the rank of M.

4.2 Approximate subalgebras of cubulated groups

This subsection is devoted to the proof of Theorem B and to a few examples of how this result can fail for automorphisms that do not preserve the coarse median structure (Example 4.13). Towards the end, we use similar techniques to prove Corollaries G and H.

Our main focus will be the following result. Recall that, if φ is a coarse-median preserving automorphism of a cocompactly cubulated group, Lemma 2.35 guarantees that the subgroup $\operatorname{Fix}\varphi$ is an approximate median subalgebra. Thus, Theorem B immediately follows from:

Theorem 4.10 *Let $G \curvearrowright X$ be a proper cocompact action on a* CAT(0) *cube complex. Let $[\mu_X]$ be the induced coarse median structure on G. If a subgroup $H \leq G$ is an approximate median subalgebra of $(G, [\mu_X])$, then*:

(1) *H is finitely generated and undistorted in G.*

(2) *H admits a proper cocompact action on a* CAT(0) *cube complex.*

As a first step, we need to show that the subgroup H in Theorem 4.10 is finitely generated. The proof of this is a straightforward adaptation of an argument due to Cooper and Paulin [33; 86] for fixed subgroups of automorphisms of hyperbolic groups.

Proposition 4.11 *Let $G \curvearrowright X$ be a proper cocompact action on a* CAT(0) *cube complex. If a subgroup $H \leq G$ is an approximate median subalgebra of $(G, [\mu_X])$, then H is finitely generated.*

Proof Fix a base vertex $p \in X$. The main observation is the following:

Claim *If $x_n \in H \cdot p$ is a diverging sequence, then there exists an element $h \in H$ such that $d(p, hx_n) < d(p, x_n)$ holds for infinitely many values of n.*

Proof Since H is an approximate median subalgebra of $(G,[\mu_X])$, there exists $L \geq 0$ such that all medians of points in $H \cdot p$ lie in the L–neighbourhood of $H \cdot p$ in X.

Passing to a subsequence, we can assume that the vertices x_n converge to a point in the Roller boundary $\xi \in \partial X$. Recalling that $X^{(0)}$ is discrete in the Roller compactification and that the median map is continuous, there exist integers $M(n) \geq 0$ such that, for every $m \geq M(n)$, we have

$$m(p,x_n,\xi)=m(p,x_n,x_m).$$

In particular, there exist elements $h_n \in H$ such that $h_n p$ is L–close to $m(p,x_n,\xi)$.

Now, since $x_n \to \xi$, the medians $m(p,x_n,\xi)$ diverge, and so do the points $h_n p$. In particular, there exist indices $i < j$ such that

$$d(p,h_i p)+2L < d(p,h_j p).$$

Since, for $m \geq M(j)$, the point $h_j p$ is L–close to the median $m(p,x_j,x_m)$, we also have

$$d(p,h_j p)+d(h_j p,x_m) \leq d(p,x_m)+2L.$$

In conclusion, setting $h := h_i h_j^{-1}$, we obtain, for all $m \geq M(j)$,

$$\begin{aligned} d(p,hx_m)=d(p,h_i h_j^{-1} x_m) &\leq d(p,h_i p)+d(h_i p,h_i h_j^{-1} x_m)=d(p,h_i p)+d(h_j p,x_m) \\ &\leq d(p,h_i p)+d(p,x_m)-d(p,h_j p)+2L \\ &< d(p,x_m). \end{aligned}$$ ◁

Now, suppose for the sake of contradiction that H is not finitely generated. Write H as the union of an infinite ascending chain of subgroups $H_1 \subsetneq H_2 \subsetneq \cdots$, where $H_{n+1} = \langle H_n, h_{n+1} \rangle$ for some $h_{n+1} \in H$. Possibly replacing h_{n+1}, we can assume that the point $x_{n+1} := h_{n+1} p$ minimises the distance to p within the set $H_n h_{n+1} \cdot p$.

The claim provides an element $h \in H$ such that $d(p,hx_n) < d(p,x_n)$ occurs infinitely often. Since $h \in H$, there exists $N \geq 0$ such that $h \in H_n$ for all $n \geq N$. This contradicts the fact that x_{n+1} minimises the distance to p within $H_n \cdot x_{n+1}$ for $n \geq N$. □

Along with Propositions 4.1 and 4.11, the following is the only missing ingredient in the proof of Theorem 4.10. We refer the reader to the proof sketched in the introduction.

Lemma 4.12 *Let $G \curvearrowright X$ be a proper cocompact action on a* CAT(0) *cube complex. Consider a subgroup $H \leq G$. Suppose that there exists an H–invariant median subalgebra $M \subseteq X^{(0)}$ such that the action $H \curvearrowright M$ is cofinite. Then*:

(1) *H is finitely generated and undistorted in G.*

(2) *H admits a proper cocompact action on a* CAT(0) *cube complex.*

Proof Halfspaces and hyperplanes of the cube complex X, as usually defined, correspond exactly to halfspaces and hyperplanes of the median algebra $X^{(0)}$. As customary, we write $\mathscr{H}(X)$ and $\mathscr{W}(X)$ to mean $\mathscr{H}(X^{(0)})$ and $\mathscr{W}(X^{(0)})$. By Remark 2.2(1), we have a natural surjection $\mathrm{res}_M : \mathscr{H}_M(X) \to \mathscr{H}(M)$.

Since H acts cofinitely on the subalgebra M, it is an approximate median subalgebra of $(G, [\mu_X])$ and Proposition 4.11 implies that H is finitely generated. Thus, every H–orbit in X is coarsely connected and, since $H \curvearrowright M$ is cofinite, M is coarsely connected as well. It follows that there exists a uniform upper bound m to the cardinality of the fibres of the map res_M.

As in [89, Section 10; 31, Theorem 6.1], we can construct a CAT(0) cube complex $X(M)$ such that M is naturally isomorphic to the median algebra $X(M)^{(0)}$. Given $x, y \in M$, let us denote by $d(x, y)$ and $d_M(x, y)$ their distance in the 1–skeleta of X and $X(M)$, respectively.

By construction, $d_M(x, y)$ coincides with the number of walls of M separating x and y. It follows from the above discussion on res_M that

$$d_M(x, y) \leq d(x, y) \leq m \cdot d_M(x, y)$$

for all $x, y \in M$. Thus, the identification between $X(M)^{(0)}$ and $M \subseteq X^{(0)}$ gives a quasi-isometric embedding $X(M) \to X$ that is equivariant with respect to the inclusion $H \hookrightarrow G$.

The action $H \curvearrowright (M, d_M)$ is cofinite by assumption, and it follows from the above inequalities that it is also proper. This shows that the induced action $H \curvearrowright X(M)$ is proper and cocompact, proving part (2). The Milnor–Schwarz Lemma now implies that the inclusion $H \hookrightarrow G$ is a quasi-isometric embedding, which proves part (1). □

Proof of Theorem 4.10 For any vertex $p \in X$, the orbit $H \cdot p$ is an approximate median subalgebra of X. By Proposition 4.1, the subalgebra $M := \langle H \cdot p \rangle$ is at finite Hausdorff distance from $H \cdot p$. Since X is locally finite, it follows that the action $H \curvearrowright M$ is cofinite, hence Lemma 4.12 shows that H is finitely generated, undistorted and cocompactly cubulated. □

As discussed above, this completes the proof of Theorem B. The next example shows that, even for automorphisms of RAAGs, all of the claims in the statement of Theorem B can fail if the automorphism does not preserve the coarse median structure.

Example 4.13 Here is a recipe to construct automorphisms with unpleasant fixed subgroups. Consider a group G and a homomorphism $\alpha : G \to \mathbb{Z}$. These data define an automorphism $\psi \in \mathrm{Aut}(G \times \mathbb{Z})$ by the formula

$$\psi(g, n) := (g, n + \alpha(g)).$$

It is clear that $\mathrm{Fix}\,\psi = \ker\alpha \times \mathbb{Z}$.

Now, consider the situation where G is a right-angled Artin group $\mathcal{A}_\Gamma$ and $\alpha : \mathcal{A}_\Gamma \to \mathbb{Z}$ takes all standard generators to $+1$. The resulting automorphism $\psi \in \mathrm{Aut}(\mathcal{A}_\Gamma \times \mathbb{Z})$ is a product of finitely many twists (as defined in Section 3.4) and we have $\mathrm{Fix}\,\psi = BB_\Gamma \times \mathbb{Z}$, where BB_Γ denotes the Bestvina–Brady subgroup of $\mathcal{A}_\Gamma$; see [8].

We apply this construction to obtain examples where $\mathrm{Fix}\,\psi$ fails to have the three properties provided by Theorem B.

(1) The subgroup BB_Γ is finitely generated if and only if Γ is connected [80]. For instance, there exists $\psi \in \mathrm{Aut}(F_2 \times \mathbb{Z})$ such that $\mathrm{Fix}\,\psi$ is not finitely generated.

(2) If $\mathcal{A}_\Gamma$ is freely irreducible, directly irreducible and noncyclic, then BB_Γ is finitely generated and quadratically distorted [98, Theorem 1.1]. This gives examples where $\mathrm{Fix}\,\psi$ is finitely generated, but distorted.

(3) As shown in [8, Main Theorem], the finiteness properties and homological finiteness properties of BB_Γ are governed by the homology and homotopy groups of the flag simplicial complex L_Γ determined by Γ. The same is true of $\mathrm{Fix}\,\psi = BB_\Gamma \times \mathbb{Z}$; see [79; 22]. In particular, if L_Γ is not contractible, then $\mathrm{Fix}\,\psi$ is not of type F (hence not cocompactly cubulated, since there is no torsion). This can even be achieved while ensuring that $\mathrm{Fix}\,\varphi$ is undistorted: by [98, Theorem 1.1], it suffices to make sure that $\mathcal{A}_\Gamma$ splits as a product.

We emphasise that, by embedding $\mathcal{A}_\Gamma \times \mathbb{Z}$ as a parabolic subgroup of a larger RAAG and suitably extending the automorphism ψ, we can ensure that all these bad behaviours also occur for automorphisms of *irreducible* RAAGs.

We conclude this subsection by proving Corollaries G and H. All that is required is Proposition A, Proposition 4.1 and (part of) Lemma 4.12.

Proof of Corollary G Let $H \leq G$ be a finite-index subgroup with a proper cocompact action on a CAT(0) cube complex $H \curvearrowright X$. Replacing H with a finite-index subgroup, it is not restrictive to suppose that $H \lhd G$. By our assumption, the conjugation action $G \curvearrowright H$ preserves the coarse median structure $[\mu]$ induced on H by $H \curvearrowright X$.

It is well-known that the concept of induced representation can be generalised to actions on metric spaces; see eg [21, Section 2.1] or [9, Section 2.2]. In our context, this yields a proper action $G \curvearrowright X_1 \times \cdots \times X_n$, where n is the index of H in G and each X_i is isomorphic to X. Since H is normal, each factor is preserved by H and each action $H \curvearrowright X_i$ can be made equivariantly isomorphic to $H \curvearrowright X$ by twisting it by an automorphism of H corresponding to a conjugation by an element of G.

Since $[\mu]$ is preserved by the conjugation action $G \curvearrowright H$, it is the coarse median structure induced by all the cubulations $H \curvearrowright X_i$. This implies that, for every finite subset $A \subseteq X^{(0)}$, the orbit $H \cdot A$ is an approximate median subalgebra. Since H is normal, we can choose a finite subset $A \subseteq X^{(0)}$ so that $H \cdot A$ is G–invariant. Proposition 4.1 guarantees that the G–invariant median subalgebra $M := \langle H \cdot A \rangle$ is at finite Hausdorff distance from $H \cdot A$. Since each X_i is locally finite, this implies that the action $G \curvearrowright M$ is cofinite.

Since M is a discrete median algebra, the natural CAT(0) cube complex $X(M)$ with M as its 0–skeleton (as in [89, Section 10] or [31, Theorem 6.1]) gives the required cocompact cubulation of G. Here properness of the G–action on $X(M)$ can be checked as in the proof of Lemma 4.12, using the fact that M is coarsely connected to conclude that $X(M)$ and $\prod X_i$ induce bi-Lipschitz equivalent metrics on M. □

Proof of Corollary H Let $G = \mathcal{A}_\Gamma$ or $G = \mathcal{W}_\Gamma$. Consider a finite subgroup $F \leq \operatorname{Out} G$ as in the statement. Let $\pi : \operatorname{Aut} G \to \operatorname{Out} G$ be the quotient projection. Our goal is to construct a proper, cocompact action on a CAT(0) cube complex $\pi^{-1}(F) \curvearrowright Y$. We can then take Q to be the quotient of Y by the finite-index normal subgroup $G \simeq \ker \pi \lhd \pi^{-1}(F)$.

Let $G \curvearrowright X$ be the standard action on the universal cover of the Salvetti/Davis complex. In both cases, Proposition A shows that F preserves the coarse median structure on G induced by this action. Thus, Corollary G provides the required proper cocompact action $\pi^{-1}(F) \curvearrowright Y$. □

4.3 Staircases in cube complexes

In the rest of Section 4, our goal is to obtain a *quasiconvexity* criterion for median subalgebras of cube complexes, which will lead to the proof of Theorem C. Ultimately, we will restrict to universal covers of Davis/Salvetti complexes for right-angled groups and an important point will be that they do not admit infinite *staircases*.

In this subsection, we study staircases in general CAT(0) cube complexes.

Definition 4.14 Let M be a median algebra.

(1) A *length-n staircase* in M is the data of two chains of halfspaces $\mathfrak{h}_1 \supsetneq \cdots \supsetneq \mathfrak{h}_n$ and $\mathfrak{k}_1 \supsetneq \cdots \supsetneq \mathfrak{k}_n$ such that $\mathfrak{h}_i$ is transverse to $\mathfrak{k}_j$ for $j \leq i$, while $\mathfrak{k}_{i+1} \subsetneq \mathfrak{h}_i$.

(2) The *staircase length* of M is the supremum of $n \in \mathbb{N}$ such that M has a length-n staircase.

Figure 3 depicts part of a staircase of length ≥ 5.

When speaking of staircases in relation to a CAT(0) cube complex X, we always refer to the median algebra $M = X^{(0)}$. Note that the above notion of staircase seems to be a bit more general than the one in [63, page 51]: given hyperplanes bounding halfspaces as in Definition 4.14, there might not be a convex subcomplex of X with exactly these hyperplanes.

In view of the following discussion, it is convenient to introduce a notation for gate-projections to intervals. Given a median algebra M and points $x, y \in M$, we denote by $\pi_{xy} : M \to I(x, y)$ the map $\pi_{xy}(z) = m(x, y, z)$.

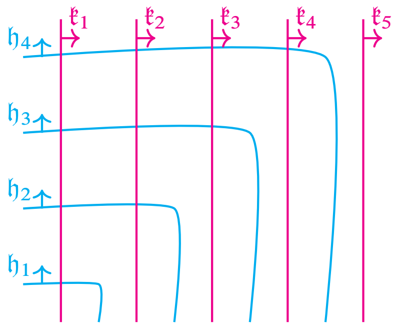


Figure 3

Lemma 4.15 *Let M be a median algebra of rank r and staircase length d. If there exist halfspaces $\mathfrak{k}_1 \supsetneq \cdots \supsetneq \mathfrak{k}_n$ and points $x, y \in \mathfrak{k}_1^*$ such that $\pi_{xy}(\mathfrak{k}_1) \supsetneq \cdots \supsetneq \pi_{xy}(\mathfrak{k}_n)$, then $n \leq 2rd$.*

Proof The sets $C_i := \pi_{xy}(\mathfrak{k}_i)$ are convex, for instance by [50, Lemma 2.2(1)]. Since $C_{i+1} \subsetneq C_i$, there exist halfspaces $\mathfrak{h}_i \in \mathscr{H}(M)$ such that $\mathfrak{h}_i \in \mathscr{H}_{C_i}(M)$ and $C_{i+1} \subseteq \mathfrak{h}_i$.

Since both $\mathfrak{h}_i$ and $\mathfrak{h}_i^*$ intersect $C_i \subseteq I(x, y)$, we have $\mathfrak{h}_i \in \mathscr{H}(x|y) \sqcup \mathscr{H}(y|x)$ for all i. Possibly swapping x and y, we can assume that at least $n/2$ of the $\mathfrak{h}_i$ lie in $\mathscr{H}(x|y)$. By Dilworth's lemma, there exist $k \geq n/2r$ and indices $i_1 < \cdots < i_k$ such that $\mathfrak{h}_{i_1}, \ldots, \mathfrak{h}_{i_k}$ lie in $\mathscr{H}(x|y)$ and no two of them are transverse. Up to reindexing, we can assume that these are $\mathfrak{h}_1, \ldots, \mathfrak{h}_k$.

Since C_j is contained in $\mathfrak{h}_i$ if and only if $j > i$, we must have $\mathfrak{h}_1 \supsetneq \cdots \supsetneq \mathfrak{h}_k$. Note that $y \in \mathfrak{h}_i \cap \mathfrak{k}_j^*$ and $x \in \mathfrak{h}_i^* \cap \mathfrak{k}_j^*$ for all i, j. If $j \leq i$, we have $\mathfrak{h}_i \in \mathscr{H}_{C_j}(X)$, hence $\mathfrak{h}_i \cap \mathfrak{k}_j$ and $\mathfrak{h}_i^* \cap \mathfrak{k}_j$ are both nonempty. This shows that $\mathfrak{h}_i$ and $\mathfrak{k}_j$ are transverse for $j \leq i$, while the fact that $C_{i+1} \subseteq \mathfrak{h}_i$ implies that $\mathfrak{k}_{i+1} \subseteq \mathfrak{h}_i$. In conclusion, the $\mathfrak{h}_i$ and $\mathfrak{k}_j$ form a length-k staircase with $k \geq n/2r$. Since M has staircase length d, we have $n \leq 2rk \leq 2rd$. □

Lemma 4.16 *Let X be a* CAT(0) *cube complex of dimension r and staircase length d. Consider vertices $x, y \in X$ and $z \in I(x, y)$. Let $\alpha \subseteq I(x, z)$ be a (combinatorial) geodesic from x to z. Then the median subalgebra $M = X^{(0)} \cap I(x, y) \cap \pi_{xz}^{-1}(\alpha)$ has staircase length $\leq d(1 + 2r^2)$.*

Proof Since $\pi_{xz}(y) = z$ and $x, z \in \alpha$, the three points x, y, z all lie in M. Since $M \subseteq I(x, y)$, every wall of M separates x and y. Recall that we use the notation $\mathscr{H}(X)$ and $\mathscr{W}(X)$ with the meaning of $\mathscr{H}(X^{(0)})$ and $\mathscr{W}(X^{(0)})$.

Claim 1 *If $\mathfrak{u}, \mathfrak{v} \in \mathscr{W}(M)$ separate x and z, then $\mathfrak{u}$ and $\mathfrak{v}$ are not transverse.*

Proof Pick halfspaces $\widehat{\mathfrak{h}}, \widehat{\mathfrak{k}} \in \mathscr{H}(X) \cap \mathscr{H}(x|z)$ such that $\mathfrak{h} := \widehat{\mathfrak{h}} \cap M \in \mathscr{H}(M)$ is bounded by $\mathfrak{u}$ and $\mathfrak{k} := \widehat{\mathfrak{k}} \cap M$ is bounded by $\mathfrak{v}$; this is possible by Remark 2.2(1). The intersections $\widehat{\mathfrak{h}} \cap \alpha$ and $\widehat{\mathfrak{k}} \cap \alpha$ are subsegments of α containing z. Without loss of generality, we have $\widehat{\mathfrak{h}} \cap \alpha \subseteq \widehat{\mathfrak{k}} \cap \alpha$. Then $\widehat{\mathfrak{h}} \cap \widehat{\mathfrak{k}}^* \cap \alpha = \varnothing$, hence $\varnothing = \widehat{\mathfrak{h}} \cap \widehat{\mathfrak{k}}^* \cap M = \mathfrak{h} \cap \mathfrak{k}^*$, proving the claim. ◁

Claim 2 *If $\widehat{\mathfrak{h}},\widehat{\mathfrak{k}}\in\mathscr{H}(z|y)$ are halfspaces of X, then $\widehat{\mathfrak{h}}$ and $\widehat{\mathfrak{k}}$ are transverse if and only if $\widehat{\mathfrak{h}}\cap M$ and $\widehat{\mathfrak{k}}\cap M$ are transverse halfspaces of M.*

Proof Since $\pi_{xz}(I(z,y))=\{z\}$, the vertex set of the interval $I(z,y)\subseteq X$ is entirely contained in M. Thus, $I(z,y)$ is a convex subset of both X and M. Remark 2.2(2) then shows that $\widehat{\mathfrak{h}}$ and $\widehat{\mathfrak{k}}$ are transverse if and only if $\widehat{\mathfrak{h}}\cap I(z,y)$ and $\widehat{\mathfrak{k}}\cap I(z,y)$ are transverse, if and only if $\widehat{\mathfrak{h}}\cap M$ and $\widehat{\mathfrak{k}}\cap M$ are transverse. ◁

Now, suppose that M contains a length-n staircase. Thus M has halfspaces $\mathfrak{h}_1\supsetneq\cdots\supsetneq\mathfrak{h}_n$ and $\mathfrak{k}_1\supsetneq\cdots\supsetneq\mathfrak{k}_n$ such that each $\mathfrak{h}_i$ is transverse to all $\mathfrak{k}_j$ with $j\le i$, while $\mathfrak{k}_{i+1}\subseteq\mathfrak{h}_i$.

Since $\mathfrak{k}_n\subseteq\mathfrak{h}_{n-1}\subseteq\mathfrak{h}_1$, we have either $\{\mathfrak{h}_1,\mathfrak{k}_n\}\subseteq\mathscr{H}(x|y)$ or $\{\mathfrak{h}_1,\mathfrak{k}_n\}\subseteq\mathscr{H}(y|x)$. If we replace all $\mathfrak{h}_i$ and $\mathfrak{k}_j$ with $\mathfrak{k}^*_{n-i+1}$ and $\mathfrak{h}^*_{n-j+1}$, respectively, we obtain another length-n staircase. Thus, we can assume that $\{\mathfrak{h}_1,\mathfrak{k}_n\}\subseteq\mathscr{H}(x|y)$. It follows that all $\mathfrak{h}_i$ and $\mathfrak{k}_j$ lie in $\mathscr{H}(x|y)$.

Let $0\le a,b\le n$ be the largest indices such that $z\in\mathfrak{h}_i$ and $z\in\mathfrak{k}_j$ hold for $1\le i\le a$ and $1\le j\le b$. Since $\mathfrak{h}_1$ and $\mathfrak{k}_1$ are transverse, Claim 1 shows that they cannot both lie in $\mathscr{H}(x|z)$. Thus $\min\{a,b\}=0$. Since $\mathfrak{k}_{a+2}\subseteq\mathfrak{h}_{a+1}$, we have $z\notin\mathfrak{k}_{a+2}$, hence $b\le a+1$. In conclusion, either $b=0$, or $(a,b)=(0,1)$.

The halfspaces $\mathfrak{h}_i,\mathfrak{k}_j$ with $i,j>\max\{a,b\}$ all lie in $\mathscr{H}(z|y)$ and form a staircase of length $n-\max\{a,b\}$. By Remark 2.2(1) and Claim 2, this determines a staircase of halfspaces of X. Since X has staircase length d, we deduce that $n-\max\{a,b\}\le d$.

If $b=1$ and $a=0$, we get $n\le d+1$ and we are done. If instead $b=0$, then $n\le a+d$ and the proof is completed with the following claim:

Claim 3 *If $b=0$, then $a\le 2r^2d$.*

Proof As a recap, M has halfspaces $\mathfrak{h}_1\supsetneq\cdots\supsetneq\mathfrak{h}_a$ in $\mathscr{H}(x|z,y)$ and $\mathfrak{k}_1\supsetneq\cdots\supsetneq\mathfrak{k}_a$ in $\mathscr{H}(x,z|y)$ forming a length-a staircase. By Remark 2.2(1), there exist halfspaces $\widehat{\mathfrak{h}}_i,\widehat{\mathfrak{k}}_j\in\mathscr{H}(X)$ such that $\mathfrak{h}_i=\widehat{\mathfrak{h}}_i\cap M$ and $\mathfrak{k}_i=\widehat{\mathfrak{k}}_i\cap M$.

By Dilworth's lemma, there exist $a'\ge a/r$ and indices $1\le j_1<\cdots<j_{a'}\le a$ such that no two among $\widehat{\mathfrak{k}}_{j_1},\dots,\widehat{\mathfrak{k}}_{j_{a'}}$ are transverse. Thus, up to reindexing, we can assume that $\widehat{\mathfrak{k}}_1\supsetneq\cdots\supsetneq\widehat{\mathfrak{k}}_{a'}$.

Now, since the $\mathfrak{h}_i$ and $\mathfrak{k}_j$ form a staircase in M and $\widehat{\mathfrak{h}}_i\in\mathscr{H}(x|z)$, we have, for every $1\le j\le a'$,

- $\varnothing=\mathfrak{h}^*_j\cap\mathfrak{k}_{j+1}=\widehat{\mathfrak{h}}^*_j\cap\widehat{\mathfrak{k}}_{j+1}\cap M$, hence $\pi_{xz}(\widehat{\mathfrak{k}}_{j+1})\cap\widehat{\mathfrak{h}}^*_j\cap\alpha=\varnothing$;
- $\varnothing\neq\mathfrak{h}^*_j\cap\mathfrak{k}_j=\widehat{\mathfrak{h}}^*_j\cap\widehat{\mathfrak{k}}_j\cap M$, hence $\pi_{xz}(\widehat{\mathfrak{k}}_j)\cap\widehat{\mathfrak{h}}^*_j\cap\alpha\neq\varnothing$.

Note moreover that $x,z\in\widehat{\mathfrak{k}}^*_1$. If we had $a'>2rd$, Lemma 4.15 would imply that there exists j with $\pi_{xz}(\widehat{\mathfrak{k}}_j)=\pi_{xz}(\widehat{\mathfrak{k}}_{j+1})$. However, $\pi_{xz}(\widehat{\mathfrak{k}}_j)$ intersects $\widehat{\mathfrak{h}}^*_j\cap\alpha$ while $\pi_{xz}(\widehat{\mathfrak{k}}_{j+1})$ does not.

We conclude that $a\le ra'\le 2r^2d$, as required. ◁

As discussed before Claim 3, this proves the lemma. □

Recall that, if Γ is a finite simplicial graph, $\mathcal{X}_\Gamma$ and $\mathcal{Y}_\Gamma$ denote the universal covers, respectively, of the Salvetti complex for $\mathcal{A}_\Gamma$ and the Davis complex for $\mathcal{W}_\Gamma$.

Lemma 4.17 *The staircase length of $\mathcal{X}_\Gamma$ and $\mathcal{Y}_\Gamma$ is at most $\#\Gamma^{(0)}$.*

Proof We only run the proof for $\mathcal{X}_\Gamma$, since the argument for $\mathcal{Y}_\Gamma$ is identical. The important property, shared by both complexes, is that there is a map $\gamma\colon \mathscr{W}(\mathcal{X}_\Gamma) \to \Gamma^{(0)}$ such that, if $\mathfrak{u}, \mathfrak{v}$ are hyperplanes with intersecting carriers, then $\mathfrak{u}$ and $\mathfrak{v}$ are transverse if and only if $\gamma(\mathfrak{u})$ and $\gamma(\mathfrak{v})$ are joined by an edge of Γ. For simplicity, let us extend the map γ to $\mathscr{H}(\mathcal{X}_\Gamma)$, simply by composing it with the two-to-one map $\mathscr{H}(\mathcal{X}_\Gamma) \to \mathscr{W}(\mathcal{X}_\Gamma)$ pairing each halfspace with its hyperplane.

Consider halfspaces $\mathfrak{h}_1 \supsetneq \cdots \supsetneq \mathfrak{h}_n$ and $\mathfrak{k}_1 \supsetneq \cdots \supsetneq \mathfrak{k}_n$ such that $\mathfrak{h}_i$ is transverse to all $\mathfrak{k}_j$ with $j \le i$, while $\mathfrak{k}_{i+1} \subsetneq \mathfrak{h}_i$. We define the following subsets of $\Gamma^{(0)}$:

$$\Gamma_j := \gamma(\mathfrak{k}_1^*) \cup \gamma(\mathscr{W}(\mathfrak{k}_1^*|\mathfrak{k}_j)) \cup \{\gamma(\mathfrak{k}_j)\}.$$

It is clear that $\Gamma_j \subseteq \Gamma_{j+1}$ for all $j \ge 1$. The lemma is immediate from the following claim:

Claim *We have $\Gamma_j \subsetneq \Gamma_{j+1}$ for all $j \ge 1$.*

Suppose for the sake of contradiction that, for some $j \ge 1$, we have $\Gamma_{j+1} = \Gamma_j$.

Given $\mathfrak{j} \in \mathscr{H}(\mathfrak{h}_j^*|\mathfrak{k}_{j+1})$, we have $\mathfrak{j} \cap \mathfrak{k}_1 \supseteq \mathfrak{k}_{j+1} \neq \varnothing$. Moreover, $\mathfrak{j}^* \cap \mathfrak{k}_1 \neq \varnothing$ and $\mathfrak{j}^* \cap \mathfrak{k}_1^* \neq \varnothing$, since $\mathfrak{j}^*$ contains $\mathfrak{h}_j^*$, which is transverse to $\mathfrak{k}_1$. Thus, for each $\mathfrak{j} \in \mathscr{H}(\mathfrak{h}_j^*|\mathfrak{k}_{j+1})$, there are only two possibilities: either

(a) $\mathfrak{j} \cap \mathfrak{k}_1^* = \varnothing$, hence $\mathfrak{j} \subseteq \mathfrak{k}_1$ and $\mathfrak{j} \in \mathscr{H}(\mathfrak{k}_1^*|\mathfrak{k}_{j+1})$; or

(b) $\mathfrak{j}$ is transverse to $\mathfrak{k}_1$.

Note that no halfspace of type (a) can contain a halfspace of type (b). Moreover, each $\mathfrak{j}$ of type (b) is also transverse to $\mathfrak{k}_j$: we have $\mathfrak{j} \cap \mathfrak{k}_j \supseteq \mathfrak{k}_{j+1} \neq \varnothing$, $\mathfrak{j} \cap \mathfrak{k}_j^* \supseteq \mathfrak{j} \cap \mathfrak{k}_1^* \neq \varnothing$, $\mathfrak{j}^* \cap \mathfrak{k}_j \supseteq \mathfrak{h}_j^* \cap \mathfrak{k}_j \neq \varnothing$ and $\mathfrak{j}^* \cap \mathfrak{k}_j^* \supseteq \mathfrak{j}^* \cap \mathfrak{k}_1^* \neq \varnothing$. Thus, every $\mathfrak{j}$ of type (b) is transverse to the set $\mathscr{H}(\mathfrak{k}_1^*|\mathfrak{k}_j) \cup \{\mathfrak{k}_1^*, \mathfrak{k}_j\}$.

Now, consider a maximal chain of halfspaces $\mathfrak{j}_1 \supsetneq \cdots \supsetneq \mathfrak{j}_m$ in $\mathscr{H}(\mathfrak{h}_j^*|\mathfrak{k}_{j+1})$ with $m \ge 0$. We can enlarge this chain by adding $\mathfrak{j}_0 := \mathfrak{h}_j$ and $\mathfrak{j}_{m+1} = \mathfrak{k}_{j+1}$, which are, respectively, of type (b) and (a). Thus, there exists an index $0 \le k \le m$ such that $\mathfrak{j}_0, \dots, \mathfrak{j}_k$ are of type (b) and $\mathfrak{j}_{k+1}, \dots, \mathfrak{j}_{m+1}$ are of type (a). Since the chain is maximal, the set $\mathscr{W}(\mathfrak{j}_k^*|\mathfrak{j}_{k+1})$ is empty. Thus, since $\mathfrak{j}_k$ and $\mathfrak{j}_{k+1}$ are not transverse, the labels $\gamma(\mathfrak{j}_k)$ and $\gamma(\mathfrak{j}_{k+1})$ are not joined by an edge of Γ.

However, since $\Gamma_{j+1} = \Gamma_j$, we have

$$\gamma(\mathfrak{j}_{k+1}) \in \gamma(\mathscr{W}(\mathfrak{k}_1^*|\mathfrak{k}_{j+1})) \cup \{\gamma(\mathfrak{k}_1^*), \gamma(\mathfrak{k}_{j+1})\} = \gamma(\mathscr{W}(\mathfrak{k}_1^*|\mathfrak{k}_j)) \cup \{\gamma(\mathfrak{k}_1^*), \gamma(\mathfrak{k}_j)\},$$

while $\mathfrak{j}_k$ is transverse to $\mathscr{H}(\mathfrak{k}_1^*|\mathfrak{k}_j) \cup \{\mathfrak{k}_1^*, \mathfrak{k}_j\}$, a contradiction. This proves the claim and lemma. □

4.4 A quasiconvexity criterion for median subalgebras

In this subsection, we provide a criterion (Proposition 4.25) for when a median subalgebra M of a CAT(0) cube complex X is quasiconvex. The subalgebra M will be required to satisfy two conditions, *edge-connectedness* and *weak quasiconvexity*, which we study separately in the next two subsections.

4.4.1 Edge-connected median subalgebras Let X be a CAT(0) cube complex.

Definition 4.18 A subset $A \subseteq X^{(0)}$ is *edge-connected* if, for all $x, y \in A$, there exists a sequence of points $x_1, \dots, x_n \in A$ such that $x_1 = x$, $x_n = y$ and, for all i, the points x_i and x_{i+1} are joined by an edge of X.

Remark 4.19 If $A \subseteq X^{(0)}$ is edge-connected, then there do not exist distinct halfspaces $\mathfrak{h}, \mathfrak{k} \in \mathscr{H}_A(X)$ with $\mathfrak{h} \cap A = \mathfrak{k} \cap A$. Indeed, the intersections $\mathfrak{h} \cap \mathfrak{k}$ and $\mathfrak{h}^* \cap \mathfrak{k}^*$ would both be nonempty, so, possibly swapping $\mathfrak{h}$ and $\mathfrak{k}$, we would either have $\mathfrak{h} \subsetneq \mathfrak{k}$ or $\mathfrak{h}$ and $\mathfrak{k}$ would be transverse. However, since A is edge connected and intersects both $\mathfrak{h} \cap \mathfrak{k}$ and $\mathfrak{h}^* \cap \mathfrak{k}^*$, we must have $A \cap \mathfrak{h}^* \cap \mathfrak{k} \neq \varnothing$ if $\mathfrak{h} \subsetneq \mathfrak{k}$, and either $A \cap \mathfrak{h}^* \cap \mathfrak{k} \neq \varnothing$ or $A \cap \mathfrak{h} \cap \mathfrak{k}^* \neq \varnothing$ if $\mathfrak{h}$ and $\mathfrak{k}$ are transverse. This contradicts the fact that $\mathfrak{h} \cap A = \mathfrak{k} \cap A$.

Lemma 4.20 *For a median subalgebra $M \subseteq X^{(0)}$, the following are equivalent:*

(1) *M is edge-connected.*

(2) *For all $x, y \in M$, there exists a geodesic $\alpha \subseteq X$ joining x and y such that $\alpha \cap X^{(0)} \subseteq M$.*

(3) *The restriction map $\mathrm{res}_M : \mathscr{H}_M(X) \to \mathscr{H}(M)$ is injective.*

Proof The implication (2) $\Longrightarrow$ (1) is clear and the implication (1) $\Longrightarrow$ (3) follows from Remark 4.19. Let us show that (3) $\Longrightarrow$ (2).

Since M is a discrete median algebra, it is isomorphic to the 0–skeleton of a CAT(0) cube complex $X(M)$; see [31, Theorem 6.1] or [89, Section 10]. Given $x, y \in M$, let $\beta \subseteq X(M)$ be a geodesic joining x and y, and let $x_1 = x, x_2, \dots, x_n = y$ be the elements of $\beta \cap M$ as they appear along β. Since the restriction map $\mathrm{res}_M : \mathscr{H}_M(X) \to \mathscr{H}(M)$ is injective, there is only one hyperplane $\mathfrak{w}_i \in \mathscr{W}(X)$ separating x_i and x_{i+1}, that is, these two points are joined by an edge of X. If $i \neq j$, then $\mathfrak{w}_i \neq \mathfrak{w}_j$, or β would cross the corresponding wall of M twice. We conclude that there exists a geodesic $\alpha \subseteq X$ with $\alpha \cap M = \{x_1, \dots, x_n\}$. □

By the 3rd characterisation in Lemma 4.20, edge-connected subalgebras can be viewed as a middle ground between general median subalgebras and convex subcomplexes; cf part (2) of Remark 2.2.

Lemma 4.21 *If $A \subseteq X^{(0)}$ is an edge-connected subset, then $\langle A \rangle$ is an edge-connected subalgebra.*

Proof Suppose for the sake of contradiction that $\langle A \rangle$ is not edge-connected. Then there exist distinct halfspaces $\mathfrak{h}, \mathfrak{k} \in \mathscr{H}_{\langle A \rangle}(X)$ with $\mathfrak{h} \cap \langle A \rangle = \mathfrak{k} \cap \langle A \rangle$ by Lemma 4.20. Note that $\mathfrak{h}, \mathfrak{k} \in \mathscr{H}_A(X)$, and $\mathfrak{h}^* \cap \mathfrak{k} \cap A = \varnothing$ and $\mathfrak{h} \cap \mathfrak{k}^* \cap A = \varnothing$. In particular, $\mathfrak{h} \cap A = \mathfrak{k} \cap A$, which violates Remark 4.19. □

Lemma 4.22 *Let $M \subseteq X^{(0)}$ be an edge-connected median subalgebra. Let $C \subseteq X$ be a convex subcomplex with gate-projection $\pi\colon X \to C$. Then:*

(1) *$\pi(M)$ is an edge-connected subalgebra of $C^{(0)}$.*

(2) *If $N \subseteq \pi(M)$ is an edge-connected subalgebra, then $M \cap \pi^{-1}(N)$ is edge-connected as well.*

Proof If vertices $x, y \in X$ are joined by an edge, then either $\pi(x)$ and $\pi(y)$ are joined by an edge or they are equal. Thus, part (1) is immediate from definitions.

Let us address part (2). Consider two points $x, y \in M \cap \pi^{-1}(N)$. Since N is edge-connected, there exists a geodesic $\alpha \subseteq C$ joining $\pi(x)$ and $\pi(y)$ with $\alpha \cap C^{(0)} \subseteq N$; see Lemma 4.20. It suffices to show that $M \cap \pi^{-1}(\alpha)$ is edge-connected.

In fact, since $\pi^{-1}(v) \cap M \neq \varnothing$ for every vertex $v \in \alpha$, it suffices to show that $M \cap \pi^{-1}(e)$ is edge-connected for every edge $e \subseteq \alpha$. In other words, we can suppose that $\pi(x)$ and $\pi(y)$ are joined by an edge $e \subseteq C$. Since M is edge-connected, there exists a geodesic $\beta \subseteq X$ joining x and y with $\beta \cap X^{(0)} \subseteq M$. Since π is a median morphism, the projection $\pi(\beta)$ is the image of a geodesic from $\pi(x)$ to $\pi(y)$, ie $\pi(\beta) = e$. Thus $\beta \cap X^{(0)} \subseteq M \cap \pi^{-1}(e)$, concluding the proof. □

4.4.2 Weakly quasiconvex median subalgebras Let X be a CAT(0) cube complex.

Definition 4.23 A subset $A \subseteq X^{(0)}$ is *weakly quasiconvex* if there exists a function $\eta\colon \mathbb{N} \to \mathbb{N}$ such that, for all $a, b, p \in X^{(0)}$ with $\mathscr{W}(p|a)$ transverse to $\mathscr{W}(p|b)$, we have

$$d(p, A) \leq \eta\big(\max\{d(a, A), d(b, A)\}\big).$$

Remark 4.24 (1) If $A \subseteq X^{(0)}$ is quasiconvex in the sense of Definition 2.30, then A is weakly quasiconvex. Indeed, suppose that $\mathcal{J}(A) \subseteq \mathcal{N}_R(A)$ and set $D = \max\{d(a, A), d(b, A)$. If $\mathscr{W}(p|a)$ and $\mathscr{W}(p|b)$ are transverse, then $p \in I(a, b)$. Thus, $p \in \mathcal{J}(\mathcal{N}_D(A))$ and Lemma 2.10 yields $d(p, A) \leq 2D + R =: \eta(D)$.

(2) If $A, B \subseteq X^{(0)}$ have finite Hausdorff distance, then A is weakly quasiconvex if and only if B is. This is straightforward, observing that η can always taken to be weakly increasing.

The following is the main result of this subsection.

Proposition 4.25 *If X has finite dimension and finite staircase length, then every edge-connected, weakly quasiconvex median subalgebra $M \subseteq X^{(0)}$ is quasiconvex.*

Proposition 4.25 fails for cube complexes of infinite staircase length, as the next example shows.

Example 4.26 Consider the standard structure of cube complex on $\mathbb{R}^2$. Let α be the geodesic line through all points (n,n) and $(n+1,n)$ with $n \in \mathbb{Z}$. Let $X \subseteq \mathbb{R}^2$ be the subcomplex that lies above α, including α itself. Note that X is a 2–dimensional CAT(0) cube complex of infinite staircase length, and $\alpha \subseteq X$ is an edge-connected median subalgebra that is not quasiconvex. It is not hard to see that α is weakly quasiconvex with $\eta(t) = 2t$.

The next lemma essentially proves the 2–dimensional case of Proposition 4.25.

Lemma 4.27 *Suppose that* $\dim X = 2$ *and that* X *has staircase length* d*. Let* $M \subseteq X^{(0)}$ *be an edge-connected median subalgebra. Consider* $x, y \in M$ *and* $z \in X^{(0)} \cap I(x,y)$*. Then there exist* $0 \le k \le d$ *and vertices* $z_0, z_1, z_2, \dots, z_k \in I(x,y)$ *and* $w_1, \dots, w_k \in I(x,y)$ *such that*

- $z_0 = z$*, while* $z_k \in M$ *and* $w_1, \dots, w_k \in M$*;*
- *the sets* $\mathscr{W}(z_i | w_{i+1}) \subseteq \mathscr{W}(X)$ *and* $\mathscr{W}(z_i | z_{i+1}) \subseteq \mathscr{W}(X)$ *are transverse for all* $0 \le i \le k-1$*.*

Proof If $z \in M$, we can simply take $k = 0$. If $z \notin M$, we begin with the following observation:

Claim *We can assume that there exist transverse hyperplanes* $\mathfrak{u} \in \mathscr{W}(x, z | y)$ *and* $\mathfrak{v} \in \mathscr{W}(y, z | x)$ *such that* x, z *lie in the carrier of* $\mathfrak{u}$ *and* y, z *lie in the carrier of* $\mathfrak{v}$*.*

Proof Up to replacing x and y with other points in the interval $I(x,y)$, we can assume that there do not exist points $x', y' \in I(x,y)$ with $z \in I(x', y')$, except for $\{x', y'\} = \{x, y\}$.

Since M is edge-connected, there exists a point $x' \in M \cap I(x,y)$ such that x and x' are separated by a single hyperplane $\mathfrak{u} \in \mathscr{W}(X)$. By the above assumption on x and y, we must have $z \notin I(x', y)$, hence $\varnothing \neq \mathscr{W}(z | x', y) = \mathscr{W}(z, x | x', y) \subseteq \{\mathfrak{u}\}$. It follows that $\mathscr{W}(z, x | x', y) = \{\mathfrak{u}\}$.

Observing that $\mathscr{W}(z | \mathfrak{u}) \subseteq \mathscr{W}(z | x', y) = \mathscr{W}(z, x | x', y) = \{\mathfrak{u}\}$, we conclude that $\mathscr{W}(z | \mathfrak{u})$ is empty. This shows that the carrier of $\mathfrak{u}$ contains z, while it is clear that it also contains x. The existence of $\mathfrak{v}$ is obtained similarly. Finally, since $\mathfrak{v} \in \mathscr{W}(y, z | x)$ and $\mathfrak{v} \neq \mathfrak{u}$, we must have $\mathfrak{v} \in \mathscr{W}(y, z | x, x')$. Recalling that $\mathfrak{u} \in \mathscr{W}(z, x | x', y)$, this shows that $\mathfrak{u}$ and $\mathfrak{v}$ are transverse. ◁

Now, the sets $\mathscr{H}(z|x)$ and $\mathscr{H}(z|y)$ are transverse, respectively, to $\mathfrak{u}$ and $\mathfrak{v}$. Since $\dim X = 2$, the set $\mathscr{H}(z|x)$ is a descending chain $\mathfrak{h}_1 \supsetneq \dots \supsetneq \mathfrak{h}_m$, and $\mathscr{H}(z|y)$ is a descending chain $\mathfrak{k}_1 \supsetneq \dots \supsetneq \mathfrak{k}_n$. Note that $\mathfrak{k}_1$ and $\mathfrak{h}_1$ are bounded, respectively, by $\mathfrak{u}$ and $\mathfrak{v}$, as depicted in Figure 4.

Since $\mathfrak{h}_1$ and $\mathfrak{k}_1$ are transverse, there exists a function $\tau : \{1, \dots, m\} \to \{1, \dots, n\}$ such that $\mathfrak{h}_i$ is transverse to $\mathfrak{k}_j$ if and only if $1 \le j \le \tau(i)$. Note that $\tau(1) = n$ and that τ is weakly decreasing.

Let $1 \le i_1 < \dots < i_{k-1} < m$ be all indices i with $\tau(i+1) < \tau(i)$. Also define $i_k := m$ and set $\tau_s := \tau(i_s)$ for simplicity. Since the halfspaces $\mathfrak{h}^*_{i_k}, \dots, \mathfrak{h}^*_{i_1}$ and $\mathfrak{k}_{\tau_k}, \dots, \mathfrak{k}_{\tau_1}$ form a length-k staircase, while X has staircase length d, we must have $k \le d$.

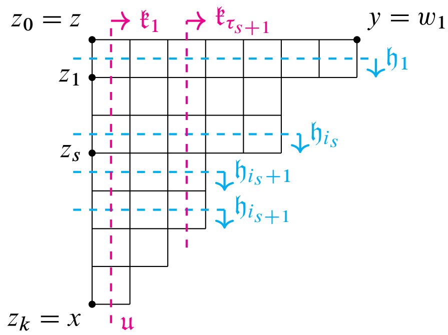


Figure 4

Set $z_0 = z$ and $w_1 = y$. For $1 \le s \le k$, let $z_s \in I(x,y)$ be the point with $\mathscr{H}(z|z_s) = \{\mathfrak{h}_1, \dots, \mathfrak{h}_{i_s}\}$. In particular, $z_k = x \in M$. Since M is edge-connected, there exist points

$$w_{s+1} \in M \cap \mathfrak{h}_{i_s} \cap \mathfrak{h}^*_{i_s+1} \cap \mathfrak{k}^*_{\tau_{s+1}+1}.$$

Observing that $\mathscr{H}(z_s|w_{s+1}) \subseteq \{\mathfrak{k}_1, \dots, \mathfrak{k}_{\tau_{s+1}}\}$ is transverse to $\mathscr{H}(z_s|z_{s+1}) = \{\mathfrak{h}_{i_s+1}, \dots, \mathfrak{h}_{i_{s+1}}\}$, this completes the proof of the lemma. □

The next lemma allows us to reduce the proof of Proposition 4.25 to the 2–dimensional case.

Lemma 4.28 *Let X have dimension r and staircase length d. Let $M \subseteq X^{(0)}$ be an edge-connected median subalgebra. For all points $x, y \in M$ and $z \in X^{(0)} \cap I(x,y)$, there exists a median subalgebra $N \subseteq X^{(0)} \cap I(x,y)$ with the following properties*:

- *$x, y, z \in N$ and $\operatorname{rk} N \le 2$.*
- *N has staircase length $\le d(1+2r^2)^2$.*
- *N and $N \cap M$ are edge-connected.*

Proof Let $\pi_{xz}\colon X \to I(x,z)$ be the gate-projection and note that $\pi_{xz}(y) = z$. By Lemma 4.22(1), the projection $\pi_{xz}(M)$ is an edge-connected median subalgebra containing x and z. Thus there exists a (combinatorial) geodesic $\alpha \subseteq I(x,z)$ joining x and z with $\alpha \cap X^{(0)} \subseteq \pi_{xz}(M)$.

By Lemma 4.22(2), the median subalgebras $N' := \pi_{xz}^{-1}(\alpha) \cap I(x,y) \cap X^{(0)}$ and $M \cap N'$ are edge-connected. Lemma 4.16 shows that N' has staircase length $\le d(1+2r^2)$, while it is clear that $\operatorname{rk} N' \le \dim X = r$.

Note that $x, y, z \in N'$. Since $\pi_{xz}(I(z,y)) = \{z\}$, the entire interval $I(z,y) \cap X^{(0)}$ is contained in N'. Consider the projection $\pi_{zy}\colon X \to I(z,y)$. Since $M \cap N'$ is edge-connected, Lemma 4.22 again shows that the projection $\pi_{zy}(M \cap N')$ is edge-connected, and we can join y and z by a geodesic β with $\beta \cap X^{(0)} \subseteq \pi_{zy}(M \cap N')$. Repeating the above argument, we see that $N := N' \cap \pi_{yz}^{-1}(\beta)$ has staircase length $\le d(1+2r^2)^2$, that N and $N \cap M$ are edge-connected, and that $x, y, z \in N$ (recall that N' is a finite median algebra, so it is naturally identified with the 0–skeleton of a CAT(0) cube complex and we can run the above argument in this cube complex).

We are left to show that $\operatorname{rk} N \le 2$. Since $x, y \in N \subseteq I(x, y)$, every wall of N either separates x from y, z, or it separates x, z from y. If two walls of N separate x and z, then they are not transverse; cf Claim 1 during the proof of Lemma 4.16. The same is true of walls separating z and y. This implies that $\operatorname{rk} N \le 2$, concluding the proof. □

Proof of Proposition 4.25 Let X have dimension r and staircase length d. Let M be an edge-connected, weakly quasiconvex subalgebra. We will show that $d_{\mathrm{Haus}}(I(x, y), M \cap I(x, y))$ remains uniformly bounded as x and y vary in M, which implies that M is quasiconvex.

Consider $x, y \in M$ and $z \in X^{(0)} \cap I(x, y)$. By Lemma 4.28, the points x, y, z lie in a median subalgebra $N \subseteq X^{(0)} \cap I(x, y)$ such that N and $N \cap M$ are edge-connected, $\operatorname{rk} N \le 2$, and N has staircase length $\le d(1 + 2r^2)^2$.

Viewing N as the vertex set of a finite CAT(0) cube complex and applying Lemma 4.27 to $M \cap N$, there exist points $z_0 = z, z_1, \dots, z_{k-1} \in N$ and $z_k, w_1, \dots, w_k \in N \cap M$ with $k \le d(1 + 2r^2)^2$, such that each wall of N separating z_i and z_{i+1} is transverse to every wall of N separating z_i and w_{i+1}. The same is true of hyperplanes of X separating these points.

Since M is weakly quasiconvex, it admits a function η as in Definition 4.23. Without loss of generality, we can take η to be weakly increasing. Then, since $d(w_i, M) = 0$, we have

$$\begin{aligned} d(z, M) \le \max\{\eta(d(z_1, M)), \eta(0)\} &\le \max\{\eta^2(d(z_2, M)), \eta^2(0), \eta(0)\} \\ &\le \cdots \le \max\{\eta^k(0), \dots, \eta^2(0), \eta(0)\}. \end{aligned}$$

The last constant only depends on d, r and η, so this proves that M is quasiconvex. □

4.5 Fixed subgroups in right-angled groups

In this subsection, we combine the results of the previous two subsections to prove Theorem C.

Let Γ be a finite simplicial graph. Our focus will be on the right-angled Artin group $\mathcal{A} = \mathcal{A}_\Gamma$ and the universal cover of its Salvetti complex $\mathcal{X} = \mathcal{X}_\Gamma$. Throughout, we will identify $\mathcal{A} \cong \mathcal{X}^{(0)}$.

However, all results and proofs in this subsection (except for Remark 4.29) immediately extend to right-angled Coxeter groups $\mathcal{W} = \mathcal{W}_\Gamma$ and Davis complexes $\mathcal{Y}_\Gamma$, without requiring any adaptations. We suggest that the reader keep track of this as they make their way through the results, in view of Corollary 4.35 below. The relevant properties shared by RAAGs and RACGs are:

- The Cayley graph of $\mathcal{A}/\mathcal{W}$ associated to the standard generators (vertices of Γ) is the 1–skeleton of a CAT(0) cube complex (the universal cover of the Salvetti/Davis complex) of finite staircase length (Lemma 4.17).

- Hyperplanes are labelled by vertices of Γ and labels of transverse hyperplanes are joined by an edge of Γ.
- Elementary automorphisms of $\mathcal{A}$ and $\mathcal{W}$ (as defined in Section 3.4) have the same form with respect to standard generators.

We are interested in the subgroups $U_0(\mathcal{A}) \leq U(\mathcal{A})$ and $\mathrm{Aut}_0\,\mathcal{W} \leq \mathrm{Aut}\,\mathcal{W}$ generated by inversions, folds and partial conjugations, as defined at the end of Section 3.4.

Given a subset $\Delta \subseteq \Gamma^{(0)}$, it is convenient to introduce the notation

$$\Delta^{\perp} = \bigcap_{v \in \Delta} \mathrm{lk}\, v.$$

Remark 4.29 It is not hard to observe that a subgroup of $\mathcal{A}$ is an intersection of stabilisers of hyperplanes of $\mathcal{X}$ if and only if it is conjugate to a subgroup of the form $\mathcal{A}_{\Delta^{\perp}}$ for some $\Delta \subseteq \Gamma$.

Although we will not be using this remark in the present paper, we find it interesting in relation to Lemma 4.30 below: elements of $U_0(\mathcal{A})$ permute hyperplane-stabilisers while preserving labels.

Statements similar to the next lemma have been widely used in the literature, eg in [24, Proposition 3.2; 27, Proposition 3.2; 28, Section 3]). Compared to these references, we get a slightly stronger result because here we are only concerned with untwisted automorphisms.

Lemma 4.30 *For every $\varphi \in U_0(\mathcal{A})$ and $\Delta \subseteq \Gamma$, the subgroups $\mathcal{A}_{\Delta^{\perp}}$ and $\varphi(\mathcal{A}_{\Delta^{\perp}})$ are conjugate.*

Proof It suffices to prove the lemma for elementary generators. It is clear that it holds for inversions, so we are left to consider folds and partial conjugations.

If $\tau_{v,w}$ is a fold, then $\tau_{v,w}(\mathcal{A}_{\Delta^{\perp}}) = \mathcal{A}_{\Delta^{\perp}}$. This is immediate if $v \notin \Delta^{\perp}$. If instead $v \in \Delta^{\perp}$, we have $\Delta \subseteq \mathrm{lk}\, v \subseteq \mathrm{lk}\, w$, hence $w \in \Delta^{\perp}$.

If $\kappa_{w,C}$ is a partial conjugation, then $\kappa_{w,C}(\mathcal{A}_{\Delta^{\perp}})$ is either $\mathcal{A}_{\Delta^{\perp}}$ or $w^{-1}\mathcal{A}_{\Delta^{\perp}}w$. This is clear if $\Delta^{\perp}$ intersects at most one connected component of $\Gamma \setminus \mathrm{st}\, w$. Suppose instead that $\Delta^{\perp}$ intersects two distinct components of $\Gamma \setminus \mathrm{st}\, w$. Then, for every $a \in \Delta$, the fact that $\Delta^{\perp} \subseteq \mathrm{lk}\, a$ implies that $a \in \mathrm{lk}\, w$. Thus, $w \in \Delta^{\perp}$ and $\kappa_{w,C}(\mathcal{A}_{\Delta^{\perp}}) = \mathcal{A}_{\Delta^{\perp}}$ in this case. □

Corollary 4.31 *For every $\varphi \in U_0(\mathcal{A})$ and $g \in \mathcal{A}$, we have $\Gamma(\varphi(g))^{\perp} = \Gamma(g)^{\perp}$.*

Proof It suffices to show that $\Gamma(\varphi(g))^{\perp} \supseteq \Gamma(g)^{\perp}$ for all $\varphi \in U_0(\mathcal{A})$ and $g \in \mathcal{A}$. Note that g has a conjugate in $\mathcal{A}_{\Gamma(g)} \leq \mathcal{A}_{\Gamma(g)^{\perp\perp}}$. Thus, Lemma 4.30 implies that a conjugate of $\varphi(g)$ lies in $\mathcal{A}_{\Gamma(g)^{\perp\perp}}$. This shows that $\Gamma(\varphi(g)) \subseteq \Gamma(g)^{\perp\perp}$, hence $\Gamma(\varphi(g))^{\perp} \supseteq \Gamma(g)^{\perp\perp\perp} = \Gamma(g)^{\perp}$, as required. □

For the next results, recall that we are identifying elements of $\mathcal{A}$ and vertices of $\mathcal{X}$.

Lemma 4.32 *For every $\varphi \in U_0(\mathcal{A})$, there exists a constant $K(\varphi)$ with the following property. For all $x, y \in \mathcal{A}$, at most $K(\varphi)$ among the hyperplanes in $\mathscr{W}(\varphi(x)|\varphi(y))$ have label outside $\gamma(\mathscr{W}(x|y))^{\perp\perp}$.*

Proof It suffices to show that, for every $g \in \mathcal{A}$, at most $K(\varphi)$ among the hyperplanes in $\mathscr{W}(1|\varphi(g))$ have label outside $\gamma(\mathscr{W}(1|g))^{\perp\perp}$.

Since Γ has only finitely many subsets, Lemma 4.30 shows that there exists a constant $K'(\varphi)$ with the following property. For every $\Delta \subseteq \Gamma$ there exists $x_\Delta \in \mathcal{A}$ with $\varphi(\mathcal{A}_{\Delta^\perp}) = x_\Delta \mathcal{A}_{\Delta^\perp} x_\Delta^{-1}$ and $|x_\Delta| \le K'(\varphi)$. Here $|\cdot|$ denotes word length with respect to the standard generators.

Now, consider $g \in \mathcal{A}$ and set $\Delta(g) := \gamma(\mathscr{W}(1|g))^\perp$. Then $g \in \mathcal{A}_{\Delta(g)^\perp}$ and the above observation shows that all but $2|x_{\Delta(g)}|$ hyperplanes in $\mathscr{W}(1|\varphi(g))$ have label in $\Delta(g)^\perp$. Taking $K(\varphi) := 2K'(\varphi)$, this concludes the proof. □

Proposition 4.33 *If $\varphi \in U_0(\mathcal{A})$, the subgroup* $\mathrm{Fix}\,\varphi$ *is a weakly quasiconvex subset of $\mathcal{X}^{(0)} \cong \mathcal{A}$.*

Proof Consider vertices $a, b, p \in \mathcal{X}$ with $\mathscr{W}(p|a)$ transverse to $\mathscr{W}(p|b)$. Set

$$D := \max\{d(a, \mathrm{Fix}\,\varphi), d(b, \mathrm{Fix}\,\varphi)\}.$$

Let $K = K(\varphi)$ be as in Lemma 4.32, let ζ_1, ζ_2 be the functions provided by Lemma 2.34 (without loss of generality, strictly increasing), and let C be a constant such that

$$\varphi(m(x,y,z)) \approx_C m(\varphi(x), \varphi(y), \varphi(z)) \quad \text{for all } x, y, z \in \mathcal{X}.$$

Let us write a', b', p' for $\varphi(a), \varphi(b), \varphi(p)$. Since $\mathscr{W}(p|a)$ and $\mathscr{W}(p|b)$ are transverse, we have $p \in I(a,b)$, so $\mathscr{W}(p|a,b) = \varnothing$. Observing that $m(a', b', p') \approx_C \varphi(m(a,b,p)) = p'$, we also have $\#\mathscr{W}(p'|a',b') \le C$. Finally, by the first inequality in Lemma 2.34, we have $a' \approx_{D'} a$ and $b' \approx_{D'} b$, where $D' := \zeta_1^{-1}(D)$.

Putting together these inequalities, we obtain

$$\begin{aligned}\#\mathscr{W}(p|p') &= \#\mathscr{W}(p|a',b',p') + \#\mathscr{W}(p,a'|b',p') + \#\mathscr{W}(p,b'|a',p') + \#\mathscr{W}(p,a',b'|p') \\ &\le \#\mathscr{W}(p|a,b) + 2D' + \#\mathscr{W}(p,a'|b,p') + D' + \#\mathscr{W}(p,b'|a,p') + D' + \#\mathscr{W}(a',b'|p') \\ &\le \#\mathscr{W}(p,a'|b,p') + \#\mathscr{W}(p,b'|a,p') + C + 4D'.\end{aligned}$$

By Lemma 4.32, at most K elements of $\mathscr{W}(a'|p')$ have label in $\gamma(\mathscr{W}(a|p))^\perp$. Since $\mathscr{W}(p|a)$ and $\mathscr{W}(p|b)$ are transverse, we deduce that $\#\mathscr{W}(p,a'|b,p') \le K$ and, similarly, $\#\mathscr{W}(p,b'|a,p') \le K$. We conclude that

$$d(p, \varphi(p)) = \#\mathscr{W}(p|p') \le 2K + C + 4D'.$$

Lemma 2.34 gives $d(p, \mathrm{Fix}\,\varphi) \le \zeta_2(2K + C + 4 \cdot \zeta_1^{-1}(D))$, as required by Definition 4.23. □

Corollary 4.34 *For every $\varphi \in U_0(\mathcal{A})$, the subgroup* $\mathrm{Fix}\,\varphi$ *is convex-cocompact in $\mathcal{A} \curvearrowright \mathcal{X}$.*

Proof Set $H := \operatorname{Fix}\varphi$. By Theorem B, H is finitely generated, so there exists $R \geq 0$ such that $\mathcal{N}_R(H)$ is edge-connected, viewed as a subset of $\mathcal{X}$. By Lemma 4.21, the median subalgebra $M := \langle \mathcal{N}_R(H) \rangle$ is edge-connected. Since H is an approximate median subalgebra by Lemma 2.35, Proposition 4.1 shows that M is at finite Hausdorff distance from H. Since H is weakly quasiconvex by Proposition 4.33, so is M.

Finally, $\mathcal{X}$ has finite staircase length by Lemma 4.17. We have shown that $M \subseteq \mathcal{X}^{(0)}$ is edge-connected and weakly quasiconvex, so Proposition 4.25 implies that M is quasiconvex. By Lemma 2.10, $\operatorname{Hull} M$ is at finite Hausdorff distance from M, which is at finite Hausdorff distance from H. This implies that H acts cocompactly on the convex subcomplex $\operatorname{Hull} M \subseteq \mathcal{X}$. □

The discussion in this subsection immediately extends to right-angled Coxeter groups $\mathcal{W}$ and the finite-index subgroup $\operatorname{Aut}_0 \mathcal{W} \leq \operatorname{Aut} \mathcal{W}$ generated by folds and partial conjugations.

Corollary 4.35 *For every $\varphi \in \operatorname{Aut}_0 \mathcal{W}$, the subgroup $\operatorname{Fix}\varphi$ is convex-cocompact in $\mathcal{W} \curvearrowright \mathcal{Y}$, where $\mathcal{Y}$ is the universal cover of the Davis complex.*

Recalling Lemma 3.2 and Remark 2.31, the previous two corollaries prove Theorem C.

5 Invariant splittings of RAAGs

This section only contains the proofs of Proposition D and Corollary E, which are independent from all other results mentioned in the introduction.

Let Γ be a finite simplicial graph and let $\mathcal{A} = \mathcal{A}_\Gamma$ be the corresponding right-angled Artin group. All results and proofs in this section immediately extend to the right-angled Coxeter group $\mathcal{W}_\Gamma$ and automorphisms in $\operatorname{Aut}_0 \mathcal{W}_\Gamma$. We encourage the reader to verify this as they go through the material, emphasising that only Lemmas 5.3 and 5.4 and Corollary 5.10 require any kind of attention, as all other results in this section are purely about the finite graph Γ.

The following is Proposition D from the introduction.

Proposition 5.1 *Let $\mathcal{A}$ be directly irreducible, freely irreducible and noncyclic. Then there exists an amalgamated product splitting $\mathcal{A} = \mathcal{A}_+ *_{\mathcal{A}_0} \mathcal{A}_-$, with $\mathcal{A}_\pm$ and $\mathcal{A}_0$ parabolic subgroups of $\mathcal{A}$, such that the corresponding Bass–Serre tree $\mathcal{A} \curvearrowright T$ is $U_0(\mathcal{A})$–invariant. That is, for every $\varphi \in U_0(\mathcal{A})$, there exists an isometry $f : T \to T$ satisfying $f \circ g = \varphi(g) \circ f$ for all $g \in \mathcal{A}$.*

Proposition 5.1 follows from Corollary 5.4 and Proposition 5.5 below. The latter will be proved right after Lemma 5.9.

Given a partition $\Gamma^{(0)} = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$, we write $\mathcal{A}_+ := \mathcal{A}_{\Lambda \sqcup \Lambda^+}$ and $\mathcal{A}_- := \mathcal{A}_{\Lambda \sqcup \Lambda^-}$ for simplicity. If $\Lambda^\pm$ are nonempty and $d(\Lambda^+, \Lambda^-) \geq 2$ (where d denotes the graph metric on Γ), then the partition corresponds to a splitting as amalgamated product,

$$\mathcal{A} = \mathcal{A}_+ *_{\mathcal{A}_\Lambda} \mathcal{A}_-.$$

We denote by $\mathcal{A} \curvearrowright T_\Lambda$ the Bass–Serre tree of this splitting. This will not cause any ambiguity related to possible different choices of the sets $\Lambda^\pm$ in the following discussion.

We are interested in partitions of $\Gamma^{(0)}$ that satisfy a certain list of properties.

Definition 5.2 A partition $\Gamma^{(0)} = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$ into three nonempty subsets is *good* if:

(i) $d(\Lambda^+, \Lambda^-) \geq 2$, where d is the graph metric on Γ.

(ii) For every $\epsilon \in \{\pm\}$ and $w \in \Lambda^\epsilon$, there does not exist $v \in \Lambda \sqcup \Lambda^{-\epsilon}$ with $\mathrm{lk}\, v \subseteq \mathrm{lk}\, w \cup \Lambda^\epsilon$.

(iii) For every $\epsilon \in \{\pm\}$ and $w \in \Lambda^\epsilon$, the subgraph of Γ spanned by $(\Lambda \sqcup \Lambda^{-\epsilon}) \setminus \mathrm{st}\, w$ is connected.

We will simply write $\Gamma = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$, rather than $\Gamma^{(0)} = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$.

The motivation for Definition 5.2 comes from the next lemma and the subsequent corollary. Definition 5.2 actually contains slightly stronger requirements than what is strictly necessary to the two results: this will facilitate the inductive construction of good partitions of graphs Γ.

Lemma 5.3 *Let $\Gamma = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$ be a good partition. For every $\psi \in U_0(\mathcal{A})$, there exists $\varphi \in U_0(\mathcal{A})$ representing the same outer automorphism and simultaneously satisfying $\varphi(\mathcal{A}_+) = \mathcal{A}_+$ and $\varphi(\mathcal{A}_-) = \mathcal{A}_-$ (hence also $\varphi(\mathcal{A}_\Lambda) = \mathcal{A}_\Lambda$).*

Proof Inversions preserve $\mathcal{A}^+$ and $\mathcal{A}^-$. Given vertices $v, w \in \Gamma$ with $\mathrm{lk}\, v \subseteq \mathrm{lk}\, w$, condition (ii) implies that either $w \in \Lambda$, or $\{v, w\} \subseteq \Lambda^+$, or $\{v, w\} \subseteq \Lambda^-$. Thus, folds also preserve $\mathcal{A}^+$ and $\mathcal{A}^-$.

We are left to prove the lemma in the case when ψ is a partial conjugation $\kappa_{w,C}$. If $w \in \Lambda$, it is clear that $\kappa_{w,C}$ preserves $\mathcal{A}^+$ and $\mathcal{A}^-$. Thus, let us assume without loss of generality that $w \in \Lambda^+$. By condition (iii), the set $\Lambda \cup \Lambda^-$ intersects a unique connected component $K \subseteq \Gamma \setminus \mathrm{st}\, w$.

If $K \neq C$, then $\kappa_{w,C}$ is the identity on $\mathcal{A}^-$, so $\mathcal{A}^\pm$ are both preserved. If $K = C$, then $\kappa_{w,C}$ represents the same outer automorphism as $\kappa_{w^{-1},K_1} \cdots \kappa_{w^{-1},K_k}$, where $K_1, \dots, K_k$ are the connected components of $\Gamma \setminus \mathrm{st}\, w$ other than K. Again, the latter is the identity on $\mathcal{A}^-$, so $\mathcal{A}^\pm$ are preserved. □

This shows that T_Λ is invariant under twisting by elements of $U_0(\mathcal{A})$:

Corollary 5.4 *Let $\Gamma = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$ be a good partition. For every $\varphi \in U_0(\mathcal{A})$, there exists an automorphism $f : T_\Lambda \to T_\Lambda$ satisfying $f \circ g = \varphi(g) \circ f$ for all $g \in \mathcal{A}$.*

Proof If φ is inner, we can take f to coincide with an element of $\mathcal{A}$. If $\varphi(\mathcal{A}_+) = \mathcal{A}_+$ and $\varphi(\mathcal{A}_-) = \mathcal{A}_-$, the statement is also clear, since the Bass–Serre tree can be defined in terms of cosets of $\mathcal{A}^{\pm}$. By Lemma 5.3, every element of $U_0(\mathcal{A})$ is a product of two automorphisms of these two types. □

Our next goal is to show that good partitions (almost) always exist. We say that Γ is *irreducible* if it does not split as a nontrivial join (equivalently, the opposite graph Γ^o is connected).

Proposition 5.5 *If Γ is connected, irreducible and not a singleton, then Γ admits a good partition.*

Proposition 5.5 and Corollary 5.4 immediately imply Proposition 5.1, as well as the analogous result for right-angled Coxeter groups.

Before proving Proposition 5.5, we need to obtain a few lemmas.

Lemma 5.6 *If Γ is connected and $\operatorname{diam}\Gamma^{(0)} \geq 3$, there exists a good partition of Γ.*

Proof Let $x, y \in \Gamma$ be arbitrary vertices with $d(x, y) \geq 3$. Let C_y be the connected component of $\Gamma \setminus \operatorname{st} x$ that contains y. Similarly, let C_x be the connected component of $\Gamma \setminus \operatorname{st} y$ that contains x.

Since $d(x, y) \geq 3$, we have $\operatorname{st} x \cap \operatorname{st} y = \varnothing$, hence $\operatorname{st} y \subseteq C_y$ and $\operatorname{st} x \subseteq C_x$. Since Γ is connected, $\Gamma \setminus C_x$ is also connected. Note that $\operatorname{st} x$ and $\Gamma \setminus C_x$ are disjoint and $y \in \Gamma \setminus C_x$. This implies that $\Gamma \setminus C_x \subseteq C_y$. In conclusion, $\Gamma = C_x \cup C_y$.

Note that, if $z \in \Gamma^{(0)}$ and $\operatorname{lk} z \cap C_y = \varnothing$, we cannot have $z \in C_y$. Indeed, this would imply that $C_y = \{z\}$ and $\operatorname{lk} z \subseteq \operatorname{st} x$. Since $y \in C_y$, we would then have $y = z$ and $\operatorname{lk} y \subseteq \operatorname{st} x$, contradicting the fact that Γ is connected and $d(x, y) \geq 3$.

Thus, we can define

$$\begin{aligned} \Lambda^+ &:= \{z \in \Gamma^{(0)} \mid \operatorname{st} z \cap C_y = \varnothing\} = \{z \in \Gamma^{(0)} \mid \operatorname{lk} z \cap C_y = \varnothing\}, \\ \Lambda^- &:= \{z \in \Gamma^{(0)} \mid \operatorname{st} z \cap C_x = \varnothing\} = \{z \in \Gamma^{(0)} \mid \operatorname{lk} z \cap C_x = \varnothing\}, \\ \Lambda &:= \Gamma^{(0)} \setminus (\Lambda^+ \sqcup \Lambda^-). \end{aligned}$$

Note that $x \in \Lambda^+$ and $y \in \Lambda^-$. If $z \in \Lambda^+$ and $w \in \Lambda^-$, we have $\operatorname{st} z \cap \operatorname{st} w = \varnothing$, since $\Gamma = C_x \cup C_y$. This shows that $d(\Lambda^+, \Lambda^-) \geq 3$. Since Γ is connected, we also conclude that $\Lambda \neq \varnothing$. We are left to verify conditions (ii) and (iii) of Definition 5.2.

If $v \in \Lambda$, then $\operatorname{lk} v$ intersects both C_x and C_y. Since C_y is disjoint from $\operatorname{lk} w \cup \Lambda^+$ for every $w \in \Lambda^+$ (and similarly with C_x and Λ^-), this implies condition (ii) when $v \in \Lambda$. On the other hand, the case with $v \in \Lambda^{-\epsilon}$ is immediate from the fact that $d(\Lambda^+, \Lambda^-) \geq 3$ and Γ is connected.

Finally, let us check condition (iii). Without loss of generality, we can suppose that $w \in \Lambda^+$. Note that C_y is connected, contained in $(\Lambda \sqcup \Lambda^-) \setminus \operatorname{st} w$, and it intersects the link of every point of Λ. Moreover, since $\Lambda^- \cap C_x = \varnothing$ and $\Gamma = C_x \cup C_y$, we have $\Lambda^- \subseteq C_y$. This shows that $(\Lambda \sqcup \Lambda^-) \setminus \operatorname{st} w$ is connected, concluding the proof. □

When the previous lemma cannot be applied, we will construct a good partition of Γ inductively, extending good partitions on subgraphs. We now prove a sequence of three lemmas aimed precisely at this, after which we will give the argument for Proposition 5.5.

For $x \in \Gamma^{(0)}$, let $\Gamma \setminus x$ be the graph obtained by removing x and all open edges incident to x.

Lemma 5.7 *Let $\Gamma \setminus x = \Delta^+ \sqcup \Delta \sqcup \Delta^-$ be a good partition. Then one of the following happens*:

1. *There exist $w \in \Delta^+$ and $z \in \Delta^-$ with* $\mathrm{lk}\, x \subseteq \mathrm{lk}\, z \cap \mathrm{lk}\, w$.
2. *The partition of Γ with $\Lambda^+ = \Delta^+ \sqcup \{x\}$, $\Lambda = \Delta$, $\Lambda^- = \Delta^-$ is good.*
3. *The partition of Γ with $\Lambda^+ = \Delta^+$, $\Lambda = \Delta \sqcup \{x\}$, $\Lambda^- = \Delta^-$ is good.*
4. *The partition of Γ with $\Lambda^+ = \Delta^+$, $\Lambda = \Delta$, $\Lambda^- = \Delta^- \sqcup \{x\}$ is good.*

Proof We begin with the following observation:

Claim *If there exists $w \in \Delta^+$ such that $\mathrm{lk}\, x \subseteq \mathrm{lk}\, w \cup \Delta^+$, we are either in case* (1) *or in case* (2).

Proof We assume that we are not in case (1) and show that the partition of Γ in case (2) is good. We need to verify conditions (i)–(iii) from Definition 5.2.

Since $d(\Delta^+, \Delta^-) \geq 2$ (both in $\Gamma \setminus x$ and in Γ), the set Δ^- is disjoint from $\mathrm{lk}\, w \cup \Delta^+$. Since $\mathrm{lk}\, x \subseteq \mathrm{lk}\, w \cup \Delta^+$, it follows that $\Delta^- \cap \mathrm{st}\, x = \varnothing$, hence $d(\Lambda^+, \Lambda^-) \geq 2$. This proves condition (i).

If condition (ii) fails, there exist $u \in \Lambda^\epsilon$ and $v \in \Lambda \sqcup \Lambda^{-\epsilon}$ with $\mathrm{lk}\, v \subseteq \mathrm{lk}\, u \cup \Lambda^\epsilon$. Since the partition of $\Gamma \setminus x$ is good, we must have either $v = x$ or $u = x$. If $v = x$, then $u \in \Delta^-$ and

$$\mathrm{lk}\, x \subseteq (\mathrm{lk}\, u \cup \Delta^-) \cap (\mathrm{lk}\, w \cup \Delta^+) = \mathrm{lk}\, u \cap \mathrm{lk}\, w,$$

which lands us in case (1). If instead $u = x$, we have $v \in \Delta \sqcup \Delta^-$ with

$$\mathrm{lk}\, v \subseteq \mathrm{lk}\, x \cup \Lambda^+ \subseteq \mathrm{lk}\, w \cup \Delta^+ \cup \{x\}.$$

This violates condition (ii) for the partition of $\Gamma \setminus x$.

Finally, suppose that condition (iii) fails. Thus, there exists $u \in \Lambda^\epsilon$ such that $(\Lambda \sqcup \Lambda^{-\epsilon}) \setminus \mathrm{st}\, u$ is disconnected. Since the partition of $\Gamma \setminus x$ is good, this can happen only in two ways: either $u = x$, or $u \in \Lambda^-$ and x is isolated in $(\Lambda \sqcup \Lambda^+) \setminus \mathrm{st}\, u$. In the latter case, we have $\mathrm{lk}\, x \subseteq \mathrm{lk}\, u \cup \Delta^-$, which again leads to case (1).

Suppose instead that $u = x$ and let us show that $(\Lambda \sqcup \Lambda^-) \setminus \mathrm{st}\, x = (\Delta \sqcup \Delta^-) \setminus \mathrm{lk}\, x$ is connected. Since $\mathrm{lk}\, x \subseteq \mathrm{lk}\, w \cup \Delta^+$, the set $(\Delta \sqcup \Delta^-) \setminus \mathrm{lk}\, x$ contains $(\Delta \sqcup \Delta^-) \setminus \mathrm{lk}\, w$. The latter is connected, as the partition of $\Gamma \setminus x$ satisfies condition (iii). Since condition (ii) is satisfied, every point of $(\Delta \sqcup \Delta^-) \cap \mathrm{lk}\, w = \Delta \cap \mathrm{lk}\, w$ is joined by an edge to a point of $\Gamma \setminus (\mathrm{lk}\, w \cup \Lambda^+) = (\Delta \sqcup \Delta^-) \setminus \mathrm{lk}\, w$. Thus, the star of every point of $(\Delta \sqcup \Delta^-) \setminus \mathrm{lk}\, x$ intersects the connected set $(\Delta \sqcup \Delta^-) \setminus \mathrm{lk}\, w$, proving that $(\Delta \sqcup \Delta^-) \setminus \mathrm{lk}\, x$ is connected. This completes the proof of the claim. ◁

By the claim, if there exist either $w \in \Delta^+$ with $\operatorname{lk} x \subseteq \operatorname{lk} w \cup \Delta^+$ or $z \in \Delta^-$ with $\operatorname{lk} x \subseteq \operatorname{lk} z \cup \Delta^-$, then we are in cases (1), (2) or (4). In order to conclude the proof of the lemma, let us suppose that neither of the two inclusions is satisfied. We will show that the partition in case (3) is good.

Condition (i) is clear. Condition (ii) is immediate from the corresponding condition for $\Gamma \setminus x$ and our assumption that $\operatorname{lk} x$ be not contained in any subsets as in the previous paragraph.

Suppose that condition (iii) fails. Then there exists $u \in \Lambda^{\epsilon}$ such that $(\Lambda \sqcup \Lambda^{-\epsilon}) \setminus \operatorname{st} u$ is disconnected. Without loss of generality, we have $u \in \Lambda^+$. Since the partition of $\Gamma \setminus x$ satisfies condition (iii), the point x must be isolated in $(\Lambda \sqcup \Lambda^-) \setminus \operatorname{st} u$. Hence $\operatorname{lk} x \subseteq \operatorname{lk} u \cup \Delta^+$, again violating our assumption. □

Lemma 5.8 *Let Γ be an irreducible graph, and let $x \in \Gamma$ be a vertex such that there does not exist $y \in \Gamma^{(0)} \setminus \{x\}$ with $\operatorname{lk} x \subseteq \operatorname{lk} y$. Suppose that $\Gamma \setminus x$ is reducible. Then the partition of Γ given by $\Lambda^+ = \{x\}$, $\Lambda = \operatorname{lk} x$, $\Lambda^- = \Gamma \setminus \operatorname{st} x$ is good.*

Proof Write $\Gamma \setminus x$ as a join of nonempty subgraphs Γ_1 and Γ_2. Since Γ is irreducible, there exist points $a_1 \in \Gamma_1 \setminus \operatorname{lk} x$ and $a_2 \in \Gamma_2 \setminus \operatorname{lk} x$. Condition (i) is clear.

In order to verify condition (ii), we need to exclude the existence of $w \in \Lambda^{\epsilon}$ and $v \in \Lambda \sqcup \Lambda^{-\epsilon}$ with $\operatorname{lk} v \subseteq \operatorname{lk} w \cup \Lambda^{\epsilon}$. If $\epsilon = -$ and $v \in \Lambda$, then x lies in $\operatorname{lk} v$, but not in $\operatorname{lk} w \cup \Lambda^-$. If $\epsilon = -$ and $v = x$, then $\operatorname{lk} x$ is disjoint from Λ^-, and it cannot be contained in the link of any point of $\Gamma \setminus x$ by our hypotheses. If $\epsilon = +$, then $\operatorname{lk} w \cup \Lambda^{\epsilon} = \operatorname{st} x$, which cannot contain the link of any point of $\Gamma \setminus x$, as it does not contain a_1 and a_2.

Finally, let us show that, for every $w \in \Lambda^{\epsilon}$, the set $(\Lambda \sqcup \Lambda^{-\epsilon}) \setminus \operatorname{st} w$ is connected. If $\epsilon = +$, this amounts to showing that $\Gamma \setminus \operatorname{st} x$ is connected. This is immediate, since every point of $\Gamma \setminus x$ is joined by an edge to either a_1 or a_2, and these two points are themselves joined by an edge. If instead $\epsilon = -$, we need to show that $\operatorname{st} x \setminus \operatorname{st} w$ is connected for every $w \in \Gamma \setminus \operatorname{st} x$. This is also clear since this set is a cone over x. □

Consider the equivalence relation on $\Gamma^{(0)}$ where $v \sim w$ if and only if $\operatorname{lk} v = \operatorname{lk} w$. We define a graph $\overline{\Gamma}$ with a vertex for every $\sim$–equivalence class $[v] \subseteq \Gamma$ and an edge joining $[v]$ and $[w]$ exactly when v and w are joined by an edge (this is independent of the chosen representatives).

It is clear that $\overline{\Gamma}$ is again a simplicial graph, with at most as many vertices as Γ. We denote by $r \colon \Gamma \to \overline{\Gamma}$ the natural morphism of graphs.

Lemma 5.9 (1) *Γ is irreducible if and only if $\overline{\Gamma}$ is irreducible.*

(2) *If Γ has at least one edge, then Γ is connected if and only if $\overline{\Gamma}$ is connected.*

(3) *If $\overline{\Gamma} = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$ is a good partition, then so is $\Gamma = r^{-1}(\Lambda^+) \sqcup r^{-1}(\Lambda) \sqcup r^{-1}(\Lambda^-)$.*

Proof Parts (1) and (2) are straightforward, so we only prove part (3).

Consider a good partition $\overline{\Gamma} = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$. It is clear that the partition of Γ satisfies condition (i), while condition (ii) follows from the observation that $\operatorname{lk} r(x) = r(\operatorname{lk} x)$ for every $x \in \Gamma$.

Finally, we verify condition (iii). Given $w \in r^{-1}(\Lambda^{\epsilon})$, observe that r maps the subgraph

$$(r^{-1}(\Lambda) \sqcup r^{-1}(\Lambda^{-\epsilon})) \setminus \operatorname{st} w$$

onto the connected graph $(\Lambda \sqcup \Lambda^{-\epsilon}) \setminus \operatorname{st} r(w)$. As in part (2), this shows that $(r^{-1}(\Lambda) \sqcup r^{-1}(\Lambda^{-\epsilon})) \setminus \operatorname{st} w$ is connected, possibly except the case when $(\Lambda \sqcup \Lambda^{-\epsilon}) \setminus \operatorname{st} r(w)$ is a singleton. The latter is ruled out by the fact that the partition of $\overline{\Gamma}$ satisfies condition (ii). □

Proof of Proposition 5.5 We proceed by induction on the number of vertices of Γ. Since no graph with at most 3 vertices satisfies the hypotheses of the proposition, the base step is trivially satisfied. For the inductive step, we consider a connected irreducible graph Γ with at least 4 vertices, and assume that the proposition is satisfied by all graphs with fewer vertices than Γ.

If $\operatorname{diam} \Gamma^{(0)} \geq 3$, we can simply appeal to Lemma 5.6. If the graph $\overline{\Gamma}$ defined above has fewer vertices than Γ, then we can use the inductive hypothesis and Lemma 5.9. Thus, we can assume that $\Gamma = \overline{\Gamma}$ and $\operatorname{diam} \Gamma^{(0)} = 2$.

Pick a vertex $x \in \Gamma$ whose link is maximal under inclusion. Since $\Gamma = \overline{\Gamma}$, there does not exist $y \in \Gamma^{(0)} \setminus \{x\}$ with $\operatorname{lk} x = \operatorname{lk} y$. If $\Gamma \setminus x$ is reducible, Lemma 5.8 then shows that Γ admits a good partition. If $\Gamma \setminus x$ were disconnected, then the fact that $\operatorname{diam} \Gamma^{(0)} = 2$ would imply that $\operatorname{lk} x = \Gamma \setminus x$, contradicting the assumption that Γ is irreducible.

In conclusion, $\Gamma \setminus x$ is connected, irreducible, not a singleton, and it has fewer vertices than Γ. We conclude by applying the inductive hypothesis and Lemma 5.7 (case (1) of the latter is ruled out by our choice of x). □

The previous results prove Proposition 5.1. The following is Corollary E from the introduction:

Corollary 5.10 *Consider $\varphi \in U_0(\mathcal{A})$.*

1. *If $\mathcal{A}$ splits as a direct product $\mathcal{A}_1 \times \mathcal{A}_2$, then $\varphi(\mathcal{A}_i) = \mathcal{A}_i$ and $\operatorname{Fix} \varphi = \operatorname{Fix} \varphi|_{\mathcal{A}_1} \times \operatorname{Fix} \varphi|_{\mathcal{A}_2}$.*
2. *If $\mathcal{A}$ is directly irreducible, then the subgroup $\operatorname{Fix} \varphi \leq \mathcal{A}$ splits as a (possibly trivial) finite graph of groups with vertex and edge groups of the form $\operatorname{Fix} \varphi|_P$, for proper parabolic subgroups $P \leq \mathcal{A}$ with $\varphi(P) = P$ and $\varphi|_P \in U_0(P)$.*

Proof For simplicity, set $H := \operatorname{Fix} \varphi$. We distinguish three cases.

Case 1 ($\mathcal{A}$ is not directly irreducible) Let us write $\mathcal{A} = A \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_m$, where A is a free abelian group and $\mathcal{A}_i$ are directly irreducible (noncyclic) right-angled Artin groups. This corresponds to a splitting of Γ as a join of a complete subgraph and irreducible subgraphs $\Gamma_1, \dots, \Gamma_m$.

Since $\varphi \in U_0(\mathcal{A})$, we have $\varphi(\mathcal{A}_k) = \mathcal{A}_k$ and $\varphi|_{\mathcal{A}_k} \in U_0(\mathcal{A}_k)$ for every $1 \leq k \leq m$, and $\varphi|_A$ is a product of inversions. Indeed, this is clear for inversions, folds and partial conjugations.

Thus $H = A' \times H_1 \times \cdots \times H_m$, where $H_i = \operatorname{Fix}(\varphi|_{\mathcal{A}_i})$ and A' is a standard direct factor of A. This proves part (1) of the corollary.

Case 2 ($\mathcal{A}$ is not freely irreducible) Write $\mathcal{A} = F * \mathcal{A}_1 * \dots * \mathcal{A}_m$, where F is a free group and $\mathcal{A}_i$ are freely irreducible (noncyclic) right-angled Artin groups of lower complexity. Since H is finitely generated by Theorem B, Kurosh's theorem guarantees that H decomposes as a free product $H = L * H_1 * \dots * H_n$, where L is a finitely generated free group and each H_i is a finitely generated subgroup of some $g_i \mathcal{A}_{k_i} g_i^{-1}$ with $g_i \in \mathcal{A}$ and $1 \le k_i \le m$.

By Grushko's theorem, the subgroup $\varphi(\mathcal{A}_k)$ is conjugate to some $\mathcal{A}_{k'}$ for every $1 \le k \le m$. Since φ fixes the nontrivial subgroup $H_i \le g_i \mathcal{A}_{k_i} g_i^{-1}$ pointwise, we must have $\varphi(g_i \mathcal{A}_{k_i} g_i^{-1}) = g_i \mathcal{A}_{k_i} g_i^{-1}$ for $1 \le i \le n$.

Consider the automorphism $\psi_i \in U_0(\mathcal{A})$ defined by $\psi_i(x) = g_i^{-1} \varphi(g_i x g_i^{-1}) g_i$. Note that $\psi_i(\mathcal{A}_{k_i}) = \mathcal{A}_{k_i}$ and $\operatorname{Fix} \psi_i|_{\mathcal{A}_{k_i}} = g_i^{-1} H_i g_i$. By Lemma 3.29, we have $\psi_i|_{\mathcal{A}_{k_i}} \in U_0(\mathcal{A}_{k_i})$. This proves part (2) of the corollary in the freely reducible case.

Case 3 ($\mathcal{A}$ is freely and directly irreducible) We can assume that $\mathcal{A} \not\simeq \mathbb{Z}$. By Proposition 5.5, Γ admits a good partition $\Gamma = \Lambda^+ \sqcup \Lambda \sqcup \Lambda^-$. By Corollary 5.4, there exists $f \in \operatorname{Aut} T_\Lambda$ satisfying $f \circ g = \varphi(g) \circ f$ for all $g \in \mathcal{A}$.

If H is elliptic in T_Λ, we have $H \le V$, where V is the $\mathcal{A}$–stabiliser of some vertex of T_Λ. The existence of the automorphism $f \in \operatorname{Aut} T_\Lambda$ guarantees that all subgroups $\varphi^n(V)$ with $n \in \mathbb{Z}$ are $\mathcal{A}$–stabilisers of vertices of T_Λ; in particular, they are all conjugate to either $\mathcal{A}_+$ or $\mathcal{A}_-$. We conclude that H is contained in the $\langle\varphi\rangle$–invariant parabolic subgroup $P := \bigcap_{n \in \mathbb{Z}} \varphi^n(V)$. Thus, we have $H = \operatorname{Fix} \varphi|_P$ and, by Lemma 3.29, $\varphi|_P \in U_0(P)$. This proves the corollary in this case, with H splitting as a trivial graph of groups.

Suppose instead that H is not elliptic in T_Λ and denote by $T_H \subseteq T_\Lambda$ the H–minimal subtree. Since H is finitely generated, the action $H \curvearrowright T_H$ is cocompact and gives a splitting of H as a (nontrivial) finite graph of groups. We are left to understand vertex-stabilisers of the action $H \curvearrowright T_H$.

As f normalises H in $\operatorname{Aut} T_\Lambda$, we have $f(T_H) = T_H$. It is convenient to distinguish two subcases.

Case 3a (f is elliptic in T_Λ) Since f commutes with every element of H, the tree T_H is fixed pointwise by f. For every $v \in T_H$, its $\mathcal{A}$–stabiliser $\mathcal{A}_v$ satisfies $\varphi(\mathcal{A}_v) = \mathcal{A}_v$ and is conjugate to either $\mathcal{A}_+$ or $\mathcal{A}_-$. By Lemma 3.29, we have $\varphi|_{\mathcal{A}_v} \in U_0(\mathcal{A}_v)$, proving the corollary in this case.

Case 3b (f is loxodromic in T_Λ) Let $\alpha \subseteq T_\Lambda$ be the axis of f. Since f commutes with every element of H, the geodesic α must be H–invariant and every nonloxodromic element of H fixes α pointwise. Note that T_H cannot be a singleton, or f would be elliptic. Thus, $T_H = \alpha$ and H contains a shortest loxodromic element $h \in H$. Moreover, $H = H_0 \rtimes \langle h\rangle$, where H_0 is the kernel of the action $H \curvearrowright \alpha$.

Let $Q \le \mathcal{A}$ be the intersection of the $\mathcal{A}$–stabilisers of the vertices of α. Being an intersection of parabolic subgroups, Q is itself a (possibly trivial) parabolic subgroup of $\mathcal{A}$. Since $f(\alpha) = \alpha$, we have $\varphi(Q) = Q$ and $H_0 = \operatorname{Fix} \varphi|_Q$. Lemma 3.29 guarantees that $\varphi|_Q \in U_0(Q)$. Thus, the HNN splitting $H = H_0 \rtimes \langle h\rangle$ is as required by the corollary. □

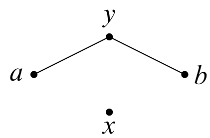


Figure 5

Remark 5.11 In Case 3b of the proof of Corollary 5.10, we can actually say more on the structure of $H = \operatorname{Fix}\varphi$. Specifically, $H = H_0 \times \langle h \rangle$ and h can be taken to be label-irreducible.

Indeed, since $h\alpha = \alpha$, the element h lies in the normaliser of Q in $\mathcal{A}$, which is a subgroup of the form $Q \times Q'$ (since Q is parabolic in $\mathcal{A}$). If $h = h_1 \cdots h_k$ is the decomposition of h into label-irreducible components, every h_i lies in either Q or Q'. Since φ is coarse-median preserving and fixes h, it must permute the h_i; Corollary 4.31 then shows that $\varphi(h_i) = h_i$ for every i. Thus, all the label-irreducible components of h that lie in Q actually lie in H_0. Up to replacing h, we can assume that all h_i lie in Q'; in particular, h lies in Q', hence it commutes with H_0. Since $H = \operatorname{Fix}\varphi$ is generated by H_0 and h, we must then have $k = 1$, ie h is label-irreducible.

In relation to Theorem C, it is natural to wonder if the proof of Corollary 5.10 can be used to give an alternative, inductive argument showing that $\operatorname{Fix}\varphi$ is convex-cocompact in $\mathcal{A}$ for every $\varphi \in U_0(\mathcal{A})$. In light of Remark 5.11, the only problematic situation is the one in Case 3a.

Unfortunately, cubical convex-cocompactness does not seem to be well-behaved with respect to graph-of-groups constructions, as the next example shows.

Example 5.12 Let Γ be the graph in Figure 5. Consider the subgroup $H = \langle ayx^{-1}, xby \rangle \leq \mathcal{A}_\Gamma$. We have an amalgamated product splitting $\mathcal{A}_\Gamma = \langle a, x, y \rangle *_{\langle x,y \rangle} \langle b, x, y \rangle$, which induces a splitting $H = \langle ayx^{-1} \rangle * \langle xby \rangle \simeq F_2$. The subgroups $\langle ayx^{-1} \rangle$ and $\langle xby \rangle$ are convex-cocompact, as they are each generated by a single label-irreducible element.

However, H is not convex-cocompact in $\mathcal{A}$: the element aby^2 lies in H, but no power of its label-irreducible components ab and y^2 does (which, for instance, violates Lemma 3.16).

6 Projectively invariant metrics on finite-rank median algebras

In this section, we initiate the lengthy proof of Theorem F, which will be completed in Section 7. Our main goal here is to formulate a criterion, for a group U and a subgroup $G \leq U$, guaranteeing that a U–action on a finite-rank median algebra admits a G–invariant compatible pseudometric for which U acts by homotheties (Corollary 6.23). An important tool will be the Lefschetz fixed point theorem for compact ANRs.

Throughout the section, M denotes a fixed median algebra of finite rank r.

6.1 Multibridges

The *bridge* of two gate-convex sets was first studied in [4; 30] for CAT(0) cube complexes and in [49, Section 2.2] for general median algebras. We will need an extension of this concept to arbitrary finite collections of gate-convex subsets: *multibridges*.

We briefly motivate why. As a recurring setup in the rest of the paper (especially in Sections 6.2.3 and 7.4), we will often find ourselves studying a group $G \leq \operatorname{Aut} M$ with a finite generating subset $S \subseteq G$ and a G–invariant compatible pseudometric η on M. It will be important to understand which points of M are moved as little as possible by all elements of S, ie which points realise the quantity $\overline{\tau}_S^{\eta}$ from Section 2.1. It turns out that the set of such points does not depend much on the specific pseudometric η, and can instead be characterised purely in terms of the median-algebra structure on M, using the notion of multibridge (Propositions 6.9 and 6.11).

Let $C_1, \dots, C_k \subseteq M$ be gate-convex subsets, with gate-projections $\pi_i : M \to C_i$. Let $\mathcal{H} \subseteq \mathscr{H}(M)$ be the set of halfspaces that contain at least one C_i and intersect each C_i. Then we have a partition

$$\mathscr{H}(M) = (\mathcal{H} \sqcup \mathcal{H}^*) \sqcup \Big(\bigcap_{1 \leq i \leq k} \mathscr{H}_{C_i}(M) \Big) \sqcup \Big(\bigcup_{1 \leq i,j \leq k} \mathscr{H}(C_i | C_j) \Big).$$

If $i \neq j$, the sets $\mathscr{H}_{C_i}(M) \cap \mathscr{H}_{C_j}(M)$ and $\mathscr{H}(C_i|C_j)$ are transverse. Thus, every halfspace in the second set of the above partition of $\mathscr{H}(M)$ is transverse to every halfspace in the third set.

Lemma 6.1 *The intersection of all halfspaces in $\mathcal{H}$ is a nonempty convex subset of M.*

Proof We will prove this by appealing to Lemma 2.4(1). It is clear that the elements of $\mathcal{H}$ intersect pairwise. Let us show that, for every chain $\mathscr{C} \subseteq \mathcal{H}$, the set $\mathfrak{k} := \bigcap \mathscr{C}$ is again an element of $\mathcal{H}$.

Note that there exist $1 \leq i_0 \leq k$ and a cofinal subset $\mathscr{C}' \subseteq \mathscr{C}$ consisting of halfspaces containing C_{i_0}. Thus, $C_{i_0} \subseteq \mathfrak{k}$ and $\mathfrak{k}$ is nonempty. Since $\mathfrak{k}$ is the intersection of a chain of halfspaces, both $\mathfrak{k}$ and $\mathfrak{k}^*$ are convex. It follows that $\mathfrak{k}$ is a halfspace of M.

For every $\mathfrak{h} \in \mathscr{C} \subseteq \mathcal{H}$, the fact that $\mathfrak{h}$ intersects each C_i implies that $\pi_i(\mathfrak{h}) = \mathfrak{h} \cap C_i$; see for instance [50, Lemma 2.2(1)]. Recalling that $\mathfrak{k} = \bigcap \mathscr{C}$, we deduce that $\pi_i(\mathfrak{k}) \subseteq \mathfrak{k} \cap C_i$ for $1 \leq i \leq k$, hence $\mathfrak{k}$ intersects all C_i. Since we have already seen that $C_{i_0} \subseteq \mathfrak{k}$, we conclude that $\mathfrak{k} \in \mathcal{H}$, as required. □

Definition 6.2 The intersection $\mathcal{B} = \mathcal{B}(C_1, \dots, C_k) \subseteq M$ of all halfspaces in $\mathcal{H}$ is the *multibridge* of the gate-convex sets $C_1, \dots, C_k$.

For every $\mathfrak{k} \in \mathscr{H}(M) \setminus \mathcal{H}^*$, the set $\mathcal{H} \sqcup \{\mathfrak{k}\}$ is again pairwise-intersecting. Hence, Lemma 2.4(1) yields

$$\mathscr{H}_{\mathcal{B}}(M) = \mathscr{H}(M) \setminus (\mathcal{H} \sqcup \mathcal{H}^*) = \Big(\bigcap \mathscr{H}_{C_i}(M) \Big) \sqcup \Big(\bigcup \mathscr{H}(C_i|C_j) \Big).$$

We have already observed that the two sets in this partition are transverse. By Remark 2.2(2) and Lemma 2.6, we obtain a natural product splitting

$$\mathcal{B}=\mathcal{B}_{/\!/}\times\mathcal{B}_{\perp},\quad \text{where } \mathscr{H}_{\mathcal{B}_{/\!/}}(M)=\bigcap\mathscr{H}_{C_i}(M)\quad\text{and}\quad \mathscr{H}_{\mathcal{B}_{\perp}}(M)=\bigcup\mathscr{H}(C_i|C_j).$$

We can view $\mathcal{B}_{/\!/}$ and $\mathcal{B}_{\perp}$ as subsets of M by identifying them with any fibre of the splitting of $\mathcal{B}$.

Lemma 6.3 *The sets $\mathcal{B}$, $\mathcal{B}_{/\!/}$ and $\mathcal{B}_{\perp}$ are gate-convex in M.*

Proof Since each C_i is gate-convex, Lemma 2.4(2) shows that, for every chain $\mathscr{C}\subseteq\bigcap\mathscr{H}_{C_i}(M)$, either $\bigcap\mathscr{C}$ is empty in M, or $\bigcap\mathscr{C}\in\bigcap\mathscr{H}_{C_i}(M)$. Hence $\mathcal{B}_{/\!/}$ is gate-convex in M.

If $\mathscr{C}\subseteq\bigcup\mathscr{H}(C_i|C_j)$ is a chain, a cofinal subset of $\mathscr{C}$ is contained in a single $\mathscr{H}(C_i|C_j)$. Hence $\bigcap\mathscr{C}\in\mathscr{H}(C_i|C_j)$. Invoking again Lemma 2.4(2), this shows that $\mathcal{B}_{\perp}$ is gate-convex.

Every chain in $\mathscr{H}_{\mathcal{B}}(M)$ has a cofinal subset contained in either $\bigcap\mathscr{H}_{C_i}(M)$ or $\bigcup\mathscr{H}(C_i|C_j)$. One last application of Lemma 2.4(2) shows that $\mathcal{B}$ is gate-convex. □

Corollary 6.4 *If $C_1,\dots,C_k\subseteq M$ are gate-convex subsets, their multibridge $\mathcal{B}=\mathcal{B}(C_1,\dots,C_k)$ is a gate-convex subset of M enjoying the following properties*:

(1) *$\mathcal{B}$ splits as a product $\mathcal{B}_{/\!/}\times\mathcal{B}_{\perp}$ with $\mathscr{H}_{\mathcal{B}_{/\!/}}(M)=\bigcap\mathscr{H}_{C_i}(M)$ and $\mathscr{H}_{\mathcal{B}_{\perp}}(M)=\bigcup\mathscr{H}(C_i|C_j)$.*

(2) *Each fibre $\{*\}\times\mathcal{B}_{\perp}$ intersects all of the C_i.*

Proof The only statement that has not already been proved is part (2). If it were false, there would exist an index i and $\mathfrak{h}\in\mathscr{H}(M)$ such that $C_i\subseteq\mathfrak{h}$ and $\{*\}\times\mathcal{B}_{\perp}\subseteq\mathfrak{h}^*$. Since $C_i\subseteq\mathfrak{h}$, we have $\mathfrak{h}\notin\mathscr{H}_{\mathcal{B}_{/\!/}}(M)$, so $\mathcal{B}\subseteq\mathfrak{h}^*$. Hence $\mathfrak{h}^*\in\mathcal{H}$, contradicting the fact that $C_i\subseteq\mathfrak{h}$. □

Recall the notation $\mathcal{PD}(M)$ and $\mathcal{D}(M)$ for compatible (pseudo)metrics, as in Section 2.3.

Remark 6.5 If $\eta\in\mathcal{PD}(M)$ and $x,y\in\mathcal{B}$ lie in the same fibre $\mathcal{B}_{/\!/}\times\{*\}$, then $\eta(x,C_i)=\eta(y,C_i)$ for all $1\le i\le k$. Indeed, since $\mathscr{H}(x|y)\subseteq\mathscr{H}_{\mathcal{B}_{/\!/}}(M)=\bigcap\mathscr{H}_{C_i}(M)$, we have $\mathscr{W}(x|C_i)=\mathscr{W}(y|C_i)$ and it follows (eg by Remark 2.9) that $\eta(x,\pi_i(x))=\eta(y,\pi_i(y))$ for every $\eta\in\mathcal{PD}(M)$.

Remark 6.6 If $\eta\in\mathcal{PD}(M)$, then $\eta(x,\mathcal{B})\le r\cdot\max_i\eta(x,C_i)$ for every $x\in M$.

In order to see this, let $\mathfrak{h}_1,\dots,\mathfrak{h}_k$ be the minimal elements of $\mathscr{H}(x|\mathcal{B})$. Since the $\mathfrak{h}_i$ are pairwise transverse and $\operatorname{rk}M=r$, we have $k\le r$. Each $\mathfrak{h}_i$ must lie in $\mathcal{H}$, hence there exists an index j_i such that $C_{j_i}\subseteq\mathfrak{h}_i$. It follows that:

$$\mathscr{H}(x|\mathcal{B})\subseteq\bigcup\mathscr{H}(x|\mathfrak{h}_i)\subseteq\bigcup\mathscr{H}(x|C_{j_i}).$$

Hence $\eta(x,\mathcal{B})\le k\cdot\max_i\eta(x,C_i)\le r\cdot\max_i\eta(x,C_i)$.

Remark 6.7 If $\delta \in \mathcal{D}(M)$ and (M,δ) is complete, then $\mathcal{B}_\perp$ is compact in (M,δ).

In order to prove this, let $x_{i,j} \in C_i$ and $x_{j,i} \in C_j$ be a pair of gates for all distinct $1 \le i, j \le k$. Let K be the convex hull of the finite set $F = \{x_{i,j} \mid 1 \le i, j \le k\}$. Recall that $K = \mathcal{J}^r(F)$ by Remark 2.5, so it follows from [50, Corollary 2.20] that K is compact.

We have $K \cap \mathcal{B} \neq \varnothing$. Otherwise, the set $\mathscr{H}(K|\mathcal{B})$ would be nonempty and contained in $\mathcal{H}$. However, each element of $\mathcal{H}$ contains some C_i, so it cannot be disjoint from K.

Finally, observing that $\mathscr{H}_K(M)$ contains the set

$$\bigcup \mathscr{H}(x_{i,j}|x_{j,i}) = \bigcup \mathscr{H}(C_i|C_j) = \mathscr{H}_{\mathcal{B}_\perp}(M),$$

we deduce that $K \cap \mathcal{B}$ must contain a fibre $\{*\} \times \mathcal{B}_\perp$. Since $\mathcal{B}_\perp$ is gate-convex, it must be a closed subset of K, hence it is compact too.

Now, let $S \subseteq \operatorname{Aut} M$ be a finite set of automorphisms acting nontransversely and stably without inversions. By Theorem 2.16(1), the reduced cores $\overline{\mathcal{C}}(s)$ of $s \in S$ are all gate-convex. Let $\mathcal{B}(S)$ be their multibridge.

Definition 6.8 We refer to $\mathcal{B}(S)$ as the *multibridge* of the finite set $S \subseteq \operatorname{Aut} M$.

Recalling the notation introduced in Section 2.1, we have:

Proposition 6.9 *Let $S \subseteq \operatorname{Aut} M$ be a finite set of automorphisms acting nontransversely and stably without inversions. The multibridge $\mathcal{B}(S)$ is gate-convex and, for all $\eta \in \mathcal{PD}(M)^{\langle S\rangle}$:*

(1) *We have $\tau_S^\eta(\pi_{\mathcal{B}}(x)) \le \tau_S^\eta(x)$ for all $x \in M$, where $\pi_{\mathcal{B}}\colon M \to \mathcal{B}(S)$ is the gate-projection.*
(2) *$\tau_S^\eta(\,\cdot\,)$ is constant on each fibre $\mathcal{B}_{/\!/}(S) \times \{*\}$.*
(3) *If $\delta \in \mathcal{D}(M)^{\langle S\rangle}$ and (M,δ) is complete, then there exists $z \in \mathcal{B}(S)$ with $\tau_S^\delta(x) = \overline{\tau}_S^\delta$.*

Proof Since the multibridge $\mathcal{B}(S)$ intersects each $\overline{\mathcal{C}}(s)$, we have $\mathscr{H}(\pi_{\mathcal{B}}(x)|\overline{\mathcal{C}}(s)) \subseteq \mathscr{H}(x|\overline{\mathcal{C}}(s))$ for all $x \in M$. Hence $\eta(\pi_{\mathcal{B}}(x), \overline{\mathcal{C}}(s)) \le \eta(x, \overline{\mathcal{C}}(s))$. Theorem 2.16(2) now implies that $\tau_S^\eta(\pi_{\mathcal{B}}(x)) \le \tau_S^\eta(x)$, proving part (1). By Remark 6.5, if $x, y \in \mathcal{B}(S)$ lie in the same fibre $\mathcal{B}_{/\!/}(S) \times \{*\}$, then $\eta(x, \overline{\mathcal{C}}(s)) = \eta(y, \overline{\mathcal{C}}(s))$. This proves part (2). Finally, part (3) follows from the previous two parts and Remark 6.7. □

Example 6.10 Let $G = \langle a, b\rangle$ be the free group over two generators. Let T be the standard Cayley graph of G, with all edges of length 1. Let (X,δ) be the (incomplete) median space obtained by removing from T all midpoints of edges. Then, taking $S = \{a, bab^{-1}\} \subseteq G \subseteq \operatorname{Isom} X$, there is no point $x \in X$ with $\tau_S^\delta(x) = \overline{\tau}_S^\delta = 2$.

Our interest in multibridges is due to the following result, which helps us understand the behaviour on M of the functions $\tau_S^\eta(\,\cdot\,)$ for $\eta \in \mathcal{PD}(M)^{\langle S\rangle}$.

Proposition 6.11 *Let $S \subseteq \operatorname{Aut} M$ be a finite set of automorphisms acting nontransversely and stably without inversions. Recall that $r = \operatorname{rk} M$. Then, the following hold for every $\eta \in \mathcal{PD}(M)^{\langle S \rangle}$:*

(1) *If $s_1, s_2 \in S$, then $\eta(\bar{\mathcal{C}}(s_1), \bar{\mathcal{C}}(s_2)) \leq \bar{\tau}_S^\eta$.*

(2) *If $s \in S$ and $x \in \mathcal{B}(S)$, then $\eta(x, \bar{\mathcal{C}}(s)) \leq r\bar{\tau}_S^\eta$.*

(3) *If $x \in \mathcal{B}(S)$, then $\tau_S^\eta(x) \leq (2r+1)\bar{\tau}_S^\eta$.*

(4) *The η–diameter of each fibre $\{*\} \times \mathcal{B}_\perp(S)$ is at most $r^2\bar{\tau}_S^\eta$.*

(5) *If $x \in M$, then $\eta(x, \mathcal{B}(S)) \leq \frac{1}{2} r \tau_S^\eta(x)$.*

(6) *For any $x \in M$ and any fibre $P = \mathcal{B}_{/\!/}(S) \times \{*\}$, we have $\eta(x, P) \leq 2r^2 \tau_S^\eta(x)$.*

Proof We begin with part (1). For every $x \in M$, we have

$$\mathscr{W}(\bar{\mathcal{C}}(s_1)|\bar{\mathcal{C}}(s_2)) = \mathscr{W}(x, \bar{\mathcal{C}}(s_1)|\bar{\mathcal{C}}(s_2)) \sqcup \mathscr{W}(\bar{\mathcal{C}}(s_1)|\bar{\mathcal{C}}(s_2), x) \subseteq \mathscr{W}(x|\bar{\mathcal{C}}(s_1)) \sqcup \mathscr{W}(x|\bar{\mathcal{C}}(s_2)).$$

Along with Theorem 2.16(2), this implies that

$$\tfrac{1}{2}\eta(\bar{\mathcal{C}}(s_1), \bar{\mathcal{C}}(s_2)) \leq \max\{\eta(x, \bar{\mathcal{C}}(s_1)), \eta(x, \bar{\mathcal{C}}(s_2))\} \leq \tfrac{1}{2}\max\{\eta(x, s_1x), \eta(x, s_2x)\} \leq \tfrac{1}{2}\tau_S^\eta(x).$$

Part (1) follows by taking an infimum over $x \in M$.

Let us prove part (2). If $x \in \mathcal{B}(S)$ and $s \in S$, Corollary 6.4(2) implies that $\mathscr{H}(x|\bar{\mathcal{C}}(s))$ is contained in the union of the sets $\mathscr{H}(\bar{\mathcal{C}}(t)|\bar{\mathcal{C}}(s))$ with $t \in S \setminus \{s\}$. The maximal halfspaces in $\mathscr{H}(x|\bar{\mathcal{C}}(s))$ are pairwise transverse, so there are at most r of them. Hence, there exist $t_1, \dots, t_r \in S$ such that $\Omega := \bigcup_i \mathscr{H}(\bar{\mathcal{C}}(t_i)|\bar{\mathcal{C}}(s))$ contains every maximal element of $\mathscr{H}(x|\bar{\mathcal{C}}(s))$. In particular, $\mathscr{H}(x|\bar{\mathcal{C}}(s)) \subseteq \Omega$ and part (1) yields $\eta(x, \bar{\mathcal{C}}(s)) \leq r\bar{\tau}_S^\eta$.

Part (3) of the proposition now follows from Theorem 2.16(2):

$$\tau_S^\eta(x) = \max_{s \in S}[\ell(s, \eta) + 2\eta(x, \bar{\mathcal{C}}(s))] \leq \max_{s \in S}[\bar{\tau}_S^\eta + 2r\bar{\tau}_S^\eta] = (2r+1)\bar{\tau}_S^\eta.$$

Regarding part (4), consider two points x, y lying in the same fibre $\{*\} \times \mathcal{B}_\perp(S)$. Let $\mathfrak{h}_1, \dots, \mathfrak{h}_k$ be the minimal elements of $\mathscr{H}(x|y)$. Since $\operatorname{rk} M = r$, we have $k \leq r$. By definition of $\mathcal{B}_\perp(S)$, there exist elements $s_i \in S$ with $\bar{\mathcal{C}}(s_i) \subseteq \mathfrak{h}_i$. Thus,

$$\mathscr{H}(x|y) \subseteq \bigcup \mathscr{H}(x|\mathfrak{h}_i) \subseteq \bigcup \mathscr{H}(x|\bar{\mathcal{C}}(s_i)).$$

Using part (2) of the proposition, it follows that $\eta(x, y) \leq k \cdot \max_s \eta(x, \bar{\mathcal{C}}(s)) \leq kr\bar{\tau}_S^\eta \leq r^2\bar{\tau}_S^\eta$.

Finally, part (5) is a consequence of Remark 6.6 and the fact, due to Theorem 2.16(2), that $\tau_S^\eta(x) \geq 2\eta(x, \bar{\mathcal{C}}(s))$ for every $s \in S$. Part (6) is obtained by combining parts (4) and (5):

$$\eta(x, P) \leq \eta(x, \mathcal{B}(S)) + r^2\bar{\tau}_S^\eta \leq \tfrac{r}{2}\tau_S^\eta(x) + r^2\bar{\tau}_S^\eta \leq 2r^2\tau_S^\eta(x). \qquad \square$$

6.2 Promoting median automorphisms to homotheties

Recall that M is a median algebra of finite rank r. In this subsection, we consider subgroups $G \lhd U \leq \operatorname{Aut} M$, with the goal of constructing G–invariant compatible pseudometrics $\eta \in \mathcal{PD}^G(M)$ with respect to which U acts by homotheties and G is nonelliptic. In general, this will only be possible after passing to a subalgebra of M. The final result in this direction is Corollary 6.23.

Our main technical tools are the notion of multibridge (exploited in Lemma 6.22) and the Lefschetz fixed point theorem applied to projectivisations of certain cones $\mathcal{C}$ in the topological vector space $\mathcal{PD}^G(M)$ (Proposition 6.17). Some extra work is required in order to ensure that our cones $\mathcal{C}$ have compact projectivisation and that they only contain pseudometrics η for which G acts nonelliptically (ie $\overline{\tau}_S^{\eta} > 0$ for some/any generating set $S \subseteq G$).

6.2.1 Preliminaries on normed spaces and ARs

Definition 6.12 Let V be a real vector space.

1. (1) A *cone* is a convex subset $\mathcal{C} \subseteq V$ that is closed under multiplication by scalars in $[0, +\infty)$.
2. (2) A *positive cone* is a cone $\mathcal{C} \subseteq V$ for which $\mathcal{C} \setminus \{0\}$ is convex. Equivalently, $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$.
3. (3) The *projectivisation* $\mathbb{P}(\mathcal{C})$ of a cone $\mathcal{C}$ is the quotient of $\mathcal{C} \setminus \{0\}$ obtained by identifying points that differ by multiplication by a scalar.

Given a countable probability space (Ω, σ) and a function $f : \Omega \to \mathbb{R}$, recall that

$$\|f\|_1 = \sum_{\omega \in \Omega} |f(\omega)| \sigma(\omega) \quad \text{and} \quad \|f\|_\infty = \sup_{\omega \in \Omega} |f(\omega)|.$$

We denote by $\ell^1(\Omega, \sigma)$ and $\ell^\infty(\Omega)$ the spaces of functions where $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are finite, respectively.

The next result collects a few simple observations that will be useful later in this subsection. In particular, part (3) will be our compactness criterion for projectivised cones: we only need to ensure that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are bi-Lipschitz equivalent on the cone. This is one of the reasons we are forced to work with both norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

Lemma 6.13 *Let (Ω, σ) be a countable set with a fully supported probability measure.*

1. (1) *We have $\ell^\infty(\Omega) \subseteq \ell^1(\Omega, \sigma)$ and $\|\cdot\|_1 \leq \|\cdot\|_\infty$.*
2. (2) *The topology of $(\ell^1(\Omega, \sigma), \|\cdot\|_1)$ is finer than the topology of pointwise convergence on Ω. The converse holds on those subsets of $\ell^1(\Omega, \sigma)$ where $\|\cdot\|_\infty$ is bounded.*
3. (3) *Let $\mathcal{C} \subseteq \ell^1(\Omega, \sigma)$ be a positive cone that is closed in the topology of $\|\cdot\|_1$. Suppose that there exists $c > 0$ such that $\|f\|_\infty \leq c \cdot \|f\|_1$ for all $f \in \mathcal{C}$. Then $\mathbb{P}(\mathcal{C})$ is compact with respect to the quotient topology induced by $\|\cdot\|_1$.*

Proof Part (1) is clear. The two halves of part (2) follow respectively from the inequalities

$$|f(\omega)|\sigma(\{\omega\}) \le \|f\|_1 \quad \text{and} \quad \|f\|_1 \le \sum_{x\in F} |f(x)|\sigma(\{x\}) + \|f\|_\infty \cdot \sigma(\Omega \setminus F),$$

which hold for all $f \in \ell^1(\Omega,\sigma)$, all $\omega \in \Omega$ and every finite subset $F \subseteq \Omega$.

Finally, let us prove part (3). If S is the unit sphere in $\ell^1(\Omega,\sigma)$, then $\mathbb{P}(\mathcal{C})$ is homeomorphic to $\mathcal{C}\cap S$. Since the latter is metrisable, it suffices to show that every sequence $(f_k)_k \subseteq \mathcal{C}\cap S$ has a converging subsequence. Since $\|f_k\|_\infty \le c\cdot\|f_k\|_1 = c$, the sequence $(f_k(\omega))_k$ takes values in the compact interval $[-c,c]$ for all $\omega\in\Omega$. Since Ω is countable, a diagonal argument allows us to replace $(f_k)_k$ with a subsequence that converges pointwise to a function $f:\Omega\to[-c,c]$. Thus, part (2) shows that $\|f_k - f\|_1 \to 0$. Since $\mathcal{C}$ is closed in $\ell^1(\Omega,\sigma)$, we have $f \in \mathcal{C}\cap S$, as required. □

Definition 6.14 A metrisable topological space X is an *absolute retract* (*AR*) if it enjoys the following property. For every metrisable topological space Y and every closed subset $A \subseteq Y$ homeomorphic to X, there exists a continuous retraction $Y \to A$.

The following summarises the key properties of ARs that we will need.

Theorem 6.15 (1) *Let X be a compact AR. Then every continuous map $f: X \to X$ has a fixed point.*

(2) *Let $(E, \|\cdot\|)$ be a normed space. If $\mathcal{C} \subseteq E$ is any positive cone, then $\mathbb{P}(\mathcal{C})$ is an AR (with the quotient of the norm topology of E).*

Proof Part (1) is a consequence of the Lefschetz fixed point theorem for compact ANRs [76; 77]. See for instance Theorem III.7.4 and Section I.6 in [70] for a clear statement.

If S is the unit sphere in the normed space E, then $\mathbb{P}(\mathcal{C})$ is homeomorphic to $\mathcal{C}\cap S$. Recall that every convex subset of a normed space is an AR; see for example [45, Corollary 4.2] or Corollary II.14.2 and Theorem III.3.1 in [70]. Every retract of an AR is again an AR; see [70, Proposition 7.7]. Thus, part (2) is immediate from the observation that $\mathcal{C}\cap S$ is a retract of the convex set $\mathcal{C}\setminus\{0\}$. □

6.2.2 Finding a projectively invariant metric Let M be a *countable*, finite-rank median algebra. Consider a finite set $S \subseteq \operatorname{Aut} M$ and let $G \le \operatorname{Aut} M$ be the subgroup that it generates. Let $\alpha \in \operatorname{Aut} M$ be an element that normalises G.

Consider the locally convex real vector space $\mathcal{E}(M) = \mathbb{R}^{M\times M}$, endowed with the topology of pointwise convergence on $M \times M$. We have a continuous linear action $\operatorname{Aut} M \curvearrowright \mathcal{E}(M)$ given by

$$(\psi\cdot f)(x,y) = f(\psi^{-1}(x),\psi^{-1}(y)) \quad \text{for all } \psi\in\operatorname{Aut} M, \text{ all } f\in\mathcal{E}(M) \text{ and all } x,y\in M.$$

Remark 6.16 The sets $\mathcal{PD}(M)$ and $\mathcal{PD}^G(M)$ (introduced in Section 2.3) are closed positive cones in $\mathcal{E}(M)$. In addition, $\mathcal{PD}(M)$ is $(\operatorname{Aut} M)$–invariant and $\mathcal{PD}^G(M)$ is $\langle\alpha\rangle$–invariant.

Although $\mathcal{D}(M)\cup\{0\}$ also is a positive cone, it is only closed when M is a single point.

Given a function $\mathfrak{c}\colon M\times M\to(0,+\infty)$, consider the (not necessarily convex) subset

$$\mathcal{PD}^G_{\mathfrak{c}}(M):=\{\eta\in\mathcal{PD}^G(M)\mid \eta(x,y)\le\mathfrak{c}(x,y)\cdot\overline{\tau}^{\eta}_S \text{ for all } x,y\in M\}.$$

As we shall see, this serves two purposes: on the one hand all closed cones in $\mathcal{PD}^G_{\mathfrak{c}}(M)$ have compact projectivisation; on the other, they only contain pseudometrics with $\overline{\tau}^{\eta}_S>0$ (except for $\eta=0$).

Our main aim in this subsection is to prove the following result:

Proposition 6.17 *Suppose that, for some $\mathfrak{c}\colon M\times M\to(0,+\infty)$, there exists a nontrivial $\langle\alpha\rangle$–invariant cone $\mathcal{C}\subseteq\mathcal{PD}^G_{\mathfrak{c}}(M)$ that is closed in $\mathcal{E}(M)$ with respect to the topology of pointwise convergence. Then there exists $\eta\in\mathcal{C}\setminus\{0\}$ such that $\overline{\tau}^{\eta}_S>0$ and $\alpha\cdot\eta=\lambda\eta$ for some $\lambda>0$.*

In order to prove the proposition, let us fix a probability measure σ on M with full support. Given a function $\mathfrak{c}\colon M\times M\to(0,+\infty)$, for $f\in\mathcal{E}(M)$ we define

$$\|f\|^{\mathfrak{c}}_1:=\sum_{x,y\in M}\frac{|f(x,y)|}{\mathfrak{c}(x,y)}\sigma(x)\sigma(y)\quad\text{and}\quad \|f\|^{\mathfrak{c}}_\infty:=\sup_{x,y\in M}\frac{|f(x,y)|}{\mathfrak{c}(x,y)}.$$

Note that $\|f\|^{\mathfrak{c}}_1$ is a norm on the subspace $\mathcal{E}^1_{\mathfrak{c}}(M)\subseteq\mathcal{E}(M)$ where it is finite. (The same is true of $\|f\|^{\mathfrak{c}}_\infty$, but this will not be relevant to us.)

Remark 6.18 Rescaling functions $f\in\mathcal{E}(M)$ by $\mathfrak{c}$, we map $\mathcal{E}^1_{\mathfrak{c}}(M)$ onto $\ell^1(M\times M,\sigma\otimes\sigma)$ linearly isometrically, while taking $\|f\|^{\mathfrak{c}}_\infty$ to $\|f\|_\infty$. Thus, we can apply Lemma 6.13 in this context.

Lemma 6.19 *Consider a function $\mathfrak{c}\colon M\times M\to(0,+\infty)$.*

(1) *The subset $\mathcal{PD}^G_{\mathfrak{c}}(M)\subseteq\mathcal{E}(M)$ is closed under pointwise convergence.*

(2) *There exists a constant $c>0$ (depending on $\mathfrak{c}$ and σ) such that, for every $\eta\in\mathcal{PD}^G_{\mathfrak{c}}(M)$,*

$$\|\eta\|^{\mathfrak{c}}_1\le\|\eta\|^{\mathfrak{c}}_\infty\le\overline{\tau}^{\eta}_S\le c\cdot\|\eta\|^{\mathfrak{c}}_1.$$

Proof We begin with part (1). First, observe that the function $\eta\mapsto\overline{\tau}^{\eta}_S$ is upper-semicontinuous. Indeed, if $\eta_n\in\mathcal{PD}^G(M)$ converge pointwise to some $\eta\in\mathcal{PD}^G(M)$, then, for every $x\in M$,

$$\max_{s\in S}\eta(x,sx)=\lim_{n\to+\infty}\max_{s\in S}\eta_n(x,sx)\ge\limsup_{n\to+\infty}\overline{\tau}^{\eta_n}_S.$$

Hence $\overline{\tau}^{\eta}_S\ge\limsup\overline{\tau}^{\eta_n}_S$, which proves upper-semicontinuity. Now, if $\eta_n\in\mathcal{PD}^G_{\mathfrak{c}}(M)$, then

$$\eta(x,y)=\lim_{n\to+\infty}\eta_n(x,y)\le\limsup_{n\to+\infty}\mathfrak{c}(x,y)\cdot\overline{\tau}^{\eta_n}_S\le\mathfrak{c}(x,y)\cdot\overline{\tau}^{\eta}_S$$

for all $x,y\in M$. Along with Remark 6.16, this yields $\eta\in\mathcal{PD}^G_{\mathfrak{c}}(M)$, proving part (1).

Regarding part (2), the first inequality is in Lemma 6.13(1) and the second is immediate from the fact that $\eta \in \mathcal{PD}^G_{\mathfrak{c}}(M)$. In order to prove the third one, choose any point $x_0 \in M$. Then

$$\overline{\tau}^{\eta}_S = \inf_{x \in M} \max_{s \in S} \eta(x, sx) \le \max_{s \in S} \eta(x_0, sx_0) \le \|\eta\|^{\mathfrak{c}}_1 \cdot \max_{s \in S} \frac{\mathfrak{c}(x_0, sx_0)}{\sigma(\{x_0\})\sigma(\{sx_0\})}.$$

The constant appearing on the rightmost side is positive and well-defined, since $\mathfrak{c}$ takes positive values and σ has full support. This concludes the proof. □

Proof of Proposition 6.17 We want to apply the Lefschetz fixed point theorem to $\alpha\colon \mathbb{P}(\mathcal{C}) \to \mathbb{P}(\mathcal{C})$.

Since $\mathcal{C} \subseteq \mathcal{PD}^G(M)$, the cone $\mathcal{C}$ is actually a positive cone. By Lemma 6.19(2), the set $\mathcal{C}$ is contained in $\mathcal{E}^1_{\mathfrak{c}}(M)$. Thus, Theorem 6.15(2) shows that the projectivisation $\mathbb{P}(\mathcal{C})$, endowed with the quotient topology induced by $\|\cdot\|^{\mathfrak{c}}_1$, is an AR.

Since $\mathcal{C} \subseteq \mathcal{E}^1_{\mathfrak{c}}(M)$ is closed in the topology of pointwise convergence, the first half of Lemma 6.13(2) guarantees that $\mathcal{C}$ is also closed in the topology of $\|\cdot\|^{\mathfrak{c}}_1$. Thus, by Lemmas 6.19(2) and 6.13(3), the projectivisation $\mathbb{P}(\mathcal{C})$ is compact.

We are left to show that the action $\langle\alpha\rangle \curvearrowright \mathcal{C}$ is continuous with respect to the topology of $\|\cdot\|^{\mathfrak{c}}_1$. Note that, by Lemma 6.19(2), α takes $\|\cdot\|^{\mathfrak{c}}_1$–bounded subsets of $\mathcal{C} \subseteq \mathcal{PD}^G_{\mathfrak{c}}(M)$ to $\|\cdot\|^{\mathfrak{c}}_1$–bounded subsets of $\mathcal{C}$:

$$\|\alpha \cdot \eta\|^{\mathfrak{c}}_1 \le \overline{\tau}^{\alpha\cdot\eta}_S = \inf_{x \in M} \max_{s \in S} \eta(\alpha^{-1}x, \alpha^{-1}sx) = \overline{\tau}^{\eta}_{\alpha^{-1}S\alpha} \le |\alpha^{-1}S\alpha|_S \cdot \overline{\tau}^{\eta}_S \le c|\alpha^{-1}S\alpha|_S \cdot \|\eta\|^{\mathfrak{c}}_1.$$

Since the topology given by $\|\cdot\|^{\mathfrak{c}}_1$ is metrisable, it suffices to show that $\alpha\colon \mathcal{C} \to \mathcal{C}$ is sequentially continuous. Let $\eta_n \in \mathcal{C}$ be a sequence that $\|\cdot\|^{\mathfrak{c}}_1$–converges to $\eta \in \mathcal{C}$. By Lemma 6.13(2), η_n converges to η pointwise. Since the action $\operatorname{Aut} M \curvearrowright \mathcal{E}(M)$ is continuous, the sequence $\alpha \cdot \eta_n$ converges to $\alpha \cdot \eta$ pointwise. Note that the set $\{\eta_n\}_{n\ge 0} \cup \{\eta\}$ is $\|\cdot\|^{\mathfrak{c}}_1$–bounded and, by the above observation, so must be $\{\alpha \cdot \eta_n\}_{n \ge 0} \cup \{\alpha\eta\}$. By Lemma 6.19(2), this set is also $\|\cdot\|^{\mathfrak{c}}_\infty$–bounded, so Lemma 6.13(2) shows that $\alpha \cdot \eta_n$ $\|\cdot\|^{\mathfrak{c}}_1$–converges to $\alpha \cdot \eta$, as required.

In conclusion, α induces a homeomorphism of the compact AR $\mathbb{P}(\mathcal{C})$. Theorem 6.15(1) yields an $\langle\alpha\rangle$–fixed point $[\eta] \in \mathbb{P}(\mathcal{C})$. The fact that $\overline{\tau}^{\eta}_S > 0$ is clear since $\eta \in \mathcal{PD}^G_{\mathfrak{c}}(M) \setminus \{0\}$. □

In fact, Proposition 6.17 can be easily generalised to extensions of G by abelian groups.

Corollary 6.20 *Let $U \le \operatorname{Aut} M$ be a countable subgroup such that $G \lhd U$, with abelian quotient U/G; let $p\colon U \to A$ be the quotient projection. Suppose that, for some $\mathfrak{c}$, there exists a nontrivial, U–invariant, closed cone $\mathcal{C} \subseteq \mathcal{PD}^G_{\mathfrak{c}}(M)$. Then there exists $\eta \in \mathcal{C} \setminus \{0\}$ with $\overline{\tau}^{\eta}_S > 0$ and a homomorphism $\lambda\colon A \to (\mathbb{R}_{>0}, *)$ such that $u \cdot \eta = \lambda(p(u))\eta$ for all $u \in U$.*

Proof Let $\{a_i\}_{i \ge 0}$ be a generating set for A. Consider the subgroups $A_n := \langle a_i \mid i < n\rangle$ and $U_n := p^{-1}(A_n)$; in particular, $A_0 = \{1\}$ and $U_0 = G$. We will show by induction on $n \ge 0$ that there exist nontrivial, U–invariant, closed cones $\mathcal{C}_n \subseteq \mathcal{PD}^G_{\mathfrak{c}}(M)$ and homomorphisms $\lambda_n \colon A_n \to (\mathbb{R}_{>0}, *)$ such that $u \cdot \eta = \lambda_n(p(u))\eta$ for all $\eta \in \mathcal{C}_n$ and $u \in U_n$. As base step, set $\mathcal{C}_0 := \mathcal{C}$.

Regarding the inductive step, suppose that we have constructed $\mathcal{C}_n$ and λ_n. By Proposition 6.17, there exists a point $[\eta_{n+1}] \in \mathbb{P}(\mathcal{C}_n)$ fixed by $p^{-1}(a_{n+1})$. In fact, since U_n acts trivially on $\mathbb{P}(\mathcal{C}_n)$, the entire group U_{n+1} fixes $[\eta_{n+1}]$ and there exists a homomorphism $\lambda_{n+1} \colon A_{n+1} \to (\mathbb{R}_{>0}, *)$ such that $u \cdot \eta_{n+1} = \lambda_{n+1}(p(u))\eta_{n+1}$ for all $u \in U_{n+1}$. We can then define $\mathcal{C}_{n+1}$ as the closed cone

$$\{\eta \in \mathcal{C}_n \mid u \cdot \eta = \lambda_{n+1}(p(u))\eta \text{ for all } u \in U_{n+1}\}.$$

Since $U \curvearrowright \mathcal{C}$ factors through the abelian group A, this cone is U–invariant, as required.

Finally, when A is not finitely generated, note that the intersection of the descending chain $\mathcal{C}_n$ is not just $\{0\}$. This is because, as we observed in the proof of Proposition 6.17, the sets $\mathbb{P}(\mathcal{C}_n)$ are compact. This concludes the proof. □

6.2.3 Universal uniform nonelementarity Let $G \curvearrowright M$ be an action by automorphisms on a median algebra of finite rank r. Consider the following strengthening of Definition 2.36 in the context of compatible metrics on median algebras.

Definition 6.21 The action $G \curvearrowright M$ is *universally uniformly nonelementary* (*WNE*) if there exists a constant $c > 0$ such that, for every $\eta \in \mathcal{PD}^G(M)$, the action $G \curvearrowright (M, \eta)$ is c–UNE.

This may seem an impossibly strong requirement to impose on $G \curvearrowright M$, but we will see in Corollary 7.24 that many actions arising from ultralimits of Salvetti complexes are WNE.

Lemma 6.22 *Let $G \leq \mathrm{Aut}\, M$ be generated by a finite set S of automorphisms acting nontransversely and stably without inversions. Let $G \lhd U \leq \mathrm{Aut}\, M$. Pick a point q in the multibridge $\mathcal{B}(S) \subseteq M$ and let $\mathfrak{M} \subseteq M$ be the median subalgebra generated by the orbit $U \cdot q$. Then:*

1. *There exists $\mathfrak{c}_1 \colon \mathfrak{M} \to (0, +\infty)$ such that $\tau^\eta_S(x) \leq \mathfrak{c}_1(x) \cdot \overline{\tau}^\eta_S$ for all $\eta \in \mathcal{PD}^G(M)$ and $x \in \mathfrak{M}$.*
2. *If $G \curvearrowright M$ is WNE, there exists $\mathfrak{c}_2 \colon \mathfrak{M} \times \mathfrak{M} \to (0, +\infty)$ such that $\eta(x, y) \leq \mathfrak{c}_2(x, y) \cdot \overline{\tau}^\eta_S$ for all $\eta \in \mathcal{PD}^G(M)$ and $x, y \in \mathfrak{M}$.*

Proof We only prove part (1), since part (2) then follows by setting $\mathfrak{c}_2(x, y) := c \cdot (\mathfrak{c}_1(x) + \mathfrak{c}_1(y))$, for a constant c as in Definition 6.21.

If part (1) holds for points $x, y, z \in \mathfrak{M}$, then it holds for their median $m(x, y, z)$. Indeed, we can take $\mathfrak{c}_1(m(x, y, z)) = \mathfrak{c}_1(x) + \mathfrak{c}_1(y) + \mathfrak{c}_1(z)$ and we have

$$\begin{aligned}\tau^\eta_S(m(x, y, z)) &= \max_{s \in S} \eta(m(x, y, z), m(sx, sy, sz)) \\ &\leq \max_{s \in S}[\eta(x, sx) + \eta(y, sy) + \eta(z, sz)] \\ &\leq \tau^\eta_S(x) + \tau^\eta_S(y) + \tau^\eta_S(z) \leq [\mathfrak{c}_1(x) + \mathfrak{c}_1(y) + \mathfrak{c}_1(z)] \cdot \overline{\tau}^\eta_S.\end{aligned}$$

Thus, it suffices to prove part (1) for $x \in U \cdot q$. Since $q \in \mathcal{B}(S)$, we have $uq \in \mathcal{B}(uSu^{-1})$ for all $u \in U$. Moreover, since U normalises G, the set uSu^{-1} is just another generating set of G. By Proposition 6.11(3), we have

$$\begin{aligned}\tau_S^\eta(uq) \le |S|_{uSu^{-1}} \cdot \tau_{uSu^{-1}}^\eta(uq) &\le |S|_{uSu^{-1}} \cdot (2r+1)\overline{\tau}_{uSu^{-1}}^\eta \\ &\le |S|_{uSu^{-1}} \cdot (2r+1) \cdot |uSu^{-1}|_S \cdot \overline{\tau}_S^\eta.\end{aligned}$$

So we can take $\mathfrak{c}_1(uq) = (2r+1) \cdot |S|_{uSu^{-1}} \cdot |uSu^{-1}|_S$. This concludes the proof. □

Corollary 6.23 *Let $G \le \operatorname{Aut} M$ be generated by a finite set S of automorphisms acting nontransversely and stably without inversions. Suppose that $G \curvearrowright M$ is WNE and that $\mathcal{D}^G(M) \neq \emptyset$. Consider a countable subgroup $U \le \operatorname{Aut} M$ such that $G \lhd U$ and U/G is abelian. Then there exist a nonempty, countable, U–invariant, median subalgebra $\mathfrak{M} \subseteq M$, a pseudometric $\eta \in \mathcal{PD}^G(\mathfrak{M}) \setminus \{0\}$ with $\overline{\tau}_S^\eta > 0$, and a homomorphism $\lambda\colon U \to (\mathbb{R}_{>0}, *)$ (trivial on G) with $u \cdot \eta = \lambda(u)\eta$ for all $u \in U$.*

Proof Define the median subalgebra $\mathfrak{M} \subseteq M$ as in the statement of Lemma 6.22. Since $\mathfrak{M}$ is generated by a countable set, it is itself countable. The restriction map

$$\operatorname{res}_{\mathfrak{M}}\colon \mathcal{PD}(M) \to \mathcal{PD}(\mathfrak{M})$$

takes $\mathcal{PD}^G(M)$ into $\mathcal{PD}^G(\mathfrak{M})$ without decreasing the value of $\overline{\tau}_S^\bullet$. Thus, in the notation of Section 6.2.2, Lemma 6.22(2) yields

$$\operatorname{res}_{\mathfrak{M}}(\mathcal{PD}^G(M)) \subseteq \mathcal{PD}^G_{\mathfrak{c}_2}(\mathfrak{M}).$$

Choose $\delta \in \mathcal{D}^G(M)$ and let $\mathcal{C} \subseteq \mathcal{D}^G(M)$ be the smallest cone containing the U–orbit of δ. In other words, $\mathcal{C}$ is the convex hull of $U \cdot \delta$, saturated under multiplication by nonnegative scalars. Then $\operatorname{res}_{\mathfrak{M}}(\mathcal{C})$ is a nontrivial U–invariant cone contained in $\mathcal{PD}^G_{\mathfrak{c}_2}(\mathfrak{M})$.

Its closure $\overline{\operatorname{res}_{\mathfrak{M}}(\mathcal{C})} \subseteq \mathcal{E}(\mathfrak{M})$ in the topology of pointwise convergence is also a U–invariant cone. By Lemma 6.19(1), this is still contained in the set $\mathcal{PD}^G_{\mathfrak{c}_2}(\mathfrak{M})$. We can thus apply Corollary 6.20, obtaining $\eta \in \overline{\operatorname{res}_{\mathfrak{M}}(\mathcal{C})} \setminus \{0\}$ with $\overline{\tau}_S^\eta > 0$, and a homomorphism $\lambda\colon U \to (\mathbb{R}_{>0}, *)$ such that $u \cdot \eta = \lambda(u)\eta$ for all $u \in U$. □

7 Ultralimits and coarse-median preserving automorphisms

In this section we prove Theorem I (Corollary 7.23) and complete the proof of Theorem F (Theorem 7.25). Both results will follow quickly once we prove Theorem 7.21 in Section 7.4, which can be viewed as the main goal of this entire section.

This theorem claims that, in many cases, if $G \curvearrowright M$ is an action of a special group on a median algebra, η is a G–invariant compatible pseudometric and C is a large k–cube in M, then any subset of G that moves all points in C by a lot less than the "size" of C must commute with a copy of $\mathbb{Z}^k$ sitting inside G. This result holds, for instance, for co-special cubulations of G, ultralimits of these, and subalgebras thereof, with uniform constants that are independent of the specific choice of η.

The case $k = 1$ thus implies that all these actions are WNE (Definition 6.21) and that centreless special groups are UNE (Definition 2.36). The cases with $k > 1$ ensure that the actions on median spaces that we will construct for Theorem F are *moderate*, as defined in the introduction.

7.1 The Bestvina–Paulin construction

As sketched in the introduction, the first step in the proof of Theorem F will involve a standard Bestvina–Paulin construction, with some additional issues caused by the lack of hyperbolicity. In this subsection, we discuss the role played by UNE groups (Definition 2.36) in addressing these issues.

Consider a group G, a geodesic metric space (X, d), and a homomorphism $\rho\colon G \to \operatorname{Isom} X$ inducing a proper cocompact action $G \curvearrowright X$ (we simply write gx rather than $\rho(g) \cdot x$).

7.1.1 The classical Bestvina–Paulin construction Fix a finite generating set $S \subseteq G$ and let $|\cdot|_S$ be the induced word length on G. Denote by $\pi : \operatorname{Aut} G \to \operatorname{Out} G$ the quotient projection. Given $g, h \in G$, we write $\mathfrak{c}[g](h) := ghg^{-1}$.

Every group automorphism $\varphi\colon G \to G$ is bi-Lipschitz with respect to $|\cdot|_S$. By the Milnor–Schwarz lemma, φ induces a quasi-isometry $\widetilde{\varphi}\colon X \to X$ satisfying $\widetilde{\varphi} \circ \rho(g) = \rho(\varphi(g)) \circ \widetilde{\varphi}$ for all $g \in G$.

Consider a sequence $\varphi_n \in \operatorname{Aut} G$ and set $\rho_n := \rho \circ \varphi_n$ for all $n \geq 0$. Pick basepoints $p_n \in X$ with

$$\tau_S^{\rho_n}(p_n) - \overline{\tau}_S^{\rho_n} \leq 1.$$

We introduce the quantities $\epsilon_n := 1/\overline{\tau}_S^{\rho_n}$ to simplify the notation.

Assumption 7.1 In the rest of Section 7.1, we assume that no two elements of the sequence $\pi(\varphi_n) \in \operatorname{Out} G$ coincide. A classical argument due to Bestvina and Paulin (see eg [7] and [87, page 338]) then guarantees that $\epsilon_n \to 0$ for $n \to +\infty$.

Fix a nonprincipal ultrafilter ω and consider the ultralimit $(X_\omega, d_\omega, p_\omega) = \lim_\omega(X, \epsilon_n d, p_n)$. We have a homomorphism $\rho_\omega\colon G \to \operatorname{Isom} X_\omega$ obtained as ultralimit of the actions ρ_n, namely

$$\rho_\omega(g) \cdot (x_n) = (\rho_n(g) \cdot x_n) = (\varphi_n(g)x_n)$$

for all $g \in G$ and $(x_n) \in X_\omega$. This is well-defined since

$$\begin{aligned}\lim_\omega \epsilon_n d(\varphi_n(g)x_n, p_n) &\leq \lim_\omega \epsilon_n[d(\varphi_n(g)x_n, \varphi_n(g)p_n) + d(\varphi_n(g)p_n, p_n)] \\ &\leq \lim_\omega \epsilon_n[d(x_n, p_n) + |g|_S \cdot \tau_S^{\rho_n}(p_n)] \\ &= d_\omega((x_n), p_\omega) + |g|_S < +\infty.\end{aligned}$$

One easily checks that $\tau_S^{\rho_\omega}(p_\omega) = \overline{\tau}_S^{\rho_\omega} = 1$, so the action $G \curvearrowright X_\omega$ induced by ρ_ω does not have a global fixed point.

7.1.2 Automorphisms of UNE groups Suppose for a moment that we are in the special case where there exists $\varphi \in \operatorname{Aut} G$ such that $\varphi_n = \varphi^n$ for all $n \geq 0$ (thus $\rho_n = \rho \circ \varphi^n$). We want to show that φ induces a map $\Phi\colon X_\omega \to X_\omega$ with the property that $\Phi \circ \rho_\omega(g) = \rho_\omega(\varphi(g)) \circ \Phi$ for all $g \in G$. A natural attempt is setting $\Phi((x_n)) = (\widetilde{\varphi}(x_n))$ for all $(x_n) \in X_\omega$. However, for this to be well-defined we need $\lim_\omega \epsilon_n d(\widetilde{\varphi}(p_n), p_n) < +\infty$.

We are actually interested in the following more general setting.

Assumption 7.2 Let $N \leq \operatorname{Out} G$ be a subgroup with infinite centre $Z(N)$. Let $\varphi_n \in \operatorname{Aut} G$ be a sequence that is mapped by the projection $\pi : \operatorname{Aut} G \to \operatorname{Out} G$ to a sequence of pairwise distinct elements in $Z(N)$. Consider again $\rho_n = \rho \circ \varphi_n$ as above.

If $\psi \in \pi^{-1}(N)$, then $\pi(\psi)$ commutes with each $\pi(\varphi_n)$. For every $n \in \mathbb{Z}$, choose $g_{n,\psi} \in G$ with

$$\varphi_n \circ \psi = \mathfrak{c}[g_{n,\psi}] \circ \psi \circ \varphi_n.$$

We are about to prove that, if G is UNE, ψ induces a well-defined map $\zeta(\psi)\colon X_\omega \to X_\omega$ given by

$$\zeta(\psi)((x_n)) = (g_{n,\psi} \widetilde{\psi}(x_n)).$$

(Recall that $\widetilde{\psi} : X \to X$ is the quasi-isometry induced by ψ.) We essentially use the same argument as [88, pages 154–156], replacing hyperbolicity with the UNE condition.

The proof of this result is quite technical. On a first read, we suggest restricting to the situation where $N \simeq \mathbb{Z}$ and the automorphisms ψ and φ_n are all powers of a given automorphism, in which case the elements $g_{n,\psi}$ can all be taken to be the identity and our strategy boils down to what is described right before Assumption 7.2. This case is sufficient for Theorem F, though not for the more general Theorem 7.25 below.

Proposition 7.3 *Suppose that G is UNE. Let $N \leq \operatorname{Out} G$ and $\varphi_n \in \operatorname{Aut} G$ be as in Assumption 7.2. Then there exists a homomorphism $\zeta : \pi^{-1}(N) \to \operatorname{Homeo} X_\omega$ that extends ρ_ω, in the sense that $\zeta(\mathfrak{c}[g]) = \rho_\omega(g)$ for every $g \in G$. Every homeomorphism in the image of ζ is bi-Lipschitz.*

Proof Consider an element $\psi \in \pi^{-1}(N)$. Let $L \geq 1$ be a constant such that $\widetilde{\psi} : X \to X$ is an (L, L)–quasi-isometry and such that $\psi : G \to G$ is L–bi-Lipschitz with respect to $|\cdot|_S$.

Step 1 *The map $\zeta(\psi)$ described above is a well-defined bi-Lipschitz homeomorphism of X_ω.*

Since $\widetilde{\psi}$ is a quasi-isometry and $\epsilon_n \to 0$, it suffices to show that $\zeta(\psi)$ is a well-defined map, ie that $\lim_\omega \epsilon_n d(g_{n,\psi} \widetilde{\psi}(p_n), p_n)$ is finite.

We begin by observing that, since $\varphi_n \circ \psi = \mathfrak{c}[g_{n,\psi}] \circ \psi \circ \varphi_n$ and $\widetilde{\psi} \circ \rho(g) = \rho(\psi(g)) \circ \widetilde{\psi}$,

$$\begin{aligned}
\tau_S^{\rho_n}(g_{n,\psi}\widetilde{\psi}(p_n)) &= \max_{s\in S} d(\varphi_n(s)g_{n,\psi}\widetilde{\psi}(p_n), g_{n,\psi}\widetilde{\psi}(p_n)) = \max_{s\in S} d((\mathfrak{c}[g_{n,\psi}]^{-1}\varphi_n)(s)\widetilde{\psi}(p_n), \widetilde{\psi}(p_n)) \\
&= \max_{s\in S} d(\widetilde{\psi}((\psi^{-1}\mathfrak{c}[g_{n,\psi}]^{-1}\varphi_n)(s)p_n), \widetilde{\psi}(p_n)) = \max_{s\in S} d(\widetilde{\psi}(\varphi_n\psi^{-1}(s)p_n), \widetilde{\psi}(p_n)) \\
&\leq L\cdot \max_{s\in S} d(\varphi_n\psi^{-1}(s)p_n, p_n) + L = L\cdot\max_{s\in S} d(\rho_n(\psi^{-1}(s))\cdot p_n, p_n) + L \\
&\leq L\cdot\max_{s\in S} |\psi^{-1}(s)|_S\cdot\tau_S^{\rho_n}(p_n) + L \leq L^2\cdot\tau_S^{\rho_n}(p_n) + L.
\end{aligned}$$

Now, since G is UNE, there exists a constant $c > 0$ such that, for every generating set $T \subseteq G$ and all $x, y \in X$, we have $d(x,y) \leq c\cdot(\tau_T^{\rho}(x) + \tau_T^{\rho}(y))$. For $T = \varphi_n(S)$, we obtain

$$\begin{aligned}
\lim_\omega \epsilon_n d(g_{n,\psi}\widetilde{\psi}(p_n), p_n) &\leq c\cdot\lim_\omega \epsilon_n(\tau^{\rho}_{\varphi_n(S)}(g_{n,\psi}\widetilde{\psi}(p_n)) + \tau^{\rho}_{\varphi_n(S)}(p_n)) \\
&= c\cdot\lim_\omega \epsilon_n(\tau_S^{\rho_n}(g_{n,\psi}\widetilde{\psi}(p_n)) + \tau_S^{\rho_n}(p_n)) \\
&\leq c(L^2+1)\cdot\lim_\omega \epsilon_n\tau_S^{\rho_n}(p_n) < +\infty.
\end{aligned}$$

Step 2 *The map ζ is a homomorphism.*

Since G is UNE, Example 2.38(3) shows that the centre $Z(G) \leq G$ is finite. Then, since G acts cocompactly on X, there exists a constant M such that $d(x, zx) \leq M$ for all $x \in X$ and $z \in Z(G)$. Given $\psi_1, \psi_2 \in N$, we can take $\widetilde{\psi_1\psi_2} = \widetilde{\psi}_1\widetilde{\psi}_2$. Moreover,

$$\begin{aligned}
\mathfrak{c}[g_{n,\psi_1\psi_2}]\psi_1\psi_2\varphi_n = \varphi_n\psi_1\psi_2 &= \mathfrak{c}[g_{n,\psi_1}]\psi_1\varphi_n\psi_2 \\
&= \mathfrak{c}[g_{n,\psi_1}]\psi_1\mathfrak{c}[g_{n,\psi_2}]\psi_2\varphi_n \\
&= \mathfrak{c}[g_{n,\psi_1}]\mathfrak{c}[\psi_1(g_{n,\psi_2})]\psi_1\psi_2\varphi_n.
\end{aligned}$$

Hence $g_{n,\psi_1\psi_2}$ and $g_{n,\psi_1}\psi_1(g_{n,\psi_2})$ differ by multiplication by an element of $Z(G)$. It follows that, for every $x \in X$, we have $d(g_{n,\psi_1\psi_2}x, g_{n,\psi_1}\psi_1(g_{n,\psi_2})x) \leq M$. Thus, for every $(x_n) \in X_\omega$,

$$\begin{aligned}
\zeta(\psi_1\psi_2)((x_n)) = (g_{n,\psi_1\psi_2}\widetilde{\psi_1\psi_2}(x_n)) &= (g_{n,\psi_1}\psi_1(g_{n,\psi_2})\widetilde{\psi}_1(\widetilde{\psi}_2(x_n))) \\
&= (g_{n,\psi_1}\widetilde{\psi}_1(g_{n,\psi_2}\widetilde{\psi}_2(x_n))) = \zeta(\psi_1)((g_{n,\psi_2}\widetilde{\psi}_2(x_n))) \\
&= \zeta(\psi_1)\zeta(\psi_2)((x_n)).
\end{aligned}$$

Step 3 *We have $\zeta(\mathfrak{c}[g]) = \rho_\omega(g)$ for all $g \in G$.*

Since $\mathfrak{c}[g]\colon G \to G$ is at bounded distance from left multiplication by g, the quasi-isometry $\widetilde{\mathfrak{c}[g]}$ is at bounded distance from $\rho(g)$. Moreover, observing that

$$\mathfrak{c}[\varphi_n(g)] \circ \varphi_n = \varphi_n \circ \mathfrak{c}[g] = \mathfrak{c}[g_{n,\mathfrak{c}[g]}] \circ \mathfrak{c}[g] \circ \varphi_n,$$

we deduce that $\mathfrak{c}[\varphi_n(g)g^{-1}] = \mathfrak{c}[g_{n,\mathfrak{c}[g]}]$, hence $g_{n,\mathfrak{c}[g]} \in Z(G)\varphi_n(g)g^{-1}$. Thus, for every $(x_n) \in X_\omega$,

$$\zeta(\mathfrak{c}[g])((x_n)) = (g_{n,\mathfrak{c}[g]}\widetilde{\mathfrak{c}[g]}(x_n)) = (g_{n,\mathfrak{c}[g]}gx_n) = (\varphi_n(g)g^{-1}gx_n) = (\varphi_n(g)x_n) = \rho_\omega(g)((x_n)).$$

This concludes the proof of the proposition. □

In the special case where there exists $\varphi \in \operatorname{Aut} G$ such that $\varphi_n = \varphi^n$ and $N = \langle \pi(\varphi) \rangle$, we have $\pi^{-1}(N) \simeq (G/Z(G)) \rtimes_\varphi \mathbb{Z}$ and we obtain:

Corollary 7.4 *Suppose that G is UNE and that $\pi(\varphi) \in \operatorname{Out} G$ has infinite order. Take $\varphi_n = \varphi^n$. Then the map $\Phi \colon X_\omega \to X_\omega$ given by $\Phi((x_n)) = (\widetilde{\varphi}(x_n))$ is a well-defined bi-Lipschitz homeomorphism of X_ω satisfying $\Phi \circ \rho_\omega(g) = \rho_\omega(\varphi(g)) \circ \Phi$ for all $g \in G$.*

7.1.3 Coarse-median preserving automorphisms of UNE groups Suppose now that X admits a coarse median μ of finite rank r. We can define a map $\mu_\omega \colon X_\omega^3 \to X_\omega$ by setting $\mu_\omega((x_n), (y_n), (z_n)) = (\mu(x_n, y_n, z_n))$. It was shown in [15, Section 9] that μ_ω is well-defined and the pair (X_ω, μ_ω) is a median algebra of rank $\leq r$.

If the coarse median structure $[\mu]$ is fixed by $G \curvearrowright X$, then the action $G \curvearrowright X_\omega$ is by automorphisms of the median algebra (X_ω, μ_ω). Moreover, if an automorphism $\psi \in \pi^{-1}(N) \leq \operatorname{Aut} G$ is such that $\widetilde{\psi}$ fixes $[\mu]$, then $\zeta(\psi) \in \operatorname{Aut}(X_\omega, \mu_\omega)$. Note that, although the metric d_ω on X_ω is G–invariant, it needs not be preserved by $\zeta(\psi)$.

Remark 7.5 If the space X is coarse median but not median, the metric d_ω may not be compatible with μ_ω (in the sense of Definition 2.8). However, it was shown by Zeidler [101, Proposition 3.3] that there always exists a metric $\delta \in \mathcal{D}^G(X_\omega, \mu_\omega)$ such that (X_ω, δ) is complete, geodesic, and bi-Lipschitz equivalent to (X_ω, d_ω). Theorem 2.14(2) and the fact that G does not fix a point in X_ω then imply that G acts on (X_ω, δ) with unbounded orbits (alternatively, one can appeal to [17]).

This is only tangentially relevant to us as we will only be interested in ultralimits of CAT(0) cube complexes in the forthcoming subsections.

Summing up the above discussion:

Corollary 7.6 *Let G be a UNE group. Let $N \leq \operatorname{Out} G$ be a subgroup with infinite centre. Let $(X, [\mu])$ be a geodesic coarse median space of finite rank r. Let $G \curvearrowright X$ be a proper cocompact action fixing the coarse median structure $[\mu]$. Suppose that the quasi-isometries of X induced by the elements of $\pi^{-1}(N)$ also preserve $[\mu]$.*

Then there exists a complete, geodesic median space X_ω of rank $\leq r$, and an action $\pi^{-1}(N) \curvearrowright X_\omega$ by bi-Lipschitz homeomorphisms that preserve the underlying median-algebra structure. The composition $G \to G/Z(G) \hookrightarrow \pi^{-1}(N) \curvearrowright X_\omega$ is an isometric G–action with unbounded orbits.

7.2 Equivariant embeddings in products of $\mathbb{R}$–trees

Let M be a median algebra and $G \curvearrowright M$ an action by median automorphisms. In the rest of Section 7, we will be interested in situations where M can be embedded G–equivariantly into a finite product of $\mathbb{R}$–trees. We reserve this subsection for a few general remarks on this setting.

Definition 7.7 An $\mathbb{R}$–*tree* is a geodesic, rank-1 median space.

This is equivalent to the usual definition of $\mathbb{R}$–trees as geodesic metric spaces where every geodesic triangle is a tripod. We stress that $\mathbb{R}$–trees are not required to be complete.

The next remark collects various simple observations for later use.

Remark 7.8 Consider isometric G–actions on $\mathbb{R}$–trees $T_1, \dots, T_k$. Equip $T_1 \times \dots \times T_k$ with the diagonal G–action. Let $f = (f_i)\colon M \hookrightarrow \prod T_i$ be a G–equivariant, injective median morphism.

(1) The image $f(M)$ is a median subalgebra of $\prod_i T_i$. The set of halfspaces of the median algebra $\prod_i T_i$ is naturally identified with the disjoint union $\bigsqcup_i \mathscr{H}(T_i)$.

Every halfspace of T_i is either open or closed. Open halfspaces are precisely the single connected components of the sets $T_i \setminus \{p\}$, as p varies through all points of T_i (including when $T_i \setminus \{p\}$ is connected). Closed halfspaces are precisely the complements of open halfspaces.

If we let $\mathscr{H}_i \subseteq \mathscr{H}(M)$ be the set of halfspaces of the form $f_i^{-1}(\mathfrak{h})$ with $\mathfrak{h} \in \mathscr{H}(T_i)$, then the $\mathscr{H}_i$ cover $\mathscr{H}(M)$ by Remark 2.2(1). However, the $\mathscr{H}_i$ are usually not pairwise disjoint.

(2) Since the sets $\mathscr{H}_i$ are G–invariant and no two halfspaces in the same $\mathscr{H}_i$ are transverse, we see that each $g \in G$ must act nontransversely on M.

(3) Suppose that, for all i, all $x \in T_i$ and all $g \in G$, we have $g^2 x = x$ if and only if $gx = x$. Then the action $G \curvearrowright M$ has no wall inversions.

Indeed, suppose instead that there exists $\mathfrak{h} \in \mathscr{H}(M)$ such that $g\mathfrak{h} = \mathfrak{h}^*$. Pick i such that $\mathfrak{h} \in \mathscr{H}_i$, and choose $\mathfrak{k} \in \mathscr{H}(T_i)$ with $f_i^{-1}(\mathfrak{k}) = \mathfrak{h}$. Then $g\mathfrak{k} \cap \mathfrak{k}$ and $g\mathfrak{k}^* \cap \mathfrak{k}^*$ are disjoint from the $\langle g \rangle$–invariant median subalgebra $f_i(M)$. Note that we cannot have $g\mathfrak{k} \subseteq \mathfrak{k}$ or $g\mathfrak{k} \supseteq \mathfrak{k}$, so, without loss of generality, $g\mathfrak{k} \cap \mathfrak{k} = \varnothing$. It follows that $f_i(M) \subseteq \mathfrak{k} \cup g\mathfrak{k}$, hence g is elliptic and fixes a unique point p in the convex hull of $\mathfrak{k} \cup g\mathfrak{k}$. We conclude that $g^2\mathfrak{k} = \mathfrak{k}$, hence the points on the arc connecting p to $\mathfrak{k}$ are fixed by g^2, but not by g. This is a contradiction.

(4) Suppose that g acts on M stably without wall inversions. By Remark 2.18(2) and Theorem 2.14(3), a halfspace $\mathfrak{h} \in \mathscr{H}(M)$ lies in the set $\mathscr{H}_{\overline{\mathcal{C}}(g)}(M)$ if and only if either $\mathfrak{h} \subsetneq g\mathfrak{h}$, or $\mathfrak{h} \subsetneq g^{-1}\mathfrak{h}$, or $\mathfrak{h} = g\mathfrak{h}$.

It follows that, for every i, either g is loxodromic in T_i and $f_i(\overline{\mathcal{C}}(g, M))$ is contained in its axis, or g is elliptic in T_i and fixes $f_i(\overline{\mathcal{C}}(g, M))$ pointwise.

Now, let us fix a nonprincipal ultrafilter ω. Let the group G be generated by a finite subset S. Consider a sequence of actions by automorphism on median algebras $G \curvearrowright M_n$, along with metrics $\delta_n \in \mathcal{D}^G(M_n)$ and basepoints $p_n \in M_n$. Suppose moreover that

$$\max_{s \in S} \sup_n \delta_n(sp_n, p_n) < +\infty.$$

Define $(M_\omega, \delta_\omega, p_\omega) := \lim_\omega (M_n, \delta_n, p_n)$. The set M_ω becomes a median algebra if we endow it with the operator $m((x_n), (y_n), (z_n)) = (m(x_n, y_n, z_n))$. We have an action by median automorphisms $G \curvearrowright M_\omega$ given by $g(x_n) = (gx_n)$. Finally, note that $\delta_\omega \in \mathcal{D}^G(M_\omega)$, and that $(M_\omega, \delta_\omega)$ is a complete median space (every ultralimit of metric spaces is complete).

Given a sequence of subsets $A_n \subseteq M_n$, we will employ the notation

$$\lim_\omega A_n := \{(x_n) \in M_\omega \mid x_n \in A_n \text{ for } \omega\text{–all } n\} = \{(y_n) \in M_\omega \mid \lim_\omega \delta_n(y_n, A_n) = 0\}.$$

Note that $\lim_\omega A_n$ is a (possibly empty) *closed* subset of $(M_\omega, \delta_\omega)$ for *any* sequence of subsets $A_n \subseteq M_n$. It is also clear that $\lim_\omega A_n \subseteq M_\omega$ is convex as soon as $A_n \subseteq M_n$ is convex for ω–all n.

Fix an integer $k \geq 1$. Suppose that each action $G \curvearrowright M_n$ is equipped with a G–equivariant, δ_n–isometric embedding $f_n = (f_n^i)\colon M_n \hookrightarrow \prod_i T_n^i$, where $\prod_i T_n^i$ is a product of k $\mathbb{R}$–trees endowed with an isometric, diagonal G–action as in Remark 7.8. (We have switched the index i from subscript to superscript to avoid confusion.)

It is straightforward to check that the ultralimits $\lim_\omega (T_n^i, f_n^i(p_n))$ yield isometric G–actions on $\mathbb{R}$–trees T_ω^i and a G–equivariant, δ_ω–isometric embedding $f_\omega = (f_\omega^i)\colon M_\omega \hookrightarrow \prod_i T_\omega^i$.

Lemma 7.9 *Consider the above setting. For every $g \in G$, we have*

(1) $\ell(g, T_\omega^i) = \lim_\omega \ell(g, T_n^i)$ *and* $\overline{\mathcal{C}}(g, T_\omega^i) = \lim_\omega \overline{\mathcal{C}}(g, T_n^i)$ *for all* $1 \leq i \leq k$.

If, in addition, (M_n, δ_n) is a geodesic space for ω–all n, then $(M_\omega, \delta_\omega)$ is geodesic and

(2) $\ell(g, \delta_\omega) = \lim_\omega \ell(g, \delta_n)$ *and* $\overline{\mathcal{C}}(g, M_\omega) = \lim_\omega \overline{\mathcal{C}}(g, M_n)$.

Proof We only prove part (2), since part (1) is a special case of it.

By Remarks 2.12 and 7.8(2), each $g \in G$ acts on M_ω stably without inversions and nontransversely; the same is true of the action on ω–all M_n. Theorem 2.16(2) shows that, for every $x = (x_n) \in M_\omega$, we have

$$\delta_\omega(x, gx) = \lim_\omega \delta_n(x_n, gx_n) = \lim_\omega [\ell(g, \delta_n) + 2\delta_n(x_n, \overline{\mathcal{C}}(g, M_n))] \geq \lim_\omega \ell(g, \delta_n).$$

Hence $\ell(g, \delta_\omega) \geq \lim_\omega \ell(g, \delta_n)$. By Theorem 2.16(1), the sets $\overline{\mathcal{C}}(g, M_n)$ are gate-convex. If y_n is the gate-projection of the basepoint $p_n \in M_n$ to $\overline{\mathcal{C}}(g, M_n)$, we have

$$\lim_\omega \delta_n(y_n, p_n) = \lim_\omega \delta_n(p_n, \overline{\mathcal{C}}(g, M_n)) \leq \lim_\omega \tfrac{1}{2}\delta_n(p_n, gp_n) < +\infty.$$

It follows that we have a well-defined point $y = (y_n) \in M_\omega$ and that $\delta_\omega(y, gy) = \lim_\omega \ell(g, \delta_n)$. This shows that $\ell(g, \delta_\omega) = \lim_\omega \ell(g, \delta_n)$.

Finally, since $\overline{\mathcal{C}}(g, M_\omega)$ is gate-convex, it is a closed subset of the complete median space $(M_\omega, \delta_\omega)$. Thus a point $x = (x_n) \in M_\omega$ lies in $\overline{\mathcal{C}}(g, M_\omega)$ if and only if $\delta_\omega(x, \overline{\mathcal{C}}(g, M_\omega)) = 0$, which happens if and only if $\delta_\omega(x, gx) = \ell(g, \delta_\omega)$ (again by Theorem 2.16). Equivalently, x lies in $\overline{\mathcal{C}}(g, M_\omega)$ if and only if $\lim_\omega \delta_n(x_n, \overline{\mathcal{C}}(g, M_n)) = 0$, ie if and only if $x \in \lim_\omega \overline{\mathcal{C}}(g, M_n)$. This concludes the proof. □

Lemma 7.10 *Consider again the above setting, with (M_n, δ_n) geodesic for ω–all n. Consider two elements $g, h \in G$ and $s \geq 1$.*

1. *Suppose that, for some $\mathfrak{w} \in \mathscr{W}(M_\omega)$, we have $\{\mathfrak{w}, g^s\mathfrak{w}\} \subseteq \mathcal{W}_1(g, M_\omega) \cap \mathcal{W}_1(h, M_\omega)$. Then, for ω–all n, there exists $\mathfrak{w}_n \in \mathscr{W}(M_n)$ such that $\{\mathfrak{w}_n, g^s\mathfrak{w}_n\} \subseteq \mathcal{W}_1(g, M_n) \cap \mathcal{W}_1(h, M_n)$.*
2. *If there exist walls $\mathfrak{u}, \mathfrak{v} \in \mathcal{W}_1(g, M_\omega)$ such that $\{\mathfrak{u}, g^s\mathfrak{u}\}$ is transverse to $\{\mathfrak{v}, g^s\mathfrak{v}\}$, then, for ω–all n, there exist $\mathfrak{u}_n, \mathfrak{v}_n \in \mathcal{W}_1(g, M_n)$ such that $\{\mathfrak{u}_n, g^s\mathfrak{u}_n\}$ is transverse to $\{\mathfrak{v}_n, g^s\mathfrak{v}_n\}$.*

Proof We begin with some general observations. We have already noted in Lemma 7.9 that $(M_\omega, \delta_\omega)$ is connected, hence g and h act stably without inversions. By parts (1) and (4) of Remark 7.8, each wall of M_ω arises from a wall of (at least) one of the trees T^i_ω. Moreover, each projection $f^i_\omega(\overline{\mathcal{C}}(g, M_\omega))$ is either fixed pointwise by g or it is a $\langle g\rangle$–invariant geodesic (and similarly for h).

We now prove part (1). By the above discussion, there exist an index i and $\mathfrak{v} \in \mathscr{W}(T^i_\omega)$ such that $\{\mathfrak{v}, g^s\mathfrak{v}\} \subseteq \mathcal{W}_1(g, T^i_\omega) \cap \mathcal{W}_1(h, T^i_\omega)$. Thus, g and h are both loxodromic in T^i_ω, which implies that they are loxodromic in ω–all T^i_n. Let α_ω, α_n and β_ω, β_n be the axes in T^i_ω, T^i_n of g and h, respectively. By Lemma 7.9, we have $\alpha_\omega = \lim_\omega \alpha_n$ and $\beta_\omega = \lim_\omega \beta_n$. Since α_ω and β_ω both cross $\mathfrak{v}$ and $g^s\mathfrak{v}$, they must share a segment of length $\epsilon + s\cdot\ell(g, T^i_\omega)$ for some $\epsilon > 0$.

If y and z are the endpoints of this segment, we can write $y = (y_n) = (y'_n)$ and $z = (z_n) = (z'_n)$ with $y_n, z_n \in \alpha_n$ and $y'_n, z'_n \in \beta_n$. Denoting by δ^i_n the metric of T^i_n, we have

$$\lim_\omega \delta^i_n(y_n, y'_n) = \lim_\omega \delta^i_n(z_n, z'_n) = 0 \quad \text{and} \quad \lim_\omega \delta^i_n(y_n, z_n) = \lim_\omega \delta^i_n(y'_n, z'_n) = \epsilon + s\cdot\lim_\omega \ell(g, T^i_n).$$

Hence α_n and β_n share a segment σ_n of length $> s\cdot\ell(g, T^i_n)$ for ω–all n. It follows that there exists a wall $\mathfrak{v}_n \in \mathscr{W}(T^i_n)$ such that σ_n crosses $\mathfrak{v}_n$ and $g^s\mathfrak{v}_n$. Hence $\{\mathfrak{v}_n, g^s\mathfrak{v}_n\} \subseteq \mathcal{W}_1(g, T^i_n) \cap \mathcal{W}_1(h, T^i_n)$, and it is clear that $\mathfrak{v}_n$ determines a wall $\mathfrak{w}_n$ of M with $\{\mathfrak{w}_n, g^s\mathfrak{w}_n\} \subseteq \mathcal{W}_1(g, M_n) \cap \mathcal{W}_1(h, M_n)$.

We now prove part (2). By Remark 7.8(4), $\mathfrak{u}$ and $\mathfrak{v}$ determine halfspaces $\mathfrak{h}, \mathfrak{k} \in \mathscr{H}(M_\omega)$ satisfying $g\mathfrak{h} \subsetneq \mathfrak{h}$ and $g\mathfrak{k} \subsetneq \mathfrak{k}$. Since $\{\mathfrak{u}, g^s\mathfrak{u}\}$ and $\{\mathfrak{v}, g^s\mathfrak{v}\}$ are transverse, Helly's lemma implies that there exist points

$$x \in g^s\mathfrak{h} \cap g^s\mathfrak{k} \cap \overline{\mathcal{C}}(g, M_\omega), \qquad y \in g^s\mathfrak{h} \cap \mathfrak{k}^* \cap \overline{\mathcal{C}}(g, M_\omega),$$
$$z \in \mathfrak{h}^* \cap g^s\mathfrak{k} \cap \overline{\mathcal{C}}(g, M_\omega), \qquad w \in \mathfrak{h}^* \cap \mathfrak{k}^* \cap \overline{\mathcal{C}}(g, M_\omega).$$

Suppose that $\mathfrak{u}$ and $\mathfrak{v}$ arise from trees T^i_ω and T^j_ω, where g has axes α^i and α^j, respectively. Then the points $f^i_\omega(x), f^i_\omega(y), f^i_\omega(z), f^i_\omega(w)$ lie on α^i, and $\{f^i_\omega(x), f^i_\omega(y)\}$ is separated from $\{f^i_\omega(z), f^i_\omega(w)\}$ by a segment of length $> s\cdot\ell(g, T^i_\omega)$. Similarly, $\{f^j_\omega(x), f^j_\omega(z)\}$ and $\{f^j_\omega(y), f^j_\omega(w)\}$ are separated by a subsegment of α^j of length $> s\cdot\ell(g, T^j_\omega)$.

Writing $x = (x_n)$, $y = (y_n)$, $z = (z_n)$ and $w = (w_n)$, it follows that, for ω–all n, there exist walls $\mathfrak{u}'_n \in \mathcal{W}_1(g, T^i_n)$ and $\mathfrak{v}'_n \in \mathcal{W}_1(g, T^j_n)$ such that

$$\{\mathfrak{u}'_n, g^s\mathfrak{u}'_n\} \subseteq \mathscr{W}(f^i_n(x_n), f^i_n(y_n) \,|\, f^i_n(z_n), f^i_n(w_n)),$$
$$\{\mathfrak{v}'_n, g^s\mathfrak{v}'_n\} \subseteq \mathscr{W}(f^j_n(x_n), f^j_n(z_n) \,|\, f^j_n(y_n), f^j_n(w_n)).$$

Thus $\mathfrak{u}'_n, \mathfrak{v}'_n$ induce $\mathfrak{u}_n, \mathfrak{v}_n \in \mathcal{W}_1(g, M_n)$ with $\{\mathfrak{u}_n, g^s\mathfrak{u}_n\}$ transverse to $\{\mathfrak{v}_n, g^s\mathfrak{v}_n\}$; cf Lemma 4.4. □

7.3 Ultralimits of convex-cocompact actions on Salvettis

Let Γ be a finite simplicial graph, $\mathcal{A} = \mathcal{A}_\Gamma$ the associated right-angled Artin group, and $\mathcal{X} = \mathcal{X}_\Gamma$ the universal cover of its Salvetti complex. Denote by d the ℓ^1 metric on $\mathcal{X}$ and set $r = \dim \mathcal{X}$. Fix a nonprincipal ultrafilter ω.

When we speak *convex-cocompactness* in $\mathcal{A}$ from now on (Definition 3.1), this is always meant with respect to the standard action $\mathcal{A} \curvearrowright \mathcal{X}$. Note that a group G is isomorphic to a convex-cocompact subgroup of a right-angled Artin group if and only if G is the fundamental group of a compact special cube complex [68]. In particular, G must be torsionfree and finitely generated.

In the rest of Section 7 we make the following assumption.

Assumption 7.11 Let $G \leq \mathcal{A}$ be a convex-cocompact subgroup. Let $Y \subseteq \mathcal{X}$ be a G–invariant, convex subcomplex on which G acts with exactly q orbits of vertices. Let $[\mu]$ be the induced coarse median structure on G. Consider a sequence $\varphi_n \in \mathrm{Aut}(G, [\mu])$. Denote by $\rho\colon G \hookrightarrow \mathcal{A}$ the standard inclusion and set $\rho_n = \rho \circ \varphi_n$.

We say for simplicity that $g \in G$ is *label-irreducible* if $\rho(g)$ is a label-irreducible element of $\mathcal{A}$.

Remark 7.12 If $g \in G$ is label-irreducible, then Corollary 3.3 and Lemma 3.11(2) show that $\rho_n(g) \in \mathcal{A}$ is label-irreducible for all $n \geq 0$.

Let $S \subseteq G$ be a finite generating set. Choose basepoints $p_n \in Y_n$ with $\tau_S^{\rho_n}(p_n) = \overline{\tau}_S^{\rho_n}$ and define $\delta_n := d/\overline{\tau}_S^{\rho_n} \in \mathcal{D}^G(\mathcal{X})$. For ease of notation, let us write $G \curvearrowright \mathcal{X}_n$ and $G \curvearrowright Y_n$ for the actions of G on $\mathcal{X}$ and Y induced by the homomorphism ρ_n.

Recall that $\gamma\colon \mathscr{W}(\mathcal{X}) \to \Gamma^{(0)}$ is the map pairing each hyperplane with its label. For every $v \in \Gamma^{(0)}$, the hyperplanes in $\gamma^{-1}(v)$ are pairwise disjoint. Hence there is a natural simplicial tree $\mathcal{T}^v$ (usually locally infinite) that is dual to the collection $\gamma^{-1}(v)$. In the terminology of Section 2.5, the tree $\mathcal{T}^v$ is the restriction quotient of $\mathcal{X}$ associated to $\gamma^{-1}(v) \subseteq \mathscr{W}(\mathcal{X})$.

In particular, we have an $\mathcal{A}$–equivariant, surjective median morphism $\pi^v\colon \mathcal{X} \to \mathcal{T}^v$ taking cubes to edges or vertices, and an $\mathcal{A}$–equivariant, isometric median morphism $(\pi^v)\colon \mathcal{X} \hookrightarrow \prod_{v\in\Gamma} \mathcal{T}^v$.

Let $\mathcal{T}_n^v$ denote the tree $\mathcal{T}^v$ equipped with the twisted G–action induced by ρ_n and with its graph metric rescaled by $\overline{\tau}_S^{\rho_n}$. We obtain a G–equivariant, δ_n–isometric embedding $(\pi_n^v)\colon \mathcal{X}_n \hookrightarrow \prod_{v\in\Gamma} \mathcal{T}_n^v$.

Thus, our setting is a special case of the one in the second part of Section 7.2 (after Remark 7.8). If the automorphisms φ_n are pairwise distinct in $\mathrm{Out}\, G$, then we are also in a special case of Section 7.1, but we do not make this assumption for the moment.

As in Section 7.2, the sequence of actions $G \curvearrowright \mathcal{X}_n$ with metrics δ_n and basepoints p_n yields a limit action $G \curvearrowright \mathcal{X}_\omega$, along with a metric $\delta_\omega \in \mathcal{D}^G(\mathcal{X}_\omega)$, a basepoint $p_\omega \in \mathcal{X}_\omega$, and a G–equivariant, δ_ω–isometric embedding $(\pi^v_\omega)\colon \mathcal{X}_\omega \hookrightarrow \prod_{v\in\Gamma} \mathcal{T}^v_\omega$. The pair $(\mathcal{X}_\omega, \delta_\omega)$ is a complete, geodesic median space of rank $\leq r$.

We now prove a sequence of fairly straightforward lemmas regarding the action of G on $\mathcal{X}_\omega$ and its median subalgebras. After that comes the most important part of this subsection, which is concerned with the notion of *cubical configurations* (Definition 7.17).

Lemma 7.13 *Consider label-irreducible elements $g, h \in G$.*

1. *If there exist walls $\mathfrak{u}$ and $\mathfrak{w}$ with $\{\mathfrak{u}, \mathfrak{w}, h^{4r}\mathfrak{u}, g^{4r}\mathfrak{w}\} \subseteq \mathcal{W}_1(g, \mathcal{X}_\omega) \cap \mathcal{W}_1(h, \mathcal{X}_\omega)$, then $\langle g, h\rangle \simeq \mathbb{Z}$.*
2. *There do not exist walls $\mathfrak{u}, \mathfrak{w} \in \mathcal{W}_1(g, \mathcal{X}_\omega)$ such that $\{\mathfrak{u}, g^{4r}\mathfrak{u}\}$ is transverse to $\{\mathfrak{w}, g^{4r}\mathfrak{w}\}$.*

Proof We begin with part (1). By Lemma 7.10(1), there exist hyperplanes $\mathfrak{u}_n, \mathfrak{w}_n \in \mathscr{W}(\mathcal{X}_n)$ for some n, such that $\{\mathfrak{u}_n, \mathfrak{w}_n, h^{4r}\mathfrak{u}_n, g^{4r}\mathfrak{w}_n\} \subseteq \mathcal{W}_1(g, \mathcal{X}_n) \cap \mathcal{W}_1(h, \mathcal{X}_n)$. Since $\rho_n(g)$ and $\rho_n(h)$ are label-irreducible by Remark 7.12, Lemma 3.13 guarantees that $\langle g, h\rangle \simeq \mathbb{Z}$.

Regarding part (2), if there existed such walls, Lemma 7.10(2) would yield hyperplanes $\mathfrak{u}_n, \mathfrak{w}_n \in \mathscr{W}(\mathcal{X}_n)$ for some n, such that the sets $\{\mathfrak{u}_n, g^{4r}\mathfrak{u}_n\}$ and $\{\mathfrak{w}_n, g^{4r}\mathfrak{w}_n\}$ were transverse and both contained in $\mathcal{W}_1(g, \mathcal{X}_n)$. This would violate Lemma 3.10, since $\rho_n(g)$ is label-irreducible. □

Lemma 7.14 *For every G–invariant median subalgebra $M \subseteq \mathcal{X}_\omega$, we have:*

1. *The action $G \curvearrowright M$ has no wall inversions.*
2. *Each element $g \in G$ is elliptic (resp. loxodromic) in M if and only if it is in $\mathcal{X}_\omega$.*

Proof Part (2) follows from part (1). Indeed, note that $\mathcal{H}_1(g, M) = \varnothing$ if and only if $\mathcal{H}_1(g, \mathcal{X}_\omega) = \varnothing$, for instance by Remark 2.18(3). Since the action $G \curvearrowright M$ has no inversions, Theorem 2.14(2) then shows that g is elliptic/loxodromic in M if and only if it is $\mathcal{X}_\omega$.

Regarding part (1), we will need the following observation:

Claim *Let an action $G \curvearrowright (T_\omega, d_\omega)$ be the ultralimit of a sequence of actions on $\mathbb{R}$–trees $G \curvearrowright (T_n, d_n)$. Suppose in addition that $g \in G$ is loxodromic in ω–all T_n. Then, for all $k \in \mathbb{Z} \setminus \{0\}$ and all $x \in T_\omega$, the point x is fixed by g^k if and only if it is fixed by g.*

Proof Let α_n be the axis of g in T_n and consider a point $y = (y_n) \in T_\omega$. Then

$$d_n(y_n, g^k y_n) = \ell(g^k, T_n) + 2d_n(y_n, \alpha_n) \geq \ell(g, T_n) + 2d_n(y_n, \alpha_n) = d_n(y_n, gy_n).$$

It follows that $d_\omega(y, g^k y) \geq d_\omega(y, gy)$ for all $k \in \mathbb{Z} \setminus \{0\}$, which proves the claim. ◁

Now, we will deduce that the action $G \curvearrowright M$ has no wall inversions from Remark 7.8(3). We need to show that, for every $v \in \Gamma$, every $x \in \mathcal{T}^v_\omega$ and every $g \in G$, we have $g^2 x = x$ if and only if $gx = x$. If $\rho_n(g)$ is loxodromic in $\mathcal{T}^v_n$ for ω–all n, this follows from the claim. If instead $\rho_n(g)$ is elliptic in $\mathcal{T}^v_n$ for ω–all n, then it follows from the observation that edge-stabilisers for the action $G \curvearrowright \mathcal{T}^v_n$ are closed under taking roots in G (since they are hyperplane-stabilisers for $G \curvearrowright \mathcal{X}_n$). $\square$

Lemma 7.15 *Consider $g \in G$ such that its label-irreducible components $g_1, \dots, g_k$ also lie in G (in general, they only lie in $\mathcal{A}$). Then, for every G–invariant median subalgebra $M \subseteq \mathcal{X}_\omega$:*

1. *We have a partition $\mathcal{W}_1(g, M) = \mathcal{W}_1(g_1, M) \sqcup \dots \sqcup \mathcal{W}_1(g_k, M)$.*
2. *Each wall in $\mathcal{W}_1(g_i, M)$ is preserved by each g_j with $j \neq i$.*
3. *The sets $\mathcal{W}_1(g_1, M), \dots, \mathcal{W}_1(g_k, M)$ are transverse to each other.*
4. *We have $\bar{\mathcal{C}}(g, M) = \bar{\mathcal{C}}(g_1, M) \cap \dots \cap \bar{\mathcal{C}}(g_k, M)$ and $\bar{\mathcal{C}}(g^m, M) = \bar{\mathcal{C}}(g, M)$ for all $m \geq 1$.*
5. *For every $\eta \in \mathcal{PD}^G(M)$, we have $\ell(g, \eta) = \ell(g_1, \eta) + \dots + \ell(g_k, \eta)$.*

Proof Let us prove parts (1) and (2) first, except for disjointness of the sets $\mathcal{W}_1(g_i, M)$, which will follow from part (3). Note that it suffices to consider the case when $M = \mathcal{X}_\omega$. Indeed, by Remark 2.2, we have a surjection $\mathrm{res}_M : \mathscr{W}_M(\mathcal{X}_\omega) \to \mathscr{W}(M)$ and, by Remark 2.18(3), a wall $\mathfrak{w} \in \mathscr{W}_M(\mathcal{X}_\omega)$ lies in $\mathcal{W}_1(g, \mathcal{X}_\omega)$ if and only if $\mathrm{res}_M(\mathfrak{w})$ lies in $\mathcal{W}_1(g, M)$.

In fact, Remark 7.8(1) shows that it suffices to prove parts (1) and (2) "for the trees $\mathcal{T}^v_\omega$", ie prove that, for every $v \in \Gamma$, we have $\mathcal{W}_1(g, \mathcal{T}^v_\omega) = \mathcal{W}_1(g_1, \mathcal{T}^v_\omega) \cup \dots \cup \mathcal{W}_1(g_k, \mathcal{T}^v_\omega)$, and that g_j fixes the set $\mathcal{W}_1(g_i, \mathcal{T}^v_\omega)$ pointwise for $j \neq i$.

Note that distinct components g_i cannot be loxodromic in the same tree $\mathcal{T}^v_\omega$. Otherwise they would have the same axis, since they commute, and Lemma 7.13(1) would give a contradiction. Thus, at most one of the sets $\mathcal{W}_1(g_1, \mathcal{T}^v_\omega), \dots, \mathcal{W}_1(g_k, \mathcal{T}^v_\omega)$ can be nonempty, for each v.

Recalling that $g = g_1 \cdots g_k$ and that the g_i commute pairwise, we conclude that either $\mathcal{W}_1(g, \mathcal{T}^v_\omega)$ is empty, or it coincides with $\mathcal{W}_1(g_{i_v}, \mathcal{T}^v_\omega)$, where g_{i_v} is the only label-irreducible component that is loxodromic in $\mathcal{T}^v_\omega$. If $j \neq i_v$, then g_j is elliptic in $\mathcal{T}^v_\omega$ and, since it commutes with g_{i_v}, it must fix pointwise its entire axis. In particular, g_j preserves every wall in the set $\mathcal{W}_1(g_{i_v}, \mathcal{T}^v_\omega)$. This proves parts (1) and (2), except for disjointness of the sets $\mathcal{W}_1(g_i, M)$.

In order to prove part (3), note that part (2) shows that $\mathcal{W}_1(g_i, M) \subseteq \mathcal{W}_0(g_j, M)$ for $i \neq j$. By Lemma 7.14(1), the action $G \curvearrowright M$ has no wall inversions. Thus $\mathcal{W}_1(g_i, M)$ and $\mathcal{W}_1(g_j, M)$ are transverse by Theorem 2.14(3). In particular, $\mathcal{W}_1(g_i, M)$ and $\mathcal{W}_1(g_j, M)$ are disjoint, which completes the proof of part (1).

Regarding part (4), it suffices to prove the statements for $M = \mathcal{X}_\omega$. Indeed, G acts nontransversely on $\mathcal{X}_\omega$ and without inversions on M, so we have $\bar{\mathcal{C}}(g, M) = M \cap \bar{\mathcal{C}}(g, \mathcal{X}_\omega)$, for instance by [51, Proposition 3.40].

The same holds for the g_i. Now, Lemma 7.9(2) implies that $\bar{\mathcal{C}}(g, \mathcal{X}_\omega)$ coincides with $\bigcap_i \bar{\mathcal{C}}(g_i, \mathcal{X}_\omega)$ and $\bar{\mathcal{C}}(g^m, \mathcal{X}_\omega)$, since this is true for convex cores in all $\mathcal{X}_n$ (recall Lemma 3.7(2) and Remark 7.12).

Finally, we prove part (5). Parts (1) and (2) imply that a $\mathscr{B}$–measurable fundamental domain for the action $\langle g\rangle \curvearrowright \mathcal{H}_1(g, M)$ can be constructed as the disjoint union of $\mathscr{B}$–measurable fundamental domains for the actions $\langle g_i\rangle \curvearrowright \mathcal{H}_1(g_i, M)$. Since $G \curvearrowright M$ has no wall inversions, translation lengths coincide with a measure of these fundamental domains (Remark 2.19) and part (5) follows. □

Lemma 7.16 *Let $M \subseteq \mathcal{X}_\omega$ be a G–invariant median subalgebra with a pseudometric $\eta \in \mathcal{PD}^G(M)$. Consider an element $g \in G$ and a point $x \in M$.*

(1) *For every $m \geq 1$, we have $\eta(x, gx) \leq \eta(x, g^m x)$.*

(2) *If $h \in G$ is a label-irreducible component of g, then $\eta(x, hx) \leq \eta(x, gx)$.*

Proof Recall that the action $G \curvearrowright M$ is nontransverse by Remark 7.8(2), and without inversions by Lemma 7.14(1). Thus, Theorem 2.16(2) guarantees that $\eta(x, gx) = \ell(g, \eta) + 2\eta(x, \bar{\mathcal{C}}(g, M))$.

Now, part (1) is obtained by observing that $\ell(g^m, \eta) = m \cdot \ell(g, \eta)$ and $\bar{\mathcal{C}}(g^m, M) = \bar{\mathcal{C}}(g, M)$, which follow from Remark 2.19 and Lemma 7.15(4), respectively. For part (2), it suffices to recall that $\bar{\mathcal{C}}(g, M) \subseteq \bar{\mathcal{C}}(h, M)$ and $\ell(h, \eta) \leq \ell(g, \eta)$, as shown in parts (4) and (5) of Lemma 7.15. □

We now introduce *cubical configurations*, which will be important in the proof of Theorem 7.21, hence in those of Theorems F and I. The idea is that large cubes in $\mathcal{X}_\omega$ that are moved very little by a subset $F \subseteq G$ will give rise to cubical configurations in $\mathcal{X}_\omega$ (Lemma 7.22).

After the definition, we will see how to transfer cubical configurations from $\mathcal{X}_\omega$ to $\mathcal{X}$ (Lemma 7.18) and how to use them to construct large abelian subgroups in the centraliser $Z_G(F)$ (Lemma 7.19).

Definition 7.17 Consider an action on a median algebra $G \curvearrowright M$ and a finite subset $F \subseteq G$. An (s, t, F)–*cubical configuration of width* $m \geq 1$ in M is the datum of nonempty subsets $\mathcal{U}_1, \dots, \mathcal{U}_s \subseteq \mathscr{W}(M)$, walls $\mathfrak{v}_1, \dots, \mathfrak{v}_t \in \mathscr{W}(M)$ and a partition $F = F_0 \sqcup \{g_1, \dots, g_t\}$ such that

(1) the sets $\mathcal{U}_1, \dots, \mathcal{U}_s, \{\mathfrak{v}_1, g_1^m \mathfrak{v}_1\}, \dots, \{\mathfrak{v}_t, g_t^m \mathfrak{v}_t\}$ are transverse to each other and their union is contained in $\mathcal{W}_0(f, M)$ for every $f \in F_0$,

(2) for each $1 \leq j \leq t$, we have $\{\mathfrak{v}_j, g_j^m \mathfrak{v}_j\} \subseteq \mathcal{W}_1(g_j, M)$, while $\mathcal{W}_0(g_j, M)$ contains $\mathcal{U}_1, \dots, \mathcal{U}_s$ and all sets $\{\mathfrak{v}_{j'}, g_{j'}^m \mathfrak{v}_{j'}\}$ with $j' \neq j$.

We refer to $\mathcal{U}_1, \dots, \mathcal{U}_s$ as the *static sets* and to $g_1, \dots, g_t$ as the *skewering elements*.

The proof of the next result is quite similar in spirit to that of Lemma 7.9, but we repeat it for the reader's convenience, since it is a bit more technical.

We denote by $Y_\omega \subseteq \mathcal{X}_\omega$ the convex subset obtained as $\lim_\omega Y_n$. A subset $\mathscr{C} \subseteq \mathscr{W}(Y)$ is a *chain* if it is the set of hyperplanes associated to a set of halfspaces that is totally ordered by inclusion.

Lemma 7.18 *Suppose that the sequence φ_n is not ω–constant. Let $F \subseteq G$ be a finite subset of label-irreducible elements such that no two of them generate a cyclic subgroup. Suppose that Y_ω admits an (s,t,F)–cubical configuration of width $\geq 4r$ with skewering elements $g_1,\dots,g_t$.*

Then, for ω–all n, there exists a $(\sigma,\tau,\varphi_n(F))$–cubical configuration of width $\geq 4r$ in Y such that $\sigma+\tau=s+t$ and the $\varphi_n(g_i)$ are skewering elements (hence $\tau \geq t$ and $\sigma \leq s$). In addition, the static sets of this configuration can be taken to be arbitrarily long chains of hyperplanes.

Proof Let the cubical configuration in Y_ω consist of static sets $\mathcal{U}_1,\dots,\mathcal{U}_s$, walls $\mathfrak{v}_1,\dots,\mathfrak{v}_t$ and the partition $F=F_0 \sqcup \{g_1,\dots,g_t\}$. It suffices to assume that each $\mathcal{U}_i$ is a singleton $\{\mathfrak{u}_i\}$. Recall that the action $G \curvearrowright Y_\omega$ is nontransverse by Remark 7.8(2), and without inversions by Remark 2.12.

By Remark 7.8(1), there exist vertices $u_1,\dots,u_s,v_1,\dots,v_t \in \Gamma$ such that the walls $\mathfrak{u}_i$ and $\mathfrak{v}_j$ arise from walls $\overline{\mathfrak{u}}_i \in \mathscr{W}(\mathcal{T}_\omega^{u_i})$ and $\overline{\mathfrak{v}}_j \in \mathscr{W}(\mathcal{T}_\omega^{v_j})$, respectively. Note that each $\overline{\mathfrak{u}}_i$ is preserved by all elements of F, while $\overline{\mathfrak{v}}_j$ and $g_j^{4r}\overline{\mathfrak{v}}_j$ are preserved by $F \setminus \{g_j\}$ and cross the axis of g_j in $\mathcal{T}_\omega^{v_j}$.

Claim *There exist nontrivial arcs $\alpha_i \subseteq \mathcal{T}_\omega^{u_i}$ and $\beta_j \subseteq \mathcal{T}_\omega^{v_j}$, with endpoints $\alpha_i^\pm$ and $\beta_j^\pm$, such that:*

(a) *Each α_i is fixed by F and each β_j is fixed by $F \setminus \{g_j\}$.*

(b) *β_j is contained in the axis of g_j in $\mathcal{T}_\omega^{v_j}$ and it has length $> 4r \cdot \ell(g_j,\mathcal{T}_\omega^{v_j})$.*

(c) *These arcs induce transverse sets of walls of Y_ω. More precisely, for every $(\epsilon,\zeta) \in \{\pm\}^s \times \{\pm\}^t$, there exists a point $x^{\epsilon,\zeta} \in Y_\omega$ such that, for all i and j, the nearest-point projection of $\pi_\omega^{u_i}(x^{\epsilon,\zeta})$ to α_i is $\alpha_i^{\epsilon_i}$, and the nearest-point projection of $\pi_\omega^{v_j}(x^{\epsilon,\zeta})$ to β_j is $\beta_j^{\zeta_j}$.*

Proof The walls $\mathfrak{u}_i,\mathfrak{v}_j \in \mathscr{W}(Y_\omega)$ correspond to halfspaces $\mathfrak{u}_i^\pm,\mathfrak{v}_j^\pm \in \mathscr{H}(Y_\omega)$, which we label so that, for each j, the halfspaces $\mathfrak{v}_j^-$ and $g_j^{4r}\mathfrak{v}_j^+$ are disjoint. Since the sets $\{\mathfrak{u}_i\}$ and $\{\mathfrak{v}_j, g_j^{4r}\mathfrak{v}_j\}$ are all transverse to each other, Helly's lemma allows us to find points $x^{\epsilon,\zeta} \in Y_\omega$ so that $x^{\epsilon,\zeta}$ lies in $\mathfrak{u}_i^{\epsilon_i}$ for all i and so that, for all j, it lies in $\mathfrak{v}_j^-$ if $\zeta_j=-$ and in $g_j^{4r}\mathfrak{v}_j^+$ if $\zeta_j=+$.

For each i, there exists a point $q_i \in \mathcal{T}_\omega^{u_i}$ such that one of the two halfspaces associated to $\overline{\mathfrak{u}}_i$ is a connected component κ_i of $\mathcal{T}_\omega^{u_i} \setminus \{q_i\}$; in particular, κ_i is open. Since G acts on Y_ω without inversions, F fixes q_i and leaves κ_i invariant. Since F is finite, it fixes nontrivial arc of $\mathcal{T}_\omega^{u_i}$ with one endpoint equal to q_i and the other lying in κ_i. We let α_i be this arc, possibly shrinking it a bit to ensure that the finitely many points $\pi_\omega^{u_i}(x^{\epsilon,\zeta})$ have the correct projections to α_i.

Finally, consider an index $1 \leq j \leq t$. The set of points $x^{\epsilon,\zeta}$ with $\zeta_j=-$ is contained in $\mathfrak{v}_j^-$, whereas the set of points $x^{\epsilon,\zeta}$ with $\zeta_j=+$ is contained in $g_j^{4r}\mathfrak{v}_j^+$. Note that either $\pi_\omega^{v_j}(\mathfrak{v}_j^-)$ or $\pi_\omega^{v_j}(g_j^{4r}\mathfrak{v}_j^+)$ is an *open* halfspace of $\mathcal{T}_\omega^{v_j}$. It follows that the set of points $\pi_\omega^{v_j}(x^{\epsilon,\zeta})$ with $\zeta_j=-$ is separated by the set of points $\pi_\omega^{v_j}(x^{\epsilon,\zeta})$ with $\zeta_j=+$ by an arc β_j that is contained in the axis of g_j in $\mathcal{T}_\omega^{v_j}$ and has length $> 4r \cdot \ell(g_j,\mathcal{T}_\omega^{v_j})$. Since $F \setminus \{g_j\}$ preserves the halfspaces $\mathfrak{v}_j^-$ and $g_j^{4r}\mathfrak{v}_j^+$, we can shrink β_j a bit to ensure that it is fixed by $F \setminus \{g_j\}$, while retaining length $> 4r \cdot \ell(g_j,\mathcal{T}_\omega^{v_j})$. This concludes the proof of the claim. ◁

Now, it is straightforward to approximate, for ω–all n, the data provided by the claim by arcs $\alpha_i(n)\subseteq\mathcal{T}_n^{u_i}$, $\beta_j(n)\subseteq\mathcal{T}_n^{v_j}$ and points $x_n^{\epsilon,\zeta}\in Y_n$ satisfying analogous conditions.

Here we need to account for the fact that some elements of F that are elliptic in one of the trees $\mathcal{T}_\omega^\bullet$ might be loxodromic in the trees $\mathcal{T}_n^\bullet$, with translation lengths converging to zero. In any case, we can ensure that the following are satisfied:

- (a′) For all $f\in F$ and $1\leq i\leq s$, either f fixes α_i pointwise, or α_i is contained in the axis of f in $\mathcal{T}_n^{u_i}$ and has length $>4r\cdot\ell(f,\mathcal{T}_n^{u_i})$; the same holds for $f\in F\setminus\{g_j\}$ and β_j.
- (b′) $\beta_j(n)$ is contained in the axis of g_j in $\mathcal{T}_n^{v_j}$ and it has length $>4r\cdot\ell(g_j,\mathcal{T}_n^{v_j})$.
- (c′) The nearest-point projection of $\pi_n^{u_i}(x_n^{\epsilon,\zeta})$ to $\alpha_i(n)$ is $\alpha_i^{\epsilon_i}(n)$, and the nearest-point projection of $\pi_n^{v_j}(x_n^{\epsilon,\zeta})$ to $\beta_j(n)$ is $\beta_j^{\zeta_j}(n)$.

Fix a value of n such that the above are satisfied. Condition (c′) implies that the subsets of $\mathscr{W}(Y_n)$ corresponding to the arcs $\alpha_i(n)$ and $\beta_j(n)$ are all transverse to each other. Hence:

- *Each $f\in F$ fixes all arcs $\alpha_i(n),\beta_j(n)$ except at most one.* Otherwise, conditions (a′) and (b′) would yield $\mathfrak{w},\mathfrak{w}'\in\mathcal{W}_1(f,Y_n)$ such that $\{\mathfrak{w},f^{4r}\mathfrak{w}\}$ and $\{\mathfrak{w}',f^{4r}\mathfrak{w}'\}$ are transverse. Along with Lemma 3.10, this would contradict label-irreducibility of $\rho_n(f)$ (Remark 7.12).
- *Each of the arcs $\alpha_i(n),\beta_j(n)$ is fixed by all elements of F except at most one.* Indeed, if neither of $f_1,f_2\in F$ fixed a given arc, then the same conditions would yield $\mathfrak{w}_1,\mathfrak{w}_2$ such that $\{\mathfrak{w}_1,f_1^{4r}\mathfrak{w}_1,\mathfrak{w}_2,f_2^{4r}\mathfrak{w}_2\}\subseteq\mathcal{W}_1(f_1,Y_n)\cap\mathcal{W}_1(f_2,Y_n)$. Since $\rho_n(f_1)$ and $\rho_n(f_2)$ are label-irreducible, Lemma 3.13 would then imply that $\langle f_1,f_2\rangle\simeq\mathbb{Z}$, contradicting our assumptions.

In conclusion, up to reordering, there exists $0\leq\sigma\leq s$ such that, for $1\leq i\leq\sigma$, the arcs $\alpha_i(n)$ are fixed by the whole F, while, for $\sigma<i\leq s$, there exists $f_i\in F$ such that f_i contains $\alpha_i(n)$ in its axis and $\alpha_i(n)$ is fixed by $F\setminus\{f_i\}$. We obtain a (σ,τ,F)–cubical configuration of width $\geq 4r$ in Y_n, where the static sets are given by the hyperplanes of Y_n originating from the arcs $\alpha_i(n)$ with $i\leq\sigma$, while the skewering elements are $f_{\sigma+1},\dots,f_s$ and $g_1,\dots,g_t$ (thus, $\tau=s+t-\sigma$).

This immediately translates into a $(\sigma,\tau,\varphi_n(F))$–cubical configuration of width $\geq 4r$ in Y. The fact that the static sets can be taken with arbitrarily large cardinality is also immediate, recalling that, since φ_n is not ω–constant, the scaling factors $\overline{\tau}_S^{\rho_n}$ diverge; cf Assumption 7.1 above. □

The next result only requires the material in Section 3.3 for its proof. However, it is best stated in terms of cubical configurations, as defined above.

Recall that q is the number of orbits of vertices for the action $G\curvearrowright Y$.

Lemma 7.19 *Let $F\subseteq G$ be a finite set of label-irreducible elements. Suppose that there is an (s,t,F)–cubical configuration of width $\geq 4r$ in Y, where all the static sets are chains, each containing $\geq q$ hyperplanes. Then the centraliser $Z_G(F)$ contains a copy of $\mathbb{Z}^k$ with $k=s+t$.*

Proof Let the cubical configuration consist of static sets $\mathcal{U}_1, \dots, \mathcal{U}_s$, hyperplanes $\mathfrak{v}_1, \dots, \mathfrak{v}_t$ and the partition $F = F_0 \sqcup \{g_1, \dots, g_t\}$.

Form a set $\mathcal{U}_i'$ by adding to $\mathcal{U}_i$ all hyperplanes of Y that separate hyperplanes of $\mathcal{U}_i$. This guarantees that there exist vertices $x_i, y_i \in Y$ such that $\mathscr{W}(x_i|y_i) = \mathcal{U}_i'$. The sets $\mathcal{U}_i'$ and $\{\mathfrak{v}_j, g_j^{4r}\mathfrak{v}_j\}$ are still all transverse to each other and the elements of F still fix each $\mathcal{U}_i'$ pointwise.

Since $\#\mathscr{W}(x_i|y_i) \geq \#\mathcal{U}_i \geq q$, any geodesic joining x_i to y_i must contain two points in the same G–orbit. Thus, there exist $z_i \in Y$ and $h_i \in G \setminus \{1\}$ with $\mathscr{W}(z_i|h_i z_i) \subseteq \mathcal{U}_i'$ for all $1 \leq i \leq s$.

Lemma 3.15 shows that the h_i commute pairwise. In addition, for every $f \in F$, we have $\mathscr{W}(z_i|h_i z_i) \subseteq \mathcal{U}_i' \subseteq \mathcal{W}_0(f)$, which is transverse to $\mathcal{W}_1(f)$. Since $\mathcal{W}_1(f)$ contains $\mathscr{W}(z|fz)$ for any $z \in \bar{\mathcal{C}}(f)$, another application of Lemma 3.15 guarantees that h_i and f commute. Finally, for $1 \leq j \leq t$, the hyperplanes $\mathfrak{v}_j$ and $g_j^{4r}\mathfrak{v}_j$ are preserved by all elements of $F \setminus \{g_j\}$. Since g_j is label-irreducible, Corollary 3.14(1) implies that g_j commutes with every element of F.

In conclusion, we have shown that the subgroup generated by $A := \{h_1, \dots, h_s, g_1, \dots, g_t\}$ is abelian and contained in $Z_G(F)$. We are left to show that A is a basis for $\langle A \rangle$.

Observe that $\mathfrak{v}_j$ is preserved by all elements of $A \setminus \{g_j\}$, but lies in $\mathcal{W}_1(g_j)$. Similarly, there exist hyperplanes $\mathfrak{u}_i \in \mathscr{W}(z_i|h_i z_i)$ that are preserved by $A \setminus \{h_i\}$, but lie in $\mathcal{W}_1(h_i)$. If a product $h_1^{m_1} \cdots h_s^{m_s} \cdot g_1^{n_1} \cdots g_t^{n_t}$ represents the identity, then it must preserve all hyperplanes $\mathfrak{u}_i$ and $\mathfrak{v}_j$, which implies that $m_i = 0$ and $n_j = 0$ for all i, j. This concludes the proof. □

7.4 Ultralimits of Salvettis and the WNE property

This subsection is devoted to the proof of Theorems F and I. We keep the exact same setting as the previous subsection:

Assumption 7.20 Let $G \leq \mathcal{A}$ be a convex-cocompact subgroup. Let $Y \subseteq \mathcal{X}$ be a G–invariant, convex subcomplex on which G acts with q orbits of vertices. Set $r = \dim \mathcal{X}$. Denote by d the ℓ^1 metric on $\mathcal{X}$ and Y. Let $[\mu]$ be the induced coarse median structure on G.

Consider a sequence $\varphi_n \in \mathrm{Aut}(G, [\mu])$. Denote by $\rho\colon G \hookrightarrow \mathcal{A}$ the standard inclusion and set $\rho_n = \rho \circ \varphi_n$. Fixing a nonprincipal ultrafilter ω, define $\mathcal{X}_\omega$, $\mathcal{X}_n$, Y_n and $Y_\omega = \lim_\omega Y_n$ as in Section 7.3.

The following result is the coronation of our efforts from Section 3.3 and the previous portion of Section 7. Its second part (with $k = 1$) proves Theorem I, while its first part is the last remaining ingredient in the proof of Theorem F (together with Corollary 6.23).

Theorem 7.21 *Let $F \subseteq G$ be a finite subset and suppose that* ***one*** *of the following holds.*

(1) *Let φ_n not be ω–constant. Let $M \subseteq Y_\omega$ be a G–invariant median subalgebra and consider $\eta \in \mathcal{PD}^G(M)$. There exists a k–cube $C \subseteq M$ such that, for any two distinct points $x, y \in C$,*

$$\eta(x, y) > 4r^2 q \cdot [\tau_F^\eta(x) + \tau_F^\eta(y)].$$

(2) *There exists a (generalised) k–cube $C \subseteq Y^{(0)}$ such that, for any two distinct points $x, y \in C$,*

$$d(x,y) > \left(2r^2q + \tfrac{1}{2}rq \cdot \max\{4r, q\}\right) \cdot [\tau_F^d(x) + \tau_F^d(y)].$$

Then the centraliser $Z_G(F)$ contains a copy of $\mathbb{Z}^k$.

The theorem will follow quickly from Lemma 7.22 below, which constructs a cubical configuration in Y_ω (in case (1)) or directly in Y (in case (2)). Indeed, we can then use Lemma 7.18 to always obtain a cubical configuration in Y, and this yields the required copy of $\mathbb{Z}^k$ in $Z_G(F)$ by Lemma 7.19.

Lemma 7.22 *Consider the setting of Theorem 7.21. There exists an (s,t,F')–cubical configuration of width $\geq 4r$ in Y_ω (in case (1)) or in Y (in case (2)), where $s+t=k$ and $F' \subseteq G$ is a finite subset with $Z_G(F') = Z_G(F)$. All elements of F' are label-irreducible and no two of them generate a cyclic subgroup.*

In addition, in case (2), the static sets are chains of hyperplanes of cardinality $\geq q$.

Proof We prove the lemma simultaneously in the two cases of the theorem. In fact, in this proof it is irrelevant whether the φ_n are ω–constant or not, so we can view case (2) as a special instance of case (1) by taking $\varphi_n \equiv \mathrm{id}_G$, $Y_\omega = Y$, $M = Y^{(0)}$ and $\eta = d$.

Recall that, by Remark 7.8(2) and Lemma 7.14(1), the action $G \curvearrowright M$ is nontransverse and without inversions. We begin by constructing the finite subset $F' \subseteq G$.

Claim 1 *There exists $F' \subseteq G$ such that $Z_G(F') = Z_G(F)$, all elements of F' are label-irreducible and no two of them generate a cyclic subgroup. In addition, $\tau_{F'}^\eta(x) \leq q \cdot \tau_F^\eta(x)$ for all $x \in M$.*

Proof Let $a_1, \dots, a_N \in \mathcal{A}$ be a choice of generator for each maximal cyclic subgroup of $\mathcal{A}$ that contains a label-irreducible component of an element of F. Let $m_i \geq 1$ be the smallest integer such that $a_i^{m_i}$ lies in G; by Lemma 3.16, m_i is well-defined and, by Remark 3.17, we have $1 \leq m_i \leq q$. Define $F' := \{a_i^{m_i} \mid 1 \leq i \leq N\}$.

It is clear that every element of F' is label-irreducible and that any two elements of F' generate a noncyclic subgroup. Since all nontrivial powers of any given element of $\mathcal{A}$ have the same centraliser, Lemma 3.7(3) implies that $Z_G(F') = Z_G(F)$.

For every a_i, there exist $n \geq 1$ and $f \in F$ such that a_i^n is a label-irreducible component of f. Thus $a_i^{nm_i}$ is a label-irreducible component of f^{m_i}. Applying Lemma 7.16, it follows that

$$\eta(x, a_i^{m_i}x) \leq \eta(x, a_i^{nm_i}x) \leq \eta(x, f^{m_i}x) \leq m_i \cdot \eta(x, fx) \leq q \cdot \eta(x, fx) \leq q \cdot \tau_F^\eta(x).$$

Hence $\tau_{F'}^\eta(x) \leq q \cdot \tau_F^\eta(x)$ for all $x \in M$, as required. $\triangleleft$

Now, consider the multibridge $\mathcal{B}(F') \subseteq M$ introduced in Definition 6.8. Pick any fibre $P = \mathcal{B}_{/\!/}(F') \times \{*\}$. Let $\pi_P : M \to P$ be the gate-projection.

Claim 2 *The set $C' := \pi_P(C)$ is again a k–cube and, for all distinct points $x', y' \in C'$, we have $\eta(x', y') \geq 4r^2 \cdot \overline{\tau}^{\eta}_{F'}$. Under the assumptions of case* (2), *we further have $\eta(x', y') \geq rq \cdot \overline{\tau}^{\eta}_{F'}$.*

Proof If $x, y \in C$ are distinct, note that we have $\eta(x, y) > 4r^2 q \cdot [\tau^{\eta}_F(x) + \tau^{\eta}_F(y)]$ under the assumptions of both case (1) and case (2). Also recall that, by Proposition 6.11(6), we have $\eta(x, \pi_P(x)) \leq 2r^2 \cdot \tau^{\eta}_{F'}(x)$ for all $x \in M$. Combining these inequalities with Claim 1, we obtain

$$\begin{aligned}\eta(\pi_P(x), \pi_P(y)) &\geq \eta(x, y) - 2r^2 \cdot [\tau^{\eta}_{F'}(x) + \tau^{\eta}_{F'}(y)] \\ &> 4r^2 q \cdot [\tau^{\eta}_F(x) + \tau^{\eta}_F(y)] - 2r^2 \cdot [\tau^{\eta}_{F'}(x) + \tau^{\eta}_{F'}(y)] \\ &\geq 2r^2 \cdot [\tau^{\eta}_{F'}(x) + \tau^{\eta}_{F'}(y)] \geq 4r^2 \cdot \overline{\tau}^{\eta}_{F'}.\end{aligned}$$

In particular, distinct points of C project to distinct points of C', which guarantees that C' is again a k–cube. Moreover, $\eta(x', y') \geq 4r^2 \cdot \overline{\tau}^{\eta}_{F'}$ whenever x', y' are distinct points of C'.

Under the assumptions of case (2), we also have $\eta(x, y) > (2r^2 q + rq^2/2) \cdot [\tau^{\eta}_F(x) + \tau^{\eta}_F(y)]$ for all $x, y \in C$. Using this instead of $\eta(x, y) > 4r^2 q \cdot [\tau^{\eta}_F(x) + \tau^{\eta}_F(y)]$ in the above chain of inequalities, we obtain $\eta(\pi_P(x), \pi_P(y)) \geq rq \cdot \overline{\tau}^{\eta}_{F'}$, as required. ◁

Let $\{C'_{i,-}, C'_{i,+}\}$ be the k pairs of opposite codimension-1 faces of C'. Setting $\mathcal{H}_i := \mathscr{H}(C'_{i,-} | C'_{i,+})$, we obtain sets of halfspaces $\mathcal{H}_1, \dots, \mathcal{H}_k \subseteq \mathscr{H}(M)$ that are transverse to each other. If ν_η is the measure introduced in Remark 2.9, we have $\nu_\eta(\mathcal{H}_i) > 4r^2 \cdot \overline{\tau}^{\eta}_{F'}$ by Claim 2.

The set $\mathcal{H}_i$ can be partitioned into at most r measurable subsets such that no two halfspaces in the same subset are transverse; this follows from Corollary A.3, proved in the appendix (note that $\mathcal{D}(M) \neq \varnothing$ since $\mathcal{D}(\mathcal{X}_\omega) \neq \varnothing$, even though η is just a pseudometric). Define $\mathcal{H}'_i \subseteq \mathcal{H}_i$ as the subset with the largest measure among those in this partition. No two halfspaces in $\mathcal{H}'_i$ are transverse, and

$$\nu_\eta(\mathcal{H}'_i) \geq \frac{1}{r} \cdot \nu_\eta(\mathcal{H}_i) > 4r \cdot \overline{\tau}^{\eta}_{F'}.$$

Let $\mathcal{U}'_i \subseteq \mathscr{W}(M)$ be the set of walls associated to $\mathcal{H}'_i$. Recall that

$$\mathcal{U}'_i \subseteq \mathscr{W}_{C'}(M) \subseteq \mathscr{W}_P(M) \subseteq \bigcap_{f \in F'} \mathscr{W}_{\overline{C}(f)}(M) = \bigcap_{f \in F'} \big(\mathcal{W}_1(f, M) \sqcup \mathcal{W}_0(f, M)\big).$$

Since the sets $\mathcal{W}_1(f, M)$ and $\mathcal{W}_0(f, M)$ are transverse, while no two walls in $\mathcal{U}'_i$ are transverse, we must have either $\mathcal{U}'_i \subseteq \mathcal{W}_1(f, M)$ or $\mathcal{U}'_i \subseteq \mathcal{W}_0(f, M)$ for every index i and element $f \in F'$. Consider the partitions $F' = F_i \sqcup F_i^{\perp}$ such that $\mathcal{U}'_i \subseteq \mathcal{W}_1(f, M)$ if $f \in F_i$ and $\mathcal{U}'_i \subseteq \mathcal{W}_0(f, M)$ if $f \in F_i^{\perp}$.

Claim 3 *We have $\# F_i \leq 1$ for all $1 \leq i \leq k$, and $F_i \cap F_j = \varnothing$ for $i \neq j$.*

Proof For every $f \in F'$, we have

$$\nu_\eta(\mathcal{H}_i') > 4r \cdot \overline{\tau}^{\eta}_{F'} \geq 4r \cdot \ell(f,\eta) = \ell(f^{4r},\eta).$$

If $\mathcal{U}_i' \subseteq \mathcal{W}_1(f,M)$, it follows that there exists a wall $\mathfrak{w} \in \mathcal{U}_i'$ such that $f^{4r}\mathfrak{w} \subseteq \mathcal{U}_i'$.

Thus, if $f,g \in F_i$, there exist walls $\mathfrak{w}$ and $\mathfrak{w}'$ such that $\{\mathfrak{w}, f^{4r}\mathfrak{w}, \mathfrak{w}', g^{4r}\mathfrak{w}'\} \subseteq \mathcal{W}_1(f,M) \cap \mathcal{W}_1(g,M)$. By Remarks 2.2(1) and 2.18(3), there is an analogous inclusion involving walls of $\mathcal{X}_\omega$, so Lemma 7.13(1) yields $\langle f,g\rangle \simeq \mathbb{Z}$. Since $f,g \in F'$, this means that $f=g$. Hence $\#F_i \leq 1$.

Suppose towards a contradiction that there exists $f \in F_i \cap F_j$. Then there are walls $\mathfrak{w}_i, \mathfrak{w}_j$ such that $\{\mathfrak{w}_i, f^{4r}\mathfrak{w}_i\} \subseteq \mathcal{U}_i'$ and $\{\mathfrak{w}_j, f^{4r}\mathfrak{w}_j\} \subseteq \mathcal{U}_j'$. In particular, the subsets $\{\mathfrak{w}_i, f^{4r}\mathfrak{w}_i\}$ and $\{\mathfrak{w}_j, f^{4r}\mathfrak{w}_j\}$ are transverse to each other and contained in $\mathcal{W}_1(f,M)$. Again, Remarks 2.2(1) and 2.18(3) give walls of $\mathcal{X}_\omega$ with the same properties, which contradicts Lemma 7.13(2). ◁

Up to permuting the sets $\mathcal{U}_i'$, we can assume that there exists an index $0 \leq s \leq k$ such that $F_i = \varnothing$ for $1 \leq i \leq s$, while $F_i = \{f_i\}$ for $s < i \leq k$ and pairwise-distinct elements $f_i \in F'$. This all follows from Claim 3. Each $f \in F'$ preserves every wall in $\mathcal{U}_i'$ except if $i > s$ and $f = f_i$. In addition, the proof of Claim 3 gives walls $\mathfrak{v}_i$ for $i > s$, such that $\{\mathfrak{v}_i, f_i^{4r}\mathfrak{v}_i\} \subseteq \mathcal{U}_i' \subseteq \mathcal{W}_1(f_i,M)$. Also recall that the sets $\mathcal{U}_i'$ are all transverse to each other.

This gives an $(s,k-s,F)$–cubical configuration of width $\geq 4r$ in M, where $\mathcal{U}_1',\dots,\mathcal{U}_s'$ are the static sets and $f_{s+1},\dots,f_k$ are the skewering elements. Since $M \subseteq Y_\omega$, it is straightforward to transfer this to a cubical configuration in Y_ω with the same parameters using Remarks 2.2(1) and 2.18(3). This completes the proof of the lemma in Case (1).

In Case (2), we are left to show that the static sets $\mathcal{U}_1',\dots,\mathcal{U}_s'$ contain at least q hyperplanes each. Recall that $M = Y^{(0)}$ and $\eta = d$, so ν_η is just the counting measure. By Claim 2, we have $\#\mathcal{H}_i' = \nu_\eta(\mathcal{H}_i') \geq (1/r)\nu_\eta(\mathcal{H}_i) \geq q \cdot \overline{\tau}^{\eta}_{F'} \geq q$, which concludes the proof since $\#\mathcal{H}_i' = \#\mathcal{U}_i'$. □

Combining Lemmas 7.18, 7.19 and 7.22, we can finally prove Theorem 7.21.

Proof of Theorem 7.21 Our goal is to construct an (s,t,F'')–cubical configuration of width $\geq 4r$ in Y, where $s+t=k$, the centraliser $Z_G(F'')$ is isomorphic to $Z_G(F)$, all elements of F'' are label-irreducible, and all static sets are chains of hyperplanes of cardinality $\geq q$. If we manage to do this, then Lemma 7.19 guarantees that $Z_G(F) \simeq Z_G(F'')$ contains the required copy of $\mathbb{Z}^k$.

In case (2) of the theorem, a cubical configuration with these properties is provided by Lemma 7.22. In case (1), we first apply Lemma 7.22 to obtain an (s,t,F')–cubical configuration of width $\geq 4r$ in Y_ω, where $Z_G(F') = Z_G(F)$. Then we obtain the required cubical configuration in Y from Lemma 7.18, with $F'' = \varphi_n(F')$ for some n; this is the only place where it is important that φ_n is not ω–constant. Since $Z_G(F'') = \varphi_n(Z_G(F')) \simeq Z_G(F)$, this proves the theorem. □

The following two corollaries collect the key takeaways from Theorem 7.21 that we will need in the rest of the paper.

Corollary 7.23 *Every special group with trivial centre is UNE (Definition 2.36).*

Proof Let G be a special group with trivial centre. Embed G as a convex-cocompact subgroup of a RAAG and apply Theorem 7.21(2), taking $k = 1$ and letting F be an arbitrary generating set for G. This shows that the proper cocompact action $G \curvearrowright Y$ is UNE, hence G is a UNE group. □

Corollary 7.24 *Consider the setting of Assumption 7.20. Suppose that the φ_n are pairwise distinct.*

1. *If $C \subseteq Y_\omega$ is a k–cube and $H \leq G$ fixes C pointwise, then $Z_G(H)$ contains a copy of $\mathbb{Z}^k$.*
2. *Let G have trivial centre. Then, for every G–invariant median subalgebra $M \subseteq Y_\omega$, the action $G \curvearrowright M$ is WNE (in the sense of Definition 6.21).*

Proof By Remark 3.8, it suffices to prove part (1) under the additional assumption that H is finitely generated. So let us suppose that H is generated by a finite set F that fixes the k–cube C. We have observed in Section 7.3 that $\mathcal{D}(\mathcal{X}_\omega)^G \neq \varnothing$. Applying Theorem 7.21(1) to any choice of $\eta \in \mathcal{D}(Y_\omega)^G$, we obtain the required copy of $\mathbb{Z}^k$ inside $Z_G(F) = Z_G(H)$.

Part (2) also follows from Theorem 7.21(1), setting $k = 1$ and letting F generate G. □

The following implies parts (1) and (2) of Theorem F as a special case; parts (3) and (4) are obtained below in Remark 7.27. Note that the essentiality requirement in Theorem 7.25(3) is equivalent to the minimality requirement in Theorem F(2), because of [51, Theorem C].

Recall that we denote by $\pi : \operatorname{Aut} G \to \operatorname{Out} G$ the quotient projection. If G has trivial centre and $A \leq \operatorname{Out} G$ is a subgroup, we have $G \lhd \pi^{-1}(A)$ and $\pi^{-1}(A)/G \simeq A$.

Theorem 7.25 *Let $G \leq \mathcal{A}$ be a convex-cocompact subgroup with trivial centre. Let $[\mu]$ be the induced coarse median structure on G. Let $A \leq \operatorname{Out}(G, [\mu])$ be an infinite abelian subgroup. Then there exists an action $\pi^{-1}(A) \curvearrowright X$ with the following properties:*

1. *X is a geodesic median space X with $\operatorname{rk} X \leq r$.*
2. *$\pi^{-1}(A) \curvearrowright X$ is an action by homotheties.*
3. *The restriction $G \curvearrowright X$ is isometric, essential and with unbounded orbits.*
4. *If $C \subseteq X$ is a k–cube and $H \leq G$ fixes C pointwise, then $Z_G(H)$ contains a copy of $\mathbb{Z}^k$.*

Proof Consider a sequence of pairwise distinct automorphisms $\varphi_n \in A$ and set $\rho_n = \rho \circ \varphi_n$. Choose a finite generating set $S \subseteq G$ and consider the action $G \curvearrowright Y_\omega$ as in Section 7.3.

Corollary 7.23 shows that G is UNE. Thus, denoting by $\operatorname{Aut} Y_\omega$ the group of automorphisms of the underlying median algebra, Proposition 7.3 yields a homomorphism $\zeta : \pi^{-1}(A) \to \operatorname{Aut} Y_\omega$ that extends the isometric action $G \curvearrowright Y_\omega$.

By Corollary 7.24(2), the action $G \curvearrowright Y_\omega$ is WNE. Thus, Corollary 6.23 yields a nonempty, countable, $\pi^{-1}(A)$–invariant, median subalgebra $\mathfrak{M} \subseteq Y_\omega$, and a pseudometric $\eta \in \mathcal{PD}^G(\mathfrak{M}) \setminus \{0\}$ for which $\overline{\tau}_S^\eta > 0$ and $\pi^{-1}(A) \curvearrowright (\mathfrak{M}, \eta)$ is homothetic.

Let $(\mathfrak{M}_\circ, \delta)$ be the quotient median space obtained by identifying points $x, y \in \mathfrak{M}$ with $\eta(x,y) = 0$. By Remark 2.1, we have $\operatorname{rk} \mathfrak{M}_\circ \leq \operatorname{rk} \mathfrak{M} \leq \operatorname{rk} X_\omega \leq r$. Since $\overline{\tau}_S^\delta = \overline{\tau}_S^\eta > 0$, the action $G \curvearrowright \mathfrak{M}_\circ$ does not have a global fixed point. Moreover, since the action $G \curvearrowright \mathfrak{M}$ has no wall inversions by Lemma 7.14(1), the action $G \curvearrowright \mathfrak{M}_\circ$ also has no inversions. Theorem 2.14(2) then guarantees that G acts on $\mathfrak{M}_\circ$ with unbounded orbits.

Theorem 7.21(1) (applied to the pseudometric η on $\mathfrak{M}$) shows that $G \curvearrowright \mathfrak{M}_\circ$ satisfies part (4). Thus, we are only left to ensure that the median space is geodesic and the action essential.

In order to make our space geodesic, note that the homothetic $\pi^{-1}(A)$–action extends to the metric completion $\overline{\mathfrak{M}}_\circ$ of $\mathfrak{M}_\circ$. This is a median space of rank $\leq r$ by [29, Proposition 2.21] and [50, Lemma 2.5]. Note that $G \curvearrowright \overline{\mathfrak{M}}_\circ$ still satisfies part (4) because of Theorem 7.21(1). Now, "filling in cubes" as in [48, Corollary 2.16], the space $\overline{\mathfrak{M}}_\circ$ embeds into a complete, connected median space Z of the same rank. By [17, Lemma 4.6], the space Z is geodesic. The isometric G–action extends to Z and one can similarly check that so does the homothetic $\pi^{-1}(A)$–action.

Summing up, we have constructed an action $\pi^{-1}(A) \curvearrowright Z$ that satisfies conditions (1)–(4), possibly except essentiality of the G–action (in addition, Z is complete). By Theorem 2.14(4), there exists a $\pi^{-1}(A)$–invariant, nonempty, convex subset $K \subseteq Z$ and a $\pi^{-1}(A)$–invariant splitting $K = K_0 \times K_1$ such that the action $G \curvearrowright K_1$ is essential. We conclude by taking $X = K_1$. (Note that K is not closed in Z in general, so we may have lost completeness along the way.) □

Remark 7.26 In Theorem 7.25, we cannot both require the space X to be complete and the action $G \curvearrowright X$ to be essential. There is a very good reason for this.

Consider the special case where G is hyperbolic. Then Y_ω is an $\mathbb{R}$–tree, which forces X to also be an $\mathbb{R}$–tree. Note that an isometric action on an $\mathbb{R}$–tree is essential if and only if it is minimal.

Let us show that, if G is a finitely generated group and $G \curvearrowright T$ is a minimal action on a complete $\mathbb{R}$–tree not isometric to $\mathbb{R}$, then no homothety $\Phi\colon T \to T$ with factor $\lambda \neq 1$ can normalise G.

If G is generated by $s_1, \dots, s_k$ and $x \in T$ is any point, the union of all segments $g[x, s_i x]$ with $g \in G$ is a G–invariant subtree. Since $G \curvearrowright T$ is minimal, T must be covered by the segments $g[x, s_i x]$. In particular, the action $G \curvearrowright T$ is cocompact. If Φ normalised G, then every orbit of $G \curvearrowright T$ would be dense; see eg [88, Proposition 3.10]. Since $T \not\simeq \mathbb{R}$, this implies that each segment $g[x, s_i x]$ is nowhere-dense. This violates Baire's theorem, since a complete metric space cannot be covered by countably many nowhere-dense subsets.

I learned this argument from [54, Example II.6].

The following proves parts (3) and (4) of Theorem F.

Remark 7.27 Consider the special case of Theorem 7.25 with $A=\mathbb{Z}$, generated by an outer automorphism $\phi\in\mathrm{Out}(G,[\mu])$. Picking a representative $\varphi\in\mathrm{Aut}(G,[\mu])$, we have $\pi^{-1}(A)=G\rtimes_\varphi\mathbb{Z}$. The theorem gives an isometric action $G\curvearrowright X$ and a homothety $H\colon X\to X$ of factor λ such that $H\circ g=\varphi(g)\circ H$ for all $g\in G$. We keep the notation of the proof of Theorem 7.25.

(1) Each $g\in\mathrm{Fix}\,\varphi$ is elliptic in X. Indeed, Lemma 7.9 shows that g is elliptic in $\mathcal{X}_\omega$, since $\ell(\varphi^n(g),\mathcal{X})$ does not diverge. Lemma 7.14(2) then implies that g is elliptic in $\mathfrak{M}$, and it is clear that a fixed point in $\mathfrak{M}$ will translate into a fixed point in X.

Recalling that $\mathrm{Fix}\,\varphi$ is finitely generated by Theorem B, Theorem 2.14(2) actually implies that $\mathrm{Fix}\,\varphi$ has a global fixed point $x_0\in X$. This proves Theorem F(3).

(2) Fix a finite generating set $S\subseteq G$. Recall from Section 2.1, that we denote conjugacy length by $\|\cdot\|_S$. Let $\Lambda(\varphi)$ be the maximal exponential growth rate of the quantity $\|\varphi^n(g)\|_S^{1/n}$,

$$\Lambda(\varphi):=\sup_{g\in G}\limsup_{n\to+\infty}\|\varphi^n(g)\|_S^{1/n}.$$

Note that $\Lambda(\varphi)$ is independent of the generating set S. For every $g\in G$, we have

$$\lambda^n\ell(g,X)=\ell(H^ngH^{-n},X)=\ell(\varphi^n(g),X)\le\|\varphi^n(g)\|_S\cdot\bar{\tau}_S^X,$$

where the last inequality follows from the identities in Section 2.1. Since there exist elements $g\in G$ with $\ell(g,X)>0$, we deduce that $\lambda\le\Lambda(\varphi)$ and, similarly, $\lambda^{-1}\le\Lambda(\varphi^{-1})$.

If φ has *subexponential growth* (in the sense that $\Lambda(\varphi)=\Lambda(\varphi^{-1})=1$), then these inequalities force $\lambda=1$. Hence the homothetic action $G\rtimes_\varphi\mathbb{Z}\curvearrowright X$ provided by Theorem 7.25 is actually isometric, which proves Theorem F(4).

Appendix Measurable partitions of halfspace-intervals

This appendix is devoted to the proof of Corollary A.3 below. This is needed in the proof of Theorem 7.21 in order to get the exact constant $4r^2q$, and could be avoided if we contented ourselves with the worse bound $4rq\cdot\#\Gamma^{(0)}$. However, Corollary A.3 is important in the general theory of median spaces and we think it is likely to prove useful elsewhere.

Let M be a median algebra. Given a subset $P\subseteq M\times M$, let us write $\mathcal{H}_P:=\bigcup_{(x,y)\in P}\mathscr{H}(x|y)$.

Lemma A.1 *Every subset $P\subseteq[0,1]^n\times[0,1]^n$ contains a countable subset $\Delta\subseteq P$ with $\mathcal{H}_\Delta=\mathcal{H}_P$.*

Proof First, we prove the case $n=1$. We can assume that $x<y$ for every $(x,y)\in P$.

Let $\Omega(P)\subseteq[0,1]$ be the union of the closed arcs $I(x,y)$ with $(x,y)\in P$. Let $\mathcal{D}(P)$ be the set of points that lie in the interior of $\Omega(P)$, but not in the interior of any arc $I(x,y)$ with $(x,y)\in P$. Thus each point of $\Omega(P)$ lies either in the frontier of $\Omega(P)$, or in the interior of some $I(x,y)$, or in the set $\mathcal{D}(P)$, and these three possibilities are disjoint. There is a unique partition of $\Omega(P)$ into maximal segments J_i (closed, open, or half-open) such that

- the interior of J_i does not intersect $\mathcal{D}(P)$, and
- if J_i intersects the interior of $I(x,y)$ for some $(x,y) \in P$, then $I(x,y) \subseteq J_i$.

Observe that $\mathcal{H}_P = \bigsqcup_i \mathscr{H}_{J_i}([0,1]) \cap \mathscr{H}(0|1)$.

It is classical to see that there exists a countable subset $\Delta \subseteq P$ with $\Omega(\Delta) = \Omega(P)$. Note that $\mathcal{D}(\Delta)$ is countable and it contains $\mathcal{D}(P)$. Adding to Δ countably many pairs in P, we can thus ensure that $\mathcal{D}(\Delta) = \mathcal{D}(P)$. Hence, P and Δ determine the same the segments J_i, and $\mathcal{H}_P = \mathcal{H}_\Delta$.

Now consider a general $n \geq 1$. Let $I_i \subseteq [0,1]^n$ be the segment where all coordinates but the i^{th} vanish. Let $\pi_i \colon [0,1]^n \to I_i$ be the coordinate projections. Setting $P_i := (\pi_i \times \pi_i)(P) \subseteq [0,1]^n \times [0,1]^n$, we have $\mathcal{H}_P = \bigcup_i \mathcal{H}_{P_i}$. By the case $n = 1$, there exist countable subsets $\Delta_i \subseteq P_i$ with $\mathcal{H}_{\Delta_i} = \mathcal{H}_{P_i}$. Choosing countable sets $\Delta_i' \subseteq P$ with $(\pi_i \times \pi_i)(\Delta_i') = \Delta_i$, we have $\mathcal{H}_{\Delta_i} \subseteq \mathcal{H}_{\Delta_i'} \subseteq \mathcal{H}_P$. Hence, taking $\Delta := \bigcup_i \Delta_i'$, we obtain $\mathcal{H}_P = \mathcal{H}_\Delta$. □

Recall that $\mathscr{B}(M)$ is the σ–algebra generated by halfspace-intervals, as in Remark 2.9.

Lemma A.2 *Suppose that $M \subseteq [0,1]^n$ is a median subalgebra containing the points $\underline{0} = (0,\dots,0)$ and $\underline{1} = (1,\dots,1)$. Let $\pi_i \colon M \to [0,1]$ denote the coordinate projections. Then the induced maps $\pi_i^* \colon \mathscr{H}([0,1]) \to \mathscr{H}(M)$ (as in Remark 2.1) map $\mathscr{B}$–measurable sets to $\mathscr{B}$–measurable sets.*

Proof Since π_i^* is injective, we have

$$\pi_i^*(\mathscr{H}([0,1]) \setminus E) = \pi_i^*(\mathscr{H}(0|1)) \cup \pi_i^*(\mathscr{H}(1|0)) \setminus \pi_i^*(E)$$

for every $E \subseteq \mathscr{H}([0,1])$. Thus, it suffices to show that, for all $0 \leq a < b \leq 1$, the set $\pi_i^*\mathscr{H}(a|b)$ is $\mathscr{B}$–measurable.

Let a' and b' be, respectively, the infimum and the maximum of $\pi_i(M) \cap [a,b]$. Pick sequences of elements $a' \leq a_{n+1} < a_n < b_n < b_{n+1} \leq b'$ so that $a_n, b_n \in \pi_i(M)$ and $a_n \to a'$, $b_n \to b'$. These sequences can be empty if $\pi_i(M) \cap [a,b] = \varnothing$, or consist of single elements if $a', b' \in \pi_i(M)$. Then

$$\pi_i^*\mathscr{H}(a|b) = \bigcup \pi_i^*\mathscr{H}(a_n|b_n) \cup \{\pi_i^{-1}((a,1])\} \cup \{\pi_i^{-1}([b,1])\}.$$

Observing that singletons are $\mathscr{B}$–measurable, it suffices to show that, for every $x, y \in M$, the set $\pi_i^*\mathscr{H}(\pi_i(x)|\pi_i(y))$ is $\mathscr{B}$–measurable.

This means that it actually suffices to prove that the sets $\pi_i^*\mathscr{H}(0|1)$ are $\mathscr{B}$–measurable. We will achieve this by showing that each set $\mathscr{H}(M) \setminus \pi_i^*\mathscr{H}(0|1)$ is a countable union of halfspace-intervals.

Note that $\mathfrak{h} \in \mathscr{H}(M)$ lies in $\pi_i^*\mathscr{H}([0,1])$ if and only if the projections $\pi_i(\mathfrak{h})$ and $\pi_i(\mathfrak{h}^*)$ are disjoint. Thus, $\mathfrak{h}$ lies in $\mathscr{H}(M) \setminus \pi_i^*\mathscr{H}(0|1)$ if and only if there exist $x, y \in M$ such that $\mathfrak{h} \in \mathscr{H}(x|y)$ and $\pi_i(x) \geq \pi_i(y)$. This gives a subset $P \subseteq M \times M$ with $\mathscr{H}(M) \setminus \pi_i^*\mathscr{H}(0|1) = \mathcal{H}_P$.

In view of Lemma A.1 and Remark 2.2(1), there exists a countable subset $\Delta \subseteq P$ with $\mathcal{H}_\Delta = \mathcal{H}_P$. This concludes the proof. □

The following would be an immediate consequence of Dilworth's lemma, were it not for the measurability requirement.

Corollary A.3 *Let X be a median space of finite rank r. For all $x, y \in X$, there exists a $\mathscr{B}$–measurable partition $\mathscr{H}(x|y) = \mathcal{H}_1 \sqcup \cdots \sqcup \mathcal{H}_r$ such that no two halfspaces in the same $\mathcal{H}_i$ are transverse.*

Proof Taking the metric completion of X and applying [50, Proposition 2.19], we obtain an isometric embedding $\iota\colon I(x,y) \hookrightarrow \mathbb{R}^r$. The image of ι is contained in a product $J_1 \times \cdots \times J_r$ of compact intervals $J_i \subseteq \mathbb{R}$, which is isomorphic to the median algebra $[0,1]^r$. Let $\pi_i\colon M \to J_i$ be the composition of ι with the projection to J_i, and set $\mathcal{H}'_i := \mathscr{H}(x|y) \cap \pi_i^*(\mathscr{H}(J_i))$. We have $\mathscr{H}(x|y) = \mathcal{H}'_1 \cup \cdots \cup \mathcal{H}'_r$, no two halfspaces in the same $\mathcal{H}'_i$ are transverse, and each $\mathcal{H}'_i$ is $\mathscr{B}$–measurable by Lemma A.2. We conclude by taking $\mathcal{H}_i := \mathcal{H}'_i \setminus (\mathcal{H}'_1 \cup \cdots \cup \mathcal{H}'_{i-1})$. □

References

[1] **H-J Bandelt**, **J Hedlíková**, *Median algebras*, Discrete Math. 45 (1983) 1–30 MR Zbl

[2] **B Beeker**, **N Lazarovich**, *Resolutions of* CAT(0) *cube complexes and accessibility properties*, Algebr. Geom. Topol. 16 (2016) 2045–2065 MR Zbl

[3] **B Beeker**, **N Lazarovich**, *Stallings' folds for cube complexes*, Israel J. Math. 227 (2018) 331–363 MR Zbl

[4] **J Behrstock**, **R Charney**, *Divergence and quasimorphisms of right-angled Artin groups*, Math. Ann. 352 (2012) 339–356 MR Zbl

[5] **J Behrstock**, **M F Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups*, Geom. Topol. 21 (2017) 1731–1804 MR Zbl

[6] **J Behrstock**, **M Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, II: Combination theorems and the distance formula*, Pacific J. Math. 299 (2019) 257–338 MR Zbl

[7] **M Bestvina**, *Degenerations of the hyperbolic space*, Duke Math. J. 56 (1988) 143–161 MR Zbl

[8] **M Bestvina**, **N Brady**, *Morse theory and finiteness properties of groups*, Invent. Math. 129 (1997) 445–470 MR Zbl

[9] **M Bestvina**, **K Bromberg**, **K Fujiwara**, *Proper actions on finite products of quasi-trees*, Ann. H. Lebesgue 4 (2021) 685–709 MR Zbl

[10] **M Bestvina**, **M Feighn**, *Stable actions of groups on real trees*, Invent. Math. 121 (1995) 287–321 MR Zbl

[11] **M Bestvina**, **M Feighn**, **M Handel**, *The Tits alternative for* Out(F_n)*, I: Dynamics of exponentially-growing automorphisms*, Ann. of Math. 151 (2000) 517–623 MR Zbl

[12] **M Bestvina**, **M Handel**, *Train tracks and automorphisms of free groups*, Ann. of Math. 135 (1992) 1–51 MR Zbl

[13] **J Beyrer**, **E Fioravanti**, *Cross-ratios on* CAT(0) *cube complexes and marked length-spectrum rigidity*, J. Lond. Math. Soc. 104 (2021) 1973–2015 MR Zbl

[14] **J Beyrer**, **E Fioravanti**, *Cross ratios and cubulations of hyperbolic groups*, Math. Ann. 384 (2022) 1547–1592 MR Zbl

[15] **B H Bowditch**, *Coarse median spaces and groups*, Pacific J. Math. 261 (2013) 53–93 MR Zbl

[16] **B H Bowditch**, *Embedding median algebras in products of trees*, Geom. Dedicata 170 (2014) 157–176 MR Zbl

[17] **B H Bowditch**, *Some properties of median metric spaces*, Groups Geom. Dyn. 10 (2016) 279–317 MR Zbl

[18] **B H Bowditch**, *Convex hulls in coarse median spaces*, preprint (2018) Available at `https://homepages.warwick.ac.uk/~masgak/papers/hulls-cms.pdf`

[19] **B H Bowditch**, *Large-scale rigidity properties of the mapping class groups*, Pacific J. Math. 293 (2018) 1–73 MR Zbl

[20] **C Bregman**, **R Charney**, **K Vogtmann**, *Outer space for RAAGs*, Duke Math. J. 172 (2023) 1033–1108 MR Zbl

[21] **M R Bridson**, *The rhombic dodecahedron and semisimple actions of* $\mathrm{Aut}(F_n)$ *on* $\mathrm{CAT}(0)$ *spaces*, Fund. Math. 214 (2011) 13–25 MR Zbl

[22] **K-U Bux**, **C Gonzalez**, *The Bestvina–Brady construction revisited: geometric computation of* Σ*–invariants for right-angled Artin groups*, J. London Math. Soc. 60 (1999) 793–801 MR Zbl

[23] **P-E Caprace**, **M Sageev**, *Rank rigidity for* $\mathrm{CAT}(0)$ *cube complexes*, Geom. Funct. Anal. 21 (2011) 851–891 MR Zbl

[24] **R Charney**, **J Crisp**, **K Vogtmann**, *Automorphisms of* 2*–dimensional right-angled Artin groups*, Geom. Topol. 11 (2007) 2227–2264 MR Zbl

[25] **R Charney**, **N Stambaugh**, **K Vogtmann**, *Outer space for untwisted automorphisms of right-angled Artin groups*, Geom. Topol. 21 (2017) 1131–1178 MR Zbl

[26] **R Charney**, **H Sultan**, *Contracting boundaries of* $\mathrm{CAT}(0)$ *spaces*, J. Topol. 8 (2015) 93–117 MR Zbl

[27] **R Charney**, **K Vogtmann**, *Finiteness properties of automorphism groups of right-angled Artin groups*, Bull. Lond. Math. Soc. 41 (2009) 94–102 MR Zbl

[28] **R Charney**, **K Vogtmann**, *Subgroups and quotients of automorphism groups of RAAGs*, from "Low-dimensional and symplectic topology" (M Usher, editor), Proc. Sympos. Pure Math. 82, Amer. Math. Soc., Providence, RI (2011) 9–27 MR Zbl

[29] **I Chatterji**, **C Druţu**, **F Haglund**, *Kazhdan and Haagerup properties from the median viewpoint*, Adv. Math. 225 (2010) 882–921 MR Zbl

[30] **I Chatterji**, **T Fernós**, **A Iozzi**, *The median class and superrigidity of actions on* $\mathrm{CAT}(0)$ *cube complexes*, J. Topol. 9 (2016) 349–400 MR Zbl

[31] **V Chepoi**, *Graphs of some* $\mathrm{CAT}(0)$ *complexes*, Adv. in Appl. Math. 24 (2000) 125–179 MR Zbl

[32] **M M Cohen**, **M Lustig**, *On the dynamics and the fixed subgroup of a free group automorphism*, Invent. Math. 96 (1989) 613–638 MR Zbl

[33] **D Cooper**, *Automorphisms of free groups have finitely generated fixed point sets*, J. Algebra 111 (1987) 453–456 MR Zbl

[34] **M Cordes**, **M G Durham**, *Boundary convex cocompactness and stability of subgroups of finitely generated groups*, Int. Math. Res. Not. 2019 (2019) 1699–1724 MR Zbl

[35] **L J Corredor**, **M A Gutierrez**, *A generating set for the automorphism group of a graph product of abelian groups*, Internat. J. Algebra Comput. 22 (2012) art. id. 1250003 MR Zbl

[36] **M Culler**, *Finite groups of outer automorphisms of a free group*, from "Contributions to group theory" (K I Appel, J G Ratcliffe, P E Schupp, editors), Contemp. Math. 33, Amer. Math. Soc., Providence, RI (1984) 197–207 MR Zbl

[37] **M Culler**, **J W Morgan**, *Group actions on* $\mathbb{R}$*–trees*, Proc. London Math. Soc. 55 (1987) 571–604 MR Zbl

[38] **P Dani**, **I Levcovitz**, *Subgroups of right-angled Coxeter groups via Stallings-like techniques*, J. Comb. Algebra 5 (2021) 237–295 MR Zbl

[39] **M W Davis**, *The geometry and topology of Coxeter groups*, London Mathematical Society Monographs Series 32, Princeton Univ. Press (2008) MR Zbl

[40] **M W Davis**, **T Januszkiewicz**, *Right-angled Artin groups are commensurable with right-angled Coxeter groups*, J. Pure Appl. Algebra 153 (2000) 229–235 MR Zbl

[41] **M B Day**, *Peak reduction and finite presentations for automorphism groups of right-angled Artin groups*, Geom. Topol. 13 (2009) 817–855 MR Zbl

[42] **M B Day**, **A W Sale**, **R D Wade**, *Calculating the virtual cohomological dimension of the automorphism group of a RAAG*, Bull. Lond. Math. Soc. 53 (2021) 259–273 MR Zbl

[43] **M B Day**, **R D Wade**, *Relative automorphism groups of right-angled Artin groups*, J. Topol. 12 (2019) 759–798 MR Zbl

[44] **C Druţu**, **M Kapovich**, *Geometric group theory*, American Mathematical Society Colloquium Publications 63, Amer. Math. Soc., Providence, RI (2018) MR Zbl

[45] **J Dugundji**, *An extension of Tietze's theorem*, Pacific J. Math. 1 (1951) 353–367 MR Zbl

[46] **J L Dyer**, **G P Scott**, *Periodic automorphisms of free groups*, Comm. Algebra 3 (1975) 195–201 MR Zbl

[47] **T Fernós**, *The Furstenberg–Poisson boundary and* $\mathrm{CAT}(0)$ *cube complexes*, Ergodic Theory Dynam. Systems 38 (2018) 2180–2223 MR Zbl

[48] **E Fioravanti**, *The Tits alternative for finite rank median spaces*, Enseign. Math. 64 (2018) 89–126 MR Zbl

[49] **E Fioravanti**, *Superrigidity of actions on finite rank median spaces*, Adv. Math. 352 (2019) 1206–1252 MR Zbl

[50] **E Fioravanti**, *Roller boundaries for median spaces and algebras*, Algebr. Geom. Topol. 20 (2020) 1325–1370 MR Zbl

[51] **E Fioravanti**, *Convex cores for actions on finite-rank median algebras*, preprint (2021) arXiv 2101.03060v1

[52] **E Fioravanti**, *On automorphisms and splittings of special groups*, Compos. Math. 159 (2023) 232–305 MR Zbl

[53] **D Gaboriau**, **A Jaeger**, **G Levitt**, **M Lustig**, *An index for counting fixed points of automorphisms of free groups*, Duke Math. J. 93 (1998) 425–452 MR Zbl

[54] **D Gaboriau**, **G Levitt**, *The rank of actions on* $\mathbb{R}$*–trees*, Ann. Sci. École Norm. Sup. 28 (1995) 549–570 MR Zbl

[55] **D Gaboriau**, **G Levitt**, **M Lustig**, *A dendrological proof of the Scott conjecture for automorphisms of free groups*, Proc. Edinburgh Math. Soc. 41 (1998) 325–332 MR Zbl

[56] **S M Gersten**, *On fixed points of automorphisms of finitely generated free groups*, Bull. Amer. Math. Soc. 8 (1983) 451–454 MR Zbl Addendum in 49 (1984) 340–342

[57] **S M Gersten**, *Fixed points of automorphisms of free groups*, Adv. in Math. 64 (1987) 51–85 MR Zbl

[58] **R Z Goldstein**, **E C Turner**, *Automorphisms of free groups and their fixed points*, Invent. Math. 78 (1984) 1–12 MR Zbl

[59] **D Groves**, **M Hull**, *Abelian splittings of right-angled Artin groups*, from "Hyperbolic geometry and geometric group theory" (K Fujiwara, S Kojima, K Ohshika, editors), Adv. Stud. Pure Math. 73, Math. Soc. Japan, Tokyo (2017) 159–165 MR Zbl

[60] **V Guirardel**, *Cœur et nombre d'intersection pour les actions de groupes sur les arbres*, Ann. Sci. École Norm. Sup. 38 (2005) 847–888 MR Zbl

[61] **V Guirardel**, **A Sale**, *Vastness properties of automorphism groups of RAAGs*, J. Topol. 11 (2018) 30–64 MR Zbl

[62] **M F Hagen**, *Cocompactly cubulated crystallographic groups*, J. Lond. Math. Soc. 90 (2014) 140–166 MR Zbl

[63] **M F Hagen**, **T Susse**, *On hierarchical hyperbolicity of cubical groups*, Israel J. Math. 236 (2020) 45–89 MR Zbl

[64] **M F Hagen**, **N W M Touikan**, *Panel collapse and its applications*, Groups Geom. Dyn. 13 (2019) 1285–1334 MR Zbl

[65] **M F Hagen**, **H Wilton**, *Guirardel cores for cube complexes*, in preparation

[66] **F Haglund**, *Finite index subgroups of graph products*, Geom. Dedicata 135 (2008) 167–209 MR Zbl

[67] **F Haglund**, *Isometries of CAT(0) cube complexes are semi-simple*, Ann. Math. Qué. 47 (2023) 249–261 MR

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

[69] **S Hensel**, **D Kielak**, *Nielsen realisation for untwisted automorphisms of right-angled Artin groups*, Proc. Lond. Math. Soc. 117 (2018) 901–950 MR Zbl

[70] **S-t Hu**, *Theory of retracts*, Wayne State Univ. Press, Detroit, MI (1965) MR Zbl

[71] **J Huang**, **B Kleiner**, *Groups quasi-isometric to right-angled Artin groups*, Duke Math. J. 167 (2018) 537–602 MR Zbl

[72] **W Jaco**, **P B Shalen**, *Surface homeomorphisms and periodicity*, Topology 16 (1977) 347–367 MR Zbl

[73] **T Koberda**, **J Mangahas**, **S J Taylor**, *The geometry of purely loxodromic subgroups of right-angled Artin groups*, Trans. Amer. Math. Soc. 369 (2017) 8179–8208 MR Zbl

[74] **M Kolibiar**, **T Marcisová**, *On a question of J Hashimoto*, Mat. Časopis Sloven. Akad. Vied 24 (1974) 179–185 MR Zbl

[75] **M R Laurence**, *A generating set for the automorphism group of a graph group*, J. London Math. Soc. 52 (1995) 318–334 MR Zbl

[76] **S Lefschetz**, *On locally connected and related sets*, Ann. of Math. 35 (1934) 118–129 MR Zbl

[77] **S Lefschetz**, *On locally connected and related sets (second paper)*, Duke Math. J. 2 (1936) 435–442 MR Zbl

[78] **G Levitt**, *Automorphisms of hyperbolic groups and graphs of groups*, Geom. Dedicata 114 (2005) 49–70 MR Zbl

[79] **J Meier**, **H Meinert**, **L VanWyk**, *Higher generation subgroup sets and the Σ–invariants of graph groups*, Comment. Math. Helv. 73 (1998) 22–44 MR Zbl

[80] **J Meier**, **L VanWyk**, *The Bieri–Neumann–Strebel invariants for graph groups*, Proc. London Math. Soc. 71 (1995) 263–280 MR Zbl

[81] **A Minasyan**, **D Osin**, *Fixed subgroups of automorphisms of relatively hyperbolic groups*, Q. J. Math. 63 (2012) 695–712 MR Zbl

[82] **W D Neumann**, *The fixed group of an automorphism of a word hyperbolic group is rational*, Invent. Math. 110 (1992) 147–150 MR Zbl

[83] **G A Niblo**, **N Wright**, **J Zhang**, *A four point characterisation for coarse median spaces*, Groups Geom. Dyn. 13 (2019) 939–980 MR Zbl

[84] **G A Niblo**, **N Wright**, **J Zhang**, *Coarse median algebras: the intrinsic geometry of coarse median spaces and their intervals*, Selecta Math. 27 (2021) art. id. 20 MR Zbl

[85] **F Paulin**, *Topologie de Gromov équivariante, structures hyperboliques et arbres réels*, Invent. Math. 94 (1988) 53–80 MR Zbl

[86] **F Paulin**, *Points fixes des automorphismes de groupe hyperbolique*, Ann. Inst. Fourier (Grenoble) 39 (1989) 651–662 MR Zbl

[87] **F Paulin**, *Outer automorphisms of hyperbolic groups and small actions on* $\mathbb{R}$*–trees*, from "Arboreal group theory" (R C Alperin, editor), Math. Sci. Res. Inst. Publ. 19, Springer (1991) 331–343 MR Zbl

[88] **F Paulin**, *Sur les automorphismes extérieurs des groupes hyperboliques*, Ann. Sci. École Norm. Sup. 30 (1997) 147–167 MR Zbl

[89] **M Roller**, *Poc sets, median algebras and group actions*, preprint (2016) arXiv 1607.07747

[90] **M Sageev**, CAT(0) *cube complexes and groups*, from "Geometric group theory" (M Bestvina, M Sageev, K Vogtmann, editors), IAS/Park City Math. Ser. 21, Amer. Math. Soc., Providence, RI (2014) 7–54 MR Zbl

[91] **M Sageev**, **D T Wise**, *Cores for quasiconvex actions*, Proc. Amer. Math. Soc. 143 (2015) 2731–2741 MR Zbl

[92] **A Sale**, **T Susse**, *Outer automorphism groups of right-angled Coxeter groups are either large or virtually abelian*, Trans. Amer. Math. Soc. 372 (2019) 7785–7803 MR Zbl

[93] **Z Sela**, *Structure and rigidity in (Gromov) hyperbolic groups and discrete groups in rank* 1 *Lie groups, II*, Geom. Funct. Anal. 7 (1997) 561–593 MR Zbl

[94] **H Servatius**, *Automorphisms of graph groups*, J. Algebra 126 (1989) 34–60 MR Zbl

[95] **J R Stallings**, *Topology of finite graphs*, Invent. Math. 71 (1983) 551–565 MR Zbl

[96] **J R Stallings**, *Foldings of G–trees*, from "Arboreal group theory" (R C Alperin, editor), Math. Sci. Res. Inst. Publ. 19, Springer (1991) 355–368 MR Zbl

[97] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. 19 (1988) 417–431 MR Zbl

[98] **H C Tran**, *Geometric embedding properties of Bestvina–Brady subgroups*, Algebr. Geom. Topol. 17 (2017) 2499–2510 MR Zbl

[99] **M L J van de Vel**, *Theory of convex structures*, North-Holland Mathematical Library 50, North-Holland, Amsterdam (1993) MR Zbl

[100] **D T Wise**, **D J Woodhouse**, *A cubical flat torus theorem and the bounded packing property*, Israel J. Math. 217 (2017) 263–281 MR Zbl

[101] **R Zeidler**, *Coarse median structures and homomorphisms from Kazhdan groups*, Geom. Dedicata 180 (2016) 49–68 MR Zbl

Universität Bonn
Bonn, Germany
elia.fioravanti@kit.edu

Proposed: Mladen Bestvina — Received: 25 February 2021
Seconded: David Fisher, Urs Lang — Revised: 10 March 2022

Geometry & Topology **28**:1 (2024) 267–351
DOI: 10.2140/gt.2024.28.267
Published: 27 February 2024

Classification results for expanding and shrinking gradient Kähler–Ricci solitons

RONAN J CONLON
ALIX DERUELLE
SONG SUN

We first show that a Kähler cone appears as the tangent cone of a complete expanding gradient Kähler–Ricci soliton with quadratic curvature decay with derivatives if and only if it has a smooth canonical model (on which the soliton lives). This allows us to classify two-dimensional complete expanding gradient Kähler–Ricci solitons with quadratic curvature decay with derivatives. We then show that any two-dimensional complete shrinking gradient Kähler–Ricci soliton whose scalar curvature tends to zero at infinity is, up to pullback by an element of $\mathrm{GL}(2,\mathbb{C})$, either the flat Gaussian shrinking soliton on $\mathbb{C}^2$ or the $U(2)$–invariant shrinking gradient Kähler–Ricci soliton of Feldman, Ilmanen and Knopf on the blowup of $\mathbb{C}^2$ at one point. Finally, we show that up to pullback by an element of $\mathrm{GL}(n,\mathbb{C})$, the only complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature on $\mathbb{C}^n$ is the flat Gaussian shrinking soliton, and on the total space of $\mathcal{O}(-k)\to\mathbb{P}^{n-1}$ for $0<k<n$ is the $U(n)$–invariant example of Feldman, Ilmanen and Knopf. In the course of the proof, we establish the uniqueness of the soliton vector field of a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature in the Lie algebra of a torus. A key tool used to achieve this result is the Duistermaat–Heckman theorem from symplectic geometry. This provides the first step towards understanding the relationship between complete shrinking gradient Kähler–Ricci solitons and algebraic geometry.

53C25; 53C55, 53E30

1 Introduction

1.1 Overview

A *Ricci soliton* is a triple (M, g, X), where M is a Riemannian manifold with a complete Riemannian metric g and a complete vector field X satisfying the equation

$$\mathrm{Ric}(g) + \tfrac{1}{2}\mathcal{L}_X g = \tfrac{1}{2}\lambda g \tag{1-1}$$

for some $\lambda \in \{-1, 0, 1\}$. If $X = \nabla^g f$ for some real-valued smooth function f on M, then we say that (M, g, X) is *gradient*. In this case, the soliton equation (1-1) reduces to

$$\mathrm{Ric}(g) + \mathrm{Hess}_g(f) = \tfrac{1}{2}\lambda g. \tag{1-2}$$

If g is complete and Kähler with Kähler form ω, then we say that (M, g, X) (or (M, ω, X)) is a *Kähler–Ricci soliton* if the vector field X is complete and real holomorphic and the pair (g, X) satisfies

$$\mathrm{Ric}(g) + \tfrac{1}{2}\mathcal{L}_X g = \lambda g \tag{1-3}$$

for λ as above. If g is a Kähler–Ricci soliton and if $X = \nabla^g f$ for some real-valued smooth function f on M, then we say that (M, g, X) is *gradient*. In this case, one can rewrite the soliton equation (1-3) as

$$\rho_\omega + i\partial\bar{\partial} f = \lambda\omega, \tag{1-4}$$

where ρ_ω is the Ricci form of ω.

For Ricci solitons and Kähler–Ricci solitons (M, g, X), the vector field X is called the *soliton vector field*. Its completeness is guaranteed by the completeness of g; see Zhang [69]. If the soliton is gradient, then the smooth real-valued function f satisfying $X = \nabla^g f$ is called the *soliton potential*. It is unique up to a constant. Finally, a Ricci soliton and a Kähler–Ricci soliton are called *steady* if $\lambda = 0$, *expanding* if $\lambda = -1$, and *shrinking* if $\lambda = 1$ in equations (1-1) and (1-3), respectively.

The study of Ricci solitons and their classification is important in the context of Riemannian geometry. For example, they provide a natural generalisation of Einstein manifolds. Also, to each soliton, one may associate a self-similar solution of the Ricci flow (see Chow and Knopf [14, Lemma 2.4]), which are candidates for singularity models of the flow. The difference in normalisations between (1-1) and (1-3) reflects the difference between the constants preceding the Ricci term in the Ricci flow and the Kähler–Ricci flow when one takes this dynamic point of view.

In this article, we are concerned with complete expanding and shrinking gradient Kähler–Ricci solitons. We consider primarily complete shrinking (resp. expanding) gradient Kähler–Ricci solitons with quadratic curvature decay (resp. with derivatives) as this assumption greatly simplifies the situation and already imposes some constraints on the solitons in question. Indeed, it is known that any complete shrinking (resp. expanding) gradient Ricci soliton whose curvature decays quadratically (resp. with derivatives) along an end must be asymptotic to a cone with a smooth link along that end; see Chen and Deruelle [12], Chow and Lu [15], Kotschwar and Wang [43] and Siepmann [62]. Furthermore, any complete shrinking (resp. expanding) Ricci soliton whose Ricci curvature decays to zero (resp. quadratically with derivatives) at infinity must have quadratic curvature decay (resp. with derivatives) at infinity (see Deruelle [27] and Munteanu and Wang [53]) and consequently must be asymptotically conical along each of its ends. Kotschwar and Wang [44] have then shown that any two complete shrinking gradient Ricci solitons asymptotic along some end of each to the same cone with a smooth link, must in fact be isometric. Thus, at least for shrinking gradient Ricci solitons, classifying those that are complete with Ricci curvature decaying to zero at infinity reduces to classifying their possible asymptotic cone models. Here we are principally concerned with the classification of complete shrinking (resp. expanding) gradient Kähler–Ricci solitons whose curvature tensor has quadratic decay (resp. with derivatives). Such examples of expanding type have been constructed in Conlon and Deruelle [19] on certain equivariant resolutions of Kähler cones, whereas such examples of shrinking type include the flat Gaussian shrinking soliton on $\mathbb{C}^n$ and those constructed by Feldman, Ilmanen and Knopf [30] on the total space of the holomorphic line bundles $\mathcal{O}(-k)$ over $\mathbb{P}^{n-1}$ for $0 < k < n$. These shrinking solitons are $U(n)$–invariant and in complex dimension two, they yield two known examples of complete shrinking gradient Kähler–Ricci solitons with scalar curvature tending to zero at infinity: the flat Gaussian shrinking soliton on $\mathbb{C}^2$ and the aforementioned $U(2)$–invariant example of Feldman, Ilmanen and Knopf on the blowup of $\mathbb{C}^2$ at the origin. One of our main results is that in fact, up to pullback by an element of $\mathrm{GL}(2,\mathbb{C})$, these are the only two examples of complete shrinking gradient Kähler–Ricci solitons with scalar curvature tending to zero at infinity in two complex dimensions. Other examples of complete (and indeed incomplete) shrinking gradient Kähler–Ricci solitons with quadratic curvature decay have been constructed on the total space of certain holomorphic vector bundles; see for example Dancer and Wang [23], Futaki and Wang [34], Li [47] and Yang [68].

1.2 Main results

1.2.1 General structure theorem Our first result concerns the structure of complete shrinking (respectively expanding) gradient Kähler–Ricci solitons (M, g, X) with quadratic curvature decay (resp. with derivatives). By "quadratic curvature decay with derivatives", we mean that the curvature $\mathrm{Rm}(g)$ of the Kähler–Ricci soliton g satisfies

$$A_k(g) := \sup_{x\in M} |(\nabla^g)^k \operatorname{Rm}(g)|_g(x) d_g(p,x)^{2+k} < \infty \quad \text{for all } k \in \mathbb{N}_0,$$

where $d_g(p,\cdot)$ denotes the distance to a fixed point $p \in M$ with respect to g.

Theorem A (general structure theorem) *Let (M, g, X) be a complete expanding (resp. shrinking) gradient Kähler–Ricci soliton of complex dimension $n \geq 2$ with complex structure J whose curvature $\mathrm{Rm}(g)$ satisfies*

$$A_k(g) := \sup_{x\in M} |(\nabla^g)^k \operatorname{Rm}(g)|_g(x)\, d_g(p,x)^{2+k} < \infty \quad \textit{for all } k \in \mathbb{N}_0 \text{ (resp. for } k = 0),$$

where $d_g(p,\cdot)$ denotes the distance to a fixed point $p \in M$ with respect to g. Then:

(a) *(M, g) has a unique tangent cone at infinity (C_0, g_0).*

(b) *There exists a Kähler resolution* $\pi\colon M \to C_0$ *of* C_0 *with exceptional set* E *with* g_0 *Kähler with respect to* $J_0 := \pi_* J$ *such that*

(i) *the Kähler form* ω *of* g *and the curvature form* Θ *of the hermitian metric on* K_M *(resp.* $-K_M$*) induced by* ω *satisfy*

$$\int_V (i\Theta)^k \wedge \omega^{\dim_{\mathbb{C}} V-k} > 0 \tag{1-5}$$

for all positive-dimensional irreducible analytic subvarieties $V \subset E$ *and for all integers* k *such that* $1 \le k \le \dim_{\mathbb{C}} V$*;*

(ii) *the real torus action on* C_0 *generated by* $J_0 r\partial_r$ *extends to a holomorphic isometric torus action of* (M, g, J)*, where* r *denotes the radial coordinate of* g_0*;*

(iii) $d\pi(X) = r\partial_r$.

(c) *With respect to* π*, we have*

$$|(\nabla^{g_0})^k(\pi_* g - g_0 - \mathrm{Ric}(g_0))|_{g_0} \le C_k r^{-4-k} \quad \textit{for all } k \in \mathbb{N}_0,$$
$$(\textit{resp.}\quad |(\nabla^{g_0})^k(\pi_* g - g_0 + \mathrm{Ric}(g_0))|_{g_0} \le C_k r^{-4-k} \quad \textit{for all } k \in \mathbb{N}_0).$$

In the expanding case, this theorem provides a converse to [19, Theorem A]. Also notice that this theorem rules out the existence of shrinking (resp. expanding) gradient Kähler–Ricci solitons with quadratic curvature decay (resp. with derivatives) on smoothings of Kähler cones in contrast to the behaviour in the Calabi–Yau case [20]. This degree of flexibility for expanding and shrinking gradient Kähler–Ricci solitons is essentially ruled out due to the requirement of having a conical holomorphic soliton vector field. Finally, it has recently been shown by Kotschwar and Wang [44, Corollary 1.3] that the isometry group of the link of the asymptotic cone of a complete shrinking gradient Kähler–Ricci soliton with quadratic curvature decay embeds into the isometry group of the soliton itself; compare with statement (b)(ii) of the theorem above.

1.2.2 Application to expanding gradient Kähler–Ricci solitons As an application of Theorem A, we exploit the uniqueness [38, Proposition 8.2.5] of canonical models of normal varieties to obtain a classification theorem for complete expanding gradient Kähler–Ricci solitons whose curvature decays quadratically with derivatives. This provides a partial answer to question 7 of [30, Section 10].

Corollary B (strong uniqueness for expanders) *Let* (C_0, g_0) *be a Kähler cone of complex dimension* $n \ge 2$ *with radial function* r*. Then there exists a unique (up to pullback by biholomorphisms) complete expanding gradient Kähler–Ricci soliton* (M, g, X)*, whose curvature* $\mathrm{Rm}(g)$ *satisfies*

$$A_k(g) := \sup_{x \in M} |(\nabla^g)^k \mathrm{Rm}(g)|_g(x) d_g(p, x)^{2+k} < \infty \quad \textit{for all } k \in \mathbb{N}_0, \tag{1-6}$$

where $d_g(p, \cdot)$ *denotes the distance to a fixed point* $p \in M$ *with respect to* g*, with tangent cone* (C_0, g_0) *if and only if* C_0 *has a smooth canonical model. When this is the case,*

(a) M *is the smooth canonical model of* C_0*, and*

(b) *there exists a resolution map* $\pi: M \to C_0$ *such that* $d\pi(X) = r\partial_r$ *and*

$$|(\nabla^{g_0})^k(\pi_* g - g_0 - \mathrm{Ric}(g_0))|_{g_0} \le C_k r^{-4-k} \quad \text{for all } k \in \mathbb{N}_0.$$

This corollary yields an algebraic description of those Kähler cones appearing as the tangent cone of a complete expanding gradient Kähler–Ricci soliton with quadratic curvature decay with derivatives — such cones are precisely those admitting a smooth canonical model. In general, the canonical model will be singular and in particular, for a two-dimensional cone, it is obtained by contracting all exceptional curves with self-intersection (-2) in the minimal resolution. Applying Corollary B to this case yields a classification of two-dimensional complete expanding gradient Kähler–Ricci solitons with quadratic curvature decay with derivatives.

Corollary C (classification of two-dimensional expanders) *Let* (C_0, g_0) *be a two-dimensional Kähler cone with radial function* r. *Then there exists a unique (up to pullback by biholomorphisms) two-dimensional complete expanding gradient Kähler–Ricci soliton* (M, g, X) *whose curvature* $\mathrm{Rm}(g)$ *satisfies*

$$A_k(g) := \sup_{x \in M} |(\nabla^g)^k \mathrm{Rm}(g)|_g(x) d_g(p, x)^{2+k} < \infty \quad \text{for all } k \in \mathbb{N}_0,$$

where $d_g(p, \cdot)$ *denotes the distance to a fixed point* $p \in M$ *with respect to* g, *with tangent cone* (C_0, g_0) *if and only if* C_0 *is biholomorphic to one of*:

(I) $\mathbb{C}^2/\Gamma$, *where* Γ *is a finite subgroup of* $U(2)$ *acting freely on* $\mathbb{C}^2 \setminus \{0\}$ *that is generated by the matrix*

$$\begin{pmatrix} e^{2\pi i/p} & 0 \\ 0 & e^{2\pi i q/p} \end{pmatrix},$$

where p *and* q *coprime integers with* $p > q > 0$, *and after writing*

$$\frac{q}{p} = r_1 - \cfrac{1}{r_2 - \cfrac{1}{\cdots - \cfrac{1}{r_k}}}$$

we have that $r_j > 2$ *for* $j = 1, \dots, k$.

(II) $L^\times$, *the blowdown of the zero section of a negative line bundle* $L \to C$ *over a proper curve* C *of genus* $g > 0$.

(III) $L^\times/G$, *where* G *is a nontrivial finite group of automorphisms of a proper curve* C *of genus* $g > 0$ *and* $L^\times$ *is the blowdown of the zero section of a* G*–invariant negative line bundle* $L \to C$ *over* C *with* G *acting freely on* $L^\times$ *except at the apex, such that the (unique) minimal good resolution* $\pi: M \to L^\times/G$ *contains no* (-1)*– or* (-2)*–curves.*

When this is the case,

(a) *there exists a resolution map* $\pi : M \to C_0$ *such that* $d\pi(X) = r\partial_r$ *and*

$$|(\nabla^{g_0})^k(\pi_* g - g_0 - \mathrm{Ric}(g_0))|_{g_0} \leq C_k r^{-4-k} \quad \textit{for all } k \in \mathbb{N}_0,$$

and

(b) $\pi : M \to C_0$ *is*

 (i) *the minimal resolution* $\pi : M \to \mathbb{C}^2/\Gamma$ *when* C_0 *is as in* (I);

 (ii) *the blowdown map* $\pi : L \to L^\times$ *when* C_0 *is as in* (II);

 (iii) *the minimal good resolution* $\pi : M \to L^\times/G$ *when* C_0 *is as in* (III).

Note that the Kähler cones of item (III) here are (up to analytic isomorphism) in one-to-one correspondence with the following data, which encodes the exceptional set of the minimal good resolution $\pi : M \to L^\times/G$:

- A weighted dual graph which is a star comprising a central vertex with n branches, with $n \geq 1$, each of finite length, with the j^{th} vertex of the i^{th} branch labelled with an integer $b_{ij} \geq 3$ and the central curve labelled with an integer $b \geq 1$. The central vertex represents the curve C of genus $g > 0$ with self-intersection $-b$ and the j^{th} vertex of the i^{th} branch represents a $\mathbb{P}^1$ with self-intersection $-b_{ij}$. The intersection matrix given by this graph must be negative-definite. There is a numerical criterion to determine when this is the case.
- The analytic type of C.
- A negative line bundle on C of degree $-b$ (which is the normal bundle of C in M).
- n marked points on C.

The algebraic equations of C_0 can be reconstructed from this data by the ansatz in [59, Section 5].

1.2.3 Application to shrinking gradient Kähler–Ricci solitons Tian and Zhu [64] showed that the soliton vector field of a compact shrinking gradient Kähler–Ricci soliton is unique up to holomorphic automorphisms of the underlying complex manifold. The method of proof there involved defining a weighted volume functional F which was strictly convex and was shown to be in fact independent of the metric structure of the soliton. The soliton vector field was then characterised as the unique critical point of this functional.

Tian and Zhu's proof breaks down in the noncompact case due to the fact that, in general, one cannot a priori guarantee that the weighted volume functional (defined analytically in terms of a certain integral) is well-defined on a noncompact Kähler manifold, let alone investigate its convexity properties. Our key observation to circumvent this difficulty is to apply the *Duistermaat–Heckman theorem* from symplectic geometry, which provides a localisation formula to express the weighted volume functional in terms of an algebraic formula involving only the fixed point set of a torus action. This latter algebraic formula is much more amenable. Combined with a version of Matsushima's theorem [50] for complete noncompact shrinking gradient Kähler–Ricci solitons (cf Theorem 5.1) and Iwasawa's theorem [39], we are able to implement Tian and Zhu's strategy of proof for the compact case to obtain the following noncompact analogue of their uniqueness result.

Theorem D (uniqueness of the soliton vector field for shrinkers) *Let M be a noncompact complex manifold with complex structure J endowed with the effective holomorphic action of a real torus T. Denote by $\mathfrak{t}$ the Lie algebra of T. Then there exists at most one element $\xi \in \mathfrak{t}$ that admits a complete shrinking gradient Kähler–Ricci soliton (M, g, X) with bounded Ricci curvature with $X = \nabla^g f = -J\xi$ for a smooth real-valued function f on M.*

We expect that the assumption of bounded Ricci curvature is superfluous in the statement of this theorem and that, given the uniqueness of the soliton vector field in the Lie algebra of the torus here, the corresponding shrinking gradient Kähler–Ricci soliton is also unique up to automorphisms of the complex structure commuting with the flow of the soliton vector field.

Not only does the Duistermaat–Heckman theorem imply the uniqueness of the soliton vector field in our case, it also provides a formula to compute the unique critical point of the weighted volume functional, the point at which the soliton vector field is achieved. Using this formula, we compute explicitly the soliton vector field of a shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature on $\mathbb{C}^n$ and on the total space of $\mathcal{O}(-k)$ over $\mathbb{P}^{n-1}$ for $0 < k < n$; see Examples A.6 and A.8, respectively. This recovers the polynomials of Feldman, Ilmanen and Knopf [30, equation (36)]. Having identified the soliton vector field on these manifolds, we then use our noncompact version of Matsushima's theorem (Theorem 5.1) together with an application of Iwasawa's theorem [39] to deduce that, up to pullback by an element of $\mathrm{GL}(n, \mathbb{C})$, the corresponding soliton metrics have to be invariant under the action of $U(n)$. Consequently, thanks to a uniqueness theorem of Feldman, Ilmanen and Knopf [30, Proposition 9.3], up to pullback by an element of $\mathrm{GL}(n, \mathbb{C})$, the shrinking gradient Kähler–Ricci soliton must be the flat Gaussian shrinking soliton if on $\mathbb{C}^n$, or the unique $U(n)$–invariant shrinking gradient Kähler–Ricci soliton constructed by Feldman, Ilmanen and Knopf [30] if on the total space of $\mathcal{O}(-k)$ over $\mathbb{P}^{n-1}$ for $0 < k < n$.

In the complex two-dimensional case, we are actually able to identify the underlying complex manifold of a shrinking gradient Kähler–Ricci soliton whose scalar curvature tends to zero at infinity as either $\mathbb{C}^2$ or $\mathbb{C}^2$ blown up at the origin using the fact that the soliton, if nontrivial, has an asymptotic cone with strictly positive scalar curvature. Using a classification theorem for Sasaki manifolds in real dimension three (see Belgun [4]) and the fact that a shrinking Kähler–Ricci soliton can only contain (-1)–curves in complex dimension two, this is enough to identify the asymptotic cone at infinity as $\mathbb{C}^2$, from which the identification of the underlying complex manifold easily follows. Combined with the above discussion, this yields a complete classification of such solitons in two complex dimensions.

These conclusions are summarised in the following theorem.

Theorem E (classification of shrinkers) *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton.*

(1) *If $M = \mathbb{C}^n$ and g has bounded Ricci curvature, then up to pullback by an element of* $\mathrm{GL}(n, \mathbb{C})$, *$(M, g, X)$ is the flat Gaussian shrinking soliton.*

(2) *If M is the total space of $\mathcal{O}(-k)$ over $\mathbb{P}^{n-1}$ for $0 < k < n$ and g has bounded Ricci curvature, then up to pullback by an element of* $\mathrm{GL}(n, \mathbb{C})$, *$(M, g, X)$ is the unique $U(n)$–invariant shrinking gradient Kähler–Ricci soliton constructed by Feldman, Ilmanen and Knopf* [30].

(3) *If* $\dim_{\mathbb{C}} M = 2$ *and the scalar curvature of g tends to zero at infinity, then up to pullback by an element of* $\mathrm{GL}(2, \mathbb{C})$, *$(M, g, X)$ is the flat Gaussian shrinking soliton on $\mathbb{C}^2$ or the unique $U(2)$–invariant shrinking gradient Kähler–Ricci soliton constructed by Feldman, Ilmanen and Knopf* [30] *on the total space of $\mathcal{O}(-1)$ over $\mathbb{P}^1$.*

As exemplified in complex dimension two, in contrast to the expanding case, not many Kähler cones appear as tangent cones of complete shrinking gradient Kähler–Ricci solitons.

Outline of paper We begin in Section 2 by recalling the basics of Kähler cones, Ricci and Kähler–Ricci solitons, metric measure spaces, and the relevant algebraic geometry that we require. We also mention some important properties of the soliton vector field and of real vector fields that commute with the soliton vector field. In Section 3, we prove Theorem A for expanding gradient Kähler–Ricci solitons. The proof for shrinking Kähler–Ricci solitons is verbatim. By a theorem of Siepmann [62, Theorem 4.3.1], under our curvature assumption, a complete expanding gradient Ricci soliton flows out of a Riemannian cone. Our starting point is to prove some preliminary lemmas before providing a refinement of Siepmann's theorem, namely Theorem 3.8, where, in the course of its proof, we construct a diffeomorphism between the cone and the end of the Ricci soliton using the flow of the soliton vector field that encapsulates the asymptotics of the soliton along the end. We then show in Proposition 3.10 that if the soliton is Kähler, then the cone is Kähler with respect to a complex structure that makes the aforementioned diffeomorphism a biholomorphism. In Theorem 3.11, this biholomorphism is then shown to extend to an equivariant resolution with the properties as stated in Theorem A.

In the first part of Section 4, we use Theorem A to prove Corollary B. This also requires an application of previous work from [19]. In the latter part of Section 4, we apply Corollary B to two-dimensional expanding gradient Kähler–Ricci solitons to conclude the statement of Corollary C, making use of the classification of two-dimensional Kähler cones, namely Theorem 2.5.

From Section 5 onwards, we turn our attention exclusively to complete shrinking gradient Kähler–Ricci solitons. We begin in Section 5.1 by proving a Matsushima-type theorem stating that the Lie algebra of real holomorphic vector fields commuting with the soliton vector field may be written as a direct sum. This is the statement of Theorem 5.1. Our proof of this theorem follows a manner similar to the proof of

Matsushima's theorem on Kähler–Einstein Fano manifolds as presented in [40, Proof of Theorem 5.1, page 95]. After deriving some properties of the automorphism groups of a complete shrinking gradient Kähler–Ricci soliton (M, g, X), we then apply Theorem 5.1 to prove the maximality of a certain compact Lie group acting on M. This is the content of Corollary 5.11.

We continue in Section 5.2 by showing in Proposition 5.12 that every real holomorphic Killing vector field on M admits a Hamiltonian potential satisfying a certain linear equation. This allows us to define a moment map in Definition 5.13, which is used in the definition of the weighted volume functional in Definition 5.14. The weighted volume functional is vital in proving the uniqueness statement of Theorem D, namely that the soliton vector field of a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature in the Lie algebra of a torus is unique. The weighted volume functional is the same as that defined by Tian and Zhu [64], although in our situation it is defined as an integral over the noncompact manifold M. This is compensated for by the fact that the domain of definition of the weighted volume functional is restricted to an open cone in the Lie algebra of the torus. Several important properties of the weighted volume functional are then derived in Lemma 5.16, including the crucial fact that it has a unique critical point in its open cone of definition given by the complex structure applied to the soliton vector field X. We conclude this subsection by taking note of the fact that the Duistermaat–Heckman formula may be used to compute the weighted volume functional. In particular, it follows that the weighted volume functional is independent of the complete shrinking gradient Kähler–Ricci soliton.

In Section 5.3, we prove the uniqueness statement of Theorem D, which has been recalled in the statement of Theorem 5.18. The proof of this theorem follows as in [64, page 322] using Iwasawa's theorem [39] and the corollary of Matsushima's theorem, namely Corollary 5.11 discussed above. Section 5.4 then comprises an application of Theorems A and D to classify complete shrinking gradient Kähler–Ricci solitons with bounded Ricci curvature on $\mathbb{C}^n$ and on the total space of $\mathcal{O}(-k) \to \mathbb{P}^{n-1}$ for $0 < k < n$. This completes the proof of items (1) and (2) of Theorem E.

In Section 6, we show that the underlying complex manifold M of a two-dimensional shrinking gradient Kähler–Ricci soliton with scalar curvature decaying to zero at infinity is either $\mathbb{C}^2$ or $\mathbb{C}^2$ blown up at the origin. This is the statement of Theorem 6.1. Combined with items (1) and (2) of Theorem E, Theorem 6.1 suffices to prove item (3) of Theorem E. The proof of Theorem 6.1 relies on first identifying the underlying complex space of the tangent cone. The fact that any nonflat shrinking gradient Ricci soliton has positive scalar curvature implies that the same property holds true on the tangent cone. From this we can identify the cone as a quotient of $\mathbb{C}^2$. Theorem A then tells us that M is a resolution of this cone which, by virtue of the shrinking Kähler–Ricci soliton equation, can only contain (-1)–curves. It turns out that the only possibility is that the cone is biholomorphic to $\mathbb{C}^2$ and M is as stated in Theorem 6.1.

We conclude the paper in Section 7 with some closing remarks and open problems. In Section A.1 in the appendix, we recall the statement of the Duistermaat–Heckmann theorem on a noncompact symplectic manifold in Theorem A.3 as presented in [60]. We also provide an outline of its proof in Section A.3

after introducing some preliminaries in Section A.2. We then use it to compute the weighted volume functional and its unique critical point on $\mathbb{C}^n$, on the total space of $\mathcal{O}(-k)$ over $\mathbb{P}^{n-1}$ for $0 < k < n$, and on certain holomorphic line bundles over Fano manifolds in Section A.4. In Section A.5, we characterise algebraically, in the setting of asymptotically conical Kähler manifolds, those elements in the Lie algebra of a torus that admit a Hamiltonian potential that is proper and bounded below. A precise statement is given in Theorem A.10. For such elements of the Lie algebra of the torus, the weighted volume functional is defined. Finally, in Section A.6, we show directly that the weighted volume functional is defined on a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature without appealing to the Duistermaat–Heckman theorem. This conclusion follows from the estimates we derive in Proposition A.13 on the growth of those Hamiltonian potentials that are proper and bounded below.

Acknowledgements The authors wish to thank Richard Bamler, John Lott, Bing Wang and Yuanqi Wang for useful discussions and for their interest in this work, and Brett Kotschwar for his interest, for useful discussions involving [43; 44] and for comments on a preliminary version of this article. Part of this work was carried out while the first two authors where visiting the Institut Henri Poincaré as part of the Research in Paris program in July 2019. They wish to thank the institute for their hospitality and for the excellent working conditions provided. Finally, the authors wish to thank the anonymous referee for helpful comments regarding the proof of Theorem 5.1.

Conlon is supported by Simons Collaboration Grant #581984. Deruelle is supported by grant ANR-17-CE40-0034 of the French National Research Agency ANR (project CCEM) and Fondation Louis D, project *Jeunes géomètres*. Sun is supported by the Simons Collaboration Grant on *Special holonomy in geometry, analysis, and physics* #488633, an Alfred P Sloan fellowship, and NSF grant DMS-1708420.

2 Preliminaries

2.1 Riemannian cones

For us, the definition of a Riemannian cone will take the following form.

Definition 2.1 Let (S, g_S) be a closed Riemannian manifold. The *Riemannian cone* C_0 with *link* S is defined to be $\mathbb{R}_{>0} \times S$ with metric $g_0 = dr^2 \oplus r^2 g_S$ up to isometry. The radius function r is then characterised intrinsically as the distance from the apex in the metric completion.

The following is a simple computation.

Lemma 2.2 *Let (S, g_S) be a closed Riemannian manifold of real dimension m and let (C_0, g_0) be the Riemannian cone with link S and radial function r. Then the Ricci curvature $\mathrm{Ric}(g_S)$ of g_S and the Ricci curvature $\mathrm{Ric}(g_0)$ of the cone metric g_0 over (S, g_S) are related by*

$$\mathrm{Ric}(g_0) = \mathrm{Ric}(g_S) - (m-1)g_S.$$

In particular, the scalar curvatures R_{g_0} and R_{g_S} of g_0 and g_S respectively are related by

$$R_{g_0} = \frac{1}{r^2}(R_{g_S} - m(m-1)).$$

2.2 Kähler cones

We may further impose that a Riemannian cone is Kähler, as the next definition demonstrates.

Definition 2.3 A *Kähler cone* is a Riemannian cone (C_0, g_0) such that g_0 is Kähler, together with a choice of g_0–parallel complex structure J_0. This will in fact often be unique up to sign. We then have a Kähler form $\omega_0(X,Y) = g_0(J_0X,Y)$, and $\omega_0 = \frac{1}{2} i\partial\bar{\partial} r^2$ with respect to J_0.

The vector field $r\partial_r$ is real holomorphic and $\xi := J_0 r\partial_r$ is real holomorphic and Killing [49, Appendix A]. This latter vector field is known as the *Reeb vector field*. The closure of its flow in the isometry group of the link of the cone generates the holomorphic isometric action of a real torus on C_0 that fixes the apex of the cone. We call a Kähler cone "quasiregular" if this action is an S^1–action (and, in particular, "regular" if this S^1–action is free), and "irregular" if the action generated is that of a real torus of rank > 1.

Every Kähler cone is affine algebraic.

Theorem 2.4 *For every Kähler cone (C_0, g_0, J_0), the complex manifold (C_0, J_0) is isomorphic to the smooth part of a normal algebraic variety $V \subset \mathbb{C}^N$ with one singular point. In addition, V can be taken to be invariant under a $\mathbb{C}^*$–action $(t, z_1, \dots, z_N) \mapsto (t^{w_1} z_1, \dots, t^{w_N} z_N)$ such that all of the w_i are positive integers.*

This can be deduced from arguments written down by van Coevering in [17, Section 3.1].

Kähler cones of complex dimension two have been classified.

Theorem 2.5 [4, Theorem 8; 59, Theorem 1.1] *Let C_0 be a Kähler cone of complex dimension two. Then C_0 is biholomorphic to either*

(i) $\mathbb{C}^2/\Gamma$, *where Γ is a finite subset of $U(2)$ acting freely on $\mathbb{C}^2 \setminus \{0\}$,*

(ii) *the blowdown $L^\times$ of the zero section of a negative line bundle $L \to C$ over a smooth proper curve C of genus g with $g > 0$, or*

(iii) *$L^\times/G$, where G is a nontrivial finite group of automorphisms of a proper curve C of genus $g > 0$ and $L^\times$ is the blowdown of the zero section of a G–invariant negative line bundle $L \to C$ over C with G acting freely on $L^\times$ except at the apex.*

In cases (ii) *and* (iii), *the corresponding Reeb vector field is quasiregular and is generated by a scaling of the standard S^1–action on the fibres of L.*

Any automorphism of a resolution of a Kähler cone preserves the exceptional set of the resolution.

Lemma 2.6 *Let $\pi : M \to C_0$ be a resolution of a Kähler cone C_0 with exceptional set E. Denote by J the complex structure on M. Then for any automorphism σ of (M, J), it holds that $\sigma(E) \subseteq E$. In particular, real holomorphic vector fields on M are tangent to E.*

Such an automorphism of (M, J) therefore descends to an automorphism of the cone fixing the apex.

Proof If σ is an automorphism of (M, J), then $\pi \circ \sigma : M \to C_0$ is also a resolution of C_0. The exceptional set of this resolution is then a compact analytic subset of M. Since E is the maximal compact analytic subset of M, we must have that $(\pi \circ \sigma)^{-1}(o) \subseteq E$, where o denotes the apex of C_0, ie $\sigma^{-1}(E) \subseteq E$. □

The holomorphic torus action on a Kähler cone leads to the notion of an *equivariant resolution*.

Definition 2.7 Let C_0 be a Kähler cone with complex structure J_0, let $\pi : M \to C_0$ be a resolution of C_0, and let G be a Lie subgroup of the automorphism group of (C_0, J_0) fixing the apex of C_0. We say that $\pi : M \to C_0$ is an *equivariant resolution with respect to* G if the action of G on C_0 extends to a holomorphic action on M in such a way that $\pi(g \cdot x) = g \cdot \pi(x)$ for all $x \in M$ and $g \in G$.

Such a resolution of a Kähler cone always exists; see [41, Proposition 3.9.1].

A closed Riemannian manifold (S, g_S) is *Sasaki* if and only if its Riemannian cone is a Kähler cone [8], in which case we identify (S, g_S) with the level set $\{r = 1\}$ of its corresponding Kähler cone, r here denoting the radial function of the cone. The restriction of the Reeb vector field to this level set induces a nonzero vector field $\xi \equiv J_0 r \partial_r|_{\{r=1\}}$ on S. Let η denote the g_S–dual one-form of ξ. Then we get a g_S–orthogonal decomposition $TS = \mathcal{D} \oplus \langle \xi \rangle$, where $\mathcal{D}$ is the kernel of η and $\langle \xi \rangle$ is the $\mathbb{R}$–span of ξ in TS, and correspondingly a decomposition of the metric g_S as $g_S = \eta \otimes \eta + g^T$, where $g^T = g_S|_{\mathcal{D}}$. The metric g^T is invariant under the flow of ξ and induces a Riemannian metric on the local leaf space of the foliation of S induced by the flow of ξ. We call g^T the *transverse metric*. We can then define the *transverse scalar curvature* R^T and the *transverse Ricci curvature* Ric^T as the corresponding curvatures of g^T. We also get an induced *transverse complex structure* J^T on the local leaf space of the foliation with respect to which g^T is Kähler given by the restriction of the complex structure of the cone to $\mathcal{D}$. In particular, Ric^T will be J^T–invariant. We have the following relationships between the various curvatures.

Lemma 2.8 [8, Theorem 7.3.12] *Let (S, g_S) be a real $(2n+1)$–dimensional Sasaki manifold. Then the following identities hold:*

(i) $\mathrm{Ric}(g_S)(X, \xi) = 2n\eta(X)$ *for any vector field* X.

(ii) $\mathrm{Ric}(g_S)(X, Y) = \mathrm{Ric}^T(X, Y) - 2g_S(X, Y)$ *for any vector fields* $X, Y \in \mathcal{D}$.

In particular, we deduce:

Corollary 2.9 *Let (S, g_S) be a real $(2n+1)$–dimensional Sasaki manifold with scalar curvature R_{g_S}. Then*

$$R^T = R_{g_S} + 2n.$$

2.3 Canonical models

Resolutions of Kähler cones that are consistent with admitting an expanding Kähler–Ricci soliton are of the following type.

Definition 2.10 [38, Definition 8.2.4] A partial resolution $\pi: M \to C_0$ of a normal isolated singularity $x \in C_0$ is called a *canonical model* if

(i) M has at worst canonical singularities, and

(ii) K_M is π–ample.

Note that the choice of partial resolution π is part of the data here. The existence of a canonical model $\pi: M \to C_0$ is guaranteed by [7] and it is unique up to isomorphisms over C_0 [38, Proposition 8.2.5]. We have the following criterion to determine when K_M is π–ample.

Theorem 2.11 (Nakai's criterion for a mapping [46, Corollary 1.7.9]) *Let $\pi: M \to C_0$ be a proper morphism of schemes. A $\mathbb{Q}$–divisor D on M is π–ample if and only if $(D^{\dim_{\mathbb{C}} V} \cdot V) > 0$ for every irreducible subvariety $V \subset M$ of positive dimension that maps to a point in C_0.*

In particular, in complex dimension two, item (ii) of Definition 2.10 implies that the exceptional set of the canonical model cannot contain any (-1)– or (-2)–curves.

In our case, C_0 will be a Kähler cone, hence is affine algebraic, and x will be the apex of the cone. As the next lemma shows, the canonical model of C_0 is quasiprojective.

Lemma 2.12 *Let $\pi: M \to C_0$ be the canonical model of a Kähler cone C_0. Then M is quasiprojective.*

Proof From Theorem 2.4, we see that C_0 admits an affine embedding that is invariant under a $\mathbb{C}^*$–action with positive integer weights. Taking the weighted projective closure of C_0 with respect to this action, we obtain a projective compactification $\overline{C}_0$ of C_0 by adding an ample divisor D at infinity. In particular, $\overline{C}_0$ will have at worst orbifold singularities along D. Let $\sigma: \overline{N} \to \overline{C}_0$ denote the canonical model of $\overline{C}_0$. By construction, $\overline{N}$ is projective and the restricted map $\sigma|_N: N \to C_0$, where $N := \overline{N} \setminus \sigma^{-1}(D)$, is a canonical model of C_0. By uniqueness of canonical models, M must be biholomorphic to N, hence $\overline{N}$ provides a projective compactification of M obtained by adjoining the set $\sigma^{-1}(D)$ to M at infinity. M is therefore quasiprojective, as claimed. □

In addition, uniqueness of the canonical model implies that when smooth, the canonical model of a Kähler cone is equivariant with respect to the torus action on C_0 generated by the flow of $J_0 r\partial_r$.

Lemma 2.13 *Let C_0 be a Kähler cone with complex structure J_0 and radial function r and let $\pi : M \to C_0$ denote the canonical model of C_0. If M is smooth, then the resolution $\pi : M \to C_0$ is equivariant with respect to the holomorphic isometric torus action on C_0 generated by $J_0 r\partial_r$. In particular, there exists a holomorphic vector field X on M such that $d\pi(X) = r\partial_r$.*

Proof Let T denote the torus generated by the flow of $J_0 r\partial_r$ and let $T_{\mathbb{C}}$ denote its complexification. We will show that the holomorphic action of $T_{\mathbb{C}}$ on C_0 lifts to a holomorphic action on M.

For $h \in T_{\mathbb{C}}$, let $\psi_h : C_0 \to C_0$ denote the corresponding automorphism. Since $\psi_h \circ \pi : M \to C_0$ is again a canonical model and since $\pi : M \to C_0$ is unique up to isomorphisms over C_0 [38, Proposition 8.2.5], there exists a unique biholomorphism $\widetilde{\psi}_h : M \to M$ such that $\pi \circ \widetilde{\psi}_h = \psi_h \circ \pi$. This is the desired lift of ψ_h. Thus, we have a well-defined map $\phi : T_{\mathbb{C}} \times M \to M$ defined by $\phi(h, x) = \widetilde{\psi}_h(x)$. Since $\widetilde{\psi}_h$ coincides with ψ_h off of the exceptional set E of the resolution $\pi : M \to C_0$, the map $\phi|_{T_{\mathbb{C}} \times (M \setminus E)}$ is holomorphic. We wish to show that ϕ is holomorphic globally.

To this end, let $h \in T_{\mathbb{C}}$, let $x \in E$, let $y = \widetilde{\psi}_h(x) \in E$, let B_x be an open ball in a chart containing x with x in its interior and let B_y be an open ball in a chart containing y with y in its interior. Since $\widetilde{\psi}_h$ is continuous, by shrinking B_x if necessary we may assume that $\widetilde{\psi}_h(B_x) \subset B_y$. Let U be a neighbourhood of h in $T_{\mathbb{C}}$ such that $\widetilde{\psi}_{h'}(B_x \setminus E) \subset B_y$ for all $h' \in U$. Again, this is possible because $\phi|_{T_{\mathbb{C}} \times (M \setminus E)}$ is continuous. Then since $\widetilde{\psi}_{h'} : M \to M$ is itself continuous and preserves E for each fixed $h' \in U$ and since $B_x \cap E$ lies in the closure of $B_x \setminus E$, we have, after shrinking B_x further if necessary, that $\widetilde{\psi}_{h'}(B_x) \subset B_y$ for all $h' \in U$.

Next, let $N_\varepsilon := \{x \in C_0 \mid r(x) < \varepsilon\}$ for $\varepsilon > 0$. Then $N_\varepsilon \setminus \{o\}$ is foliated by disjoint punctured discs, obtained as the orbits in $N_\varepsilon \setminus \{o\}$ of a $\mathbb{C}^*$–action from within $T_{\mathbb{C}}$. The open set $\widehat{N}_\varepsilon := \pi^{-1}(N_\varepsilon)$ will then be a neighbourhood of E in M with $\widehat{N}_\varepsilon \setminus E$ foliated by disjoint punctured discs. Let $B'_x \subset B_x$ be an open ball containing x strictly contained in B_x such that $\partial B_x \cap \partial B'_x = \varnothing$, and let $\varepsilon > 0$ be sufficiently small that each point of $(\widehat{N}_\varepsilon \cap B'_x) \setminus E$ is contained in a punctured holomorphic disk of radius ε which itself is contained in $(\widehat{N}_\varepsilon \cap B_x) \setminus E$. Let $V := \widehat{N}_\varepsilon \cap B'_x$. Then this is an open neighbourhood of x in M and each point $z \in V \setminus E$ will lie on a unique punctured holomorphic disk, which we shall denote by D_z. We have that $D_z \subseteq (\widehat{N}_\varepsilon \cap B_x) \setminus E$ and that $\partial D_z \subseteq B_x \cap \partial\widehat{N}_\varepsilon \subset M \setminus E$. Let $\overline{D}_z$ denote the closure of D_z in M. Since $\widetilde{\psi}_{h'} : M \to M$ is holomorphic for each fixed $h' \in U$, after localising, we see from the maximum principle that for all $z \in V \setminus E$ and all $h' \in U$,

$$|\widetilde{\psi}_{h'}(z)| \le \sup_{w \in \overline{D}_z} |\widetilde{\psi}_{h'}(w)| \le \sup_{\{w \in \partial D_z\}} |\widetilde{\psi}_{h'}(w)| \le \sup_{\{w \in B_x \cap \partial\widehat{N}_\varepsilon\}} |\phi(h', w)| \le C$$

for some constant $C > 0$, where the last inequality follows from the fact that $\phi|_{T_{\mathbb{C}} \times (M \setminus E)}$ is holomorphic, hence continuous. Thus, $\phi|_{U \times (V \setminus E)}$ is a bounded holomorphic function. Since $E \cap V$ is an analytic subset of V, it follows from the Riemann extension theorem that $\phi|_{U\times(V\setminus E)}$ has a unique extension to a holomorphic function $\widetilde{\phi}\colon U\times V\to B_y$. Due to the fact that $\phi(h',\cdot)\colon M\to M$ is holomorphic for each fixed $h'\in T_{\mathbb{C}}$, we have from uniqueness of holomorphic extensions that $\widetilde{\phi}(h',\cdot)=\phi(h',\cdot)\colon V\to B_y$ for all $h'\in U$ so that in fact $\widetilde{\phi}=\phi|_{U\times V}$. Thus, we see that $\phi|_{U\times V}$ is holomorphic. Since being holomorphic is a local property, this suffices to show that $\phi\colon T_{\mathbb{C}}\times M\to M$ is holomorphic, as desired. □

2.4 Minimal models

We consider two types of resolution of a normal isolated singularity of complex dimension two, the first being the minimal resolution.

Definition 2.14 [38, Definition 7.1.14] A resolution $\pi: M\to C_0$ of a normal isolated singularity $x\in C_0$ is called a *minimal resolution* if for every resolution $\pi'\colon M'\to C_0$ of C_0 there exists a unique morphism $\varphi\colon M'\to M$ such that π' factors as $\pi'=\pi\circ\varphi$.

By definition, if there exists a minimal resolution, then it is unique up to isomorphisms over C_0. The following shows that there exists a minimal resolution for two-dimensional isolated normal singularities.

Theorem 2.15 [38, Theorem 7.1.15] *Assume that* $\dim_{\mathbb{C}} C_0=2$. *A resolution* $\pi\colon M\to C_0$ *of an isolated normal singularity* $x\in C_0$ *is the minimal resolution if and only if* $\pi^{-1}(\{x\})$ *does not contain a* (-1)*–curve. In particular, there exists a minimal resolution.*

We also consider "good" resolutions. We henceforth follow [56; 59].

Definition 2.16 A resolution $\pi\colon M\to C_0$ of a normal isolated surface singularity $x\in C_0$ is called *good* if

(i) all of the components of the exceptional divisor of $\pi: M\to C_0$ are smooth and intersect transversely;

(ii) not more than two components pass through any given point;

(iii) two different components intersect at most once.

It is known that there is a unique resolution which is minimal among all good resolutions for two-dimensional isolated normal singularities [45, Theorem 5.12], which we henceforth refer to as the "minimal good resolution" (not to be confused with the "minimal resolution"). In general, the minimal resolution of a two-dimensional isolated normal singularity will not be the minimal good resolution of the singularity; by Theorem 2.15, the two coincide precisely when the minimal good resolution does not contain any (-1)–curves.

For the Kähler cones of Theorem 2.5(ii), the minimal good resolution is given by $\pi: L\to L^{\times}$, that is, by contracting the zero section of L. By adjunction, this resolution will be the canonical model of the singularity.

As for the Kähler cones of Theorem 2.5(iii), the situation is slightly more complicated. The minimal good resolution of $L^{\times}/G$ is obtained as follows. A partial resolution is given by the induced map $\pi\colon L/G \to L^{\times}/G$ between the quotient spaces. The variety L/G will only have isolated cyclic quotient singularities along the exceptional set of π [59, Lemma 3.5], which comprises a single curve C of genus $g > 0$. Each of these cyclic quotient singularities has a minimal resolution with exceptional set a string of $\mathbb{P}^1$'s. Resolving them with this resolution then yields the minimal good resolution of $L^{\times}/G$, the exceptional set of which will then comprise the curve C (of genus $g > 0$) with branches of $\mathbb{P}^1$'s stemming from finitely many points of C. The original singularity is determined up to analytic isomorphism by this data which can be succinctly stored in a "weighted dual graph". This we now explain.

The *weighted dual graph* of a good resolution is a graph each vertex of which represents a component of the exceptional divisor, weighted by self-intersection. Two vertices are connected if the corresponding components intersect.

In our case, the weighted dual graph of the minimal good resolution of $L^{\times}/G$ is represented by a *star*, that is, a connected tree where at most one vertex is connected to no more than two other vertices. C itself is contained in the exceptional set of the minimal good resolution, hence one of the vertices of the star will represent C. We call C the *central curve*. The connected components of the graph minus the central curve are called the *branches* of the graph and are indexed by i, where $1 \le i \le n$. The curves of the i^{th} branch are denoted by C_{ij} for $1 \le j \le r_i$, where C_{i1} intersects C and C_{ij} intersects $C_{i,j+1}$. Let $b = -C.C$ and $b_{ij} = -C_{ij}.C_{ij}$. Then $b_{ij} \ge 2$ and $b \ge 1$. Finally, set

$$\frac{d_i}{e_i} = b_{i1} - \cfrac{1}{b_{i2} - \cfrac{1}{\ddots \cfrac{1}{b_{ir_i}}}}$$

with $e_i < d_i$ and e_i and d_i relatively prime. Then one has:

Theorem 2.17 [59, Theorem 2.1] *The singularity $L^{\times}/G$ with G nontrivial is determined up to analytic isomorphism by the following data*:

(i) *The weighted dual graph of the minimal good resolution.*

(ii) *The analytic type of the central curve C (of genus $g > 0$).*

(iii) *The conormal bundle of C in the resolution.*

(iv) *The n points $P_i = C \cap C_{i1}$ on C with $n \ge 1$.*

Conversely, given any data as above, there exists a unique singularity of the form $L^{\times}/G$ having this data, provided that the intersection matrix given by the graph in (i) *is negative-definite; this condition can be written as*

$$b - \sum_{i=1}^{n} \frac{e_i}{d_i} > 0.$$

Indeed, the algorithm that recovers the algebraic equations cutting out $L^\times/G$ is laid out in [59, Section 5]. However, it does not identify the group G.

For the Kähler cones of Theorem 2.5(iii), the minimal good resolution does not contain any (-1)–curves, hence it coincides with the minimal model of the singularity. Moreover, since the central curve has trivial or negative anticanonical bundle, adjunction tells us that the canonical model is obtained from the minimal good resolution by further contracting all of its (-2)–curves. However, the result of this will be singular unless the minimal good resolution does not contain any (-2)–curves. Thus, the canonical model will be smooth and coincide with the minimal good resolution if the minimal good resolution does not contain any (-2)–curves. Conversely, if the canonical model of the singularity is smooth, then, since it cannot contain any (-1)–curves, it coincides with the minimal model, which itself coincides with the minimal good resolution for the cones in question, so that the minimal good resolution does not contain any (-2)–curves since the canonical model cannot contain any (-2)–curves. Combining this observation with Theorem 2.17, we are able to characterise those cones of Theorem 2.5(iii) that admit a smooth canonical model.

Proposition 2.18 *A Kähler cone of Theorem 2.5*(iii) *admits a smooth canonical model if and only if the minimal good resolution does not contain any* (-2)*–curves. These cones are in one-to-one correspondence with the data* (i)–(iv) *listed in Theorem 2.17 with the intersection matrix of the graph in* (i) *being negative-definite and with the labels* b_{ij} *of this graph being* ≥ 3. *Moreover, the canonical model and the minimal resolution of such a cone are given by the minimal good resolution.*

2.5 Ricci solitons

Recall the definitions given at the beginning of Section 1.1. For a gradient Kähler–Ricci soliton (M, g, X) with complex structure J, the vector field JX is Killing by [32, Lemma 2.3.8]. We also have the following asymptotics on the soliton potential of a complete expanding gradient Ricci soliton with quadratic Ricci curvature decay.

Proposition 2.19 *Let* $(M^n, g, \nabla^g f)$ *be a complete expanding gradient Ricci soliton of real dimension* n, *ie* $2\,\mathrm{Ric}(g) - \mathcal{L}_{\nabla^g f}(g) = -g$. *If* $|\mathrm{Ric}(g)|_g = O(d_g(p,\cdot)^{-2})$, *where* $d_g(p,\cdot)$ *denotes the distance to a fixed point* $p \in M$, *then the function* $(-f)$ *is equivalent to* $\frac{1}{4}d_g(p,\cdot)^2$ *as* $d_g(p,\cdot)$ *tends to* $+\infty$.

Proof See [12] or [62, Lemma 4.2.1]. □

Because of Proposition 2.19, we prefer to deal with an asymptotically positive soliton potential. Henceforth, an *expanding* gradient Ricci soliton will be a triple (M, g, X), where $X = \nabla^g f$ for some real-valued smooth function f on M, such that the equation

$$2\,\mathrm{Ric}(g) - \mathcal{L}_X g = -g \tag{2-1}$$

is satisfied. When the Ricci curvature of g decays quadratically, the bound of Proposition 2.19 on f may then be given as

$$\tfrac{1}{4}d_g^2(p,x)-c_1d_g(p,x)-c_2\le f(x)\le \tfrac{1}{4}d_g^2(p,x)+c_1d_g(p,x)+c_2, \tag{2-2}$$

where $p\in M$ is fixed and c_1 and c_2 are positive constants depending on p. In particular, f is proper under the assumption of quadratic Ricci curvature decay on the expanding soliton metric.

In the case of a shrinking gradient Ricci soliton, the quadratic growth of the soliton potential is always satisfied without further conditions on the decay of the Ricci tensor at infinity. More precisely, one has the following.

Theorem 2.20 *Let (M,g,X) be a complete noncompact shrinking gradient Ricci soliton satisfying* (1-1) *with $\lambda=1$, with soliton vector field $X=\nabla^g f$ for a smooth real-valued function $f\colon M\to\mathbb{R}$. Then the following properties hold true.*

(i) ***Growth of the soliton potential*** [11, Theorem 1.1] *For $x\in M$, f satisfies the estimates*

$$\tfrac{1}{4}(d_g(p,x)-c_1)^2-C\le f(x)\le \tfrac{1}{4}(d_g(p,x)+c_2)^2$$

for some $C>0$, where $d_g(p,\cdot)$ denotes the distance to a fixed point $p\in M$ with respect to g. Here, c_1 and c_2 are positive constants depending only on the real dimension of M and the geometry of g on the unit ball $B_p(1)$ based at p.

(ii) ***Polynomial volume growth*** [11, Theorem 1.2] *For each $x\in M$, there exists a positive constant $C>0$ such that*

$$\mathrm{vol_g(B_r(x))\le Cr^n}\quad \textit{for } \mathrm{r>0} \textit{ sufficiently large,}$$

where $n=\dim_{\mathbb{R}} M$.

(iii) ***Regularity at infinity*** *If the curvature tensor decays quadratically, ie if $A_0(g)<+\infty$, then the soliton metric has quadratic curvature decay with derivatives, ie $A_k(g)<+\infty$ for all $k\in\mathbb{N}$.*

Proof References for items (i) and (ii) have been provided above. Item (iii) concerning the covariant derivatives of the curvature tensor follows from Shi's estimates for ancient solutions of the Ricci flow; see [43, Section 2.2.3] for a proof. □

Remark 2.21 The regularity at infinity stated in Theorem 2.20(iii) does not hold for expanding gradient Ricci solitons; see [26] for examples of expanding gradient Ricci solitons coming out of metric cones with a finite amount of regularity at infinity.

The next lemma collects together some well-known Ricci soliton identities concerning expanding gradient Ricci solitons and shrinking gradient Kähler–Ricci solitons that we require. Their proofs are standard.

Lemma 2.22 (Ricci soliton identities) *Let (M, g, X) be a shrinking gradient Kähler–Ricci soliton of complex dimension n satisfying* (1-4) *with $\lambda = 1$ (resp. an expanding gradient Ricci soliton of real dimension n satisfying* (2-1)) *with soliton vector field $X = \nabla^g f$ for a smooth real-valued function $f : M \to \mathbb{R}$. Then the trace and first-order soliton identities are*:

$$\begin{aligned} &\Delta_\omega f + \tfrac{1}{2} R_g = n && (\textit{resp.}\ -\Delta_g f + R_g = -\tfrac{1}{2} n),\\ &\nabla^g R_g - 2\,\mathrm{Ric}(g)(X) = 0 && (\textit{resp.}\ \nabla^g R_g + 2\,\mathrm{Ric}(g)(X) = 0),\\ &|\nabla^g f|^2 + R_g - 2f = \mathrm{const.} && (\textit{resp.}\ |\nabla^g f|^2 + R_g - f = \mathrm{const.}), \end{aligned}$$

where R_g denotes the scalar curvature of g and $|\nabla^g f|^2 := g^{ij}\partial_i f \partial_j f$.

Remark 2.23 We henceforth normalise the soliton potential f of a shrinking gradient Kähler–Ricci soliton of complex dimension n satisfying (1-4) with $\lambda = 1$ so that $|\nabla^g f|^2 + R_g - 2f = 2n$. The choice of constant $2n$ is dictated by the following equation satisfied by f:

$$\Delta_\omega f - \tfrac{1}{2} X \cdot f = -f.$$

This choice of constant also implies that $f + n$ is nonnegative on M, since the scalar curvature R_g of g is necessarily nonnegative.

Kähler cones are quasiprojective. This property is inherited by complete expanding and shrinking gradient Kähler–Ricci solitons on resolutions of Kähler cones.

Proposition 2.24 *Let (M, g, X) be a complete expanding or shrinking gradient Kähler–Ricci soliton on a resolution $\pi : M \to C_0$ of a Kähler cone C_0. Then M is quasiprojective.*

Proof We prove this proposition in the case that (M, g, X) is an expanding gradient Kähler–Ricci soliton. The proof for the shrinking case is similar.

As explained in the proof of Lemma 2.12, by adding an appropriate ample divisor D to C_0 at infinity, we obtain a projective compactification $\overline{C}_0$ of C_0 so that $\overline{C}_0$ will have at worst orbifold singularities along D. Using D, we then compactify M at infinity to obtain a compact complex orbifold $\overline{M}$ such that $M = \overline{M} \setminus D$. We claim that $\overline{M}$ admits an ample line bundle, hence is projective.

Indeed, since the normal orbibundle of D in $\overline{C}_0$ is positive, the normal orbibundle of D in $\overline{M}$ will also be positive, hence by the proof of [21, Lemma 2.3], we may endow the line orbibundle $[D]$ on $\overline{M}$ with a nonnegatively curved hermitian metric with strictly positive curvature on some tubular neighbourhood U of D in $\overline{M}$. Next note that the curvature of the hermitian metric h induced on K_M by the expanding gradient Kähler–Ricci soliton metric g is $-\rho_\omega$, where ρ_ω is the Ricci form of the Kähler form ω associated to g. Let f denote the soliton potential so that $X = \nabla^g f$. Then by virtue of the expanding soliton equation, the curvature of the hermitian metric $e^f h$ on K_M is precisely the Kähler form ω of g. In particular, the curvature of $e^f h$ on K_M is a positive form. Extend the hermitian metric $e^f h$ on K_M to a hermitian metric on $K_{\overline{M}}$ by amalgamating $e^f h$ with an arbitrary hermitian metric on $K_{\overline{M}}|_U$ using an appropriate bump function supported on U. Then the line orbibundle $K_{\overline{M}} + p[D]$ will be ample for p sufficiently large. A high tensor power of the resulting line orbibundle will then be an ample line bundle on $\overline{M}$ so that $\overline{M}$ is projective and M is quasiprojective, as claimed. □

Finally, note that to each complete gradient Ricci soliton, one can associate a Ricci flow that evolves via diffeomorphisms and scaling. We describe this picture for an expanding gradient Ricci soliton next.

For a complete expanding gradient Ricci soliton (M, g, X) with soliton potential f, set

$$g(t) := t\varphi_t^* g \quad \text{for } t > 0,$$

where φ_t is a family of diffeomorphisms generated by the gradient vector field $-\frac{1}{t}X$ with $\varphi_1 = \mathrm{id}$, ie

$$\frac{\partial \varphi_t}{\partial t}(x) = -\frac{\nabla^g f(\varphi_t(x))}{t} \quad \text{with } \varphi_1 = \mathrm{id}. \tag{2-3}$$

Then $\partial_t g(t) = -2\,\mathrm{Ric}(g(t))$ for $t > 0$, $g(1) = g$, and if we define $f(t) = \varphi_t^* f$ so that $f(1) = f$, then $g(t)$ satisfies

$$\mathrm{Ric}(g(t)) - \mathrm{Hess}_{g(t)} f(t) + \frac{g(t)}{2t} = 0 \quad \text{for all } t > 0. \tag{2-4}$$

Taking the divergence of this equation and using the Bianchi identity yields

$$R_{g(t)} + |\nabla^{g(t)} f(t)|^2_{g(t)} - \frac{f(t)}{t} = \frac{C_1}{t} \tag{2-5}$$

for some constant C_1, where $R_{g(t)}$ denotes the scalar curvature of $g(t)$.

Similarly, for a complete expanding gradient Kähler–Ricci soliton with Kähler form ω, one obtains a solution of the Kähler–Ricci flow $\partial_t \omega(t) = -\rho_{\omega(t)}$, where $\rho_{\omega(t)}$ denotes the Ricci form of $\omega(t)$. The difference in normalisations between (1-2) and (1-4) is accounted for by the fact that the constant preceding the Ricci term in the Ricci flow is -2 and that preceding the Ricci term in the Kähler–Ricci flow is -1. In the same way that a Kähler–Ricci flow yields a solution of the Ricci flow and, vice versa, a solution of the Ricci flow which is Kähler yields a solution of the Kähler–Ricci flow, the same holds true for gradient Ricci solitons and gradient Kähler–Ricci solitons. Indeed, a solution (M, g, X) of (1-4) yields a solution of (1-2) by replacing g with $2g$ and composing (1-4) with the complex structure in the first arguments. Conversely, a solution (M, g, X) of (1-2) for which g is Kähler and X is real holomorphic defines a solution of (1-4) after replacing g with $\frac{1}{2}g$ and composing (1-2) with the complex structure in the first arguments.

2.6 Properties of the soliton vector field

In this subsection, we provide sufficient conditions for which the zero set of the soliton vector field of a complete shrinking gradient Kähler–Ricci soliton is compact. We begin with the following simple observation.

Lemma 2.25 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with bounded scalar curvature. Then the zero set of X is compact.*

Proof With f denoting the soliton potential, the boundedness of the scalar curvature R_g of g together with the properness of f as a consequence of Theorem 2.20(i) imply that $2f - R_g$ is proper. From the soliton identity $|\nabla^g f|^2 + R_g = 2f$ (see Lemma 2.22), we then see that the function $|\nabla^g f|^2$ is proper. The compactness of the zero set of X is now immediate. □

In the case that M is in addition "1–convex", meaning that M carries a plurisubharmonic exhaustion function which is strictly plurisubharmonic outside of a compact set, we can be more precise. Since a 1–convex space is in particular holomorphically convex, M in this case will admit a "Remmert reduction" $p\colon M \to M'$ [36], ie a proper holomorphic map $p\colon M \to M'$ onto a normal Stein space M' with finitely many isolated singularities obtained by contracting the maximal compact analytic subset E of M. As a Stein space with only finitely many isolated singularities, [3, Theorem 3.1] asserts that M' admits an embedding $h\colon M' \to \mathbb{C}^P$ into $\mathbb{C}^P$ for some $P \in \mathbb{N}$. We have:

Proposition 2.26 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton of complex dimension n with bounded scalar curvature. Assume that M is 1–convex with maximal compact analytic subset E. Then the zero set of X is compact and*

(i) *if $E = \varnothing$, then the zero set of X comprises a single point and M is biholomorphic to $\mathbb{C}^n$;*

(ii) *if $E \neq \varnothing$, then the zero set of X is contained in E.*

Before we prove this proposition, an auxiliary result is required, which will be used several times throughout.

Proposition 2.27 *Let (N, g) be a complete Riemannian manifold and let $u\colon N \to \mathbb{R}$ be a C^2 function that is proper and bounded below. Assume that the flow $\phi_x(t)$ of $\nabla^g u$ with $\phi_x(0) = x \in N$ exists for all $t \in (-\infty, 0]$. Then, for any $x \in N$, the orbit $(\phi_x(t))_{t\leq 0}$ accumulates in the critical set of u, ie for all sequences $(t_i)_i$ diverging to $-\infty$, there exists a subsequence $(t_i')_i$ such that $(\phi_{t_i'}(x))_i$ converges to a point $x_\infty \in N$ satisfying $(\nabla^g u)(x_\infty) = 0$.*

Proof Let $x \in N$ and let $(\phi_x(t))_{t\leq 0}$ denote the flow of $\nabla^g u$ which passes through x at $t = 0$ and is defined for all nonpositive times. Since $\partial_t \phi_x(t) = (\nabla^g u)(\phi_x(t))$, the function $t \in (-\infty, 0] \mapsto u(\phi_x(t)) \in \mathbb{R}$ is a nondecreasing function and for all nonpositive times t,

$$u(x) - u(\phi_x(t)) = \int_t^0 |\nabla^g u|_g^2(\phi_x(\tau))\, d\tau \geq 0. \tag{2-6}$$

In particular, the orbit $(\phi_x(t))_{t\leq 0}$ lies in the sublevel set $\{y \in M \mid u(y) \leq u(x)\}$ of u, which is compact since u is proper and bounded below. Moreover, since $u(\phi_x(t))$ is bounded from below, the estimate

(2-6) implies that the function $\tau \in (-\infty, 0] \mapsto |\nabla^g u|^2(\phi_x(\tau)) \in \mathbb{R}$ is integrable on $(-\infty, 0]$; that is,

$$\int_{-\infty}^{0} |\nabla^g u|^2(\phi_x(\tau))\, d\tau < +\infty. \tag{2-7}$$

Now, since u is C^2 and the orbit $(\phi_x(t))_{t\leq 0}$ lies in a compact subset of M, the function $\tau \in (-\infty, 0] \mapsto |\nabla^g u|^2(\phi_x(\tau)) \in \mathbb{R}$ is Lipschitz, ie there is a positive constant C such that

$$\big| |\nabla^g u|^2(\phi_x(t)) - |\nabla^g u|^2(\phi_x(s)) \big| \leq C|t-s| \quad \text{for all } s, t \in (-\infty, 0].$$

This fact, together with (2-7), implies that $\lim_{\tau\to-\infty} |\nabla^g u|(\phi_x(\tau)) = 0$. This allows us to conclude that any accumulation point of $(\phi_x(t))_{t\leq 0}$ lies in the critical set of u. □

Proof of Proposition 2.26 Let f denote the soliton potential and let $M_0(X)$ denote the zero set of X, a set which is compact by Lemma 2.25. Our first claim encapsulates the structure of $M_0(X)$.

Claim 2.28 *Each connected component of $M_0(X)$ is a smooth compact complex submanifold of M contained in a level set of f.*

Proof Let J denote the complex structure of M and let F be a connected component of $M_0(X)$. Then since F is locally the zero set of the holomorphic vector field $X^{1,0} = \frac{1}{2}(X - iJX)$, it is a complex-analytic subvariety of M. Furthermore, as a connected component of the zero set of the Killing vector field JX, it is a totally geodesic submanifold by [40, Theorem 5.3, page 60]. Hence F is a smooth complex submanifold of M.

Next observe that along any geodesic $\gamma(t)$ in F, we have for the soliton potential f,

$$\frac{d}{dt} f(\gamma(t)) = df(\gamma'(t)) = g(X, \gamma'(t)) = 0,$$

so that f is constant on F. Consequently, F is contained in a level set of f. From Theorem 2.20(i), we know that f is proper so that the level sets of f are compact. Thus, as a closed subset of a compact set, F is compact. ◁

Now note that, by [31, Proof of Lemma 1], f is a Morse–Bott function on M. The critical submanifolds of f are precisely the connected components of $M_0(X)$. Since M is Kähler, the Morse indices — ie the number of negative eigenvalues of Hess(f) — of the critical submanifolds are all even [31]. Write

$$M_0(X) = M^{(0)} \cup \bigcup_{k=1}^{n} M^{(2k)},$$

where $M^{(j)}$ denotes the disjoint union of the critical submanifolds of $M_0(X)$ of index j. As a consequence of Claim 2.28, we see that each connected component of $M_0(X)$, being a compact complex submanifold of M, is either contained in the maximal compact analytic subset E of M or is an isolated point contained in $M \setminus E$. Suppose that there exists an isolated point $x \in M^{(j)} \cap (M \setminus E)$ for some $j \geq 2$.

Using ideas from [13, page 3332] in this paragraph, we see from [10] that the holomorphic vector field $X^{1,0}$ is linearisable at each of its critical points, meaning in particular that there exist local holomorphic coordinates $(z_1, \dots, z_n)$ centred at x such that $X^{1,0} = \sum_{j=1}^{n} a_j z_j \partial_{z_j}$ with $a_j \in \mathbb{R}$ for all $j = 1, \dots, n$. Since $\mathrm{Hess}(f)$ has at least one negative eigenvalue at x, we have that $a_i < 0$ for some i. Without loss of generality, we may assume that $i = n$ so that $a_n < 0$. Now, clearly the orbits of JX on the z_n–axis are all periodic. Fix one such orbit $\theta\colon S^1 \to M$. Then we can construct a map $R\colon S^1(\simeq \mathbb{R}/T\mathbb{Z}) \times \mathbb{R} \to M$ by defining $R(s,t)$ to be $\phi_t(\theta(s))$, where ϕ_t is the integral curve of the negative gradient flow of f and T is the period of the orbit of θ. Since $[X, JX] = 0$, R is holomorphic and after reparametrising, extends to a nontrivial holomorphic map $\bar{R}\colon \mathbb{C} \to M$ with $\bar{R}(0) = x$ by the Riemann removable singularity theorem.

Since f is decreasing along its negative gradient flow and is bounded from below, we see that $f(\bar{R}(z))$ is bounded for all $z \in \mathbb{C}$. Hence, by properness of f, the set $\{\bar{R}(z) \mid z \in \mathbb{C}\}$ is contained in a compact subset of M. Letting $p\colon M \to M'$ denote the Remmert reduction of M and recalling that M' admits an embedding $h\colon M' \to \mathbb{C}^P$ into $\mathbb{C}^P$ for some $P \in \mathbb{N}$, we therefore obtain a bounded nontrivial holomorphic map $h \circ p \circ \bar{R}\colon \mathbb{C} \to \mathbb{C}^P$. By Liouville's theorem, such a map is constant. This is a contradiction. Thus $M_0(X) \cap (M \setminus E)$, if nonempty, is contained in $M^{(0)}$.

The next claim concerns the structure of $M^{(0)}$.

Claim 2.29 *$M^{(0)}$ is a nonempty, connected, compact complex submanifold of M that comprises the global minima of f.*

Proof $M^{(0)}$ is clearly nonempty since f attains a global minimum and, as a closed subset of the compact set $M_0(X)$, comprises finitely many connected, compact, complex submanifolds of M by Claim 2.28. To see that $M^{(0)}$ comprises one connected component only, recall that the soliton vector field X is complete. Then by Proposition 2.27, for any point $x \in M$, the forward orbit of the negative gradient flow of f beginning at x converges to a point of $M_0(X)$. This gives rise to a stratification of M, namely $M = \bigsqcup_{k=0}^{n} W^s(M^{(2k)})$, where

$$W^s(M^{(2k)}) = \bigl\{x \in M \mid \lim_{t \to -\infty} \phi_x(t) \in M^{(2k)}\bigr\},$$

$\phi_x\colon \mathbb{R} \to M$ here denoting the gradient flow of f beginning at x. Note that

$$W^s(M^{(0)}) = M \setminus \bigsqcup_{k=1}^{n} W^s(M^{(2k)}).$$

Now, since $M_0(X)$ is compact, for each k, $W^s(M^{(2k)})$ comprises finitely many connected components, each of which is an open submanifold of M of real dimension $2n - 2k$; see [1, Proposition 3.2]. The complement of finitely many submanifolds of real codimension at least two in a connected manifold is still connected. Hence $W^s(M^{(0)})$, and consequently $M^{(0)}$, is connected. Finally, since $M^{(0)}$ contains all of the local minima of f and, comprising only one connected component, is contained in a level set of f by Claim 2.28, it must be the set of global minima of f. ◁

Now, we have already established the fact that $M_0(X) \cap (M \setminus E)$, if nonempty, is contained in $M^{(0)}$. Thus, if $E = \varnothing$, then, since $M_0(X)$ is nonempty as f attains a global minimum, we must have that $M_0(X) = M^{(0)}$, so that $M_0(X)$ is a nonempty, connected, compact complex submanifold of M by Claim 2.29. Since M is affine if $E = \varnothing$, we deduce that $M_0(X)$ must comprise a single point if $E = \varnothing$. It then follows from [10] that M is biholomorphic to $\mathbb{C}^n$ if $E = \varnothing$. This is case (i) of the proposition.

Next consider the case when $E \neq \varnothing$. If $M_0(X) \cap (M \setminus E) = \varnothing$, then $M_0(X) \subseteq E$ and we are in case (ii) of the proposition. So, to derive a contradiction, suppose that $E \neq \varnothing$ and $M_0(X) \cap (M \setminus E) \neq \varnothing$. In light of the above, we must have that $M^{(0)} \cap E = \varnothing$ and that $M_0(X) \cap (M \setminus E) = M^{(0)}$, which comprises a single point x, say. Moreover, $\left(\bigcup_{j=1}^{n} M^{(2j)}\right) \cap E \neq \varnothing$ since otherwise M would be biholomorphic to $\mathbb{C}^n$ by [10], thereby yielding a contradiction. Thus, noting that $f(M^{(0)})$ is the global minimum value of f by Claim 2.29, let A be the smallest critical value of f with $A > f(M^{(0)})$ and let $y \in f^{-1}(\{A\})$. Then we must have that $y \in M^{(k)} \subseteq E$ for some $k \geq 2$ by what we have just said. As before, we can construct a holomorphic map $\bar{R}\colon \mathbb{C} \to M$ with $\bar{R}(0) = y$. Since f is decreasing along its negative gradient flow and since there are no critical values of f in the open interval $(f(M^{(0)}), A)$, we see from Proposition 2.27 that necessarily $\lim_{z \to +\infty} \bar{R}(z) = x$. The Riemann removable singularity theorem then applies and allows us to extend $\bar{R}$ to a holomorphic map $\bar{R}'\colon \mathbb{P}^1 \to M$. Since $x \neq y$ and $x \notin E$, what we have constructed is a nontrivial holomorphic curve in M that is not contained in E. This contradicts the maximality of E. Thus, cases (i) and (ii) of the proposition are the only two possibilities that can occur. This completes the proof. □

2.7 Properties of real vector fields commuting with the soliton vector field

In this subsection, we mention some properties of real vector fields that commute with the soliton vector field on a complete shrinking gradient Kähler–Ricci soliton. As the next proposition demonstrates, a bound on the Ricci curvature yields control on the growth of the norm of these vector fields.

Proposition 2.30 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature, and let $d_g(p,\cdot)$ denote the distance to a fixed point $p \in M$ with respect to g. Then there exists $a > 0$ such that $|Y|_g^2(x) = O(d_g(p,x)^a)$ for every real vector field Y on M with $[X, Y] = 0$.*

Remark 2.31 The growth rate obtained in Proposition 2.30 will be sharpened in Claim A.14 for real *holomorphic* vector fields commuting with X, as exemplified by shrinking cylinders of the form $\mathbb{C}^{n-k} \times N^{2k}$, where $\mathbb{C}^{n-k}$ is endowed with its Gaussian soliton metric and where N^{2k} supports a closed Kähler–Einstein metric of positive scalar curvature.

Proof Let $|\cdot|$ denote the norm with respect to g and let Ric denote the Ricci curvature of g. Since $|\mathrm{Ric}|$ is bounded so that the scalar curvature of g is bounded, it follows from Lemma 2.25 that the zero set of X is contained in a compact subset of M. For $A > 0$, let $K := f^{-1}([2A, 4A])$ and $N = f^{-1}((-\infty, 3A])$.

Since f is proper and bounded below as a consequence of Theorem 2.20(i), K and N are compact subsets of M. Choose A sufficiently large so that all of the critical points of f are contained in $f^{-1}((-\infty, A])$ and so that $A > \sup_M |R_g|$. Let $\gamma_x(t)$ denote the integral curve of X with $\gamma_x(0) = x \in M$. We begin with the following claim.

Claim 2.32 *Let $y \in M \setminus N$. Then there exists $x \in K$ and $t_0 > 0$ such that $y = \gamma_x(t_0)$.*

That is to say, every point of $M \setminus N$ lies on an integral curve of X passing through K.

Proof For $y \in M \setminus K$, we see from the soliton identity $|\nabla^g f|^2 + R_g = 2f$ that

$$\frac{d}{dt} f(\gamma_y(t)) = |\nabla^g f|^2_g(\gamma_y(t)) = 2f(\gamma_y(t)) - R_g(\gamma_y(t)).$$

Using the upper bound on $|R_g|$, we deduce that

$$\left|\frac{d}{dt} f(\gamma_y(t)) - 2f(\gamma_y(t))\right| \leq 2A.$$

Integrating this differential inequality for $t < 0$ then yields the inequalities

$$(f(y) + A)e^{2t} - A \leq f(\gamma_y(t)) \leq (f(y) - A)e^{2t} + A \quad \text{for } t < 0. \tag{2-8}$$

Set

$$t_0 = -\frac{1}{2}\ln\left(\frac{3A}{f(y) + A}\right) > 0.$$

Then from (2-8) we see that

$$2A \leq f(\gamma_y(-t_0)) \leq 3A\left(\frac{f(y) - A}{f(y) + A}\right) + A = 3A\left(1 - \frac{2A}{f(y) + A}\right) + A \leq 4A.$$

Thus, $y = \gamma_x(t_0)$ where $x = \gamma_y(-t_0) \in K$. This proves the claim. ◁

Next observe that

$$\mathcal{L}_X(|Y|^2) = (\mathcal{L}_X g)(Y, Y) = g(Y, Y) - \mathrm{Ric}(Y, Y) = |Y|^2 - \mathrm{Ric}(Y, Y).$$

For $x \in M$ a point where $X \neq 0$, let $h(t) := |Y|^2(\gamma_x(t))$. Then we can rewrite the previous equation as

$$h'(t) = h(t) - \mathrm{Ric}(Y, Y)(\gamma_x(t)),$$

so that

$$\frac{h'(t)}{h(t)} = 1 - \frac{\mathrm{Ric}(Y, Y)(\gamma_x(t))}{h(t)}. \tag{2-9}$$

Analysing the error term here, we have that

$$\frac{\mathrm{Ric}(Y, Y)(\gamma_x(t))}{h(t)} = \frac{\mathrm{Ric}(Y, Y)}{|Y|^2}.$$

Since |Ric| is bounded by assumption, we then have that

$$\left|\frac{\mathrm{Ric}(Y,Y)}{|Y|^2}\right| \leq C$$

for a constant $C > 0$, so that (2-9) gives us the bound

$$\left|\frac{h'(t)}{h(t)}\right| \leq 2a$$

for some $a > 0$. Solving this for $t > 0$ yields

$$-2at \leq \ln(h(t)) - \ln(h(0)) \leq 2at,$$

so that, in particular,

$$h(t) \leq |Y|^2(x)e^{2at} \quad \text{for all } t > 0.$$

Hence,

$$|Y|^2(\gamma_x(t)) \leq |Y|^2(x)e^{2at} \quad \text{for all } t > 0. \tag{2-10}$$

Let $y \in M \setminus N$. Then by Claim 2.32, there is an $x \in K$ and $t_0 > 0$ such that $y = \gamma_x(t_0)$. Applying the above inequality to this choice of x and t_0, we deduce that

$$|Y|^2(y) \leq |Y|^2(x)e^{2at_0}.$$

Now, as in the proof of Claim 2.32, we have that

$$\left|\frac{d}{dt}f(\gamma_x(t)) - 2f(\gamma_x(t))\right| \leq 2A.$$

Integrating this for $t > 0$ yields the fact that

$$(f(x) - A)e^{2t} + A \leq f(\gamma_x(t)) \leq (f(x) + A)e^{2t} - A \quad \text{for all } t > 0.$$

Since $x \in K$ so that $f(x) \geq 2A$, we see from the left-hand side of this inequality that $f(\gamma_x(t)) \geq A(1+e^{2t})$, so that

$$e^{2t} \leq \frac{f(\gamma_x(t))}{A} - 1 \quad \text{for all } t > 0. \tag{2-11}$$

Plugging this into (2-10) and setting $t = t_0$ results in the bound

$$|Y|^2(y) \leq |Y|^2(x)e^{2at_0} \leq |Y|^2(x)\left(\frac{f(y)}{A} - 1\right)^a \leq (\sup_K |Y|^2)\left(\frac{f(y)}{A} - 1\right)^a.$$

Since K is compact and f grows quadratically with respect to the distance to a fixed point $p \in M$ by Theorem 2.20(i), we arrive at the estimate

$$|Y|(z) \leq c_1 d_g(p,z)^a + c_2 \quad \text{for all } z \in M$$

for some positive constants $c_1, c_2 > 0$. This leads to the desired conclusion. □

Remark 2.33 In the case that the Ricci curvature decays quadratically at infinity, the constant a may be taken to be equal to 2 in Proposition 2.30.

We can also show that such vector fields are complete when the zero set of the soliton vector field is compact.

Lemma 2.34 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton. Assume that the zero set of X is compact (which, by Lemma 2.25, is the case when the scalar curvature of g is bounded). Then every real vector field Y on M with $[X, Y] = 0$ is complete.*

Proof Let Y be as in the statement of the lemma and let K be any compact subset of M containing the zero set of X in its interior. Then note the following:

- Since K is compact, there exists $\varepsilon_0 > 0$ such that the flow of Y beginning at any point of K exists on the open interval $(-\varepsilon_0, \varepsilon_0)$.
- By Proposition 2.27, for any $p \in M$, there exists $T(p) > 0$ such that the image of p under the forward flow of $-X$ for time T will be contained in K.

Consequently, for any point $p \in M$, by flowing first along $-X$ into K for time $T(p)$, then flowing along Y, then flowing along X for time T, one sees from the fact that $[X, Y] = 0$ that the flow of Y beginning at any point of M exists on the interval $(-\varepsilon_0, \varepsilon_0)$. This observation suffices to prove the completeness of Y. □

2.8 Basics of metric measure spaces

We take the following from [33]; the notions introduced in this section will be used in Section 5.1.

A smooth metric measure space is a Riemannian manifold endowed with a weighted volume.

Definition 2.35 A *smooth metric measure space* is a triple $(M, g, e^{-f}\,dV_g)$, where (M, g) is a complete Riemannian manifold with Riemannian metric g, dV_g is the volume form associated to g, and $f : M \to \mathbb{R}$ is a smooth real-valued function.

A shrinking gradient Ricci soliton (M, g, X) with $X = \nabla^g f$ naturally defines a smooth metric measure space $(M, g, e^{-f}\,dV_g)$. On such a space, we define the weighted Laplacian Δ_f by

$$\Delta_f u := \Delta u - g(\nabla^g f, \nabla u)$$

on smooth real-valued functions $u \in C^\infty(M, \mathbb{R})$. There is a natural L^2–inner product $\langle \cdot, \cdot \rangle_{L^2_f}$ on the space L^2_f of square-integrable smooth real-valued functions on M with respect to the measure $e^{-f}\,dV_g$, defined by

$$\langle u, v \rangle_{L^2_f} := \int_M uv\, e^{-f}\,dV_g \quad \text{for } u, v \in L^2_f.$$

As one can easily verify, the operator Δ_f is self-adjoint with respect to $\langle \cdot, \cdot \rangle_{L^2_f}$.

In the Kähler case, we have:

Definition 2.36 If $(M, g, e^{-f}\,dV_g)$ is a smooth metric measure space and (M, g) is Kähler, we say that $(M, g, e^{-f}\,dV_g)$ is a *Kähler metric measure space*.

A shrinking gradient Kähler–Ricci soliton naturally defines such a space.

Unlike the real case, on a Kähler metric measure space we have the weighted $\bar{\partial}$–Laplacian Δ_f defined on smooth complex-valued functions $u \in C^\infty(M, \mathbb{C})$ by

$$\Delta_f u := \Delta_{\bar{\partial}} u - (\nabla^{1,0} u) f = g^{i\bar{j}} \partial^2_{i\bar{j}} u - g^{i\bar{j}} (\partial_i f)(\partial_{\bar{j}} u).$$

This may be a complex-valued function even if u is real-valued. We define a hermitian inner product on the space $C_0^\infty(M, \mathbb{C})$ by

$$\langle u, v \rangle_{L^2_f} := \int_M u \bar{v}\, e^{-f}\, dV_g \quad \text{for } u, v \in C_0^\infty(M, \mathbb{C}).$$

Then Δ_f is symmetric with respect to this inner product. In fact, we have that

$$\int_M (\Delta_f u) \bar{v}\, e^{-f}\, dV_g = \int_M u \overline{\Delta_f v}\, e^{-f}\, dV_g = -\int_M g(\bar{\partial} u, \bar{\partial} v)\, e^{-f}\, dV_g = -\langle \bar{\partial} u, \bar{\partial} v \rangle_{L^2_f},$$

where

$$g(\bar{\partial} u, \bar{\partial} v) = g^{i\bar{j}} (\partial_{\bar{j}} u)(\partial_i \bar{v}).$$

See [33] for further details.

3 Proof of Theorem A

We first consider Theorem A in the expanding case.

3.1 Construction of a map to the tangent cone

By a result of Siepmann [62, Theorem 4.3.1], a complete expanding gradient Ricci soliton (M, g, X) with quadratic curvature decay with derivatives has a unique tangent cone along each end. We first prove a series of lemmas before providing a refinement of Siepmann's result in Theorem 3.8 by using the flow of the soliton vector field X to construct a diffeomorphism between each end of the expanding Ricci soliton and its tangent cone (C_0, g_0) along that end, with respect to which $r\,\partial_r$ pushes forward to $2X$ (here r denotes the radial coordinate of g_0), and with respect to which $g - g_0 - \mathrm{Ric}(g_0) = O(r^{-4})$ with derivatives.

Our set-up in this section is as follows:

> (M, g, X) is a complete expanding gradient Ricci soliton with soliton vector field $X = \nabla^g f$ for a smooth real-valued function $f : M \to \mathbb{R}$ such that for some point $p \in M$ and all $k \in \mathbb{N}_0$,
>
> $$A_k(g) := \sup_{x \in M} |(\nabla^g)^k \,\mathrm{Rm}(g)|_g(x)\, d_g(p, x)^{2+k} < \infty,$$
>
> where $\mathrm{Rm}(g)$ denotes the curvature of g and $d_g(p, x)$ denotes the distance between p and x with respect to g.

The diffeomorphisms $(\varphi_t)_{t \in (0,1]}$ will be as in (2-3). We begin with:

Lemma 3.1 *The one-parameter family of functions $(tf\circ\varphi_t)_{t>0}$ converges to a nonnegative continuous real-valued function $q(x) := \lim_{t\to 0^+} tf(\varphi_t(x))$ on M as $t\to 0^+$.*

Proof Since
$$\frac{\partial}{\partial t}f(\varphi_t(x)) = -\frac{|\nabla^g f|_g^2(\varphi_t(x))}{t},$$
so that
$$\frac{\partial}{\partial t}(tf(\varphi_t(x))) = (f-|\nabla^g f|_g^2)(\varphi_t(x)) = R_g(\varphi_t(x)),$$
by the soliton identities for expanding gradient Ricci solitons (see Lemma 2.22), where R_g denotes the scalar curvature of g, we see after integrating that

$$tf(\varphi_t(x)) - sf(\varphi_s(x)) = \int_s^t R_g(\varphi_\tau(x))\,d\tau \quad \text{for } 0<s\le t. \tag{3-1}$$

Since R_g is bounded on M, it follows that for all $x\in M$,

$$|tf(\varphi_t(x)) - sf(\varphi_s(x))| \le C(t-s) \quad \text{for } 0<s\le t \tag{3-2}$$

for some positive constant C. Thus, $\{t(f\circ\varphi_t)\}_{t\in(0,1]}$ is a Cauchy sequence in $C^0(M)$ and hence converges uniformly as $t\to 0^+$ to a continuous real-valued function q on M as in the statement of the lemma.

To see that $q(x)\ge 0$ for all $x\in M$, note from the soliton identities that
$$tf(\varphi_t(x)) = t|X|^2(\varphi_t(x)) + tR_g(\varphi_t(x)) \ge tR_g(\varphi_t(x)) \ge t\inf_M R_g$$
since $t\in(0,1]$ and R_g is bounded from below. Letting $t\to 0^+$ in this inequality yields the desired conclusion. □

For $a>0$, set
$$M_a := \{x\in M \mid \lim_{t\to 0^+} tf(\varphi_t(x)) > a\}, \quad M_0 := \{x\in M \mid \lim_{t\to 0^+} tf(\varphi_t(x)) > 0\}.$$
In view of (3-2) and the consequences discussed thereafter, the sets M_0 and M_a are well-defined for $a>0$. Furthermore, note that

- M_a and M_0 are preserved by φ_t for all $t\in(0,1]$ and $a>0$, since $\varphi_{ts} = \varphi_t\circ\varphi_s$ for all positive times s and t.

Hence for any $a\ge 0$, $t(f\circ\varphi_t)$ defines a family of smooth functions $t(f\circ\varphi_t)\colon M_a\to\mathbb{R}$ for $t\in(0,1]$.

Lemma 3.2 *As $t\to 0^+$, $t(f\circ\varphi_t)$ converges to q in $C^\infty_{\mathrm{loc}}(M_0)$. In particular, q is smooth on M_0.*

Proof For each $a>0$, [62, Lemma 4.3.3] ensures that along the Ricci flow $g(t)$ defined by g, the norm of the curvature tensor $\mathrm{Rm}_{g(t)}$ of $g(t)$ is bounded with respect to $g(t)$ when restricted to M_a. In particular, there exists a positive constant C (depending on a) such that for all $t\in(0,1]$,
$$\sup_{x\in M_a} |\mathrm{Ric}(g(t))|_{g(t)}(x) \le C.$$

By definition of a Ricci flow, this implies that the metrics $(g(t))_{t\in(0,1]}$ are uniformly equivalent, ie there exists a positive constant C (that may vary from line to line) such that

$$C^{-1}g(x) \le g(t)(x) \le Cg(x) \quad \text{for all } t \in (0,1],\ x \in M_a. \tag{3-3}$$

An induction argument (see [62, Lemma 4.3.6]) then shows that for all $x \in M_a$, $t \in (0,1]$ and $k \ge 0$,

$$|(\nabla^g)^k(g(t))|_g(x) \le C(k,a).$$

Similarly, one obtains that

$$|(\nabla^g)^k(\mathrm{Rm}_{g(t)})|_g(x) \le C(k,a) \tag{3-4}$$

for all $x \in M_a$, $t \in (0,1]$ and $k \ge 0$. As a consequence,

$$|(\nabla^g)^k(tf(\varphi_t(x)))|_g \le C(k,a) \tag{3-5}$$

for all $x \in M_a$, $t \in (0,1]$ and $k \ge 0$. Indeed, by (3-1) with $t=1$ and $s=t$,

$$f(x) - tf(\varphi_t(x)) = \int_t^1 R_g(\varphi_s(x))\,ds = \int_t^1 sR_{g(s)}(x)\,ds \quad \text{for all } t \in (0,1],\ x \in M.$$

In particular, by deriving k times at a point $x \in M_a$, we see that

$$(\nabla^g)^k(t\varphi_t^* f) = (\nabla^g)^k f - \int_t^1 s(\nabla^g)^k R_{g(s)}\,ds \quad \text{for all } t \in (0,1],$$

which implies the desired inequality (3-5) after invoking (3-4). As a result, $t(f \circ \varphi_t)$ converges in $C^\infty_{\mathrm{loc}}(M_0)$ as $t \to 0^+$, so that q is smooth on M_0, as claimed. □

Since φ_t preserves M_a for every $a > 0$, we also have that for any $a \ge 0$, the Ricci flow $g(t)$ determined by g, namely $g(t) := t\varphi_t^* g$, defines a family of smooth metrics on M_a for all $t \in (0,1]$. This family converges in $C^\infty_{\mathrm{loc}}(M_0)$ as $t \to 0^+$ as well.

Lemma 3.3 *The family of metrics $g(t)$ converges to a Riemannian metric $\tilde{g}_0$ in $C^\infty_{\mathrm{loc}}(M_0)$ as $t \to 0^+$. Moreover, $\tilde{g}_0 = 2\,\mathrm{Hess}_{\tilde{g}_0} q$.*

Proof From the definition of the Ricci flow, one deduces from the curvature bounds (3-4) that $g(t)$ is a Cauchy sequence in $C^k(M_a)$ for every $k \ge 0$ and $a > 0$, hence converges uniformly locally as $t \to 0^+$ in $C^k(M_0)$ for every $k \ge 0$ to a Riemannian metric $\tilde{g}_0$ on M_0. To see that $\tilde{g}_0 = 2\,\mathrm{Hess}_{\tilde{g}_0} q$, multiply (2-4) across by t and take the limit as $t \to 0^+$, recalling that $\lim_{t\to 0^+} tf(t) = q$ in $C^\infty_{\mathrm{loc}}(M_0)$ by Lemma 3.2. □

We have the following properties of q.

Lemma 3.4 *The function q is proper and bounded below.*

Proof Lemma 3.1 already implies that q is nonnegative. In particular, it is bounded from below. Now, one sees from the quadratic growth of the soliton potential f given by (2-2) that for p in the critical set of the soliton potential f,

$$\tfrac{1}{4}d^2_{g(t)}(p,x) - c_1\sqrt{t}\,d_{g(t)}(p,x) - c_2 t \le tf(\varphi_t(x)) \le \tfrac{1}{4}d^2_{g(t)}(p,x) + c_1\sqrt{t}\,d_{g(t)}(p,x) + c_2 t$$

for all $t > 0$ and $x \in M$, for some constants $c_1, c_2 > 0$, where $d_{g(t)}$ denotes the distance with respect to $g(t)$. Using this inequality and taking the limit as $s \to 0^+$ in (3-2), one finds that

$$\tfrac{1}{4}d^2_{g(t)}(p,x) - c_1\sqrt{t}\,d_{g(t)}(p,x) - c_2 t \le q(x) \le \tfrac{1}{4}d^2_{g(t)}(p,x) + c_1\sqrt{t}\,d_{g(t)}(p,x) + c_2 t \tag{3-6}$$

for all $t > 0$ and $x \in M$, for some constants $c_1, c_2 > 0$ that may now vary from line to line. Thus, $q(x) \to +\infty$ as $x \to \infty$ and the result follows. □

Moreover, we have:

Lemma 3.5 *On M_0,*

(i) $|\nabla^{\tilde{g}_0} q|^2_{\tilde{g}_0} = q$, *so that the integral curves of* $\nabla^{\tilde{g}_0}(2\sqrt{q})$ *on* M_0 *are geodesics, and*

(ii) $\nabla^{\tilde{g}_0} q = X$.

In particular, X *is nowhere vanishing on* M_0.

Proof To prove the first part of (i), namely that $|\nabla^{\tilde{g}_0} q|^2_{\tilde{g}_0} = q$, we multiply equation (2-5) across by t^2 and take the limit as $t \to 0^+$. Next, let $\gamma_x(t)$ denote the integral curve of $\nabla^{\tilde{g}_0}(2\sqrt{q})$ with $\gamma_x(0) = x \in M_0$. Then $\gamma_x(t)$ remains in M_0 for $t \ge 0$. Indeed, so long as $t \ge 0$ is such that $\gamma_x(t)$ lies in M_0, we have that

$$\partial_t(2\sqrt{q})(\gamma_x(t)) = \tilde{g}_0\big(\nabla^{\tilde{g}_0}(2\sqrt{q}), \nabla^{\tilde{g}_0}(2\sqrt{q})\big)(\gamma_x(t)) = |\nabla^{\tilde{g}_0}(2\sqrt{q})|^2_{\tilde{g}_0}(\gamma_x(t)) = 1.$$

After integrating, we deduce that $(2\sqrt{q})(\gamma_x(t)) > 0$ for $t \ge 0$, as desired. To see that these integral curves are in fact geodesics with respect to the metric $\tilde{g}_0$, we compute: for any $t \ge 0$ and for any tangent vector v to M_0 at $\gamma_x(t)$, we have that

$$\tilde{g}_0(\nabla^{\tilde{g}_0}_{\dot{\gamma}_x(t)}\dot{\gamma}_x(t), v) = \tilde{g}_0\big(\nabla^{\tilde{g}_0}_{\nabla^{\tilde{g}_0}(2\sqrt{q})}\nabla^{\tilde{g}_0}(2\sqrt{q}), v\big)|_{\gamma_x(t)} = \mathrm{Hess}_{\tilde{g}_0}(2\sqrt{q})\big(\nabla^{\tilde{g}_0}(2\sqrt{q}), v\big)|_{\gamma_x(t)}. \tag{3-7}$$

Since $\tilde{g}_0 = 2\,\mathrm{Hess}_{\tilde{g}_0} q$ on M_0 by Lemma 3.3, we also have the identity

$$\tfrac{1}{2}\tilde{g}_0 = \mathrm{Hess}_{\tilde{g}_0} q = \mathrm{Hess}_{\tilde{g}_0}(\sqrt{q})^2 = (2\sqrt{q})\,\mathrm{Hess}_{\tilde{g}_0}(\sqrt{q}) + 2d(\sqrt{q}) \otimes d(\sqrt{q}).$$

Plugging this into (3-7) then leads to the following sequence of equalities on M_0:

$$\begin{aligned}(2\sqrt{q})(\gamma_x(t)) \cdot \tilde{g}_0\big(\nabla^{\tilde{g}_0}_{\dot{\gamma}_x(t)}\dot{\gamma}_x(t), v\big) &\\ &= \tilde{g}_0\big(\nabla^{\tilde{g}_0}(2\sqrt{q}), v\big)|_{\gamma_x(t)} - 4\big(d(\sqrt{q}) \otimes d(\sqrt{q})\big)\big(\nabla^{\tilde{g}_0}(2\sqrt{q}), v\big)|_{\gamma_x(t)}\\ &= \tilde{g}_0\big(\nabla^{\tilde{g}_0}(2\sqrt{q}), v\big)|_{\gamma_x(t)} - |\nabla^{\tilde{g}_0}(2\sqrt{q})|^2_{\tilde{g}_0}(\gamma_x(t)) \cdot \tilde{g}_0\big(\nabla^{\tilde{g}_0}(2\sqrt{q}), v\big)|_{\gamma_x(t)} = 0,\end{aligned}$$

where we have used in the last line the already established fact that $|\nabla^{\tilde{g}_0}(2\sqrt{q})|^2_{\tilde{g}_0} = 1$ on M_0.

As for part (ii), let φ_t be as in (2-3). Then on compact subsets of M_0, we have that

$$\begin{aligned}\nabla^{\tilde{g}_0} q &= \lim_{t \to 0^+} \nabla^{g(t)}(tf(t)) = \lim_{t \to 0^+} \nabla^{t\varphi_t^* g}(t\varphi_t^* f)\\ &= \lim_{t \to 0^+} \frac{1}{t}\nabla^{\varphi_t^* g}(t\varphi_t^* f) = \lim_{t \to 0^+} \nabla^{\varphi_t^* g}(\varphi_t^* f)\\ &= \lim_{t \to 0^+} \varphi_t^*(\nabla^g f) = \lim_{t \to 0^+} \varphi_t^* X = \lim_{t \to 0^+} X = X,\end{aligned}$$

where the penultimate equality follows from the fact that φ_t is generated by the flow of $-(1/t)X$ and $\mathcal{L}_X X = 0$. □

The above observations then imply:

Lemma 3.6 *M has only finitely many ends.*

Proof For any $a > 0$, q is a smooth function on M_a by Lemma 3.2. Furthermore, by Lemma 3.5, q has no critical points in M_a. Consequently, using the Morse flow $(\psi_t^q)_{t\geq 0}$ associated to q, one sees that all the level sets of q of the form $q^{-1}(\{b\})$ with $b \geq b_0$ for some $b_0 \in \mathbb{R}$ large enough are diffeomorphic. Since q is proper by Lemma 3.1, such level sets are compact and the map

$$\psi^q : (0, +\infty) \times q^{-1}(\{b_0\}) \to q^{-1}((b_0, +\infty)), \quad (t, x) \mapsto \psi_t^q(x),$$

is a diffeomorphism of a neighbourhood of M at infinity. Again, since q is proper, the level set $q^{-1}(\{b_0\})$ is compact hence has a finite number of connected components. Thus, M has a finite number of ends. □

Our final lemma is then:

Lemma 3.7 *There exists $A > 0$ such that for all $c > A$, the intersection of each end of M with $q^{-1}(\{c\})$ is compact, connected, and nonempty.*

Proof By Lemma 3.6, M has only a finite number of ends. Thus, since q is proper and bounded below by Lemma 3.4, there exists $A > 0$ such that all of the ends of M are contained in $M \setminus q^{-1}((-\infty, A])$. For any $c > A$, the intersection of $q^{-1}(\{c\})$ with each end of M is then compact and nonempty and comprises one connected component only since, as a consequence of Lemma 3.5, q is strictly increasing along the flow lines of the (nowhere vanishing) vector field X on M_0. □

Using the above lemmas, we can now construct our map to the tangent cone at infinity.

Theorem 3.8 (map to the tangent cone for expanding Ricci solitons) *Let $A > 0$ be as in Lemma 3.7, let $\rho\colon M_0 \to \mathbb{R}_+$ be defined by $\rho := 2\sqrt{q}$, and let $S := \rho^{-1}(\{c\})$ for any $c > 0$ with $\frac{1}{4}c^2 > A$ (so that the intersection of each end of M with S is compact, connected, and nonempty). Then there exists a diffeomorphism $\iota\colon (c, \infty) \times S \to M_{c^2/4}$ such that $g_0 := \iota^* \tilde{g}_0 = dr^2 + r^2 g_S/c^2$ and $d\iota(r\partial_r) = 2X$, where r is the coordinate on the (c, ∞)–factor and g_S is the restriction of $\tilde{g}_0$ to S. Moreover, along $M_{c^2/4}$, we have that*

$$|(\nabla^{g_0})^k(\iota^* g - g_0 - 2\,\mathrm{Ric}(g_0))|_{g_0} = O(r^{-4-k}) \quad \text{for all } k \in \mathbb{N}_0. \tag{3-8}$$

In particular, (M, g) has a unique tangent cone along each end.

Proof To prove the first part of this statement, we follow the proof of [22, Theorem 1.7.2].

We have that $\operatorname{Hess}_{\widetilde{g}_0}(\rho^2)=2\widetilde{g}_0$ from Lemma 3.3 and we know from Lemma 3.5(i) that $|\nabla^{\widetilde{g}_0}\rho^2|^2_{\widetilde{g}_0}=4\rho^2$ is constant along the level sets of ρ and that the integral curves of $\nabla^{\widetilde{g}_0}\rho$ are geodesics. Then we have that

$$\nabla^{\widetilde{g}_0}\rho^2=2\rho\nabla^{\widetilde{g}_0}\rho \quad\text{and}\quad \operatorname{Hess}_{\widetilde{g}_0}(\rho^2)=2d\rho^2+2\rho\operatorname{Hess}_{\widetilde{g}_0}(\rho),$$

so that

$$2\widetilde{g}_0=\operatorname{Hess}_{\widetilde{g}_0}(\rho^2)=2d\rho^2+2\rho\operatorname{Hess}_{\widetilde{g}_0}(\rho).$$

Hence,

$$\operatorname{Hess}_{\widetilde{g}_0}(\rho)=\frac{\widetilde{g}_0}{\rho}\quad\text{on the } \widetilde{g}_0\text{–orthogonal complement of } \nabla^{\widetilde{g}_0}\rho.$$

On the other hand, $\widetilde{g}_0=d\rho^2+\widetilde{g}_\rho$ with $\widetilde{g}_\rho$ the restriction of $\widetilde{g}_0$ to the level set of ρ, and

$$\mathcal{L}_{\nabla^{\widetilde{g}_0}\rho}\widetilde{g}_\rho=\mathcal{L}_{\nabla^{\widetilde{g}_0}\rho}(\widetilde{g}_0-d\rho^2)=2\operatorname{Hess}_{\widetilde{g}_0}(\rho)-16\rho d\rho^2=\frac{2\widetilde{g}_\rho}{\rho}+\left(\frac{2}{\rho}-16\rho\right)d\rho^2,$$

so that

$$\mathcal{L}_{\nabla^{\widetilde{g}_0}\rho}\widetilde{g}_\rho=\frac{2\widetilde{g}_\rho}{\rho}\quad\text{on the } \widetilde{g}_0\text{–orthogonal complement of } \nabla^{\widetilde{g}_0}\rho.$$

Thus,

$$\mathcal{L}_{\rho\nabla^{\widetilde{g}_0}\rho}\widetilde{g}_\rho=2\widetilde{g}_\rho\quad\text{on the } \widetilde{g}_0\text{–orthogonal complement of } \nabla^{\widetilde{g}_0}\rho. \tag{3-9}$$

Next define a map $\iota\colon(c,\infty)\times S\to M_{c^2/4}$ by

$$(r,x)\mapsto\Phi_x(r-c),$$

where $\Phi_x(\cdot)$ denotes the flow of $\nabla^{\widetilde{g}_0}\rho$ with $\Phi_x(0)=x$. By choice of c, this map is well-defined. Moreover, $d\iota(\partial_r)=\nabla^{\widetilde{g}_0}\rho$ by construction, and since $\rho(\Phi_x(t))=t+c$, we have that $\iota^*\rho=r$ so that $d\iota(r\partial_r)=\rho\nabla^{\widetilde{g}_0}\rho=2X$. In this new frame, we thus have that

$$\iota^*\widetilde{g}_0=dr^2+\iota^*\widetilde{g}_\rho,$$

where we find from (3-9) that

$$\mathcal{L}_{r\partial_r}\iota^*\widetilde{g}_\rho=2\iota^*\widetilde{g}_\rho\quad\text{on the } \iota^*\widetilde{g}_0\text{–orthogonal complement of } \partial_r.$$

Hence, $\iota^*\widetilde{g}_\rho=r^2g_S/c^2$ so that $\iota^*\widetilde{g}_0=dr^2+r^2g_S/c^2$, as claimed.

As for the fact that (3-8) holds true along $M_{c^2/4}$, we have from Young's inequality applied to (3-6) that

$$C^{-1}d_{g(t)}(p,x)-C\sqrt{t}\le\rho(x)\le Cd_{g(t)}(p,x)+C\sqrt{t}\quad\text{for all } t\in(0,1],\ x\in M_0. \tag{3-10}$$

Using this, we can now prove an estimate less sharp than (3-8).

Claim 3.9 *For all $x\in M_{c^2/4}$ and $k\in\mathbb{N}_0$,*

$$|(\nabla^g)^k(g-\widetilde{g}_0)|_g(x)\le C_k\rho(x)^{-2-k}. \tag{3-11}$$

Proof Let us prove the claim first for $k=0$. Since the curvature tensor of g (and hence that of $g(s)$) decays quadratically with derivatives, we have that for any p in the critical set of f and for any $x\in M_{c^2/4}$,

$$|g-\tilde{g}_0|_g(x)\le\int_0^1|\partial_s g(s)|_g(x)\,ds\le C\int_0^1|\mathrm{Ric}(g(s))|_g(x)\,ds\le C\int_0^1 d_{g(s)}(p,x)^{-2}\,ds\le C\rho(x)^{-2},$$

where we have used (3-3) and (3-10) after increasing C if necessary.

As for the case $k=1$, we must work slightly harder. Recall that if T is a tensor on M, then $\nabla^{g(t)}T=\nabla^g T+g(t)^{-1}*\nabla^g(g(t)-g)*T$, since at the level of Christoffel symbols, one has that

$$\Gamma(g(t))_{ij}^k=\Gamma(g)_{ij}^k+\tfrac{1}{2}g(t)^{km}\bigl(\nabla_i^g(g(t)-g)_{jm}+\nabla_j^g(g(t)-g)_{im}-\nabla_m^g(g(t)-g)_{ij}\bigr).$$

Thus, for all $x\in M_{c^2/4}$ and $t\in(0,1]$, we have that

$$\begin{aligned}\partial_t|\nabla^g(g(t)-g)|_g^2(x)&\ge-4|\nabla^g\,\mathrm{Ric}(g(t))|_g(x)|\nabla^g(g(t)-g)|_g(x)\\&\ge-4\bigl(|\nabla^{g(t)}\,\mathrm{Ric}(g(t))|_g(x)+\bigl(|(\nabla^g-\nabla^{g(t)})\,\mathrm{Ric}(g(t))|_g(x)\bigr)\bigr)|\nabla^g(g(t)-g)|_g(x)\\&\ge-C\bigl(d_{g(t)}(p,x)^{-3}+|(\nabla^g-\nabla^{g(t)})\,\mathrm{Ric}(g(t))|_g(x)\bigr)|\nabla^g(g(t)-g)|_g(x)\\&\ge-C\bigl(\rho(x)^{-3}+|\nabla^g(g(t)-g)|_g(x)|\mathrm{Ric}(g(t))|_g(x)\bigr)|\nabla^g(g(t)-g)|_g(x)\\&\ge-C\bigl(\rho(x)^{-3}+|\nabla^g(g(t)-g)|_g(x)\bigr)|\nabla^g(g(t)-g)|_g(x)\\&\ge-C|\nabla^g(g(t)-g)|^2(x)-C\rho(x)^{-6},\end{aligned}$$

where C denotes a positive constant that may vary from line to line, and where we used Young's inequality in the last line. Recalling that $|\nabla^g(g(t)-g)|_g^2=0$ when $t=1$, one can integrate the previous differential inequality between a time $t\in(0,1)$ and $t=1$ to obtain

$$|\nabla^g(g(t)-g)|_g^2(x)\le C\rho(x)^{-6}\quad\text{for }x\in M_{c^2/4},\ t\in(0,1],$$

for some positive constant C uniform in time. This fact implies the desired estimate (3-11) for $k=1$ by letting $t\to0^+$.

The cases $k\ge2$ are proved by induction on k. ◁

It follows from Claim 3.9 that

$$|(\nabla^{\tilde{g}_0})^k(g-\tilde{g}_0)|_{\tilde{g}_0}\le C_k\rho^{-2-k}\quad\text{for all }k\in\mathbb{N}_0,$$

so that, after pulling back by ι, we have that

$$|(\nabla^{g_0})^k(\iota^*g-g_0)|_{g_0}\le C_k r^{-2-k}\quad\text{for all }k\in\mathbb{N}_0.$$

We now prove (3-8). To this end, recall that $\varphi_t(x)$ satisfies

$$\frac{\partial\varphi_t}{\partial t}(x)=-\frac{\nabla^g f(\varphi_t(x))}{t}\quad\text{and}\quad\varphi_1=\mathrm{id}.$$

Since $4\nabla^g f = \nabla^{g_0}\rho^2 = 2\rho\nabla^{g_0}\rho$ by Lemma 3.5(ii), we have that

$$\begin{aligned}\frac{d}{dt}\rho(\varphi_t(x)) &= d\rho|_{\varphi_t(x)}\left(\frac{\partial\varphi_t}{\partial t}(x)\right) = d\rho|_{\varphi_t(x)}\left(-\frac{\nabla^g f(\varphi_t(x))}{t}\right)\\ &= -\frac{\rho(\varphi_t(x))}{2t}d\rho|_{\varphi_t(x)}(\nabla^{g_0}\rho|_{\varphi_t(x)}) = -\frac{\rho(\varphi_t(x))}{2t},\end{aligned}$$

so that

$$\rho(\varphi_t(x)) = \frac{\rho(x)}{t^{1/2}}. \tag{3-12}$$

Let $\widehat{\varphi}_t(x)$ satisfy

$$\frac{\partial\widehat{\varphi}_t}{\partial t}(x) = -\frac{r}{2t}\frac{\partial}{\partial r}(\widehat{\varphi}_t(x)) \quad \text{and} \quad \widehat{\varphi}_1 = \mathrm{id}.$$

Then since $d\iota(r\partial_r) = 2X$, we have that $\iota\circ\widehat{\varphi}_t = \varphi_t\circ\iota$, and in light of (3-12), we see that $\widehat{\varphi}_t^*r = r/t^{1/2}$, so that $t\widehat{\varphi}_t^*g_0 = g_0$. Recall that the Ricci flow $g(t)$ defined by g is given by $g(t) = t\varphi_t^*g$ for $t\in(0,1]$. Together with the scaling properties of the norm induced on tensors by g_0 and the invariance of the Levi-Civita connection under rescalings, these observations imply that

$$\begin{aligned}|(\nabla^{\widehat{\varphi}_t^*g_0})^k(\iota^*g(t)-g_0)|_{\widehat{\varphi}_t^*g_0}(x) &= |(\nabla^{t^{-1}g_0})^k(\iota^*g(t)-g_0)|_{t^{-1}g_0}(x)\\ &= t^{1+(k/2)}|(\nabla^{g_0})^k(\iota^*g(t)-g_0)|_{g_0}(x),\end{aligned}$$

so that

$$\begin{aligned}|(\nabla^{g_0})^k(\iota^*g(t)-g_0)|_{g_0}(x) &= t^{-1-k/2}|(\nabla^{\widehat{\varphi}_t^*g_0})^k(\iota^*g(t)-g_0)|_{\widehat{\varphi}_t^*g_0}(x)\\ &= t^{-1-k/2}\cdot t|(\nabla^{\widehat{\varphi}_t^*g_0})^k(\widehat{\varphi}_t^*\iota^*g-\widehat{\varphi}_t^*g_0)|_{\widehat{\varphi}_t^*g_0}(x)\\ &= t^{-1-k/2}\cdot t|(\nabla^{g_0})^k(\iota^*g-g_0)|_{g_0}(\widehat{\varphi}_t(x))\\ &\le C_k t\cdot t^{-1-k/2}\cdot(r(\widehat{\varphi}_t(x)))^{-2-k}\\ &= C_k t\cdot t^{-1-k/2}\cdot t^{1+k/2}(r(x))^{-2-k} \quad \text{for all } k\in\mathbb{N}_0;\end{aligned}$$

that is,

$$|(\nabla^{g_0})^k(\iota^*g(t)-g_0)|_{g_0}(x) \le C_k t r^{-2-k} \quad \text{for all } k\in\mathbb{N}_0.$$

In particular,

$$|(\nabla^{g_0})^k(\mathrm{Ric}(\iota^*g(t))-\mathrm{Ric}(g_0))|_{g_0} \le C_k t r^{-4-k} \quad \text{for all } k\in\mathbb{N}_0,$$

which is clear from the expression of the components of the Ricci curvature in local coordinates. Consequently, we have the improved estimate

$$|(\nabla^{g_0})^k(\iota^*g-g_0-2\,\mathrm{Ric}(g_0))|_{g_0}(x) \le C_k\int_0^1 \big|(\nabla^{g_0})^k\big(\mathrm{Ric}(\iota^*g(s))-\mathrm{Ric}(g_0)\big)\big|_{g_0}(x)\,ds \le C_k r(x)^{-4-k}.$$

This is precisely (3-8). □

Thus, an expanding gradient Ricci soliton M with quadratic curvature decay with derivatives has a unique tangent cone C_0 along each of its ends V. Moreover, there is a diffeomorphism

$$\iota\colon C_0\setminus K\to V,$$

where $K \subset C_0$ is a compact subset containing the apex of C_0, induced by the flow of the vector field $2X/\rho$. The statement of Theorem 3.8 is verbatim the same for expanding gradient Kähler–Ricci solitons, except that $d\iota(r\partial_r) = X$ rather than $2X$ and $2\,\mathrm{Ric}(g_0)$ is replaced by $\mathrm{Ric}(g_0)$ in (3-8), accounting for the difference in normalisation between Ricci solitons and Kähler–Ricci solitons.

3.2 Existence of a resolution map to the tangent cone

In the case that (M, g, X) is a complete expanding gradient Kähler–Ricci soliton with quadratic curvature decay with derivatives, the soliton potential is proper [12], hence M has only one end V [52] with tangent cone C_0 along the end. (Also note from [52] that any complete shrinking gradient Kähler–Ricci soliton has only one end without any curvature assumption on the metric.) Along V, we have from Theorem 3.8 the diffeomorphism $\iota\colon C_0 \setminus K \to V$ for $K \subset C_0$ a compact subset containing the apex of C_0. As we will now see, the inverse of this map actually extends to define a resolution $\pi : M \to C_0$ with respect to which $d\pi(X) = r\partial_r$. We first show that ι is a biholomorphism with respect to a complex structure on C_0 that makes the cone metric g_0 Kähler. As the next proposition demonstrates, the Kählerity of the soliton implies this fact.

Proposition 3.10 *Let (M, g, X) be a complete expanding gradient Kähler–Ricci soliton with complex structure J such that for some point $p \in M$ and all $k \in \mathbb{N}_0$,*

$$A_k(g) := \sup_{x \in M} |(\nabla^g)^k \,\mathrm{Rm}(g)|_g(x) d_g(p, x)^{2+k} < \infty,$$

where $\mathrm{Rm}(g)$ denotes the curvature of g and $d_g(p, x)$ denotes the distance between p and x with respect to g. For the unique end V of M, let $\tilde{g}_0 = \lim_{t\to 0^+} g(t)$ be the limit of the Kähler–Ricci flow $g(t)$ defined by (M, g, X), let (C_0, g_0) be the unique tangent cone along V with radial function r and let $\iota\colon (C_0 \setminus K, g_0) \to (V, \tilde{g}_0)$, for $K \subset C_0$ compact containing the apex of C_0, be the isometry of Theorem 3.8. Then (C_0, g_0) is a Kähler cone with respect to $\iota^* J$. In particular, $\iota\colon (C_0 \setminus K, \iota^* J) \to (V, J)$ is a biholomorphism.

Proof Since $\lim_{t\to 0^+} g(t) = \tilde{g}_0$ smoothly on compact subsets of V and $g(t)$ is Kähler with respect to J, we have that on V,

$$\nabla^{\tilde{g}_0} J = \lim_{t\to 0^+} \nabla^{g(t)} J = 0,$$

so that $\tilde{g}_0$ is Kähler with respect to J. The metric g_0 is therefore Kähler with respect to $\iota^* J$ away from a compact subset of C_0. Recall that the radial vector field on a Kähler cone is holomorphic with respect to its complex structure. Thus, $r\partial_r$ is holomorphic on the subset of C_0 for which $\iota^* J$ is defined. Flowing along $-r\partial_r$ then extends $\iota^* J$ to a global complex structure on C_0, with respect to which g_0 is Kähler. □

In fact the converse of Proposition 3.10 holds true for shrinking gradient Kähler–Ricci solitons; see [42] for details.

The previous proposition implies that M is 1–convex. This property is what allows us to extend the biholomorphism ι^{-1} to a resolution $\pi: M \to C_0$ that is equivariant with respect to the torus action on C_0 generated by the flow of $J_0 r\partial_r$. The details are contained in the next theorem.

Theorem 3.11 *Let (M, g, X) be a complete expanding gradient Kähler–Ricci soliton with complex structure J such that for some point $p \in M$ and all $k \in \mathbb{N}_0$,*

$$A_k(g) := \sup_{x\in M} |(\nabla^g)^k \operatorname{Rm}(g)|_g(x) d_g(p,x)^{2+k} < \infty,$$

and let (C_0, g_0) be its unique tangent cone with radial function r and complex structure J_0. Then there exists a holomorphic map $\pi: M \to C_0$ that is a resolution of C_0 with the property that $d\pi(X) = r\partial_r$. Furthermore, the holomorphic isometric real torus action on (C_0, g_0, J_0) generated by $J_0 r\partial_r$ extends to a holomorphic isometric torus action of (M, g, J).

Proof The proof of Theorem 3.11 comprises several steps. From Proposition 3.10, we know that along the unique end V of M, there is a biholomorphism $\iota: C_0 \setminus K \to V$ for $K \subset C_0$ a compact subset containing the apex of C_0. Thus, by [20, Lemma 2.15], this in particular implies that M is 1–convex, hence holomorphically convex, so that there is a Remmert reduction $p: M \to M'$ of M. Recall that this is a proper holomorphic map $p: M \to M'$ from M onto a normal Stein space M' with finitely many isolated singularities obtained by contracting the maximal compact analytic subset of M. By construction, M' is biholomorphic to M outside compact sets, therefore we have a biholomorphism given by $F := p \circ \iota: \{x \in C_0 \mid r(x) > R\} \to M' \setminus K'$ for some compact subset $K' \subset M'$ and for some $R > 0$. We claim that this biholomorphism extends globally.

Claim 3.12 *The biholomorphism $F: \{x \in C_0 \mid r(x) > R\} \to M' \setminus K'$ extends to a biholomorphism $\bar{F}: C_0 \to M'$.*

Proof Since M', as a Stein space with finitely many isolated singularities, admits an embedding $h: M' \to \mathbb{C}^P$ for some P by [3, Theorem 3.1], we have a holomorphic function

$$h \circ F: \{x \in C_0 \mid r(x) > R\} \to \mathbb{C}^P.$$

Since C_0 is in particular an example of a Stein space, this holomorphic function extends to a unique holomorphic function $\bar{F}: C_0 \to \mathbb{C}^P$ by Hartogs' theorem for Stein spaces [61, Theorem 6.6]. The fact that $\bar{F}(C_0) \subseteq M'$ follows from Hartogs' theorem. To show that $\bar{F}$ is in fact a biholomorphism, we construct an inverse map $\bar{F}^{-1}: M' \to C_0$ as an extension of the map

$$F^{-1}: M' \setminus K' \to \{x \in C_0 \mid r(x) > R\}$$

by applying the previous argument beginning with the fact that C_0 is affine algebraic. ◁

Thus, the Remmert reduction of M is actually C_0, ie the composition $\pi := \bar{F}^{-1} \circ p: M \to C_0$ is a proper holomorphic map contracting the maximal compact analytic subset E of M to obtain the cone C_0. Denote the connected components of E by $E_1, \dots, E_k$. Then π contracts each E_i to a point $p_i \in C_0$ and restricts to a biholomorphism $\pi\colon M\setminus E\to C_0\setminus\{p_1,\dots,p_k\}$. We next show that π defines a resolution of C_0 for which $d\pi(X)=r\partial_r$. Note that at infinity, $\pi=(p\circ\iota)^{-1}\circ p=\iota^{-1}$.

Claim 3.13 *The map $\pi:=\bar{F}^{-1}\circ p\colon M\to C_0$ is a resolution of C_0 with respect to which $d\pi(X)=r\partial_r$.*

Proof Consider the biholomorphism $\pi\colon M\setminus E\to C_0\setminus\{p_1,\dots,p_k\}$. This map allows us to lift the holomorphic vector field $r\partial_r$ to a holomorphic vector field $Y:=(d\pi)^{-1}(r\partial_r)$ on $M\setminus E$. Since at infinity $\pi=\iota^{-1}$, and so identifies the vector field $r\partial_r$ on C_0 with the vector field X on M outside compact subsets of each, Y will agree with X outside of a compact subset of M. Thus, analyticity implies that $X=Y$ on $M\setminus E$. The next observation is that since the flow lines of Y (and hence X) foliate $M\setminus E$, the flow of X must preserve E. Via π therefore, the flow of X induces a flow on C_0 that fixes the points $p_1,\dots,p_k$, where as before each p_i denotes the image of a connected component E_i of $E\subset M$ under π. The result of this induced flow on C_0 is a holomorphic vector field $\hat{X}$ that coincides with $r\partial_r$ on $C_0\setminus\{p_1,\dots,p_k\}$ and which is equal to zero at each p_i. By analyticity again, $\hat{X}=r\partial_r$, so that E comprises one connected component only, which is mapped to the apex of the cone by π. Thus, $\pi\colon M\to C_0$ is a resolution of the singularity of the cone, and the vector field X on M is an extension of $(d\pi)^{-1}(r\partial_r)$ from $M\setminus E$ to M so that $d\pi(X)=r\partial_r$, as claimed. ◁

The resolution $\pi\colon M\to C_0$ is clearly equivariant with respect to the flow of $J_0r\partial_r$ on C_0 and the flow of JX on M. We wish to show next that $\pi\colon M\to C_0$ is in fact equivariant with respect to the holomorphic isometric torus action on C_0 induced by the flow of $J_0r\partial_r$ and that the lift of this torus action to M acts isometrically on g. This will conclude the proof of Theorem 3.11.

Claim 3.14 *The holomorphic isometric torus action on (C_0,g_0) generated by $J_0r\partial_r$ extends to a holomorphic isometric action of (M,g,J) so that, in particular, $\pi\colon M\to C_0$ is equivariant with respect to this torus action.*

Proof Consider the isometry group of (M,g) that fixes E endowed with the topology induced by uniform convergence on compact subsets of M. By the Arzelà–Ascoli theorem, this is a compact Lie group. Taking the closure of the flow of JX in this group therefore yields the holomorphic isometric action of a torus T on (M,J,g). Since the action of T preserves E, this action pushes down via π to a holomorphic action of T on C_0 fixing the apex o of C_0. Now, by Theorem 3.8, after noting again that $\pi=\iota^{-1}$ at infinity, we see that the soliton metric g and the cone metric g_0 are asymptotic at infinity. Therefore these metrics are quasi-isometric on $C_0\setminus K$, where $K\subset C_0$ is any compact subset of C_0 containing the apex o of C_0, so that uniform convergence on compact subsets of $C_0\setminus\{o\}$ measured with respect to g and g_0 are equivalent. Recall that $d\pi(X)=r\partial_r$, so that the flow of $J_0r\partial_r$ is dense in T and the flow of $J_0r\partial_r$ is isometric with respect to g_0. Consequently, every automorphism of (C_0,J_0) induced by T is obtained as a limit of automorphisms of $(C_0\setminus\{o\},g_0,J_0)$ with respect to uniform convergence on compact subsets measured using g_0. Since a uniform limit of isometries is itself an isometry, it follows that T acts isometrically with respect to g_0 on $C_0 \setminus \{o\}$, so that the action of T on C_0 preserves the slices of C_0 and defines a torus in the isometry group of the link of C_0 in which the flow of $J_0 r\partial_r$ is dense. This final observation concludes the proof of the claim and Theorem 3.11. □

3.3 Conclusion of the proof of Theorem A

We now conclude the proof of Theorem A for complete expanding gradient Kähler–Ricci solitons. Conclusion (a) follows from [62, Theorem 4.3.1], whereas the Kählerity of (C_0, g_0) as stated in conclusion (b) follows from Proposition 3.10. The remainder of conclusion (b), apart from (b)(i), then follows from Theorem 3.11. Conclusion (c) follows from Theorem 3.8 after noting that $\pi = \iota^{-1}$ at infinity as above.

As for conclusion (b)(i), the Kähler form ω of the expanding gradient Kähler–Ricci soliton satisfies the expanding soliton equation $\rho_\omega + i\partial\bar{\partial} f = -\omega$ on M, where ρ_ω is the Ricci form of ω and f is the soliton potential. In $H^2(M)$, this equation yields $[-\rho_\omega] = [\omega]$. Since $i\rho_\omega$ is the curvature form Θ resulting from the hermitian metric on K_M induced by ω, we have that $[i\Theta] = [-\rho_\omega] = [\omega]$, so that (1-5) is seen to hold true for the expanding soliton Kähler form ω and the curvature form $i\Theta$ it induces on K_M.

For a complete shrinking gradient Ricci soliton (M, g, X) with soliton potential f, we define a Kähler–Ricci flow via

$$g(t) = -t\varphi_t^* g \quad \text{for } t < 0,$$

where φ_t is a family of diffeomorphisms generated by the gradient vector field $-(1/t)X$ with $\varphi_{-1} = \mathrm{id}$, ie

$$\frac{\partial \varphi_t}{\partial t}(x) = -\frac{\nabla^g f(\varphi_t(x))}{t} \quad \text{with} \quad \varphi_{-1} = \mathrm{id}.$$

Then $(\partial g/\partial t)(t) = -2\operatorname{Ric}(g(t))$ for $t < 0$ and $g(-1) = g$. Such a soliton with quadratic curvature decay has quadratic curvature decay with derivatives by Theorem 2.20(iii), and hence, as proved in [43, Sections 2.2–2.3], has a unique tangent cone at infinity. These observations provide the starting point for the proof of Theorem A for complete shrinking gradient Kähler–Ricci solitons with quadratic curvature decay. The proof then follows the proof for the expanding case, verbatim.

4 Classification results for expanding gradient Kähler–Ricci solitons with quadratic curvature decay with derivatives

4.1 Proof of Corollary B

Let (M, g, X) be a complete expanding gradient Kähler–Ricci soliton satisfying (1-6) with tangent cone (C_0, g_0), as in Corollary B. Let ω denote the Kähler form of g.

To see that M is the canonical model of C_0, note first that Theorem A asserts that there is a Kähler resolution $\pi\colon M \to C_0$ with exceptional set E such that

$$\int_V (i\Theta)^k \wedge \omega^{\dim_{\mathbb{C}} V - k} > 0 \tag{4-1}$$

for all positive-dimensional irreducible analytic subvarieties $V \subset E$ of $\pi : M \to C_0$ and for all integers k such that $1 \le k \le \dim_{\mathbb{C}} V$, where Θ denotes the curvature form of the hermitian metric on K_M induced by ω. In particular, (4-1) implies that

$$\int_V (i\Theta)^k \wedge \omega^{\dim_{\mathbb{C}} V - k} > 0$$

for all positive-dimensional irreducible *algebraic* subvarieties $V \subset E$ and for all integers k such that $1 \le k \le \dim_{\mathbb{C}} V$. Setting $k = \dim_{\mathbb{C}} V$, we then see that

$$\int_V (i\Theta)^{\dim_{\mathbb{C}} V} > 0 \quad \text{for every irreducible algebraic subvariety } V \subset E \text{ of positive dimension.}$$

But since M is quasiprojective by Proposition 2.24, this is the same as saying that

$$(D^{\dim_{\mathbb{C}} V} \cdot V) > 0 \quad \text{for every irreducible algebraic subvariety } V \subset E \text{ of positive dimension,}$$

where D is now a canonical divisor of M. Nakai's criterion for a mapping (Theorem 2.11) now tells us that K_M is π–ample, so that by definition, $\pi : M \to C_0$ is the canonical model of C_0. Hence C_0 has a smooth canonical model, namely M.

Conversely, suppose that (C_0, g_0) is a Kähler cone with radial function r and with a smooth canonical model $\pi : M \to C_0$. We begin by explaining that [19, Theorem A] holds true without hypothesis (b) of that theorem. This hypothesis was required in the proof of [19, Proposition 3.2] to show that $\mathcal{L}_X\omega = i\partial\bar{\partial}\theta_X$, where ω is the Kähler form of [19, Proposition 3.1], X is the lift of the radial vector field on the cone, and θ_X is a smooth real-valued function. The following claim asserts that this in fact always holds true.

Claim 4.1 *Let (C_0, g_0) be a Kähler cone with complex structure J_0 and radial function r and let $\pi : M \to C_0$ be an equivariant resolution with respect to the real torus action on C_0 generated by $J_0 r\partial_r$. Let X be the unique holomorphic vector field on M with $d\pi(X) = r\partial_r$ and let ω be the Kähler form of* [19, Proposition 3.1]. *Then $\mathcal{L}_X\omega = i\partial\bar{\partial}\theta_X$ for a smooth real-valued function $\theta_X : M \to \mathbb{R}$.*

Proof Denote the complex structure of M by J and let $X^{1,0} = \frac{1}{2}(X - iJX)$. Then since $\mathcal{L}_{JX}\omega = 0$ by construction, we have that

$$\tfrac{1}{2}\mathcal{L}_X\omega = \tfrac{1}{2}d(\omega \lrcorner X) = d(\omega \lrcorner X^{1,0}). \tag{4-2}$$

Now by construction, ω takes the form $\omega = i\widetilde{\Theta}_h + i\partial\bar{\partial}u$, where $u : M \to \mathbb{R}$ is a smooth real-valued function and $\widetilde{\Theta}_h$ is the average over the action of the torus on M of the curvature form Θ_h of a hermitian metric h on K_M. Thus,

$$\omega \lrcorner X^{1,0} = i\widetilde{\Theta}_h \lrcorner X^{1,0} + i\bar{\partial}(X^{1,0} \cdot u). \tag{4-3}$$

Studying the term $i\widetilde{\Theta}_h \lrcorner X^{1,0}$, let $n = \dim_{\mathbb{C}} M$, let Ω be a local holomorphic volume form on M, ie a nowhere vanishing locally defined holomorphic $(n,0)$–form (defined in some local holomorphic coordinate chart, for example), and set

$$v := X^{1,0} \cdot \log(\|\Omega\|_h^2) - \frac{\mathcal{L}_{X^{1,0}}\Omega}{\Omega},$$

where $\|\cdot\|_h$ denotes the norm with respect to h. We claim that v is independent of the choice of Ω and hence is globally defined. Indeed, any other local holomorphic volume form takes the form $q\Omega$ for some holomorphic function q. Then

$$\begin{aligned} X^{1,0}\cdot\log(\|q\Omega\|_h^2)-\frac{\mathcal{L}_{X^{1,0}}(q\Omega)}{q\Omega} &= X^{1,0}\cdot\log|q|^2+X^{1,0}\cdot\log(\|\Omega\|_h^2)-\frac{(X^{1,0}\cdot q)\Omega+q\mathcal{L}_{X^{1,0}}\Omega}{q\Omega}\\ &= X^{1,0}\cdot\log(\|\Omega\|_h^2)-\frac{\mathcal{L}_{X^{1,0}}\Omega}{\Omega}+\underbrace{X^{1,0}\cdot\log|q|^2-\frac{X^{1,0}\cdot q}{q}}_{=0}\\ &= X^{1,0}\cdot\log(\|\Omega\|_h^2)-\frac{\mathcal{L}_{X^{1,0}}\Omega}{\Omega}, \end{aligned}$$

as required. Next observe that

$$i\Theta_h\lrcorner X^{1,0}=-i\bar\partial\big(X^{1,0}\cdot\log(\|\Omega\|_h^2)\big)=-i\bar\partial\left(X^{1,0}\cdot\log(\|\Omega\|_h^2)-\frac{\mathcal{L}_{X^{1,0}}\Omega}{\Omega}\right)=-i\bar\partial v,$$

since $(\mathcal{L}_{X^{1,0}}\Omega)/\Omega$ is a holomorphic function. Averaging this equation over the action of T then yields the fact that $i\widetilde{\Theta}_h\lrcorner X^{1,0}=i\bar\partial\tilde v$ for a smooth function $\tilde v$ on M. Plugging this into (4-2), we thus see from (4-3) that

$$\mathcal{L}_X\omega=2d(i\bar\partial(\tilde v+X^{1,0}\cdot u))=i\partial\bar\partial(2(\tilde v+X^{1,0}\cdot u)).$$

Hence $\mathcal{L}_X\omega=i\partial\bar\partial\theta_X$, where $\theta_X:=2\operatorname{Re}(\tilde v+X^{1,0}\cdot u)$, because $\mathcal{L}_X\omega$ is a real $(1,1)$–form and $i\partial\bar\partial$ is a real operator. ◁

Remark 4.2 The existence of the function v satisfying $i\Theta_h\lrcorner X^{1,0}=-i\bar\partial v$ is essentially due to the fact that X has a canonical lift to the total space of K_M and Θ_h is the curvature form of a hermitian metric on K_M.

Returning now to our smooth canonical model $\pi\colon M\to C_0$ of C_0, we will verify the hypotheses of [19, Theorem A] (apart from the redundant hypothesis (b) of this theorem) for this resolution to show that M admits a complete expanding gradient Kähler–Ricci soliton g with the desired asymptotics. By Lemma 2.13, the radial vector field $r\partial_r$ on C_0 lifts to a holomorphic vector field X on M with $d\pi(X)=r\partial_r$, and by Lemma 2.12, M is quasiprojective, hence Kähler. Moreover, there exists a Kähler form σ on M and a hermitian metric on K_M with curvature form Θ such that

$$\int_V (i\Theta)^k\wedge\sigma^{\dim_{\mathbb{C}}V-k}>0 \tag{4-4}$$

for all positive-dimensional irreducible analytic subvarieties V contained in the exceptional set E of $\pi\colon M\to C_0$ and for all integers k such that $1\le k\le\dim_{\mathbb{C}}V$. Indeed, proceeding as in [25], let σ be the curvature form of a very ample line bundle L on the projective variety which contains M as an open subset and let Θ be the curvature form of the hermitian metric induced on K_M by σ. Then observe that for any analytic subvariety $V\subset E$ of dimension k,

$$\int_V (i\Theta)^k\wedge\sigma^{\dim_{\mathbb{C}}V-k}=\int_{V\cap H_1\cap\cdots\cap H_{\dim_{\mathbb{C}}V-k}}(i\Theta)^k \tag{4-5}$$

for generic members $H_1, \dots, H_{\dim_{\mathbb{C}} V-k}$ of the linear system $|L|$, so that $V \cap H_1 \cap \dots \cap H_{\dim V-k} \subset E$ is an irreducible subvariety of dimension k. Since E is projective (as M is quasiprojective), this intersection is a projective algebraic variety by Chow's theorem. The right-hand side of (4-5) may therefore be written as $D^k \cdot (V \cap H_1 \cap \dots \cap H_{\dim V-k})$, where D is a canonical divisor of M. By definition of the canonical model, K_M is π–ample, which by Nakai's criterion for a mapping (see Theorem 2.11) implies that this intersection is strictly positive. Thus, we have that (4-4) holds true for the Kähler form σ and the curvature form Θ that it induces on K_M. The hypotheses required for the application of [19, Theorem A] are therefore satisfied and so M admits a complete expanding gradient Kähler–Ricci soliton (M, g, X) with

$$|(\nabla^{g_0})^k(\pi_* g - g_0 - \mathrm{Ric}(g_0))|_{g_0} \le C_k r^{-4-k} \quad \text{for all } k \in \mathbb{N}_0,$$

as required.

As for the uniqueness of (M, g, X), let (M_i, g_i, X_i) for $i = 1, 2$ be two complete expanding gradient Kähler–Ricci solitons satisfying (1-6) with tangent cone (C_0, g_0). As initially proved, both M_1 and M_2 are equal to the unique (smooth) canonical model M of C_0. Moreover, Theorem A asserts that for $i = 1, 2$ there exists a resolution map $\pi_i \colon M \to C_0$ with $d\pi_i(X_i) = r\partial_r$ such that

$$|(\nabla^{g_0})^k((\pi_i)_* g_i - g_0 - \mathrm{Ric}(g_0))|_{g_0} \le C_k r^{-4-k} \quad \text{for all } k \in \mathbb{N}_0. \tag{4-6}$$

The composition $H := \pi_2 \circ \pi_1^{-1} \colon C_0 \to C_0$ induces an automorphism of C_0 fixing the vertex. As in the proof of Lemma 2.13, uniqueness of the canonical model implies that there exists a unique biholomorphism $F \colon M \to M$ such that $\pi_1 \circ F = H \circ \pi_1$. Unravelling the definition of H, this yields the fact that $\pi_1 \circ F = \pi_2$. Consequently, $d\pi_2((dF)^{-1}(X_1)) = d\pi_1(X_1) = r\partial_r$ so that $(dF)^{-1}(X_1) = X_2$. Furthermore, in light of (4-6), we have that

$$|(\nabla^{g_0})^k((\pi_2)_*(F^* g_1) - g_0 - \mathrm{Ric}(g_0))|_{g_0} \le C_k r^{-4-k} \quad \text{for all } k \in \mathbb{N}_0. \tag{4-7}$$

Thus, $(M, F^* g_1, X_2)$ and (M, g_2, X_2) are two expanding gradient Kähler–Ricci solitons with the same soliton vector field which from (4-6) for $i = 2$ and (4-7) in addition satisfy $|F^* g_1 - g_2| = O(r^{-4})$. The uniqueness theorem [19, Theorem C(ii)] therefore applies (where, in studying the proof of [19, Theorem C(ii)], one sees that finite fundamental group is not actually required) and asserts that $F^* g_1 = g_2$. Thus, (M, g, X) is unique up to pullback by biholomorphisms of M, as claimed.

As for the remainder of Corollary B, item (a) is now clear and item (b) follows from Theorem A.

4.2 Proof of Corollary C

Corollary C follows from Corollary B once we identify the two-dimensional Kähler cones that admit smooth canonical models as those stated in Corollary C(I)–(III) and realise their respective smooth canonical models as those stated in Corollary C(b)(i)–(iii).

To this end, let C_0 be a two-dimensional Kähler cone with a smooth canonical model M. By adjunction, M cannot contain any (-1)– or (-2)–curves. In particular, by Theorem 2.15, M coincides with the minimal model of C_0. Using this information, we can identify C_0 and M as follows.

Since C_0 is a two-dimensional Kähler cone, it must be prescribed as in Theorem 2.5. We henceforth work on a case-by-case basis. If C_0 is as in Theorem 2.5(i), then Γ must be as prescribed in Corollary C(I) since M cannot contain any (-2)–curves; indeed, see [48, Figure 2.1 and Theorem 4.1] for details. In this case, M will be the minimal model of C_0 as stated in Corollary C(b)(i). Otherwise, C_0 may be as in Theorem 2.5(ii) which is precisely the statement of Corollary C(II). In this case, the minimal model M is given as in the statement of Corollary C(b)(ii). Finally, C_0 may be as in Theorem 2.5(iii). Those cones of Theorem 2.5(iii) that admit a smooth canonical model have been identified in Proposition 2.18, which yields the statement of Corollary C(III). For these cones, the minimal resolution is the minimal good resolution which identifies M as in the statement of Corollary C(b)(iii).

5 A volume-minimising principle for complete shrinking gradient Kähler–Ricci solitons

We now focus our attention solely on shrinking gradient Kähler–Ricci solitons for the remainder of the article. The set-up of this section is as follows. Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton of complex dimension n with complex structure J, Kähler form ω, and with soliton vector field $X = \nabla^g f$ for a smooth real-valued function $f : M \to \mathbb{R}$. We assume that a real torus T with Lie algebra $\mathfrak{t}$ acts holomorphically, effectively and isometrically on (M, g, J). Then $\mathfrak{t}$ can be identified with real holomorphic Killing vector fields on M. We furthermore assume that $JX \in \mathfrak{t}$.

The goal of this section is to prove the uniqueness of the soliton vector field JX in $\mathfrak{t}$ by characterising JX as the unique critical point of a soon-to-be-defined weighted volume functional.

5.1 A Matsushima-type theorem

Let $\mathfrak{aut}^X(M)$ denote the Lie algebra of real holomorphic vector fields on M that commute with X and hence JX, and let $\mathfrak{g}^X$ denote the Lie algebra of real holomorphic g–Killing vector fields on M that commute with X and hence JX. Clearly $\mathfrak{g}^X$ is a Lie subalgebra of $\mathfrak{aut}^X(M)$. In order to prove the uniqueness of the soliton vector field X, we need to show that the connected component of the identity of the Lie group of holomorphic isometries of (M, g, J) commuting with the flow of X is maximal compact in the connected component of the identity of the Lie group of automorphisms of (M, J) commuting with the flow of X. This fact will follow from the next theorem, an analogue of Matsushima's theorem [50] for shrinking gradient Kähler–Ricci solitons stating that the Lie algebra $\mathfrak{aut}^X(M)$ is reducible, after we prove that the aforementioned groups are indeed Lie groups.

Theorem 5.1 (a Matsushima theorem for shrinking Kähler–Ricci solitons) *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with complex structure J endowed with the holomorphic, effective, isometric action of a real torus T with Lie algebra $\mathfrak{t}$ with $JX \in \mathfrak{t}$. If $|\mathrm{Ric}(g)|_g$ is bounded, then we have that*

$$\mathfrak{aut}^X(M) = \mathfrak{g}^X \oplus J\mathfrak{g}^X.$$

We expect this theorem to hold true without the assumption of bounded Ricci curvature.

The proof of Theorem 5.1 consists of several steps. Beginning with any real holomorphic vector field $Y \in \mathfrak{aut}^X(M)$, Hörmander's L^2–estimates allow for a complex-valued potential, that is, a smooth complex-valued function u_Y such that $Y^{1,0} = \nabla^{1,0} u_Y$, where $Y^{1,0}$ is the $(1,0)$–part of Y. Thanks to the defining equation of a shrinking gradient Kähler–Ricci soliton, we can then modify u_Y by a holomorphic function if necessary so that $\Delta_\omega u_Y + u_Y - Y^{1,0} \cdot f = 0$, where f is the soliton potential with $X = \nabla^g f$. We further average u_Y over the action of T so that $\mathcal{L}_{JX} u_Y = 0$, which results in the commutator relation $(u_Y)_{\bar{k}} f_k = (u_Y)_k f_{\bar{k}}$. Using this, we then apply a Bochner formula followed by an integration by parts argument to deduce that $\nabla^{0,2} \bar{u}_Y = 0$ so that $\nabla^{1,0} \bar{u}_Y$ is a holomorphic vector field. The bound on the norm of the Ricci curvature is required to control the boundary term in the integration by parts argument. The gradient of the real and imaginary parts of u_Y will therefore be real holomorphic vector fields so that, once one applies the complex structure to these vector fields, they become real holomorphic and Killing. From this, the stated decomposition follows. To conclude that the sum is direct, we make use of a splitting theorem for shrinking gradient Ricci solitons.

Proof Write $\langle \cdot , \cdot \rangle$, $|\cdot|$ and ∇, respectively, for the inner product, norm and Levi-Civita connection determined by g, and let $Y \in \mathfrak{aut}^X(M)$. Then Y defines a real holomorphic vector field on M with $[X, Y] = 0$. Take the $(1,0)$–part $Y^{1,0}$ of Y, ie let $Y^{1,0} = \frac{1}{2}(Y - iJY)$. Then $\bar\partial Y^{1,0} = 0$, so that $\omega \lrcorner Y^{1,0}$ is a $\bar\partial$–closed $(0,1)$–form, where ω denotes the Kähler form of g. We first claim that $\omega \lrcorner Y^{1,0}$ admits a smooth complex potential.

Claim 5.2 *There exists a smooth complex-valued function u_Y on M such that $-i\omega \lrcorner Y^{1,0} = \bar\partial u_Y$, or equivalently, such that $Y^{1,0} = \nabla^{1,0} u_Y$.*

Note that u_Y is unique up to the addition of a holomorphic function.

Proof Let h denote the metric on $-K_M$ induced by ω. Then the curvature of the metric $e^{-f} h$ on $-K_M$ is precisely ω by virtue of the defining equation of a shrinking gradient Kähler–Ricci soliton. Treat $\omega \lrcorner Y^{1,0}$ as a $-K_M$–valued $(n,1)$–form. Then since the norm of $\mathrm{Ric}(g)$ is bounded so that $|Y^{1,0}|$ grows at most polynomially by Proposition 2.30, we see from the growth on f dictated by Theorem 2.20(i) that the L^2–norm of $\omega \lrcorner Y^{1,0}$ measured with respect to $e^{-f} h$ is finite. An application of Hörmander's L^2–estimates [24, Theorem 6.1, page 376] now yields the desired conclusion. ◁

Next, contracting (1-4) with $\lambda = 1$ with $Y^{1,0}$ and using the Bochner formula, we see that

$$-i\bar\partial \Delta_\omega u_Y + i\bar\partial (Y^{1,0} \cdot f) = i\bar\partial u_Y,$$

so that

$$\bar\partial (\Delta_\omega u_Y + u_Y - Y^{1,0} \cdot f) = 0.$$

By adding a holomorphic function to u_Y if necessary, we may therefore assume that

$$\Delta_\omega u_Y + u_Y - Y^{1,0} \cdot f = 0. \tag{5-1}$$

Furthermore, by averaging u_Y over the action of T, we may assume that $\mathcal{L}_{JX}u_Y = 0$. These two operations normalise u_Y. Notice that

$$(u_Y)_{\bar{k}} f_k = Y^{1,0} \cdot f = \nabla^{1,0} u_Y \cdot f = \tfrac{1}{2}\langle \nabla u_Y - iJ\nabla u_Y, X\rangle = \tfrac{1}{2}\langle \nabla u_Y, X\rangle,$$

by virtue of the fact that $\mathcal{L}_{JX}u_Y = 0$. For the same reason, we also have that

$$(u_Y)_k f_{\bar{k}} = Y^{0,1} \cdot f = (\nabla^{0,1} u_Y) \cdot f = \tfrac{1}{2}\langle \nabla u_Y, X\rangle.$$

Hence,

$$(u_Y)_{\bar{k}} f_k = (u_Y)_k f_{\bar{k}}.$$

In particular, from (5-1) we deduce that

$$\Delta_\omega u_Y + u_Y - (u_Y)_k f_{\bar{k}} = 0. \tag{5-2}$$

Before continuing, we need to establish some estimates on u_Y together with its covariant derivatives. We will divide these estimates up into three claims.

Claim 5.3 *There exists a positive constant A such that $u_Y(x) = O(d_g(p,x)^A)$ as $d_g(p,x)$ tends to $+\infty$.*

Proof By Proposition 2.30, $Y^{1,0}$ grows polynomially, ie $|Y^{1,0}|(x) = O(d_g(p,x)^a)$ for some $a > 0$, where $d_g(p,\cdot)$ denotes the distance with respect to g to a fixed point $p \in M$, so that $|\bar{\partial} u_Y|(x) = O(d_g(p,x)^a)$. Then

$$\bar{\partial} u_Y(X) = \tfrac{1}{2}(du_Y(X) + i\underbrace{du_Y(JX)}_{=0}) = \tfrac{1}{2} X \cdot u_Y.$$

Thus,

$$|X \cdot u_Y| = 2|\bar{\partial} u_Y(X)| = O(d_g(p,x)^{a+1}). \tag{5-3}$$

Let $\gamma_x(t)$ be an integral curve of X with $\gamma_x(0) = x \in M$. Then

$$u_Y(\gamma_x(t)) = u_Y(\gamma_x(0)) + \int_0^t \frac{d}{ds} u_Y(\gamma_x(s))\,ds = u_Y(\gamma_x(0)) + \int_0^t (X \cdot u_Y)(\gamma_x(s))\,ds = C + O(e^{(a+1)t}),$$

so that, by (2-11) and Theorem 2.20(i),

$$|u_Y(x)| = O(d_g(p,x)^{a+1}). \qquad \triangleleft$$

The next claim concerns the weighted L^2–integrability of the total gradient and second covariant derivatives of u.

Claim 5.4 *The gradient ∇u_Y and the second covariant derivatives $\nabla^2 u_Y$ of u_Y belong to $L^2(e^{-f}\omega^n)$.*

Proof Since $\Delta_\omega u_Y = Y^{1,0} \cdot f - u_Y$, the estimate established in Claim 5.3 together with the polynomial growth of X and Y at infinity show that $\Delta_\omega u_Y$ is growing at most polynomially at infinity as well. By (5-3), the same holds true for the drift term $X \cdot u_Y$. Therefore the drift Laplacian $\Delta_\omega u_Y - \frac{1}{2} X \cdot u_Y$ is growing at most polynomially at infinity ensuring its weighted L^2–integrability, ie $\Delta_{\omega,X} u_Y := \Delta_\omega u_Y - \frac{1}{2} X \cdot u_Y \in L^2(e^{-f}\omega^n)$. This implies in turn that $\nabla \operatorname{Re}(u_Y)$ and $\nabla \operatorname{Im}(u_Y)$ belong to $L^2(e^{-f}\omega^n)$. Indeed, by the previous arguments, it suffices to show that if a smooth real-valued function $v\colon M \to \mathbb{R}$ satisfies $v \in L^2(e^{-f}\omega^n)$ and $\Delta_{\omega,X} v \in L^2(e^{-f}\omega^n)$, then $\nabla v \in L^2(e^{-f}\omega^n)$.

To this end, let R be a positive real number and let $\phi_R\colon M\to[0,1]$ be a cut-off function with compact support in the geodesic ball $B_g(p,2R)$ such that $\phi_R=1$ on $B_g(p,R)$ and $|\nabla\phi_R|_g\le c/R$. Then since $(\Delta_{\omega,X})v^2=2|\nabla v|^2+2\langle\Delta_{\omega,X}v,v\rangle$, integration by parts leads to the inequality

$$\begin{aligned}2\int_M|\nabla v|^2\phi_R^2\,e^{-f}\omega^n&=\int_M\Delta_{\omega,X}v^2\phi_R^2e^{-f}\omega^n-2\int_M\langle\Delta_{\omega,X}v,v\rangle\phi_R^2\,e^{-f}\omega^n\\&=-\int_M\langle\nabla v^2,\nabla\phi_R^2\rangle e^{-f}\omega^n-2\int_M\langle\Delta_{\omega,X}v,v\rangle\phi_R^2\,e^{-f}\omega^n\\&\le\int_M|\nabla v|^2\phi_R^2\,e^{-f}\omega^n+\frac{c}{R^2}\int_M v^2e^{-f}\omega^n+\int_M(|\Delta_{\omega,X}v|^2+|v|^2)e^{-f}\omega^n,\end{aligned}$$

which yields

$$\int_M|\nabla v|^2\phi_R^2\,e^{-f}\omega^n\le\frac{c}{R^2}\int_M v^2e^{-f}\omega^n+\int_M(|\Delta_{\omega,X}v|^2+|v|^2)e^{-f}\omega^n.$$

One then obtains the expected result for the gradient by letting R tend to $+\infty$.

Similarly, for the second covariant derivatives, it suffices to show that if a smooth real-valued function $v\colon M\to\mathbb{R}$ satisfies $\nabla v\in L^2(e^{-f}\omega^n)$ and $\Delta_{\omega,X}v\in L^2(e^{-f}\omega^n)$, then $\nabla^2v\in L^2(e^{-f}\omega^n)$. To this end, we apply the Bochner formula and use the soliton equation as follows:

$$\begin{aligned}\Delta_{\omega,X}|\nabla v|^2&=|\nabla^2v|^2+\big(\mathrm{Ric}(g)+\tfrac12\mathcal{L}_Xg\big)(\nabla v,\nabla v)+\langle\nabla(\Delta_{\omega,X}v),\nabla v\rangle\\&=|\nabla^2v|^2+|\nabla v|^2+\langle\nabla(\Delta_{\omega,X}v),\nabla v\rangle\\&\ge|\nabla^2v|^2+\langle\nabla(\Delta_{\omega,X}v),\nabla v\rangle.\end{aligned}\tag{5-4}$$

Next, let $\phi_R\colon M\to[0,1]$ be the cut-off function defined as above. Then using integration by parts, the identity (5-4) leads to the inequalities

$$\begin{aligned}2\int_M|\nabla^2v|^2\phi_R^2e^{-f}\omega^n&\le2\int_M\Delta_{\omega,X}|\nabla v|^2\phi_R^2e^{-f}\omega^n-2\int_M\langle\nabla(\Delta_{\omega,X}v),\nabla v\rangle\phi_R^2e^{-f}\omega^n\\&=-\int_M\langle\nabla|\nabla v|^2,\nabla\phi_R^2\rangle e^{-f}\omega^n+2\int_M(2|\Delta_{\omega,X}v|^2\phi_R^2+\Delta_{\omega,X}v\langle\nabla v,\nabla\phi_R^2\rangle)e^{-f}\omega^n\\&\le\int_M|\nabla^2v|^2\phi_R^2e^{-f}\omega^n+\frac{c}{R^2}\int_M|\nabla v|^2e^{-f}\omega^n+c\int_M(|\Delta_{\omega,X}v|^2+|\nabla v|^2)e^{-f}\omega^n\end{aligned}$$

for some positive constant c independent of R. Thus,

$$\int_M|\nabla^2v|^2\phi_R^2\,e^{-f}\omega^n\le\frac{c}{R^2}\int_M|\nabla v|^2\,e^{-f}\omega^n+\int_M(|\Delta_{\omega,X}v|^2+|\nabla v|^2)\,e^{-f}\omega^n.$$

The desired result for ∇^2v now follows by letting R tend to $+\infty$. ◁

Finally, we show that some components of the Hessian of $\bar{u}_Y$ vanish identically.

Claim 5.5 *The $(0,2)$–part $\nabla^{0,2}\bar{u}_Y$ of the Hessian of $\bar{u}_Y$ vanishes identically on M.*

Proof For clarity, we suppress the dependence of the potential u_Y on the vector field Y in what follows.

Let $R>0$ and let ϕ_R be a cut-off function as in the proof of Claim 5.4. Reminiscent of [63, equation (2.7)], from (5-2) we then find, in normal holomorphic coordinates at a point where the Ricci form ρ_ω of ω has components $\rho_{i\bar{j}}$, that

$$\begin{aligned}
0 &= \int_M (\Delta_\omega u + u - f_{\bar{k}}u_k)_i \overline{u}_{\bar{\imath}}\, \phi_R^2\, e^{-f}\,\omega^n \\
&= \int_M (u_{k\bar{k}i} + u_i - f_{i\bar{k}}u_k - f_{\bar{k}}u_{ik}) \overline{u}_{\bar{\imath}}\, \phi_R^2\, e^{-f}\,\omega^n \\
&= \int_M (u_{ki\bar{k}} - \rho_{i\bar{s}}u_s + u_i - f_{i\bar{k}}u_k - f_{\bar{k}}u_{ik}) \overline{u}_{\bar{\imath}}\, \phi_R^2\, e^{-f}\,\omega^n \quad \text{since } u_{ji\bar{j}} = u_{j\bar{j}i} + \rho_{i\bar{s}}u_s, \\
&= \int_M (u_{ik\bar{k}} + \underbrace{(-\rho_{i\bar{s}}u_s + u_i - f_{i\bar{k}}u_k)}_{=0} - f_{\bar{k}}u_{ik}) \overline{u}_{\bar{\imath}}\, \phi_R^2\, e^{-f}\,\omega^n \\
&= \int_M (u_{ik\bar{k}} - f_{\bar{k}}u_{ik}) \overline{u}_{\bar{\imath}}\, \phi_R^2\, e^{-f}\,\omega^n = \int_M (\overline{\Delta_{\omega,X}\overline{u}_{\bar{\imath}}}) \overline{u}_{\bar{\imath}}\, \phi_R^2\, e^{-f}\,\omega^n \\
&= -\int_M \overline{u}_{\bar{\imath}\bar{j}} u_{ij}\, \phi_R^2\, e^{-f}\,\omega^n - \frac{1}{2}\int_M \langle \bar{\partial}\phi_R^2, \bar{\partial}|\bar{\partial}\overline{u}|^2\rangle\, e^{-f}\,\omega^n.
\end{aligned}$$

Therefore, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
\int_M |\nabla^{0,2}\overline{u}|^2_\omega\, \phi_R^2\, e^{-f}\,\omega^n &\le c\int_M (|\bar{\partial}\phi_R||\bar{\partial}\overline{u}|)\cdot(\phi_R|\nabla^2\overline{u}|_\omega)\, e^{-f}\,\omega^n \\
&\le c\left(\int_M |\nabla\phi_R|^2|\nabla\overline{u}|^2 e^{-f}\,\omega^n\right)^{\frac{1}{2}}\left(\int_M |\nabla^2\overline{u}|^2_\omega \phi_R^2 e^{-f}\,\omega^n\right)^{\frac{1}{2}}
\end{aligned}$$

for some positive constant c independent of R that may vary from line to line. By Claim 5.4, the previous inequality leads to the bound

$$\int_M |\nabla^{0,2}\overline{u}|^2_\omega\, \phi_R^2\, e^{-f}\,\omega^n \le \frac{c}{R}$$

for some positive constant c independent of R. Letting R tend to $+\infty$, this shows that

$$\int_M |\nabla^{0,2}\overline{u}|^2_\omega\, e^{-f}\,\omega^n = 0,$$

as desired. ◁

Consequently, $\nabla^{0,2}\overline{u}_Y = 0$, from which it follows that $\nabla^{1,0}\overline{u}_Y$ is a holomorphic vector field.

Thus, $\nabla^{1,0}u_Y$ and $\nabla^{1,0}\overline{u}_Y$ are holomorphic vector fields. Write $u_Y = v_Y + i w_Y$, where v_Y and w_Y are smooth real-valued functions on M. Then we deduce that $\nabla^{1,0}v_Y = \frac{1}{2}(\nabla v_Y - iJ\nabla v_Y)$ and $\nabla^{1,0}w_Y = \frac{1}{2}(\nabla w_Y - iJ\nabla w_Y)$ are holomorphic. In particular, ∇v_Y and ∇w_Y are real holomorphic vector fields on M so that by [32, Lemma 2.3.8], $J\nabla v_Y$ and $J\nabla w_Y$ are real holomorphic g–Killing vector fields. Therefore we have the decomposition

$$\tfrac{1}{2}(Y - iJY) = Y^{1,0} = \nabla^{1,0}u_Y = \nabla^{1,0}(v_Y + iw_Y) = \tfrac{1}{2}(\nabla v_Y + J\nabla w_Y) - \tfrac{1}{2}i(J\nabla v_Y - \nabla w_Y),$$

so that

$$Y = \nabla v_Y + J\nabla w_Y = J\nabla w_Y + J(-J\nabla v_Y). \tag{5-5}$$

Moreover, since $\mathcal{L}_{JX}u_Y = 0$, we have that $\mathcal{L}_{JX}v_Y = \mathcal{L}_{JX}w_Y = 0$ so that $[JX, \nabla v_Y] = [JX, \nabla w_Y] = 0$, and consequently $[X, J\nabla v_Y] = [X, J\nabla w_Y] = 0$. Hence $J\nabla v_Y$ and $J\nabla w_Y$ lie in $\mathfrak{g}^X$, leaving (5-5) as the desired decomposition.

To show that this decomposition is direct, suppose that $Z \in \mathfrak{g}^X \cap J\mathfrak{g}^X$. Then $Z = JW$, where W and JW are real holomorphic and Killing. Since W is holomorphic and JW is Killing, ∇W is symmetric. Since W is Killing, ∇W is skew-symmetric. Thus, W is parallel. If W is nontrivial, then by [29, Corollary 3.2], (M, g) splits off a line, with W the generator of this line. In particular, we may write $M = N \times \mathbb{R}$ for a manifold N with $g = g_N \oplus dt^2$ and $W = \partial_t$, where t is the coordinate on the $\mathbb{R}$–direction and g_N a shrinking Ricci soliton on N. Now the soliton vector field X must split as a direct sum with the summand in the $\mathbb{R}$–direction necessarily $t\partial_t$. Since $[W, X] = 0$ as $Z \in \mathfrak{g}^X$, this yields a contradiction, so that $W = 0$. Hence the stated decomposition of $\mathfrak{aut}^X(M)$ is direct. □

Since M is noncompact, we need to verify that the various automorphism groups in question are indeed Lie groups. This is necessary for the applications of Theorem 5.1 that we have in mind. We begin with:

Proposition 5.6 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature. Then there exists a unique connected Lie group $\mathrm{Aut}_0^X(M)$ (endowed with the compact–open topology) of diffeomorphisms acting effectively on M with Lie algebra $\mathfrak{aut}^X(M)$.*

$\mathrm{Aut}_0^X(M)$ is of course the connected component of the identity of the holomorphic automorphisms of M that commute with the flow of X.

The fact that there is a unique Lie group $\mathrm{Aut}_0^X(M)$ with the stated properties follows from Palais' integrability theorem [57] (see also [40, Theorem 3.1, page 13]), once we establish the completeness and finite-dimensionality of $\mathfrak{aut}^X(M)$. However, this theorem only asserts that $\mathrm{Aut}_0^X(M)$ is a Lie group with respect to the "modified" compact–open topology. In order to see that it is a Lie group with respect to the compact–open topology, we must appeal to [35, Theorem 5.14], using the fact that $\mathrm{Aut}_0^X(M)$ is closed with respect to the compact–open topology. Now, the completeness of the vector fields in $\mathfrak{aut}^X(M)$ is clear from Lemma 2.34. As for their finite-dimensionality, we have the following.

Proposition 5.7 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature and let $\mathfrak{aut}^X(M)$ denote the space of all real holomorphic vector fields Y on M with $[X, Y] = 0$. Then $\mathfrak{aut}^X(M)$ is finite-dimensional.*

Proof We provide an analytic proof of this fact. Letting $|\cdot|$ denote the norm with respect to g and writing f for the soliton potential, we have a natural norm $\|\cdot\|^2_{L^2_f}$ on $\mathfrak{aut}^X(M)$ defined by

$$\|Y\|^2_{L^2_f} := \int_M |Y|^2 e^{-f}\omega^n.$$

It suffices to show that the unit ball is compact with respect to this norm. To this end, suppose that we have a sequence $(Y_i)_{i\geq 0}$ with $\|Y_i\|_{L^2_f}=1$. Then by elliptic estimates, we get uniform C^k–bounds on $|Y_i|$ over a fixed ball $B_g(p,R)\subset M$ once a C^0–estimate is established. Now, by a Nash–Moser iteration applied to the norm of Y_i, one obtains the estimate

$$\sup_{B_g(p,R/2)}|Y_i|\leq C(n,R)\|Y_i\|_{L^2(B_g(p,R))}.$$

Since $\|Y_i\|_{L^2_f}\leq 1$, then, a fortiori,

$$\sup_{B_g(p,R/2)}|Y_i|\leq C(n,R)e^{cR^2}\|Y_i\|_{L^2_f(B_g(p,R))}\leq C'(n,R).$$

Finally, according to (the proof) of Proposition 2.30, there is some large radius $R_0>0$ such that

$$|Y_i|(x)\leq C\Big(n,\sup_M|\mathrm{Ric}(g)|,\sup_{B_g(p,R_0)}|Y_i|\Big)\cdot(d_g(p,x)+1)^a\quad\text{for all } x\in M,\tag{5-6}$$

for some uniform positive constant a, where $d_g(p,x)$ denotes the distance between p and x with respect to g. Since $\sup_{B_g(p,R_0)}|Y_i|\leq C(n,R_0)$, passing to a subsequence if necessary we may assume that $(Y_i)_{i\geq 0}$ converges to some Y_∞ on the whole of M in the $C^\infty_{\mathrm{loc}}(M)$–topology. The question is whether this convergence is strong in the above norm. Thanks to (5-6), given $\varepsilon>0$, there exists some positive radius R such that for all indices $i\geq 0$,

$$\|Y_i\|_{L^2_f(M\setminus B_g(p,R))}\leq\varepsilon,$$

since the soliton potential grows quadratically by Theorem 2.20(i), and the volume growth of geodesic balls is at most polynomial by Theorem 2.20(ii). This shows that if R is chosen sufficiently large, then the remainder of the norm outside $B_g(p,R)$ is uniformly small; hence we do indeed have strong convergence. □

Remark 5.8 Munteanu and Wang [51, Theorem 1.4] proved that the space of polynomial growth holomorphic functions of a fixed degree on a shrinking gradient Kähler–Ricci soliton is finite-dimensional without assuming a Ricci curvature bound. We therefore expect that the above proposition holds true in more generality. We also expect that the ring of holomorphic functions of polynomial growth on M is finitely generated and that M is algebraic, at least under a Ricci bound assumption.

Recall that $\mathfrak{g}^X$ denotes the Lie algebra comprising real holomorphic g–Killing vector fields that commute with X and hence with JX. We next consider the existence of a Lie group with Lie algebra $\mathfrak{g}^X$.

Proposition 5.9 *Let (M,g,X) be a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature. Then there exists a unique connected Lie group G^X_0 (endowed with the compact–open topology) of diffeomorphisms acting effectively on M with Lie algebra $\mathfrak{g}^X$.*

Proof Since $\mathfrak{g}^X$ is a Lie subalgebra of the Lie algebra of g–Killing vector fields on M, $\mathfrak{g}^X$ is a finite-dimensional Lie algebra. Furthermore, vector fields induced by $\mathfrak{g}^X$ on M are complete by Lemma 2.34.

By Palais' integrability theorem [57] therefore, there exists a unique connected Lie group G_0^X of diffeomorphisms acting effectively on M with Lie algebra $\mathfrak{g}^X$. G_0^X is precisely the connected component of the identity of the Lie group of holomorphic isometries on M that commute with the flow of X. Since G_0^X is closed with respect to the compact–open topology, [35, Theorem 5.14] guarantees that G_0^X is a Lie group with respect to this topology, as stated. □

We next prove that G_0^X is a compact Lie subgroup of $\mathrm{Aut}_0^X(M)$.

Lemma 5.10 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with complex structure J and with soliton vector field $X = \nabla^g f$ for a smooth real-valued function $f : M \to \mathbb{R}$. Then elements of $\mathfrak{g}^X$ are tangent to the level sets of f. Moreover, if g has bounded Ricci curvature, then G_0^X is a compact Lie subgroup of $\mathrm{Aut}_0^X(M)$ (with respect to the compact–open topology).*

Proof Let $Y \in \mathfrak{g}^X$. For the first part of the lemma, we will show that $\mathcal{L}_Y f = 0$, so that the flow of Y preserves the level sets of f, thereby forcing Y to be tangent to the level sets of f.

Applying $\mathcal{L}_Y$ to the shrinking Kähler–Ricci soliton equation, we find that $i\partial\bar{\partial}(\mathcal{L}_Y f) = 0$. Notice that since $[JX, Y] = 0$, we have that $\mathcal{L}_{JX}(\mathcal{L}_Y f) = \mathcal{L}_Y(\mathcal{L}_{JX} f) = 0$. The function $X \cdot (\mathcal{L}_Y f)$ is therefore holomorphic. It is also real-valued, hence must be equal to a constant, say $X \cdot (\mathcal{L}_Y f) = c_0$. Since $X = \nabla^g f$ and f has a minimum, we deduce that in fact $c_0 = 0$, so that $X \cdot (\mathcal{L}_Y f) = 0$.

Next, deriving with respect to the Killing vector field Y the soliton identity from Lemma 2.22, namely

$$\mathcal{L}_X f + R_g = X \cdot f + R_g = |\nabla^g f|^2 + R_g = 2f,$$

making use of the fact that X and Y commute, we obtain

$$2\mathcal{L}_Y f = \mathcal{L}_Y(\mathcal{L}_X f) + \underbrace{\mathcal{L}_Y R_g}_{=0} = \mathcal{L}_X(\mathcal{L}_Y f) = X \cdot (\mathcal{L}_Y f),$$

where we have just seen that this last term vanishes. Hence $\mathcal{L}_Y f = 0$, as desired.

As for the second part of the lemma, note that under the assumption of bounded Ricci curvature of g, both G_0^X and $\mathrm{Aut}_0^X(M)$ are Lie groups endowed with the compact–open topology, by Propositions 5.6 and 5.9, respectively. In addition, G_0^X is a subgroup of $\mathrm{Aut}_0^X(M)$ since $\mathfrak{g}^X$ is a Lie subalgebra of $\mathfrak{aut}^X(M)$. Compactness of G_0^X with respect to the compact–open topology follows from the Arzelà–Ascoli theorem because the level sets of f are compact by properness of f and, as we have just seen, are preserved by G_0^X. Being compact, G_0^X is then a closed subgroup of $\mathrm{Aut}_0^X(M)$, hence is a compact Lie subgroup of $\mathrm{Aut}_0^X(M)$, with everything being relative to the compact–open topology. □

Finally, we can now deduce from Theorem 5.1 that G_0^X is a *maximal* compact Lie subgroup of $\mathrm{Aut}_0^X(M)$.

Corollary 5.11 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature. Then G_0^X is a maximal compact Lie subgroup of $\mathrm{Aut}_0^X(M)$.*

Proof First note that G_0^X is a compact Lie subgroup of $\mathrm{Aut}_0^X(M)$ by Lemma 5.10. Now suppose that G_0^X is not maximal in $\mathrm{Aut}_0^X(M)$. Then there exists a compact Lie subgroup K of $\mathrm{Aut}_0^X(M)$ strictly containing G_0. In particular, the real dimension of K must be strictly greater than G_0^X. On the Lie algebra level, since the decomposition of Theorem 5.1 is direct, there exists a nonzero real holomorphic vector Z in the Lie algebra of K which is not contained in $\mathfrak{g}^X$, yet is contained in $J\mathfrak{g}^X$. Since K is compact, the closure of the flow of Z in K will define a real torus T^k of real dimension k in K in which the flow of Z is dense.

Consider the vector field Z. This is a real holomorphic vector field with JZ Killing. Since a shrinking soliton has finite fundamental group [67], we have that $H^1(M) = 0$. Hence JZ admits a Hamiltonian potential $u\colon M \to \mathbb{R}$ so that $Z = \nabla^g u$. If the real dimension k of T is equal to one, then the orbits of Z are all closed, but a gradient flow has no nontrivial closed integral curves since

$$\frac{d}{dt}u(\gamma_x(t)) = |\nabla^g u|^2 \geq 0,$$

where $\gamma_x(t)$ denotes the integral curve of Z with $\gamma_x(0) = x \in M$. Hence k is strictly greater than one. But this is impossible as well. Indeed, let x be any point of M where $Z(x) \neq 0$. Then $u(\gamma_x(t))$ is an increasing function of t, so that $u(\gamma_x(t)) - u(x) > c$ for some constant $c > 0$ say, for all $t > 1$. On the other hand, since the flow of Z is dense in T, $\gamma_x(t)$ intersects any neighbourhood of x in M for some $t > 1$. This yields another contradiction. Thus, G_0^X is maximal in $\mathrm{Aut}_0^X(M)$, as claimed. □

5.2 The weighted volume functional

Recall that (M, g, X) is a complete shrinking gradient Kähler–Ricci soliton of complex dimension n with complex structure J, Kähler form ω and soliton vector field $X = \nabla^g f$ for $f\colon M \to \mathbb{R}$ smooth, and that by assumption, we have a real torus T with Lie algebra $\mathfrak{t}$ acting holomorphically, effectively and isometrically on (M, g, J) with $JX \in \mathfrak{t}$.

In order to make sense of the weighted volume functional of a shrinking gradient Kähler–Ricci soliton, we need to define a moment map for the action of T on M. This comes down to showing that every element of $\mathfrak{t}$ admits a real Hamiltonian potential, as demonstrated in the next proposition. Such a potential exists essentially because T acts by isometries and $H^1(M) = 0$. However, a Hamiltonian potential is only defined up to a constant. Therefore a normalisation is required to determine the potential uniquely. We normalise so that the potential lies in the kernel of a certain linear operator, precisely the condition required to show that JX is the unique critical point of the weighted volume functional.

Proposition 5.12 *In the above situation, let $Y \in \mathfrak{t}$ so that Y defines a real holomorphic g–Killing vector field on M with $[X, Y] = 0$. Then there exists a unique smooth real-valued function $u_Y\colon M \to \mathbb{R}$ with $\mathcal{L}_{JX} u_Y = 0$ such that $\Delta_\omega u_Y + u_Y + \frac{1}{2}(JY) \cdot f = 0$ and $du_Y = -\omega \lrcorner Y$.*

Proof Let $Z := -JY$. Then Z is real holomorphic and JZ is g–Killing. Since a shrinking soliton has finite fundamental group [67], we have that $H^1(M) = 0$. This implies that there exists a smooth real-valued function $u_Y : M \to \mathbb{R}$ such that $Z = \nabla^g u_Y$. Then $du_Y \circ J = -\omega \lrcorner Z$. Let $Z^{1,0} = \frac{1}{2}(Z - iJZ)$. Then we have that

$$\omega \lrcorner Z^{1,0} = \tfrac{1}{2}\omega \lrcorner Z - \tfrac{1}{2} i\omega \lrcorner JZ = -\tfrac{1}{2} du_Y \circ J + \tfrac{1}{2} i\, du_Y = \tfrac{1}{2} i(du_Y + i\, du_Y \circ J) = i\bar{\partial} u_Y,$$

ie $\bar{\partial}(-iu_Y) = -\omega \lrcorner Z^{1,0}$. After noting from the proof of Lemma 5.10 that $Y \in \mathfrak{t}$ implies that $Y \cdot f = 0$, we automatically have that $\mathcal{L}_{JX} u_Y = 0$. Using the Bochner formula, contracting (1-4) with $\lambda = 1$ with $Z^{1,0}$ then results in

$$-i\bar{\partial}\Delta_\omega u_Y + i\bar{\partial}(Z^{1,0} \cdot f) = i\bar{\partial} u_Y.$$

In other words,

$$\bar{\partial}(\Delta_\omega u_Y + u_Y - Z^{1,0} \cdot f) = 0.$$

Now, the fact that $\mathcal{L}_{JX} u_Y = 0$ implies that

$$0 = du_Y(JX) = g(\nabla^g u_Y, JX) = g(Z, JX) = -g(JZ, X) = -df(JZ) = -(JZ) \cdot f.$$

In particular, $\Delta_\omega u_Y + u_Y - Z^{1,0} \cdot f$ is a real-valued holomorphic function, hence is equal to a constant. By subtracting this constant from u_Y and plugging in the definition of Z, we arrive at our desired normalisation of u_Y, namely

$$\Delta_\omega u_Y + u_Y + \tfrac{1}{2}(JY) \cdot f = 0.$$

Since u_Y is defined up to a constant, this condition determines u_Y uniquely. □

With this proposition, we can now define our moment map for the action of T on M.

Definition 5.13 Let $\langle \cdot, \cdot \rangle$ denote the natural pairing between $\mathfrak{t}$ and $\mathfrak{t}^*$. Then we define the *moment map* $\mu : M \to \mathfrak{t}^*$ *for the action of* T *on* M as follows: for $x \in M$, $\mu(x)$ is defined by the equation

$$u_Y(x) = \langle \mu(x), Y \rangle \quad \text{for all } Y \in \mathfrak{t},$$

where u_Y is such that $\nabla^g u_Y = -JY$, $\mathcal{L}_{JX} u_Y = 0$ and $\Delta_\omega u_Y + u_Y + \frac{1}{2}(JY) \cdot f = 0$.

We next define the weighted volume functional for complete shrinking gradient Kähler–Ricci solitons.

Definition 5.14 (weighted volume functional, see [64, equation (2.3)]) Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton of complex dimension n with complex structure J, Kähler form ω, and with soliton vector field $X = \nabla^g f$ for a smooth real-valued function $f : M \to \mathbb{R}$, endowed with the holomorphic, effective, isometric action of a real torus T with Lie algebra $\mathfrak{t}$ and with a compact fixed point set. Let μ denote the moment map of the action as prescribed in Definition 5.13 and assume that $JX \in \mathfrak{t}$. Let $Y \in \mathfrak{t}$ and let $u_Y := \langle \mu, Y \rangle$ be the Hamiltonian potential of Y, so that $\mathcal{L}_{JX} u_Y = 0$ and

$$\Delta_\omega u_Y + u_Y + \tfrac{1}{2}(JY) \cdot f = 0. \tag{5-7}$$

Finally, let

$$\Lambda := \{Y \in \mathfrak{t} \mid u_Y \text{ is proper and bounded below}\} \subseteq \mathfrak{t}.$$

Then the *weighted volume functional* $F\colon \Lambda \to \mathbb{R}_{>0}$ is defined by

$$F(Y) = \int_M e^{-\langle \mu, Y\rangle} \omega^n = \int_M e^{-u_Y} \omega^n.$$

The set Λ is an open cone in $\mathfrak{t}$ which is determined by the image of M under μ; see Proposition A.4 for details. Since f grows quadratically at infinity by Theorem 2.20(i), we know that it is also proper. Hence JX, which by assumption lies in $\mathfrak{t}$, lies in Λ, so that Λ is nonempty. Thus, by the Duistermaat–Heckman theorem (Theorem A.3), F is seen to be well-defined. (Proposition A.13 provides an alternative argument, without using the Duistermaat–Heckman theorem, for why F is well-defined.) As the next lemma shows, the value of F is also independent of the choice of shrinking gradient Kähler–Ricci soliton. This relies on the normalisation (5-7) of the Hamiltonian potentials.

Lemma 5.15 *Let (M, g_i, X_i) for $i = 1, 2$ be two shrinking gradient Kähler–Ricci solitons, both satisfying the hypotheses of Definition 5.14 with respect to a fixed real torus T. Let F_i denote the weighted volume functional of (M, g_i, X_i) and let Λ_i denote the domain of F_i. Then $F_1 = F_2$ on $\Lambda_1 \cap \Lambda_2$.*

Proof Let ω_i denote the Kähler form of g_i and let $Y \in \Lambda_1 \cap \Lambda_2$. Write $u_Y^{(i)}$ for the Hamiltonian potential of Y with respect to ω_i. Then analysing the expression given in Theorem A.3 for each F_i, namely (A-1), one sees that the right-hand side depends only on the value of $u_Y^{(i)}$ on the (compact) zero set $M_0(Y)$ of Y and integrals over this set with respect to ω_i. Now, the normalisation condition (5-7) infers that on $M_0(Y)$, $u_Y^{(i)} = -\Delta_{\omega_i} u_Y^{(i)} = -\operatorname{div}(Y)$, a quantity that, on $M_0(Y)$, is independent of the choice of metric. Moreover, $du_Y^{(i)} = -\omega_i \lrcorner Y$ so $u_Y^{(i)}$ is evidently constant on each connected component of $M_0(Y)$. Thus, we deduce that $u_Y^{(1)} = u_Y^{(2)}$ on $M_0(Y)$, both being equal to a fixed constant on each connected component of $M_0(Y)$. As for the integrals on the right-hand side of (A-1), these involve integrating a closed form ω_i over the compact boundary-less set $M_0(Y)$. Both being shrinking Kähler–Ricci solitons, ω_1 and ω_2 lie in the same cohomology class, hence integrating over $M_0(Y)$ with respect to either ω_1 or ω_2 does not change the value of the integral. This brings us to the desired conclusion. □

We next list some more elementary properties of F, in particular the desired property that characterises JX as the unique critical point of F. Here our normalisation of the Hamiltonian potentials also comes into play.

Lemma 5.16 (volume-minimising principle) *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton of complex dimension n with Kähler form ω and with soliton vector field $X = \nabla^g f$ for a smooth real-valued function $f\colon M \to \mathbb{R}$, endowed with the holomorphic, effective, isometric action of a real torus T with Lie algebra $\mathfrak{t}$. Assume that $JX \in \mathfrak{t}$ and that the Ricci curvature of g is bounded. Then*

(i) *F is strictly convex on Λ, and*

(ii) *JX is the unique critical point of F in Λ.*

Remark 5.17 The boundedness of the scalar curvature of g guarantees here that F is well-defined on Λ. Indeed, this is clear from the Duistermaat–Heckman theorem (Theorem A.3) after noting that the zero set of X, which contains the fixed point set of T as a closed subset, is compact by Lemma 2.25.

Proof of Lemma 5.16 (i) Let $Y_1, Y_2 \in \Lambda$. Then the line segment $tY_1+(1-t)Y_2, t\in[0,1]$, is contained in Λ because Λ is convex, as one sees from its definition. Moreover, by the linearity of the moment map, we have that

$$u_{tY_1+(1-t)Y_2} = tu_{Y_1} + (1-t)u_{Y_2} \quad \text{for all } t\in[0,1].$$

Thus, since the function $x\in\mathbb{R}\mapsto e^{-x}\in\mathbb{R}$ is strictly convex, we find that

$$F(t\cdot Y_1+(1-t)\cdot Y_2) < t\cdot F(Y_1)+(1-t)\cdot F(Y_2) \quad \text{for all } t\in(0,1), \text{ unless } Y_1=Y_2.$$

(ii) As a strictly convex function on the convex set Λ, F has at most one critical point. The claim is that this critical point is obtained at JX. Indeed, let $Y\in\mathfrak{t}$ and let u_Y denote the Hamiltonian potential of Y, normalised so that $\Delta_\omega u_Y + u_Y + \frac{1}{2}(JY)\cdot f = 0$. Recall that $-J(JX) = \nabla^g f$, so that

$$d_{JX}F(Y) = -\int_M u_Y e^{-f}\omega^n.$$

Let R be a positive real number and let $\phi_R\colon M\to[0,1]$ be a cut-off function with compact support in the geodesic ball $B_g(p,2R)$ such that $\phi_R=1$ on $B_g(p,R)$ and $|\nabla\phi_R|_g\le c/R$ for some $c>0$. Then, using integration by parts, we have that

$$\begin{aligned}\left|\int_M u_Y\phi_R^2 e^{-f}\omega^n\right| &= \left|\int_M (\Delta_\omega u_Y + \tfrac{1}{2}(JY)\cdot f)\phi_R^2 e^{-f}\omega^n\right| = \left|\frac{1}{2}\int_M (\Delta_g - \nabla^g f\,\cdot)u_Y\phi_R^2 e^{-f}\omega^n\right| \\ &= \frac{1}{2}\left|\int_M g\big(\nabla u_Y, \nabla(\phi_R^2)\big)e^{-f}\omega^n\right| \\ &\le \frac{1}{2}\left(\int_M |\nabla u_Y|^2 e^{-f}\omega^n\right)^{\frac{1}{2}}\left(\int_M |\nabla(\phi_R^2)|^2 e^{-f}\omega^n\right)^{\frac{1}{2}} \\ &= \left(\int_M |\nabla u_Y|^2 e^{-f}\omega^n\right)^{\frac{1}{2}}(|\nabla\phi_R|^2|\phi_R|^2 e^{-f}\omega^n)^{\frac{1}{2}} \\ &\le \frac{c^2}{R^2}\left(\int_M |\nabla u_Y|^2 e^{-f}\omega^n\right)^{\frac{1}{2}}\left(\int_M e^{-f}\omega^n\right)^{\frac{1}{2}},\end{aligned}$$

where the fact that $\nabla u_Y\in L^2(e^{-f}\omega^n)$ follows as in the proof of Claim 5.4. Letting $R\to+\infty$, we see that $d_{JX}F(Y)=0$, as required. □

The main tool we use to compute the weighted volume functional is the Duistermaat–Heckman theorem. The statement of this theorem and a discussion have been relegated to the appendix. It expresses the weighted volume functional in terms of data determined by the induced action on M of the element $Y\in\Lambda$. In particular, this data is independent of the metric ω. Consequently, F is independent of the particular shrinking gradient Kähler–Ricci soliton. It is this observation that will allow us to ascertain the uniqueness of the soliton vector field X under certain assumptions. This is the content of the next subsection.

5.3 A general uniqueness theorem

As an application of Corollary 5.11, we prove the uniqueness statement of Theorem D for the soliton vector field X of a shrinking gradient Kähler–Ricci soliton, the precise statement of which we now recall below.

Theorem 5.18 (Theorem D) *Let M be a noncompact complex manifold with complex structure J, endowed with the effective holomorphic action of a real torus T. Denote by $\mathfrak{t}$ the Lie algebra of T. Then there exists at most one element $\xi \in \mathfrak{t}$ that admits a complete shrinking gradient Kähler–Ricci soliton (M, g, X) with bounded Ricci curvature, with $X = \nabla^g f = -J\xi$ for a smooth real-valued function f on M.*

The outline of the proof of this theorem is as follows. Suppose that M admitted two soliton vector fields X_1 and X_2. Then the maximal tori in the Lie groups $\mathrm{Aut}_0^{X_i}(M)$ for $i = 1, 2$ will be conjugate to T by Iwasawa's theorem [39]. After choosing an appropriate gauge, JX_1 and JX_2 will then be contained in the Lie algebra $\mathfrak{t}$ of T and both vector fields will be critical points of their respective weighted volume functional. But since the weighted volume functional is independent of the shrinking Kähler–Ricci soliton by the Duistermaat–Heckman theorem, both weighted volume functionals must coincide, so that $JX_1 = JX_2$ by uniqueness of the critical point.

Proof Suppose that M admitted two complete shrinking gradient Kähler–Ricci solitons, (M, g_i, X_i) for $i = 1, 2$, with bounded Ricci curvature and with $X_i = \nabla^{g_i} f_i$ for $f_i : M \to \mathbb{R}$ smooth such that $X_i = -J\xi_i$ for $\xi_i \in \mathfrak{t}$. Let $G_0^{X_i}$ denote the connected component of the identity of the group of holomorphic isometries of (M, g_i, J) that commute with the flow of X_i. Corollary 5.11 then asserts that $G_0^{X_i}$ is a maximal compact Lie subgroup of the Lie group $\mathrm{Aut}_0^{X_i}(M)$, the connected component of the identity of the group of automorphisms of (M, J) that commute with the flow of X_i. Denote by T_i the maximal real torus in $G_0^{X_i}$. Then T_i is maximal in $\mathrm{Aut}_0^{X_i}(M)$. For each $v \in \mathfrak{t}$, we have that $[v, \xi_i] = 0$, so that $[v, X_i] = 0$. Hence each element of T commutes with the flow of X_i and so T itself is a Lie subgroup of $\mathrm{Aut}_0^{X_i}(M)$. Without loss of generality, we may assume that T is maximal in $\mathrm{Aut}_0^{X_i}(M)$. Then, by Iwasawa's theorem [39], there exists an element $\alpha_i \in \mathrm{Aut}_0^{X_i}(M)$ such that $\alpha_i T_i \alpha_i^{-1} = T$. Since α_i commutes with the flow of X_i, necessarily $d\alpha_i^{-1}(X_i) = X_i$. Moreover, $\alpha_i^* g_i$ is invariant under T. Thus, $(M, \tilde{g}_i, \tilde{X}_i)$ with $\tilde{g}_i = \alpha_i^* g_i$ and $\tilde{X}_i = d\alpha_i^{-1}(X_i)$ is a T–invariant shrinking gradient Kähler–Ricci soliton with soliton vector field $\tilde{X}_i = X_i = -J\xi_i$ as before. Hence, by considering this pullback, we may assume that each (M, g_i, X_i) is invariant under T.

Now, by assumption we have that $\xi_1, \xi_2 \in \mathfrak{t}$. Since the corresponding Hamiltonian potentials are the soliton potentials, which themselves are proper and bounded below, we have that $\xi_i \in \Lambda_i \subset \mathfrak{t}$, where Λ_i denotes the open cone of elements of $\mathfrak{t}$ admitting Hamiltonian potentials with respect to the Kähler form ω_i of g_i that are proper and bounded below. We wish to show that $\xi_1, \xi_2 \in \Lambda_1 \cap \Lambda_2 \neq \varnothing$. The result will then follow from an application of the Duistermaat–Heckman theorem. So let u_1 denote the

Hamiltonian potential of $\xi_1 = JX_1$ with respect to g_2, that is, $\nabla^{g_2}u_1 = X_1$, and for $x \in M$, let $\gamma_1(t)$ denote the integral curve of X_1 through x at $t = 0$. Then we have that

$$\begin{aligned} u_1(\gamma_1(t)) &= u_1(\gamma_1(0)) + \int_0^t du_1(\dot{\gamma}_1(s))\,ds \\ &= u_1(x) + \int_0^t g_2(X_1, X_1)(\gamma_1(s))\,ds \\ &= u_1(x) + \int_0^t |X_1|^2_{g_2}(\gamma_1(s))\,ds. \end{aligned}$$

Since g_1 has bounded Ricci curvature, so that the zero set of X_1 is compact by Lemma 2.25, and since each forward orbit of the negative gradient flow of f_1 converges to a point in the zero set of X_1 by Proposition 2.27, it is clear that every point of M lies on an integral curve of X_1 passing through a fixed compact set. Thus, we see that u_1 is bounded from below. For $t \in [0,1]$, consider the vector field $Y_t := t\xi_1 + (1-t)\xi_2$. The Hamiltonian potential of Y_0 with respect to g_2 is f_2, whereas that of Y_1 is u_1. By linearity of the moment map, the Hamiltonian potential of Y_t with respect to g_2 is $h_t := tu_1 + (1-t)f_2$. Since u_1 is bounded from below and f_2 is proper, h_t is proper and bounded below for $t \in [0,1)$, so that $Y_t \in \Lambda_2$ for $t \in [0,1)$. In a similar manner, one can show that $Y_t \in \Lambda_1$ for $t \in (0,1]$. The upshot is that $Y_t \in \Lambda_1 \cap \Lambda_2 \neq \varnothing$ for $t \in (0,1)$ with $\xi_1, \xi_2 \in \overline{\Lambda_1 \cap \Lambda_2}$.

Define a real-valued function F on $[0,1]$ as follows: $F(t) := F_2(Y_t)$ if $t \in [0,1)$ and $F(t) := F_1(Y_t)$ if $t \in (0,1]$, where F_i is the weighted volume functional with respect to ω_i. Then F is well-defined as both F_1 and F_2 are well-defined because of the Ricci curvature bound (see Remark 5.17) and by Lemma 5.15 they are equal on $\Lambda_1 \cap \Lambda_2$. Moreover, F is convex and continuous on $[0,1]$ and strictly convex on $(0,1)$. Finally, observe that

$$F(0) = F_2(\xi_2) \leq \min_{[0,1)} F_2 \leq F_2(Y_t) = F_1(Y_t)$$

for every $t \in (0,1)$. By letting t tend to 1, one sees that $F(0) \leq F(1)$. By symmetry, one also sees that $F(1) \leq F(0)$, which implies that $F(1) = F(0) = \min_{[0,1]} F$. Since F is convex, F must be constant on $[0,1]$, which contradicts the fact that F is strictly convex on $(0,1)$ unless $(Y_t)_{t\in(0,1)}$ is reduced to a single point, ie unless $\xi_1 = \xi_2$. This concludes the proof. □

5.4 Shrinking gradient Kähler–Ricci solitons on $\mathbb{C}^n$ and $\mathcal{O}(-k) \to \mathbb{P}^{n-1}$ for $0 < k < n$

Using Theorem D, we are now able to classify shrinking gradient Kähler–Ricci solitons with bounded Ricci curvature on $\mathbb{C}^n$ and on the total space of the line bundle $\mathcal{O}(-k) \to \mathbb{P}^{n-1}$ for $0 < k < n$, and in doing so, prove items (1) and (2) of Theorem E.

Theorem 5.19 (items (1) and (2) of Theorem E) *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature.*

(1) *If $M = \mathbb{C}^n$, then up to pullback by an element of $\mathrm{GL}(n,\mathbb{C})$, (M, g, X) is the flat Gaussian shrinking soliton.*

(2) *If M is the total space of the line bundle $\mathcal{O}(-k) \to \mathbb{P}^{n-1}$ for $0 < k < n$, then up to pullback by an element of* $\mathrm{GL}(n, \mathbb{C})$, (M, g, X) *is the unique $U(n)$–invariant shrinking gradient Kähler–Ricci soliton constructed by Feldman, Ilmanen and Knopf* [30] *on this complex manifold.*

Proof Let f denote the soliton potential of X, so that $f: M \to \mathbb{R}$ is a smooth real-valued function with $X = \nabla^g f$, and let M be as in item (1) or (2) of the theorem. We make no distinction as of yet. Denote the complex structure of M by J and let G_0^X denote the connected component of the identity of the holomorphic isometries of (M, J, g) that commute with the flow of X. Since g has bounded Ricci curvature, G_0^X is a compact Lie group by Lemma 5.10, hence the closure of the flow of JX in G_0^X yields the holomorphic isometric action of a real torus T on (M, J, g) with Lie algebra $\mathfrak{t}$ containing JX. Since M is 1–convex by [20, Lemma 2.15] and the Ricci curvature of g is bounded, Proposition 2.26 tells us that the zero set of X, and correspondingly the fixed-point set of T, comprises a single point in item (1) and is contained in the zero section of the line bundle in item (2). Furthermore, Proposition 2.27 implies that each forward orbit of the negative gradient flow of f converges to a point in this fixed-point set. By contracting the zero section of the line bundle in item (2), we see that the action of T on M induces an action of T on $\mathbb{C}^n/\mathbb{Z}_k$ for $k = 1, \dots, n-1$, as appropriate with fixed-point set the apex, and that this action further lifts to an action of T on $\mathbb{C}^n$ with an isolated fixed point. The lift of X to $\mathbb{C}^n$ then defines a holomorphic vector field on $\mathbb{C}^n$ with $J_0 X \in \mathfrak{t}$, where J_0 denotes the standard complex structure on $\mathbb{C}^n$, and with each forward orbit of $-X$ converging to this isolated fixed point.

By [17, Section 3.1], we may choose global holomorphic coordinates $(z_1, \dots, z_n)$ on $\mathbb{C}^n$ with respect to which the action of T on $\mathbb{C}^n$ is linear, that is, lies in $\mathrm{GL}(n, \mathbb{C})$. These coordinates descend to coordinates on $\mathbb{C}^n/\mathbb{Z}_k$, then lift to coordinates on M with respect to which the action of T on M lies in $\mathrm{GL}(n, \mathbb{C})$. Without loss of generality, we may assume that T is maximal in $\mathrm{GL}(n, \mathbb{C})$. Then we still have that $JX \in \mathfrak{t}$. Since any two maximal tori in $\mathrm{GL}(n, \mathbb{C})$ are conjugate by Iwasawa's theorem [39], there exists $\alpha \in \mathrm{GL}(n, \mathbb{C})$ such that $\alpha T \alpha^{-1}$ is equal to $\{\mathrm{diag}(e^{i\eta_1}, \dots, e^{i\eta_n}) \mid \eta_i \in \mathbb{R}\}$. By considering $\alpha^*\omega$, we can therefore assume that JX lies in the Lie algebra $\mathfrak{t}$ of a torus of the form $T = \{\mathrm{diag}(e^{i\eta_1}, \dots, e^{i\eta_n}) \mid \eta_i \in \mathbb{R}\}$ acting on M. We will then have induced coordinates $(\eta_1, \dots, \eta_n)$ on $\mathfrak{t}$, where $(1, 0, \dots, 0) \in \mathfrak{t}$ will generate the vector field $\mathrm{Im}(z_1 \partial_{z_1})$ on M, and so on.

Since the fixed-point set of T is compact, we can now apply Theorem D, which tells us that there is at most one element of $\mathfrak{t}$ that admits a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature. On $\mathbb{C}^n$, we have the flat Gaussian shrinking soliton and Feldman, Ilmanen and Knopf [30] have constructed a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature on M for M as in item (2) of the theorem. In all cases, the soliton vector field X of these solitons satisfies $JX \in \mathfrak{t}$, and each is proportional to $(1, \dots, 1)$ in our coordinates on $\mathfrak{t}$. Therefore we deduce that $JX = \lambda(1, \dots, 1)$ for some $\lambda > 0$, so that on M, we have $\frac{1}{2}(X - iJX) = \lambda z_i \partial_{z_i}$. The automorphism group of (M, J) commuting with the flow of this vector field is precisely the Lie group $\mathrm{GL}(n, \mathbb{C})$. Thus, Corollary 5.11 asserts that G_0^X is maximal compact in $\mathrm{GL}(n, \mathbb{C})$ and so, by Iwasawa's theorem [39] again, there exists

$\beta \in \mathrm{GL}(n, \mathbb{C})$ such that $\beta G_0^X \beta^{-1} = U(n)$. The $(1,1)$–form $\beta^*\omega$ will then be $U(n)$–invariant and by [30, Proposition 9.3], the only such complete shrinking gradient Kähler–Ricci soliton on M is the flat Gaussian shrinking soliton if $M = \mathbb{C}^n$, and that constructed by Feldman, Ilmanen and Knopf if M is as in item (2) of the theorem. □

6 The underlying manifold of a two-dimensional shrinking gradient Kähler–Ricci soliton

Item (3) of Theorem E will result from items (1) and (2) of Theorem E once we establish the following theorem.

Theorem 6.1 *Let (M, g, X) be a two-dimensional complete shrinking gradient Kähler–Ricci soliton whose scalar curvature decays to zero at infinity. Then C_0 is biholomorphic to $\mathbb{C}^2$, and M is biholomorphic to either $\mathbb{C}^2$ or $\mathbb{C}^2$ blown up at one point.*

The key observation in proving this theorem is that the scalar curvature of the asymptotic cone is strictly positive if the shrinking soliton is not flat. Since we are working in complex dimension two, this allows us to identify the tangent cone at infinity as a quotient singularity using a classification theorem of Belgun [4, Theorem 8] for 3–dimensional Sasaki manifolds. The fact that M is a resolution of C_0 by Theorem A, combined with the fact that the exceptional set of this resolution must contain only (-1)–curves as imposed by the shrinking Kähler–Ricci soliton equation, then allows us to identify M and C_0.

6.1 Properties of shrinking Ricci solitons

We begin by noting some important features of shrinking Ricci solitons that we require in this section. We have the following condition on the scalar curvature of a shrinking gradient Ricci soliton.

Theorem 6.2 [16] *Let (M, g, X) be a complete noncompact nonflat shrinking gradient Ricci soliton with scalar curvature R_g. Then for any given point $o \in M$, there exists a constant $C > 0$ such that $R_g(x) d_g(x, o)^2 > C^{-1}$ wherever $d_g(x, o) > C$, where d_g denotes the distance function with respect to g.*

This yields the following condition on the scalar curvature of an asymptotic cone of a shrinking Ricci soliton.

Corollary 6.3 *Let (M, g, X) be a complete noncompact nonflat shrinking gradient Ricci soliton with tangent cone (C_0, g_0) along an end. Then the scalar curvature R_{g_0} of the cone metric g_0 is strictly positive.*

Proof The tangent cone at infinity is obtained as a Gromov–Hausdorff limit of a pointed sequence $(M, g_k, o) := (M, \lambda_k^{-2} g, o)$ for $o \in M$ fixed, where $\lambda_k \to \infty$ as $k \to \infty$. By our asymptotic assumption, the tangent cone is unique and this process recovers the asymptotic cone (C_0, g_0). Indeed, an arbitrary point $p \in C_0$ with $r(p) = r_0 > 0$ is associated with a sequence $p_k \to p$, where $d_g(o, p_k) = \lambda_k r_0 \to \infty$ as $k \to \infty$, where r denotes the radial coordinate of g_0 and d_g denotes the distance measured with respect to g. In particular, we see that

$$R_{g_0}(p) = \lim_{k\to\infty} R_{\lambda_k^{-2}g}(p_k) = \lim_{k\to\infty} \lambda_k^2 R_g(p_k),$$

where $R_{\lambda_k^{-2}g}$ denotes the scalar curvature of the rescaled metric $\lambda_k^{-2}g$. Using the lower bound of Theorem 6.2, we then have that

$$R_{g_0}(p) = \lim_{k\to\infty} \lambda_k^2 R_g(p_k) > \lim_{k\to\infty} \frac{\lambda_k^2 C^{-1}}{d_g(o, p_k)^2} = \lim_{k\to\infty} \frac{\lambda_k^2 C^{-1}}{(\lambda_k r_0)^2} = \frac{1}{C r_0^2} > 0$$

for some positive constant C. Since p is arbitrary, it follows that $R_{g_0} > 0$ away from the apex of C_0, as claimed. □

6.2 Proof of Theorem 6.1

Let (M, g, X) be a complete noncompact shrinking gradient Kähler–Ricci soliton of complex dimension $n+1$ with quadratic curvature decay and with tangent cone along its end the Kähler cone (C_0, g_0) given by Theorem A. Let r denote the radial function of the cone. Then the link of the cone $\{r = 1\}$, which we denote by (S, g_S), is a Sasaki manifold of real dimension $2n+1$ foliated by the orbits of the flow of ξ, the restriction of the Reeb vector field of the cone to its link.

We know from [58, Theorem 3] that if the scalar curvature R_g of g is zero at a point, then (M, g) is isometric to Euclidean space. So we henceforth assume that $R_g \neq 0$ everywhere, so that (M, g) is nonflat. Then Corollary 6.3 tells us that the scalar curvature of the cone R_{g_0} is strictly positive. Next we see from Lemma 2.2 that $R_{g_S} > 2n(2n+1)$ and so it follows from Corollary 2.9 that

$$R^T > 2n(2n+1) + 2n = 4n(n+1).$$

Identification of C_0 In our case, (M, g, X) is of complex dimension two and the scalar curvature of g decays to zero at infinity. By [54], the scalar curvature decay implies that the norm of the curvature tensor of g decays quadratically. Thus, the above applies with $n = 1$ and we have the lower bound $R^T > 8$. From the classification of 3–dimensional Sasaki manifolds by Belgun [4, Theorem 8], it then follows that C_0 is biholomorphic to $\mathbb{C}^2/\Gamma$, with Γ a finite subgroup of $U(2)$ acting freely on $\mathbb{C}^2 \setminus \{0\}$. We next wish to show that $\Gamma = \{\mathrm{id}\}$.

Recall from Theorem A that there is a resolution $\pi : M \to C_0$ of the singularity of C_0 with $d\pi(X) = r\partial_r$. Since C_0 is biholomorphic to $\mathbb{C}^2/\Gamma$ for $\Gamma \subset U(2)$ a finite subgroup acting freely on $\mathbb{C}^2 \setminus \{0\}$, it is in particular a rational singularity. It is well-known that the exceptional set of a resolution of such a singularity contains a string of $\mathbb{P}^1$'s [9, Lemma 1.3]. Since g is a shrinking Kähler–Ricci soliton, each of these $\mathbb{P}^1$'s must have self-intersection (-1) by adjunction. Moreover, since C_0 is obtained from M by blowing down all of these (-1)–curves, C_0 must in fact be smooth at the apex, so that $\Gamma = \{\mathrm{id}\}$ and C_0 is biholomorphic to $\mathbb{C}^2$.

Identification of M It follows that M is then an iterated blowup of $\mathbb{C}^2$ at the origin containing only (-1)–curves. The only iterated blowups of $\mathbb{C}^2$ at the origin containing complex curves of this type are $\mathbb{C}^2$ and $\mathbb{C}^2$ blown up at one point, since any further iterated blowup would introduce at least one $\mathbb{P}^1$ with self-intersection $(-k)$ for some $k \geq 2$. The conclusion is then that M must be biholomorphic to either $\mathbb{C}^2$ or $\mathbb{C}^2$ blown up at one point.

7 Concluding remarks

We conclude with a discussion of future directions of research emanating from the results within this paper.

7.1 The conjectural picture

The results on shrinking gradient Kähler–Ricci solitons presented here allow us to speculate on possible deeper connections between such metrics and algebraic geometry. In the compact case, Berman, Witt and Nystrom [6] gave an algebraic formula for the weighted volume functional and its derivative. We generalise this result to the noncompact case under suitable assumptions, making use of the results of Wu [66]. We begin with the definition of an anticanonically polarised Kähler manifold, the underlying complex manifold of a shrinking Kähler–Ricci soliton.

Definition 7.1 An *anticanonically polarised Kähler manifold* is a Kähler manifold M admitting a Kähler form ω together with a hermitian metric on $-K_M$ with curvature form Θ such that

$$\int_V (i\,\Theta)^k \wedge \omega^{\dim_{\mathbb{C}} V-k} > 0$$

for all positive-dimensional irreducible compact analytic subvarieties V of M and for all integers k such that $1 \leq k \leq \dim_{\mathbb{C}} V$.

By [25, Theorem 4.2], a compact anticanonically polarised Kähler manifold is a Fano manifold. Moreover, any shrinking Kähler–Ricci soliton naturally lives on an anticanonically polarised Kähler manifold.

Under certain criteria, we can write an algebraic formula for the weighted volume functional.

Proposition 7.2 *Let (M,ω) be a (possibly noncompact) Kähler manifold of complex dimension n with Kähler form ω on which there is a Hamiltonian action of a real torus T with moment map $\mu\colon M \to \mathfrak{t}^*$, where $\mathfrak{t}$ is the Lie algebra of T and $\mathfrak{t}^*$ its dual. Assume that the fixed-point set of T is compact and that*

(i) *$H^p(M,\mathcal{O}(-kK_M)) = 0$ for all $p > 0$ and for all k sufficiently large, and*

(ii) *ω is the curvature form of a hermitian metric on $-K_M$.*

If there exists an element $\zeta_0 \in \mathfrak{t}$ such that the component of the moment map $u_{\zeta_0} = \langle \mu, \zeta_0 \rangle$ is proper and bounded below, then

$$\int_M e^{-\langle \mu,\zeta\rangle} \frac{\omega^n}{n!} = \lim_{k\to\infty} \frac{1}{k^n} \operatorname{char} H^0(M, \mathcal{O}(-kK_M))\left(\frac{\zeta}{k}\right) \tag{7-1}$$

for all ζ in an open cone $\Lambda \subset \mathfrak{t}$.

In this situation, the character $\operatorname{char} H^0(M, \mathcal{O}(-kK_M))$ is well-defined by [66]. Moreover, it follows from [55, Theorem 4.5] that the vanishing condition (i) holds true for any 1–convex anticanonically polarised Kähler manifold and condition (ii) holds true for any shrinking gradient Kähler–Ricci soliton. In particular, if (M, ω, X) is a complete shrinking gradient Kähler–Ricci soliton with Ricci curvature decaying to zero at infinity, endowed with the holomorphic, effective, isometric action of a real torus T with Lie algebra $\mathfrak{t}$ containing JX, then the above theorem applies. The volume minimising principle (Lemma 5.16) then tells us that for such a soliton, the unique minimum of the weighted volume functional is obtained at JX.

Before we present the proof of equation (7-1), it is necessary to introduce some notation. Our notation will mostly follow [66]. We denote by M^T the fixed-point set of T in M. By assumption, this is compact. If nonempty, it is a complex submanifold of M. Let F be the set of connected components of M^T. Then $M^T = \bigcup_{\alpha\in F} M^T_\alpha$, where M^T_α is the component labelled by $\alpha \in F$. Let $n_\alpha = \dim_{\mathbb{C}} M^T_\alpha$ and let $N_\alpha \to M^T_\alpha$ be the holomorphic normal bundle of M^T_α in M. T acts on N_α preserving the base M^T_α pointwise. The weights of the isotropy representation on the normal fibre remain constant within any connected component. Let ℓ be the integral lattice in the Lie algebra $\mathfrak{t}$ of T, let $\ell^* \subset \mathfrak{t}^*$ denote the dual lattice, and let $\lambda_{\alpha,i} \in \ell^* \setminus \{0\}$ for $1 \le i \le n - n_\alpha$ be the isotropy weights on N_α. The hyperplanes $(\lambda_{\alpha,i})^\perp \subset \mathfrak{t}$ cut $\mathfrak{t}$ into open polyhedral cones called *action chambers* [60]. Choose an action chamber C. We define ν^C_α as the number of weights $\lambda_{\alpha,i} \in C^*$, where C^* is the dual cone in $\mathfrak{t}^*$ defined by $C^* = \{\xi \in \mathfrak{t}^* \mid \langle \xi, C\rangle > 0\}$. Let N^C_α be the direct sum of the subbundles corresponding to the weights $\lambda_{\alpha,i} \in C^*$. Then $N_\alpha = N^C_\alpha \oplus N^{-C}_\alpha$. The rank of the holomorphic vector bundle N^C_α is ν^C_α; that of N^{-C}_α is $\nu^{-C}_\alpha = n - n_\alpha - \nu^C_\alpha$.

Proof of Proposition 7.2 For $k \in \mathbb{N}$ sufficiently large, the vanishing assumption (i) together with [66, equation (3.41)] implies that

$$\operatorname{char} H^0(M, \mathcal{O}(-kK_M)) = \sum_{\alpha\in F} (-1)^{n-n_\alpha-\nu^C_\alpha} \int_{M^T_\alpha} \operatorname{ch}^T\left(\frac{-kK_M|_{M^T_\alpha} \otimes \det(N^{-C}_\alpha)}{\det(1-(N^C_\alpha)^*) \otimes \det(1-N^{-C}_\alpha)}\right) \operatorname{td}(M^T_\alpha),$$

where, if R is a finite-dimensional representation of T,

$$\frac{1}{\det(1-R)} := \bigoplus_{m=0}^{\infty} \operatorname{Sym}^m(R),$$

and where ch^T denotes the equivariant Chern character. For a fixed $\alpha \in F$, we therefore have that

$$\mathrm{ch}^T\left(\frac{-kK_M|_{M_\alpha^T} \otimes \det(N_\alpha^{-C})}{\det(1-(N_\alpha^C)^*) \otimes \det(1-N_\alpha^{-C})}\right)\mathrm{td}(M_\alpha^T)$$
$$= \mathrm{ch}^T(-kK_M|_{M_\alpha^T})\,\mathrm{ch}^T\left(\frac{1}{\det(1-(N_\alpha^C)^*)}\right)\mathrm{ch}^T\left(\frac{1}{\det(1-N_\alpha^{-C})}\right)\mathrm{ch}^T(\det(N_\alpha^{-C}))\,\mathrm{td}(M_\alpha^T).$$

Now,

$$\mathrm{td}(M_\alpha^T) = 1 + \tfrac{1}{2}c_1(-K_{M_\alpha^T}) + \cdots.$$

Analysing the term $\mathrm{ch}^T(-kK_M|_{M_\alpha^T})$, we have by adjunction that

$$K_M|_{M_\alpha^T} = K_{M_\alpha^T} - \det(N_\alpha),$$

so that

$$-kK_M|_{M_\alpha^T} = (-kK_{M_\alpha^T}) + k\det(N_\alpha).$$

Now, M_α^T is fixed under the action of T and so the action of T on $-kK_{M_\alpha^T}$ is trivial. The torus T therefore acts on $-kK_M|_{M_\alpha^T}$ as multiplication by $e^{k\sum_{i=1}^{n-n_\alpha}\lambda_{\alpha,i}}$, where n_α is the dimension of M_α^T. Thus, we have that

$$c_1^T(-kK_M|_{M_\alpha^T}) = kc_1(-K_M|_{M_\alpha^T}) + k\sum_{i=1}^{n-n_\alpha}\lambda_{\alpha,i},$$

where c_1^T is the equivariant first Chern class, so that

$$\mathrm{ch}^T(-kK_M|_{M_\alpha^T})\left(\frac{\zeta}{k}\right) = e^{c_1^T(-kK_M|_{M_\alpha^T})(\zeta/k)} = e^{kc_1(-K_M|_{M_\alpha^T})}e^{\sum_{i=1}^{n-n_\alpha}\lambda_{\alpha,i}(\zeta)}.$$

Next analysing the second term, we may write $N_\alpha^C = \bigoplus_{\{i|\lambda_{\alpha,i}\in C^*\}} L_{\alpha,i}$, where $L_{\alpha,i}$ is the line subbundle of N_α with isotropy weight $\lambda_{\alpha,i}$. Then we have that

$$\frac{1}{\det(1-(N_\alpha^C)^*)} = \bigoplus_{m=0}^{\infty}\mathrm{Sym}^m((N_\alpha^C)^*) = \bigotimes_{\{i|\lambda_{\alpha,i}\in C^*\}}\frac{1}{\det(1-L_{\alpha,i}^*)},$$

so that

$$\mathrm{ch}^T\left(\frac{1}{\det(1-(N_\alpha^C)^*)}\right) = \mathrm{ch}^T\left(\bigotimes_{\{i|\lambda_{\alpha,i}\in C^*\}}\frac{1}{\det(1-L_{\alpha,i}^*)}\right) = \prod_{\{i|\lambda_{\alpha,i}\in C^*\}}\mathrm{ch}^T\left(\frac{1}{\det(1-L_{\alpha,i}^*)}\right).$$

Now observe that, for each i,

$$\mathrm{ch}^T\left(\frac{1}{\det(1-L_{\alpha,i}^*)}\right) = \mathrm{ch}^T\left(\bigoplus_{m=0}^{\infty}\mathrm{Sym}^m(L_{\alpha,i}^*)\right) = \mathrm{ch}^T\left(\bigoplus_{m=0}^{\infty}(L_{\alpha,i}^*)^m\right)$$
$$= \sum_{m=0}^{\infty}(\mathrm{ch}^T(L_{\alpha,i}^*))^m = \sum_{m=0}^{\infty}(e^{-\lambda_{\alpha,i}+c_1(L_{\alpha,i}^*)})^m$$
$$= \frac{1}{1-e^{-\lambda_{\alpha,i}+c_1(L_{\alpha,i}^*)}}.$$

Hence,

$$\mathrm{ch}^T\left(\frac{1}{\det(1-(N_\alpha^C)^*)}\right)=\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{1}{1-e^{-\lambda_{\alpha,i}+c_1(L^*_{\alpha,i})}}.$$

Consequently,

$$\begin{aligned}
\mathrm{ch}^T&\left(\frac{1}{\det(1-(N_\alpha^C)^*)}\right)\left(\frac{\zeta}{k}\right)\\
&=\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{1}{1-e^{-(1/k)\lambda_{\alpha,i}(\zeta)-c_1(L_{\alpha,i})}}\\
&=\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{1}{1-\Bigl(1+\bigl(-\frac{1}{k}\lambda_{\alpha,i}(\zeta)-c_1(L_{\alpha,i})\bigr)+\sum_{l=2}^{\infty}\frac{1}{l!}\bigl(-\frac{1}{k}\lambda_{\alpha,i}(\zeta)-c_1(L_{\alpha,i})\bigr)^l\Bigr)}\\
&=\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{1}{\frac{1}{k}\lambda_{\alpha,i}(\zeta)+c_1(L_{\alpha,i})-\sum_{l=2}^{\infty}\frac{1}{l!}\bigl(-\frac{1}{k}\lambda_{\alpha,i}(\zeta)-c_1(L_{\alpha,i})\bigr)^l}\\
&=\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{k}{\lambda_{\alpha,i}(\zeta)+kc_1(L_{\alpha,i})-k\sum_{l=2}^{\infty}\frac{1}{l!}\bigl(-\frac{1}{k}\lambda_{\alpha,i}(\zeta)-c_1(L_{\alpha,i})\bigr)^l}\\
&=\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{k}{\lambda_{\alpha,i}(\zeta)\Bigl(1+\frac{kc_1(L_{\alpha,i})}{\lambda_{\alpha,i}(\zeta)}+\frac{k}{\lambda_{\alpha,i}(\zeta)}\sum_{l=2}^{\infty}\frac{(-1)^{l+1}}{l!}\bigl(\frac{1}{k}\lambda_{\alpha,i}(\zeta)+c_1(L_{\alpha,i})\bigr)^l\Bigr)}.
\end{aligned}$$

Now,

$$\Bigl(\frac{1}{k}\lambda_{\alpha,i}(\zeta)+c_1(L_{\alpha,i})\Bigr)^l=c_1(L_{\alpha,i})^l+\frac{l}{k}c_1(L_{\alpha,i})^{l-1}\lambda_{\alpha,i}(\zeta)+O(k^{-2}),$$

so that

$$k\Bigl(\frac{1}{k}\lambda_{\alpha,i}(\zeta)+c_1(L_{\alpha,i})\Bigr)^l=kc_1(L_{\alpha,i})^l+lc_1(L_{\alpha,i})^{l-1}\lambda_{\alpha,i}(\zeta)+O(k^{-1}).$$

Since $l\geq 2$, we have that

$$\frac{k}{\lambda_{\alpha,i}(\zeta)}\sum_{l=2}^{\infty}\frac{(-1)^{l+1}}{l!}\Bigl(\frac{1}{k}\lambda_{\alpha,i}(\zeta)+c_1(L_{\alpha,i})\Bigr)^l=O(k)c_1(L_{\alpha,i})^2P_1+c_1(L_{\alpha,i})P_2+O(k^{-1}),$$

where P_1 and P_2 are polynomials in $c_1(L_{\alpha,i})$. Therefore, we see that

$$\begin{aligned}
\mathrm{ch}^T&\left(\frac{1}{\det(1-(N_\alpha^C)^*)}\right)\left(\frac{\zeta}{k}\right)\\
&=\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{k}{\lambda_{\alpha,i}(\zeta)\Bigl(1+\frac{kc_1(L)}{\lambda_{\alpha,i}(\zeta)}\Bigr)+O(k)c_1(L_{\alpha,i})^2P_1+c_1(L_{\alpha,i})P_2+O(k^{-1})}\\
&=k^{\nu_\alpha^C}\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{1}{\lambda_{\alpha,i}(\zeta)\Bigl(1+\frac{kc_1(L)}{\lambda_{\alpha,i}(\zeta)}\Bigr)+O(k)c_1(L_{\alpha,i})^2P_1+c_1(L_{\alpha,i})P_2+O(k^{-1})}.
\end{aligned}$$

A similar argument also shows that

$$\operatorname{ch}^T\left(\frac{1}{\det(1-N_\alpha^{-C})}\right)\left(\frac{\zeta}{k}\right) = (-k)^{n-n_\alpha-\nu_\alpha^C} \prod_{\{i\mid\lambda_{\alpha,i}\in -C^*\}} \frac{1}{\lambda_{\alpha,i}(\zeta)\left(1+\dfrac{kc_1(L_{\alpha,i})}{\lambda_{\alpha,i}(\zeta)}\right)+O(k)c_1(L_{\alpha,i})^2Q_1+c_1(L_{\alpha,i})Q_2+O(k^{-1})}$$

for polynomials Q_1 and Q_2 in $c_1(L_{\alpha,i})$.

Finally,

$$\operatorname{ch}^T(\det(N_\alpha^{-C})) = e^{c_1^T(\det(N_\alpha^{-C}))} = e^{c_1(\det(N_\alpha^{-C}))+\sum_{\{i\mid\lambda_{\alpha,i}\in -C^*\}}\lambda_{\alpha,i}},$$

so that

$$\operatorname{ch}^T(\det(N_\alpha^{-C}))\left(\frac{\zeta}{k}\right) = e^{c_1(N_\alpha^{-C})+(1/k)\sum_{\{i\mid\lambda_{\alpha,i}\in -C^*\}}\lambda_{\alpha,i}(\zeta)}.$$

Putting all of the above observations together, we find that

$$\begin{aligned}
&\frac{1}{k^n}\operatorname{char} H^0(M,\mathcal{O}(-kK_M))\left(\frac{\zeta}{k}\right)\\
&= \sum_{\alpha\in F}\frac{(-1)^{n-n_\alpha-\nu_\alpha^C}}{k^n}\int_{M_\alpha^T}\operatorname{ch}^T\left(\frac{-kK_M|_{M_\alpha^T}\otimes\det(N_\alpha^{-C})}{\det(1-(N_\alpha^C)^*)\otimes\det(1-N_\alpha^{-C})}\right)\operatorname{td}(M_\alpha^T)\left(\frac{\zeta}{k}\right)\\
&= \sum_{\alpha\in F}\frac{1}{k^{n_\alpha}}\int_{M_\alpha^T} e^{kc_1(-K_M|_{M_\alpha^T})}e^{\sum_{i=1}^{n-n_\alpha}\lambda_{\alpha,i}(\zeta)}e^{c_1(N_\alpha^{-C})+(1/k)\sum_{\{i\mid\lambda_{\alpha,i}\in -C^*\}}\lambda_{\alpha,i}(\zeta)}\\
&\quad\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{1}{\lambda_{\alpha,i}(\zeta)\left(1+\dfrac{kc_1(L_{\alpha,i})}{\lambda_{\alpha,i}(\zeta)}\right)+O(k)c_1(L_{\alpha,i})^2P_1+c_1(L_{\alpha,i})P_2+O(k^{-1})}\\
&\quad\prod_{\{i\mid\lambda_{\alpha,i}\in -C^*\}}\frac{1}{\lambda_{\alpha,i}(\zeta)\left(1+\dfrac{kc_1(L_{\alpha,i})}{\lambda_{\alpha,i}(\zeta)}\right)+O(k)c_1(L_{\alpha,i})^2Q_1+c_1(L_{\alpha,i})Q_2+O(k^{-1})}\operatorname{td}(M_\alpha^T)\\
&= \sum_{\alpha\in F}\frac{1}{k^{n_\alpha}}\int_{M_\alpha^T} e^{kc_1(-K_M|_{M_\alpha^T})}e^{\sum_{i=1}^{n-n_\alpha}\lambda_{\alpha,i}(\zeta)}e^{c_1(N_\alpha^{-C})+(1/k)\sum_{\{i\mid\lambda_{\alpha,i}\in -C^*\}}\lambda_{\alpha,i}(\zeta)}\prod_{i=1}^{n-n_\alpha}\frac{1}{\lambda_{\alpha,i}(\zeta)}\\
&\quad\prod_{\{i\mid\lambda_{\alpha,i}\in C^*\}}\frac{1}{1+\dfrac{kc_1(L_{\alpha,i})}{\lambda_{\alpha,i}(\zeta)}+O(k)c_1(L_{\alpha,i})^2P_1+c_1(L_{\alpha,i})P_2+O(k^{-1})}\\
&\quad\prod_{\{i\mid\lambda_{\alpha,i}\in -C^*\}}\frac{1}{1+\dfrac{kc_1(L_{\alpha,i})}{\lambda_{\alpha,i}(\zeta)}+O(k)c_1(L_{\alpha,i})^2Q_1+c_1(L_{\alpha,i})Q_2+O(k^{-1})}\left(1+\tfrac{1}{2}c_1(-K_{M_\alpha^T})+\cdots\right)\\
&\xrightarrow[k\to\infty]{}\sum_{\alpha\in F}\int_{M_\alpha^T} e^{c_1(-K_M|_{M_\alpha^T})}e^{\sum_{i=1}^{n-n_\alpha}\lambda_{\alpha,i}(\zeta)}\prod_{i=1}^{n-n_\alpha}\frac{1}{\lambda_{\alpha,i}(\zeta)\left(1+\dfrac{c_1(L_{\alpha,i})}{\lambda_{\alpha,i}(\zeta)}\right)},
\end{aligned}$$

where, in taking the limit, we use the fact that any integrand involving terms not of the form $k^j\sigma_j$ for σ_j a real (j, j)–form vanishes. The result now follows from an application of Theorem A.3, making use of assumption (ii) of the proposition. □

Given this proposition, it is tempting to define a notion of K–stability that characterises algebraically the existence of a shrinking gradient Kähler–Ricci soliton on a complete anticanonically polarised Kähler manifold M endowed with a complete holomorphic vector field following the strategy as implemented in the Fano case. For this purpose, we make the following definition.

Definition 7.3 Let M be a quasiprojective manifold endowed with the effective holomorphic action of a real torus T whose fixed-point set is compact. Denote by $\mathfrak{t}$ the Lie algebra of T, let $\mathcal{O}_M(M)$ denote the global algebraic sections of the structure sheaf of M, and write

$$\mathcal{O}_M(M) = \bigoplus_{\alpha\in\mathfrak{t}^*} \mathcal{H}_\alpha$$

for the weight decomposition under the action of T. Then we say that a vector field $Y \in \mathfrak{t}$ on M is *positive* if $\alpha(Y) > 0$ for all $\alpha \in \mathfrak{t}^*$ such that $\mathcal{H}_\alpha \neq \varnothing$ and $\alpha \neq 0$.

Remark 7.4 If $\pi : M \to C_0$ is a quasiprojective equivariant resolution of a Kähler cone (C_0, g_0) with respect to the holomorphic isometric torus action on (C_0, g_0) generated by the flow of the Reeb vector field of g_0, and g is a Kähler metric on M that is asymptotic to g_0 and with respect to which the induced torus action on (M, g) is isometric and Hamiltonian, then in the terminology just introduced, the weighted volume functional F for (M, g) is defined on the open cone of positive vector fields in the Lie algebra of the torus if this open cone is nonempty. This fact follows from Theorem A.3 after noting Theorem A.10.

Roughly speaking, one considers equivariant degenerations (or test configurations) of the pair (M, X), where M is a quasiprojective anticanonically polarised Kähler manifold with complex structure J endowed with the holomorphic effective action of a real torus T whose fixed-point set is compact, and where X is a vector field on M with JX a positive vector field lying in the Lie algebra of T. Then one defines a Futaki invariant in the usual manner as the derivative of the algebraic realisation of the weighted volume functional which is given by the right-hand side of (7-1). Of course, one must verify that this formula is well-defined in general. One subsequently defines (M, X) as above to be K–stable if and only if the Futaki invariant is nonnegative on all test configurations, and positive if and only if the test configuration is nontrivial. This then allows one to make the following conjecture generalising the Yau–Tian–Donaldson conjecture for Fano manifolds.

Conjecture 7.5 *Let M be a quasiprojective anticanonically polarised Kähler manifold endowed with the holomorphic effective action of a real torus T whose fixed-point set is compact. Denote by $\mathfrak{t}$ the Lie algebra of T and let X be a vector field on M such that $JX \in \mathfrak{t}$ is a positive vector field. Then M admits a complete shrinking gradient Kähler–Ricci soliton with soliton vector field X if and only if (M, X) is K–stable.*

Thus, in light of this conjecture, one may view an anticanonically polarised Kähler manifold as a "noncompact Fano manifold". (In a similar manner, one may also define a noncompact manifold of general type, etc.) We expect that the well-developed machinery in the study of Kähler–Einstein metrics may be suitably adapted to study this conjecture. We leave this for future work.

7.2 Open problems

There are also various other interesting open problems that we raise here.

(1) Is a complete expanding or shrinking gradient Kähler–Ricci soliton necessarily algebraic (or quasiprojective)? In particular, is the canonical ring of an expanding gradient Kähler–Ricci soliton finitely generated? Is the anticanonical ring of a shrinking gradient Kähler–Ricci soliton finitely generated? What we can say here is that if the curvature tensor of a shrinking gradient Kähler–Ricci soliton decays quadratically, or if that of an expanding gradient Kähler–Ricci soliton decays quadratically with derivatives, then the soliton lives on a resolution of a Kähler cone by Theorem A, hence is quasiprojective by Proposition 2.24.

(2) Is there at most one complete shrinking Kähler–Ricci soliton for a given holomorphic vector field on an anticanonically polarised Kähler manifold up to automorphisms of the complex structure commuting with the flow of the vector field? More speculatively, is a complete shrinking Kähler–Ricci soliton on such a manifold unique up to automorphisms of the complex structure? A noncompact Kähler manifold may admit many nonisometric complete expanding gradient Kähler–Ricci solitons even for a fixed holomorphic soliton vector field, as demonstrated by [19, Theorem A].

(3) What are the constraints on a Kähler cone to appear as the tangent cone of a complete shrinking gradient Kähler–Ricci soliton with quadratic curvature decay? Is the underlying complex manifold of the shrinking soliton then determined uniquely by its tangent cone? By Theorem A, we know that the shrinking soliton must live on a resolution of its tangent cone that is, moreover, an anticanonically polarised Kähler manifold. For complete expanding gradient Kähler–Ricci solitons with quadratic curvature decay with derivatives, we know from Corollary B that a Kähler cone appears as the tangent cone if and only if the Kähler cone has a smooth canonical model (on which the soliton lives).

(4) Related to the previous question, modulo automorphisms of the complex structure, how many shrinking gradient Kähler–Ricci solitons with quadratic curvature decay have a given affine cone appearing as the underlying complex space of the tangent cone? For $\mathbb{C}^2$, we have shown in Theorem E that the answer is two; $\mathbb{C}^2$ only appears as the underlying complex space of the tangent cone of the flat Gaussian shrinking soliton on $\mathbb{C}^2$ and of the $U(2)$–invariant shrinking gradient Kähler–Ricci soliton of Feldman, Ilmanen and Knopf on $\mathbb{C}^2$ blown up at a point [30]. In general, we expect the answer to be finitely many for any given affine cone. By Corollary B, the answer to this question for complete expanding gradient Kähler–Ricci solitons with quadratic curvature decay with derivatives is infinitely many for any given affine cone admitting a smooth canonical model.

(5) In Corollary B, we have seen that when the canonical model of a Kähler cone is smooth, it admits a complete expanding gradient Kähler–Ricci soliton. Is this also true when the canonical model is singular? (We thank John Lott for raising this question.)

(6) Let M be a complete quasiprojective Kähler manifold endowed with the holomorphic Hamiltonian action of a real torus T with Lie algebra $\mathfrak{t}$ whose fixed-point set is compact. Are the elements of $\mathfrak{t}$ admitting Hamiltonian potentials that are proper and bounded below precisely those elements in $\mathfrak{t}$ that are positive in the sense of Definition 7.3? Equivalently, does the open cone of positive vector fields in $\mathfrak{t}$ coincide with the open cone $\operatorname{int}(\mathcal{C}(\mu(M))') \subset \mathfrak{t}$ of Proposition A.4? These questions have an affirmative answer in the setting of asymptotically conical Kähler manifolds; see Theorem A.10 for a precise statement.

(7) Does Theorem D still hold true without the assumption of bounded Ricci curvature?

(8) Given a complete shrinking gradient Kähler–Ricci soliton (M, g, X), is the zero set of X always compact? As demonstrated in Lemma 2.25, this is the case if g has bounded scalar curvature.

(9) Let M be a complete Kähler manifold endowed with the holomorphic Hamiltonian action of a real torus T with Lie algebra $\mathfrak{t}$ whose fixed-point set is compact. By the Duistermaat–Heckman theorem (Theorem A.3), the weighted volume functional F is defined on the open cone Λ of elements of $\mathfrak{t}$ admitting Hamiltonian potentials that are proper and bounded below. Is F necessarily proper on Λ? If so, then it would have a unique minimiser on Λ. Properness of the volume functional on the set of normalised Reeb vector fields of a Sasaki manifold was shown in [37, Proposition 3.3].

Appendix The Duistermaat–Heckman theorem

A.1 Statement of the theorem

The material in this section has been taken verbatim from various sources in the literature, including [5; 28; 49; 60]. We begin with the definition of a moment map, which is required for the statement of the Duistermaat–Heckman theorem.

Definition A.1 Let (M, ω) be a symplectic manifold and let T be a real torus acting on (M, ω) by symplectomorphisms. Denote by $\mathfrak{t}$ the Lie algebra of T and by $\mathfrak{t}^*$ its dual. Then we say that the action of T is *Hamiltonian* if there exists a smooth map $\mu\colon M \to \mathfrak{t}^*$ such that for all $\zeta \in \mathfrak{t}$, $-\omega \lrcorner \zeta = du_\zeta$, where $u_\zeta(x) = \langle \mu(x), \zeta \rangle$ for all $\zeta \in \mathfrak{t}$ and $x \in M$. We call μ the *moment map* of the T–action and we call u_ζ the *Hamiltonian* (*potential*) of ζ.

Notice that u_ζ is invariant under the flow of ζ. Indeed, we have that

$$\mathcal{L}_\zeta u_\zeta = du_\zeta \lrcorner \zeta = -\omega(\zeta, \zeta) = 0.$$

Consequently, each integral curve of ζ must be contained in a level set of u_ζ.

Now consider the Hamiltonian action of a real torus T of rank s on a symplectic manifold (M,ω) of real dimension $2n$. Identify T with $(S^1)^s \subset (\mathbb{C}^*)^s$ and introduce complex coordinates $(\phi_1,\dots,\phi_s)$ on T via this identification. This induces coordinates $(\eta_1,\dots,\eta_s)\in\mathbb{R}^s$ on the Lie algebra $\mathfrak{t}$ of T, where $(\eta_1,\dots,\eta_s)$ corresponds to the vector $\sum_{i=1}^{s}\eta_i(\partial/\partial\phi_i)$, each $\partial/\partial\phi_i$ the vector field on M induced by the coordinate ϕ_i on T. For $\zeta\in\mathfrak{t}$ with coordinates $(b_1,\dots,b_s)$, say, the flow on M generated by ζ will have a fixed-point set $M_0(\zeta)$ corresponding to the zero set of the vector field ζ. This set has the following properties.

Proposition A.2 [5, Proposition 7.12] *The connected components $\{F_i\}$ of $M_0(\zeta)$ are smooth submanifolds of M. The dimensions of different connected components do not have to be the same. The normal bundles $\mathcal{E}_i$ of the F_i in M are orientable vector bundles with even-dimensional fibres.*

For a disconnected component F of $M_0(\zeta)$ of real codimension $2k$ in M, let $\iota\colon F\to M$ denote the inclusion. Then $\iota^*\omega$ is a symplectic form on F so that F is a symplectic submanifold of M. The normal bundle $\mathcal{E}$ of F in M has the structure of a symplectic vector bundle and will have real dimension $2k$. We denote this induced symplectic form on $\mathcal{E}$ by τ. The flow of ζ will generate a fibre-preserving linear action $L\zeta:\mathcal{E}\to\mathcal{E}$ on $\mathcal{E}$, which is an automorphism of $\mathcal{E}$ leaving τ invariant in the infinitesimal sense. We introduce an almost complex structure $I:\mathcal{E}\to\mathcal{E}$, ie an automorphism of $\mathcal{E}$ such that $I^2=-\mathrm{id}$, commuting with $L\zeta$ and compatible with τ in the sense that $\tau(I\,\cdot,\cdot)$ defines an inner product on $\mathcal{E}$. This gives $\mathcal{E}$ the structure of a complex vector bundle over F with $L\zeta$ an automorphism of $\mathcal{E}$ preserving the complex structure.

Next, denote by $u_1,\dots,u_R\in\mathbb{Z}^s\subset\mathfrak{t}^*$ the weights of the induced representation of $\mathfrak{t}$ on $\mathcal{E}$. Then we have a direct sum decomposition of vector bundles

$$\mathcal{E}=\bigoplus_{\lambda=1}^{R}\mathcal{E}_\lambda,$$

where

$$\mathcal{E}_\lambda:=\{v\in\mathcal{E}\mid(e^{i\eta_1},\dots,e^{i\eta_s})\cdot v=u_\lambda(\eta_1,\dots,\eta_s)Iv\ \text{ for all }\ (\eta_1,\dots,\eta_n)\in\mathfrak{t}\}.$$

Each $\mathcal{E}_\lambda$ is a vector bundle of rank $2n_\lambda$, say. Clearly we must have $k=\sum_{\lambda=1}^{R}n_\lambda$. Consider now the complex vector bundle $\mathcal{E}^{1,0}$ of complex dimension k, endowed with the action of $L\zeta$ extended by $\mathbb{C}$–linearity. Then we have an induced decomposition

$$\mathcal{E}^{1,0}=\bigoplus_{\lambda=1}^{R}\mathcal{E}_\lambda^{1,0},$$

where $L\zeta$ acts on the λ^{th} factor by $iu_\lambda(b_1,\dots,b_n)$, and so the action of $L\zeta$ on $\mathcal{E}^{1,0}$ will take the form

$$L\zeta=i\,\mathrm{diag}(1_{n_1}u_1(b),\dots,1_{n_R}u_R(b)),$$

where 1_{n_λ} denotes the $n_\lambda\times n_\lambda$ identity matrix and $b=(b_1,\dots,b_s)$ are the coordinates of ζ. Thus,

$$\det\left(\frac{L\zeta}{i}\right)=\prod_{\lambda=1}^{R}u_\lambda(b)^{n_\lambda}.$$

Note that this is homogeneous of degree k in b.

We next choose any $L\zeta$–invariant connection on $\mathcal{E}^{1,0}$ with curvature matrix Ω. Finally, for a polyhedral set U of a vector space V, we define the *asymptotic cone*

$$\mathcal{C}(U) := \{v \in V \mid \text{there is a } v_0 \in V \text{ such that } v_0 + tv \in U \text{ for } t > 0 \text{ sufficiently large}\},$$

and for a subset W of V, we define the *dual cone* $W' := \{\alpha \in V^* \mid \alpha(W) \subseteq \mathbb{R}_{\geq 0}\}$. Now we can state the Duistermaat–Heckman theorem.

Theorem A.3 (the Duistermaat–Heckman theorem [60, Theorem 2.2]) *Let (M, ω) be a (possibly noncompact) symplectic manifold of real dimension $2n$ with symplectic form ω on which there is a Hamiltonian action of a real torus T with moment map $\mu\colon M \to \mathfrak{t}^*$, where $\mathfrak{t}$ is the Lie algebra of T and $\mathfrak{t}^*$ its dual. Assume that the fixed-point set of T is compact. If there exists an element $\zeta_0 \in \mathfrak{t}$ such that the component of the moment map $u_{\zeta_0} = \langle \mu, \zeta_0 \rangle$ is proper and bounded below, then*

$$\int_M e^{-\langle \mu, \zeta \rangle} \frac{\omega^n}{n!} = \sum_{F \in M_0(\zeta)} \int_F \frac{e^{-\iota^* \langle \mu, \zeta \rangle} e^{\iota^* \omega}}{\det\left(\dfrac{L\zeta - \Omega}{2\pi i} \right)} \tag{A-1}$$

for all ζ in the open cone $\operatorname{int}(\mathcal{C}(\mu(M))') \subset \mathfrak{t}$, where the sum on the right-hand side is taken over the connected components F of the zero set $M_0(\zeta)$ of ζ.

Under the assumptions on the moment map μ as in the theorem, $\mu(M)$ is a proper polyhedral set in $\mathfrak{t}^*$ and the elements of $\operatorname{int}(\mathcal{C}(\mu(M))') \subset \mathfrak{t}$ are characterised as follows.

Proposition A.4 [60, Proposition 1.4] *Under the assumptions on T and μ as in Theorem A.3, $u_\zeta = \langle \mu, \zeta \rangle$ is proper if and only if $\zeta \in \pm \operatorname{int}(\mathcal{C}(\mu(M))') \subset \mathfrak{t}$. Moreover, if $\zeta \in \operatorname{int}(\mathcal{C}(\mu(M))') \subset \mathfrak{t}$, then $u_\zeta(M) = [m_\zeta, +\infty)$ for a suitable $m_\zeta \in \mathbb{R}$.*

That is, elements of $\operatorname{int}(\mathcal{C}(\mu(M))')$ are precisely those elements of $\mathfrak{t}$ whose Hamiltonian is proper and bounded below. Notice that this cone is nonempty because it contains ζ_0 by assumption. Then for each $\zeta \in \operatorname{int}(\mathcal{C}(\mu(M))')$, each connected component of the zero set $M_0(\zeta)$ of ζ must be compact because $u_\zeta = \langle \mu, \zeta \rangle$ is proper, and moreover, it must contain a fixed point of the torus action by [60, Proposition 1.2]. Hence, since the fixed-point set of T is assumed to be compact in Theorem A.3, the sum on the right-hand side of (A-1) is over a finite set and so is itself finite for all such ζ.

Now, the sum on the right-hand side of (A-1) is over each connected component F of the zero set $M_0(\zeta)$ of ζ. The determinant is a $k \times k$ determinant and should be expanded formally into a differential form of mixed degree. Moreover, the inverse is understood to mean one should expand this formally in a Taylor series, as is standard in index theory. We next study the right-hand side of (A-1) in more detail.

Under the decomposition

$$\mathcal{E}^{1,0} = \bigoplus_{\lambda=1}^{R} \mathcal{E}^{1,0}_{\lambda},$$

let Ω_λ be the component of the curvature matrix of $\Omega_{\mathcal{E}}$ corresponding to $\mathcal{E}_\lambda$. Then we have that

$$\det\left(\frac{L\zeta-\Omega}{2\pi i}\right)=\det\left(\frac{L\zeta}{2\pi i}\right)\det(1-(L\zeta)^{-1}\Omega)=\det\left(\frac{L\zeta}{2\pi i}\right)\prod_{\lambda=1}^{R}\det(1-(L\zeta)^{-1}\Omega_\lambda).$$

Fix one of the bundles $\mathcal{E}_\lambda$. Then

$$\det(1-(L\zeta)^{-1}\Omega_\lambda)=\det(1+wi\,\Omega_\lambda)=\sum_{a\geq 0}c_a(\mathcal{E}_\lambda)w^a\in H^*(F,\mathbb{R}),$$

where $w=1/u_\lambda(b)$ and $c_a(\mathcal{E}_\lambda)$ are the Chern classes of $\mathcal{E}_\lambda$ for $0\leq a\leq n_\lambda$ with $c_0=1$. Thus,

$$\det\left(\frac{L\zeta-\Omega}{2\pi i}\right)=\prod_{\lambda=1}^{R}u_\lambda(b)^{n_\lambda}\left(\sum_{a\geq 0}c_a(\mathcal{E}_\lambda)w^a\right).$$

In particular, if F is an isolated fixed point, in which case $k=n$ and $\mathcal{E}$ is the trivial bundle, then we may write the n (possibly indistinct) weights as $u_1,\dots,u_n$. The Chern classes and the measure $e^{\iota^*\omega}$ contribute nontrivially, and we arrive at the contribution

$$e^{-\iota^*\langle\mu,\zeta\rangle}\prod_{\lambda=1}^{R}\frac{1}{u_\lambda(b)^{n_\lambda}}$$

of an isolated fixed point to the Duistermaat–Heckman formula.

We next wish to sketch the proof of Theorem A.3. Before we do so, however, we must first discuss invariant forms on a symplectic manifold.

A.2 Invariant forms

Consider a symplectic manifold (M,ω) of real dimension $2n$ endowed with the Hamiltonian action of a real torus T. For ζ in the Lie algebra $\mathfrak{t}$ of T, denote by $\Omega^k_\zeta(M)$ the space of smooth k–forms on M which are invariant under the flow of ζ, ie $\alpha\in\Omega^k_\zeta(M)$ if and only if $\mathcal{L}_\zeta\alpha=0$. The wedge product of two invariant forms is also invariant, therefore we have an algebra $\Omega^*_\zeta(M)$ of invariant forms on M. We define the *equivariant derivative* d_ζ on $\Omega^*_\zeta(M)$ by

$$d_\zeta\alpha=d\alpha-\alpha\lrcorner\zeta.$$

This derivative has the properties that $d_\zeta^2=0$ and

$$d_\zeta(\alpha\wedge\beta)=d_\zeta\alpha\wedge\beta+(-1)^p\alpha\wedge d_\zeta\beta$$

for α a p–form and β another differential form.

For $\alpha\in\Omega^*_T(M)$, we can write

$$\alpha=\alpha_{[0]}+\alpha_{[1]}+\cdots+\alpha_{[2n]},$$

with $\alpha_{[i]}$ a differential form of degree i in $\Omega^*_T(M)$. Then integration of invariant forms is defined by integrating over the highest-degree part of the form, ie

$$\int : \Omega^*_T(M) \to \mathbb{R}, \quad \int \alpha := \int_M \alpha_{[2n]}.$$

This leads to a version of Stokes' theorem for invariant forms: if an invariant form α is d_ζ–exact, ie if $\alpha = d_\zeta \beta$ for another form β, then $\alpha_{[2n]} = d\beta_{[2n-1]}$, since contracting with ζ decreases the degree of a form. We then have that

$$\int_M \alpha := \int_M \alpha_{[2n]} = \int_M d\beta_{[2n-1]} = \int_{\partial M} \beta_{[2n-1]}.$$

Recall that $M_0(\zeta)$ denotes the zero locus on M of $\zeta \in \mathfrak{t}$.

Lemma A.5 *Let $\alpha \in \Omega^*_\zeta(M)$ be d_ζ–closed. Then $\alpha_{[2n]}$ is exact on $M \setminus M_0(\zeta)$.*

Proof Let θ be a one-form on $M \setminus M_0(\zeta)$ such that

$$\mathcal{L}_\zeta \theta = 0 \quad \text{and} \quad \theta \lrcorner \zeta \neq 0. \tag{A-2}$$

Such a one-form can be constructed explicitly. Indeed, let g be a T–invariant Riemannian metric on M, let $\tilde{\zeta}$ be any nonzero positive smooth function times ζ, and define

$$\theta(v) = g(\tilde{\zeta}, v) \quad \text{for any vector field } v \text{ on } M.$$

This is well-defined on $M \setminus M_0(\zeta)$ as ζ, and hence $\tilde{\zeta}$, are nonzero on this set, and is easily seen to satisfy (A-2). We can then invert $d_\zeta\theta$ on $M \setminus M_0(\zeta)$ using a geometric series:

$$(d_\zeta\theta)^{-1} = \frac{1}{(d\theta - \theta\lrcorner\zeta)} = \frac{1}{(\theta\lrcorner\zeta)((\theta\lrcorner\zeta)^{-1}d\theta - 1)} = -(\theta\lrcorner\zeta)^{-1} - (\theta\lrcorner\zeta)^{-2}d\theta - (\theta\lrcorner\zeta)^{-3}(d\theta)^2 - \cdots.$$

Note that this geometric series is finite because the $(d\theta)^k$ vanish if $2k > 2n = \dim M$, and we have that

$$d_\zeta\theta \wedge (d_\zeta\theta)^{-1} = 1.$$

Applying d_ζ to this yields

$$d_\zeta\theta \wedge d_\zeta((d_\zeta\theta)^{-1}) = 0.$$

Further taking the wedge product with $(d_\zeta\theta)^{-1}$ on the left then leaves us with

$$d_\zeta[(d_\zeta\theta)^{-1}] = 0.$$

Define ν by

$$\nu := \theta \wedge (d_\zeta\theta)^{-1} \wedge \alpha.$$

Then, since $d_\zeta\alpha = 0$ by assumption, we have that

$$d_\zeta\nu = d_\zeta\theta \wedge (d_\zeta\theta)^{-1} \wedge \alpha = \alpha.$$

Taking the highest-degree part of each side of this equality, we obtain the result. □

A.3 Sketch of the proof of Theorem A.3

Since the left-hand side of (A-1) is analytic on $\operatorname{int}(\mathcal{C}(\mu(M))') \subset \mathfrak{t}$, it suffices to prove (A-1) for rational elements in this open cone. So let $\zeta \in \operatorname{int}(\mathcal{C}(\mu(M))')$ be rational and recall that the zero set $M_0(\zeta)$ of ζ is compact because the fixed-point set of T is compact by assumption; see the discussion after Proposition A.4. Write $\zeta = t\eta$ for some integral point $\eta \in \operatorname{int}(\mathcal{C}(\mu(M))')$ and some $t > 0$, and let $H := \langle \mu, \eta \rangle$ denote the Hamiltonian of η, which serves as a moment map of the induced S^1–action of $\{e^{i\eta}\}$ on M. Recall that H is a proper function bounded from below and so must tend to infinity as $x \to \infty$ in M.

Next, observe that

$$e^{\omega - tH} = e^{-tH}\left(1 + \omega + \cdots + \frac{\omega^n}{n!}\right) \in \Omega^*_\zeta(M),$$
$$d_\zeta e^{\omega - tH} = e^{\omega - tH}(d_\zeta(\omega - tH)) = e^{\omega - tH}(-d(tH) - \omega \lrcorner \zeta) = te^{\omega - tH}(-dH - \omega \lrcorner \eta) = 0,$$

so that $e^{\omega - tH}$ is d_ζ–closed. An immediate consequence of Lemma A.5 is therefore that $e^{-tH}\omega^n/n!$ is exact off of the zero set $M_0(\zeta)$ of ζ. Indeed, fix a T–invariant metric g on M. Then tracing through the proof of Lemma A.5, we see that

$$e^{-tH}\frac{\omega^n}{n!} = d\nu_{[2n-1]}, \quad \text{where } \nu = \theta \wedge (d_\zeta \theta)^{-1} \wedge e^{\omega - tH} \text{ and } \theta = g(\tilde{\zeta}, \cdot),$$

with $\tilde{\zeta}$ denoting any nonzero positive function times ζ. We take $\tilde{\zeta} = \eta / g(\eta, \eta)$ in what follows.

Let F denote each of the connected components of $M_0(\zeta)$ and recall that each is a smooth submanifold of M. Using the exponential map of the T–invariant metric g on M, we obtain a diffeomorphism ψ from a neighbourhood U of the zero section of the normal bundle $\mathcal{E}$ of F in E onto a neighbourhood $\psi(U)$ of F in M. For $\varepsilon > 0$, denote by B_ε the ε–ball bundle in $\mathcal{E}$ and by S_ε its boundary. Since $M_0(\zeta)$ is compact and $H(x) \to +\infty$ as $x \to \infty$ in M, we have by Stokes' theorem that

$$\text{(A-3)} \qquad \begin{aligned} \int_M e^{-tH}\frac{\omega^n}{n!} &= \lim_{a \to +\infty} \lim_{\varepsilon \to 0} \int_{H^{-1}((-\infty, a]) \setminus \cup_{F \in M_0(\zeta)} \psi(B_\varepsilon)} e^{-tH}\frac{\omega^n}{n!} \\ &= \lim_{a \to +\infty} \lim_{\varepsilon \to 0} \int_{H^{-1}((-\infty, a]) \setminus \cup_{F \in M_0(\zeta)} \psi(B_\varepsilon)} d\nu_{[2n-1]} \\ &= \lim_{\varepsilon \to 0} \sum_{F \in M_0(\zeta)} \int_{\psi(S_\varepsilon)} \nu_{[2n-1]} + \lim_{a \to +\infty} \int_{H^{-1}(a)} \nu_{[2n-1]}, \end{aligned}$$

where we recall the fact that H is proper and that a is a regular value of H for all a sufficiently large by [60, Proposition 1.2] because the fixed-point set of the torus action is compact, so that $H^{-1}(a)$ is a smooth compact submanifold of M for all such values of a.

Now, we have that

$$\nu_{[2n-1]} = -e^{-tH}\theta \wedge \sum_{j=0}^{n-1} \frac{(d\theta)^j}{(\theta \lrcorner \zeta)^{j+1}} \wedge \frac{\omega^{n-1-j}}{(n-1-j)!} = -e^{-tH}\theta \wedge \sum_{j=0}^{n-1} \frac{(d\theta)^j}{t^{j+1}} \wedge \frac{\omega^{n-1-j}}{(n-1-j)!},$$

where again $\theta = g(\eta,\cdot)/g(\eta,\eta)$. For one connected component $F \in M_0(\zeta)$ of codimension k, say, as in the proof of the Duistermaat–Heckman formula in the compact case [28], the only summand contributing to $\int_{\psi(S_\varepsilon)} \nu_{[2n-1]}$ in the limit as $\varepsilon \to 0$ is the one with $j = k-1$. Therefore, computing as in [28], one sees that

$$\text{(A-4)}\qquad \begin{aligned}\lim_{\varepsilon\to 0}\int_{\psi(S_\varepsilon)} \nu_{[2n-1]} &= -\lim_{\varepsilon\to 0}\int_F e^{-tH}\theta\wedge\frac{(d\theta)^{k-1}}{t^k}\wedge\frac{\omega^{n-k}}{(n-k)!}\\ &= -\lim_{\varepsilon\to 0}\int_F e^{-tH}\widetilde{\theta}\wedge(d\widetilde{\theta})^{k-1}\wedge\frac{\omega^{n-k}}{(n-k)!}\\ &= \frac{e^{-\iota^*(tH)}e^{\iota^*\omega}}{\det\Bigl(\frac{1}{2\pi i}(L\zeta-\Omega)\Bigr)},\end{aligned}$$

where $\widetilde{\theta} = g(\zeta,\cdot)/g(\zeta,\zeta)$ on the second line.

We finally deal with the term $\lim_{a\to+\infty}\int_{H^{-1}(a)} \nu_{[2n-1]}$. Since every a sufficiently large is a regular value of H, the moment map of the S^1–action $\{e^{i\eta}\}$, the set $H^{-1}(a)$ is a connected compact submanifold of M on which the S^1–action is locally free. Let $M_a = H^{-1}(a)/S^1$ be the symplectic quotient with canonical symplectic form ω_a. The preimage $H^{-1}(a)\to M_a$ then has the structure of a orbibundle over M_a. Moreover, since a is a regular value of H, there exists a number $\delta > 0$ such that $H^{-1}((a-\delta,a+\delta))$ is diffeomorphic to $H^{-1}(a)\times(-\delta,\delta)$. With respect to this diffeomorphism, the symplectic form ω on $H^{-1}(a)\times(-\delta,\delta)$ is, up to an exact form, equal to

$$\alpha\wedge dH - (H-a)F_a + \omega_a$$

for one (and hence any) connection one-form α on the orbibundle $H^{-1}(a)\to M_a$ with curvature F_a. Now, when restricted to $H^{-1}(a)$, one can verify that $\omega|_{H^{-1}(a)} = \omega_a$, that $\theta|_{H^{-1}(a)} =: \alpha$ defines a connection 1–form, and that $d\theta|_{H^{-1}(a)} =: F_a$ is the curvature form of α, so that $d_\zeta\theta|_{H^{-1}(a)} = F_a - t$. So we have that

$$\begin{aligned}\int_{H^{-1}(a)} \nu_{[2n-1]} = \int_{H^{-1}(a)} \theta\wedge(d_\zeta\theta)^{-1}\wedge e^{\omega-tH} &= \int_{H^{-1}(a)} \alpha\wedge(F_a-t)^{-1}\wedge e^{\omega_a-ta}\\ &= -\frac{e^{-ta}}{t}\int_{M_a}\Bigl(1-\frac{F_a}{t}\Bigr)^{-1}\wedge e^{\omega_a}\\ &= -\sum_{j=1}^{n-1}\frac{e^{-ta}}{t^{j+1}}\int_{M_a}\frac{\omega_a^{n-1-j}}{(n-1-j)!}\wedge F_a^j.\end{aligned}$$

As $a\to+\infty$, the cohomology class of ω_a depends linearly on a [28; 65], whereas that of F_a remains fixed since the topology of the bundle $H^{-1}(a)\to M_a$ does not change as a runs through a set of regular values. So the integral over M_a here is a polynomial in a. Consequently, $\int_{H^{-1}(a)} \nu_{[2n-1]} \to 0$ exponentially as $a\to+\infty$. Thus, combining this fact with (A-3) and (A-4), and noting that $tH = \langle\mu,\zeta\rangle$, we arrive at the desired conclusion.

A.4 Examples

We next consider some simple examples and see what formula (A-1) yields for the weighted volume functional F.

Example A.6 Let $M=\mathbb{C}^n$ and consider the action of the maximal torus

$$T=\{\operatorname{diag}(e^{i\eta_1},\dots,e^{i\eta_n})\mid \eta_i\in\mathbb{R}\}$$

in $\mathrm{GL}(n,\mathbb{C})$ acting on M with induced coordinates $(\eta_1,\dots,\eta_n)$ on the Lie algebra $\mathfrak{t}$ of T, where $(1,0,\dots,0)\in\mathfrak{t}$ generates the vector field $\operatorname{Im}(z_1\partial_{z_1})$ on M, etc. The fixed-point set of T is clearly compact.

For any $Y\in\{(\eta_1,\dots,\eta_n)\in\mathfrak{t}\mid\eta_i>0\}$ and for any T–invariant complete shrinking gradient Kähler–Ricci soliton (M,ω,X) with $X=\nabla^g f$ for $f\colon M\to\mathbb{R}$ smooth, let u_Y be the Hamiltonian potential of Y normalised as in Definition 5.13, so that, in particular, $\Delta_\omega u_Y+u_Y+\frac{1}{2}(JY)\cdot f=0$. Then

$$-u_Y(0)=(\Delta_\omega u_Y)(0)+\underbrace{\tfrac{1}{2}((JY)\cdot f)(0)}_{=0}=\operatorname{div}(Y)=\sum_j\eta_j,$$

and so the Duistermaat–Heckman theorem yields

$$F(\eta_1,\dots,\eta_n)=\int_M e^{-u_Y}\omega^n=\prod_j\eta_j^{-1}\cdot e^{\sum_j\eta_j}.$$

Since this function is symmetric in its components, its unique critical point must be of the form $\lambda(1,\dots,1)$ for some $\lambda>0$. It is then easy to show that $\lambda=1$. The corresponding shrinking gradient Kähler–Ricci soliton is the flat Gaussian shrinking soliton on $\mathbb{C}^n$.

Example A.7 Let M be $\mathbb{C}^2$ blown up at the origin and again consider the action of the maximal torus $T=\{\operatorname{diag}(e^{i\eta_1},e^{i\eta_2})\mid\eta_1,\eta_2\in\mathbb{R}\}$ in $\mathrm{GL}(2,\mathbb{C})$ acting on M, with induced coordinates (η_1,η_2) on the Lie algebra $\mathfrak{t}$ of T, where $(1,0)\in\mathfrak{t}$ generates the vector field $\operatorname{Im}(z_1\partial_{z_1})$ on M, etc. In this case, the weighted volume functional is given by

$$F\colon\{(\eta_1,\eta_2)\in\mathfrak{t}\mid\eta_1,\eta_2>0\}\to\mathbb{R}_{>0},\quad F(\eta_1,\eta_2)=\begin{cases}\dfrac{e^{\eta_1}}{(\eta_1-\eta_2)\eta_2}+\dfrac{e^{\eta_2}}{(\eta_2-\eta_1)\eta_1} & \text{if } \eta_1\neq\eta_2,\\ e^{\eta_1}(\eta_1^{-1}+\eta_1^{-2}) & \text{if } \eta_1=\eta_2.\end{cases}$$

Again by symmetry, the unique critical point of F here must have $\eta_1=\eta_2$, and a computation shows that $\eta_1=\eta_2=\sqrt{2}$ in this case. The corresponding shrinking gradient Kähler–Ricci soliton is that of Feldman, Ilmanen and Knopf [30] on this space.

Example A.8 More generally, let M be the total space of the line bundle $\mathcal{O}(-k)$ over $\mathbb{P}^{n-1}$ for $0<k<n$ and consider the induced action of the maximal torus $T=\{\operatorname{diag}(e^{i\eta_1},\dots,e^{i\eta_n})\mid\eta_i\in\mathbb{R}\}$ in $\mathrm{GL}(n,\mathbb{C})$ acting on M, with induced coordinates $(\eta_1,\dots,\eta_n)$ on the Lie algebra $\mathfrak{t}$ of T, where $(1,0,\dots,0)\in\mathfrak{t}$ generates the vector field $\operatorname{Im}(z_1\partial_{z_1})$ on M, etc. In this case, the weighted volume

functional $F\colon \{(\eta_1,\dots,\eta_n)\in\mathfrak{t} \mid \eta_i>0\}\to\mathbb{R}_{>0}$ is given by

$$F(\eta_1,\dots,\eta_n)=\sum_{i=1}^{n}\frac{e^{(k+1-n)\eta_i+\sum_{j\neq i}\eta_j}}{k\eta_i\prod_{j\neq i}(\eta_j-\eta_i)}$$

$$=\frac{\sum_{i=1}^{n}(-1)^{i+1}\prod_{j\neq i}\eta_j\prod_{k,l\neq i,\ k>l}(\eta_k-\eta_l)e^{(k+1-n)\eta_i+\sum_{j\neq i}\eta_j}}{k\prod_{i=1}^{n}\eta_i\prod_{i<j}(\eta_i-\eta_j)}\quad\text{if } \eta_k\neq\eta_l \text{ for } k\neq l.$$

Again, by symmetry, the unique critical point of F here must satisfy $\eta_1=\dots=\eta_n$. By taking limits, one can write down an expression for F when this is the case. Differentiating the resulting expression and setting it equal to zero, one obtains the polynomials of [30, equation (36)]. For example, in low dimensions, when $\eta_1=\dots=\eta_n=:\eta$, we obtain the formulae for F in Table 1. The corresponding shrinking gradient Kähler–Ricci solitons are those of Feldman, Ilmanen and Knopf [30] on these spaces.

Example A.9 Let L be the total space of a negative holomorphic line bundle over a Fano manifold D of complex dimension n. By adjunction, in order for L to admit a shrinking gradient Kähler–Ricci soliton, we must have $c_1(-K_D\otimes L)>0$. Assuming that this is the case, consider the action of the torus T given by rotating the fibres of L. We have an induced coordinate w on the Lie algebra $\mathfrak{t}$ of T, where $1\in\mathfrak{t}$ will generate the vector field $\operatorname{Im}(z_i\partial_{z_i})$ in a local trivialising chart of L. The zero set of every element of $\mathfrak{t}$ will be D, the zero section of L, and in this case the weighted volume functional F on the domain $\{\eta\in\mathfrak{t}\mid\eta>0\}$ is given by

$$\text{(A-5)}\quad F(\eta)=\int_{D^n}\frac{e^{\eta}e^{\iota^*\omega}}{\eta\left(1+\frac{c_1(L)}{\eta}\right)}=\frac{e^{\eta}}{\eta}\int_{D^n}e^{\iota^*\omega}\left(1+\frac{c_1(L)}{\eta}\right)^{-1}=\frac{e^{\eta}}{\eta}\int_{D^n}e^{\iota^*\omega}\left(1-\frac{c_1(L^*)}{\eta}\right)^{-1}$$

$$=\frac{e^{\eta}}{\eta}\int_{D^n}\left(1+\omega+\frac{\omega^2}{2!}+\cdots\right)\left(1+\frac{c_1(L^*)}{\eta}+\frac{c_1(L^*)^2}{\eta^2}+\cdots\right)$$

$$=\frac{e^{\eta}}{\eta}\int_{D^n}\sum_{i=0}^{n}\frac{\omega^i}{i!}\wedge\left(\frac{c_1(L^*)^{n-i}}{\eta^{n-i}}\right)=\frac{e^{\eta}}{\eta}\sum_{i=0}^{n}\frac{1}{\eta^{n-i}i!}\int_{D^n}\omega^i\wedge c_1(L^*)^{n-i}$$

$$=\frac{e^{\eta}}{\eta^{n+1}}\sum_{i=0}^{n}\frac{\eta^i}{i!}\int_{D^n}c_1(K_D^{-1}\otimes L)^i\wedge c_1(L^*)^{n-i}.$$

This formula in particular applies to the total space of the line bundle $\mathcal{O}(-k)$ over $\mathbb{P}^{n-1}$ for $0<k<n$. Its relationship to the formulae of Example A.8 is as follows. On the total space of $\mathcal{O}(-k)$, we have two torus actions, one given by the standard action of a torus T_1 rotating the fibres of $\mathcal{O}(-k)$, and another given by the action of a torus T_2 induced from the standard torus action on $\mathcal{O}(-1)$ that rotates the fibres. The formulae of Example A.8 with $\eta_1=\dots=\eta_n$ are given with respect to the action of T_2, whereas formula (A-5) is with respect to T_1. Consequently, the formulae of Example A.8 with $\eta_1=\dots=\eta_n$ are given by $F(k\eta)$, where F is as in (A-5).

line bundle	$F(\eta)$
$\mathcal{O}(-1) \to \mathbb{P}^1$	$\dfrac{(\eta+1)e^{\eta}}{\eta^2}$
$\mathcal{O}(-1) \to \mathbb{P}^2$	$\dfrac{(2\eta^2+2\eta+1)e^{\eta}}{\eta^3}$
$\mathcal{O}(-2) \to \mathbb{P}^2$	$\dfrac{(\eta^2+2\eta+2)e^{2\eta}}{\eta^3}$
$\mathcal{O}(-1) \to \mathbb{P}^3$	$\dfrac{(9\eta^3+9\eta^2+6\eta+2)e^{\eta}}{\eta^4}$
$\mathcal{O}(-2) \to \mathbb{P}^3$	$\dfrac{(4\eta^3+6\eta^2+6\eta+3)e^{2\eta}}{\eta^4}$
$\mathcal{O}(-3) \to \mathbb{P}^3$	$\dfrac{(\eta^3+3\eta^2+6\eta+6)e^{3\eta}}{\eta^4}$

Table 1

A.5 The domain of definition of the weighted volume functional

By Theorem A.3 and Proposition A.4, we see that the weighted volume functional F is defined on the open cone Λ of elements of the Lie algebra of the torus admitting Hamiltonian potentials that are proper and bounded below if Λ is nonempty. In this subsection, we characterise Λ algebraically in the setting of asymptotically conical Kähler manifolds.

Our precise set-up is as follows. Let (C_0, g_0) be a Kähler cone with apex o, complex structure J_0, and radial function r such that $g_0 = dr^2 + r^2 g_S$ for a Riemannian metric g_S on the link $S = \{r = 1\}$ of C_0. Let $\pi\colon M \to C_0$ be a quasiprojective resolution of C_0 that is equivariant with respect to the holomorphic isometric action on C_0 of the torus T with Lie algebra $\mathfrak{t}$ generated by $\xi := J_0 r\partial_r$, and let g be a Kähler metric on M with

$$|\pi_* g - g_0|_{g_0} = O(r^{-2}), \tag{A-6}$$

with respect to which T acts isometrically in a Hamiltonian fashion with moment map $\mu\colon M \to \mathfrak{t}^*$. Write $u_Y(x) := \langle \mu(x), Y\rangle$ for $x \in M$ for the Hamiltonian potential of $Y \in \mathfrak{t}$, so that $du_Y = -\omega \lrcorner Y$, where ω is the Kähler form of g, and set

$$\Lambda := \{Y \in \mathfrak{t} \mid u_Y \text{ is proper and bounded below}\}.$$

Next, let $\mathcal{O}_M(M)$ (resp. $\mathcal{O}_{C_0}(C_0)$) denote the global algebraic sections of the structure sheaf of M (resp. of C_0), and write

$$\mathcal{O}_M(M) = \bigoplus_{\alpha \in \mathfrak{t}^*} \mathcal{H}_\alpha$$

for the weight decomposition under the action of T. Then we have:

Theorem A.10 *In the above situation*,

$$\Lambda = \{Y \in \mathfrak{t} \mid \alpha(Y) > 0 \text{ for all } \alpha \in \mathfrak{t}^* \text{ such that } \mathcal{H}_\alpha \neq \varnothing \text{ and } \alpha \neq 0\}.$$

Proof Let E denote the exceptional set of the resolution $\pi\colon M \to C_0$. In what follows, we will identify $M \setminus E$ with $C_0 \setminus \{0\}$ via π. Let us begin by making some useful observations. Let X be the unique vector field on M such that $d\pi(X) = r\partial_r$. Then $d\pi(JX) = \xi$, where J denotes the complex structure on M, so that $JX \in \mathfrak{t}$ and $[X, Y] = 0$ for every $Y \in \mathfrak{t}$. Then we have:

Lemma A.11 *Let $Y \in \mathfrak{t}$, so that Y defines a real holomorphic g–Killing vector field on M with $[X, Y] = 0$. Then Y is tangent to the level sets of r on $C_0 \setminus \{o\}$.*

Proof Since T acts isometrically on g and g_0, Y will define a holomorphic g_0–Killing vector field on C_0. We claim that such a vector field is tangent to the level sets of r. Indeed, simply note that

$$0 = \mathcal{L}_Y g_0 = d(Y \cdot r) \otimes dr + dr \otimes d(Y \cdot r) + 2r\,dr(Y) g_S + r^2 \mathcal{L}_Y g_S.$$

Then, plugging ξ into both arguments on the right-hand side and observing that

$$[Y, \xi] = [Y, Jr\partial_r] = [Y, JX] = J[Y, X] = 0$$

along the end of C_0 since Y is holomorphic, we arrive at the fact that

$$-dr(Y) = \tfrac{1}{2} r (\mathcal{L}_Y g_S)(\xi, \xi) = \tfrac{1}{2} r \big(\mathcal{L}_Y (g_S(\xi, \xi)) - 2 g_S([Y, \xi], \xi)\big) = 0,$$

as required. □

We now demonstrate that

$$\Lambda \subseteq \{Y \in \mathfrak{t} \mid \alpha(Y) > 0 \text{ for all } \alpha \in \mathfrak{t}^* \text{ such that } \mathcal{H}_\alpha \neq \varnothing \text{ and } \alpha \neq 0\}.$$

To this end, let $Y \in \Lambda$, so that the Hamiltonian potential u_Y of Y is proper and bounded below, let f be a nonconstant holomorphic function on M on which Y acts with weight λ so that $JY(f) = -\lambda f$, let $x \in M \setminus E$ be a point where $f(x) \neq 0$, and denote by $\gamma_x(t)$ the flow line of $-JY$ with $\gamma_x(0) = x$. Then

$$\frac{d}{dt} f(\gamma_x(t)) = \lambda f(\gamma_x(t)),$$

so that

(A-7) $$f(\gamma_x(t)) = f(x) e^{-\lambda t} \quad \text{for all } t < 0.$$

Now, by definition, we have that $-JY = \nabla^g u_Y$ and so from Proposition 2.27 we deduce that there is a sequence $t_i \to -\infty$ as $i \to +\infty$ such that $(\gamma_x(t_i))_i$ converges to a point $x_\infty \in M$ satisfying $\nabla^g u(x_\infty) = 0$. Since the fixed-point set of T is contained in E, we must have that $x_\infty \in E$. Let $x_i := \gamma_x(t_i)$. Then plugging t_i into (A-7) yields the fact that

$$|f(x_i)| = |f(x)| e^{-\lambda t_i} \xrightarrow[i \to \infty]{} \begin{cases} +\infty & \text{if } \lambda < 0, \\ 0 & \text{if } \lambda > 0. \end{cases}$$

Since $x_i \to x_\infty \in E$ as $i \to \infty$, we conclude from the maximum principle that $\lambda > 0$, as required.

Next we show that

$$\{Y \in \mathfrak{t} \mid \alpha(Y) > 0 \text{ for all } \alpha \in \mathfrak{t}^* \text{ such that } \mathcal{H}_\alpha \neq \varnothing \text{ and } \alpha \neq 0\} \subseteq \Lambda. \tag{A-8}$$

By [20, Lemma 2.15], M is 1–convex. By construction, then, $\pi : M \to C_0$ will be the Remmert reduction of M. In particular, we have that $\pi^*\mathcal{O}_{C_0}(C_0) = \mathcal{O}_M(M)$ by the properties of the Remmert reduction. Since $\pi : M \to C_0$ is equivariant with respect to the action of T, we thus see that $Y \in \mathfrak{t}$ acts with weight λ on $f \in \mathcal{O}_{C_0}(C_0)$ if and only if it acts with weight λ on the unique lift $\pi^* f$ of f to $\mathcal{O}_M(M)$. Applying [18, Proposition 2.7], we therefore deduce that

$$\{Y \in \mathfrak{t} \mid g_S(Y,\xi) > 0\} = \{Y \in \mathfrak{t} \mid \alpha(Y) > 0 \text{ for all } \alpha \in \mathfrak{t}^* \text{ such that } \mathcal{H}_\alpha \neq \varnothing \text{ and } \alpha \neq 0\}.$$

Consequently, in order to prove the inclusion (A-8), it suffices to show that

$$\{Y \in \mathfrak{t} \mid g_S(Y,\xi) > 0\} \subseteq \Lambda.$$

This inclusion is established by the following proposition.

Proposition A.12 *Let $Y \in \{Z \in \mathfrak{t} \mid g_S(Z,\xi)(x) > 0$ for all $x \in S\}$ with Hamiltonian potential u_Y. Then $u_Y \geq cr^2$ along the end of C_0 for some $c > 0$. In particular, u_Y is proper and bounded below.*

Proof Let $x \in \{r = 1\}$ and let $\gamma_x(t)$ denote the integral curve of X, the vector field on M satisfying $d\pi(X) = r\partial_r$, with $\gamma_x(0) = x$. Then we have that

$$\begin{aligned} u_Y(\gamma_x(t)) &= u_Y(\gamma_x(0)) + \int_0^t du_Y(\dot{\gamma}_x(s))\, ds \\ &= u_Y(x) + \int_0^t g(-JY, X)(\gamma_x(s))\, ds \\ &= u_Y(x) + \int_0^t g(Y, JX)(\gamma_x(s))\, ds. \end{aligned} \tag{A-9}$$

Next observe that since $Y \in \mathfrak{t}$, Y is tangent to the level sets of r by Lemma A.11. Hence, the asymptotics (A-6) give us that

$$\begin{aligned} g(Y, JX)(\gamma_x(s)) &= g_0(Y, JX) + \overbrace{O(r^{-2})|Y|_{g_0}|JX|_{g_0}}^{=O(1)} \\ &= O(1) + r(\gamma_x(s))^2 g_S(Y,\xi) \\ &= O(1) + r(x)^2 e^{2s} g_S(Y,\xi), \end{aligned}$$

where the final equality follows from the fact that $r(\gamma_x(s)) = r(x)e^s$ because

$$\frac{\partial}{\partial s}(r(\gamma_x(s))) = r(\gamma_x(s)) \quad \text{and} \quad \gamma_x(0) = x.$$

Plugging this into (A-9) yields

$$\begin{aligned} u_Y(\gamma_x(t)) &= u_Y(x) + \tfrac{1}{2} r(x)^2 g_S(Y,\xi)(e^{2t} - 1) + O(t) \\ &= u_Y(x) - \tfrac{1}{2} g_S(Y,\xi) r(x)^2 + \tfrac{1}{2} g_S(Y,\xi) r(\gamma_x(t))^2 + O(\ln r(\gamma_x(t))) \\ &\geq c r(\gamma_x(t))^2 \end{aligned}$$

along the end of C_0 for some $c > 0$, since $g_S(Y,\xi) > 0$. From this, the assertion follows. □

A.6 Coercive estimates on Hamiltonian potentials

The goal of this subsection is to prove sharp positive bounds on the growth of the Hamiltonian potential of a real holomorphic Killing vector field on a complete shrinking gradient Kähler–Ricci soliton (M, g, X) that commutes with the soliton vector field X under certain conditions. Since $H^1(M) = 0$ by [67], such a vector field always admits a Hamiltonian potential. Let ω denote the Kähler form of g and recall that for each real holomorphic Killing vector field Y on M commuting with X, the Hamiltonian u_Y of Y is normalised so that $\Delta_\omega u_Y + u_Y + \frac{1}{2} JY \cdot f = 0$. Since $du_Y = -\omega \lrcorner Y$ by definition, one sees that

$$\Delta_\omega u_Y + u_Y = -\tfrac{1}{2} g(JY, X) = \tfrac{1}{2} g(\nabla u_Y, X) = \tfrac{1}{2} X \cdot u_Y, \tag{A-10}$$

an identity that shall prove useful in what follows. We will prove:

Proposition A.13 *Let (M, g, X) be a complete shrinking gradient Kähler–Ricci soliton with bounded Ricci curvature with soliton vector field $X = \nabla^g f$ for a smooth real-valued function $f : M \to \mathbb{R}$. Let Y be a real holomorphic Killing vector field on M commuting with X and assume that the Hamiltonian potential u_Y of Y is proper and bounded below. Then there exist positive constants c_1 and c_2 such that $c_1 f \leq u_Y \leq c_2 f$ outside of a compact set.*

As a consequence of this proposition, we see, without appealing to the Duistermaat–Heckman theorem, that the weighted volume functional is defined on a complete shrinking gradient Kähler–Ricci soliton (M, g, X) with bounded Ricci curvature on those elements admitting a Hamiltonian potential that is proper and bounded below in the Lie algebra of any torus that acts in a holomorphic Hamiltonian fashion on M and contains the flow of JX, where J denotes the complex structure of M.

Proof Let $|\cdot|_g$ denote the norm with respect to g. The following claim provides a sharp growth rate on $|Y|_g$ which improves Proposition 2.30 for real holomorphic vector fields commuting with X.

Claim A.14 *There exists a positive constant c such that $|Y|_g^2 \leq cf$ outside a compact set.*

Proof We start by taking note of the following crucial equation that holds pointwise on M and is independent of the soliton structure. For a real holomorphic vector field Y, it holds that

$$\Delta_g Y + \mathrm{Ric}(Y) = 0.$$

This observation is due to Lichnerowicz and its proof can be found for instance in [2, Proposition 4.79]. Since we allow Y to commute with X, Y satisfies the elliptic equation

$$\Delta_g Y - \nabla_X Y + Y = \Delta_g Y - \nabla_X Y + \mathrm{Ric}(Y) + \nabla_Y X = 0,$$

where we have used the soliton equation (1-4) with $\lambda = 1$. Recall also that $\Delta_g f - X \cdot f = -2f$.

Using these facts, we compute

$$(\text{A-11})\quad (\Delta_g - X\cdot)\left(\frac{|Y|_g^2}{f}\right) = 2f^{-1}|\nabla Y|_g^2 - 2\frac{|Y|_g^2}{f} + |Y|_g^2(\Delta_g - X\cdot)f^{-1} + 2g(\nabla^g|Y|_g^2, \nabla^g f^{-1})$$
$$\geq -2\frac{|Y|_g^2}{f} + \frac{|Y|_g^2}{f}\left(-\frac{(\Delta_g - \nabla^g f\cdot)f}{f} + 2\frac{|\nabla^g f|^2}{f^2}\right)$$
$$- 2g\left(\nabla^g\left(\frac{|Y|_g^2}{f}\right), \nabla^g \ln f\right) - 2|\nabla^g \ln f|^2\frac{|Y|_g^2}{f}$$
$$\geq -2g\left(\nabla^g\left(\frac{|Y|_g^2}{f}\right), \nabla^g \ln f\right),$$

a computation valid on the set where f is strictly positive. For $\varepsilon > 0$ a given constant and $a > 0$ a constant such that $|Y|_g^2 = O(f^a)$ (see Proposition 2.30), we consider the function $|Y|_g^2 \cdot f^{-1} - \varepsilon f^a$ defined on the complement of a compact set of the form $\{f \leq C(n)\}$ such that f is strictly positive on $\{f \geq C(n)\}$. By Proposition 2.30, we know that

$$\lim_{f\to+\infty}(|Y|_g^2\cdot f^{-1} - \varepsilon f^a) = -\infty,$$

so that this function attains a maximum on the set $\{f \geq C(n)\}$. Now, if $f \geq 2(a+1)$, then using (A-11), we compute that

$$(\Delta_g - X\cdot)\left(\frac{|Y|_g^2}{f} - \varepsilon f^a\right) \geq -2g\left(\nabla^g\left(\frac{|Y|_g^2}{f} - \varepsilon f^a\right), \nabla^g \ln f\right) + \varepsilon a f^a\left(2 - (a+1)\frac{|\nabla^g f|^2}{f^2}\right)$$
$$\geq -2g\left(\nabla^g\left(\frac{|Y|_g^2}{f} - \varepsilon f^a\right), \nabla^g \ln f\right) + \varepsilon a f^a\left(2 - (a+1)\frac{2}{f}\right)$$
$$> -2g\left(\nabla^g\left(\frac{|Y|_g^2}{f} - \varepsilon f^a\right), \nabla^g \ln f\right),$$

where we used the first-order soliton identity from Lemma 2.22 in the penultimate inequality. As already noted, $|Y|_g^2 \cdot f^{-1} - \varepsilon f^a$ attains a maximum on $\{f \geq C(n)\}$ and so the maximum principle applied to this function implies that it must attain its maximum at a point on the boundary of the set $\{f \geq \max\{2(a+1), C(n)\}\}$. In particular, we find that

$$(\text{A-12})\quad |Y|_g^2\cdot f^{-1} - \varepsilon f^a \leq \max_{f=C(a,n)}(|Y|_g^2\cdot f^{-1} - \varepsilon f^a) \leq \max_{f=C(a,n)}(|Y|_g^2\cdot f^{-1}) = C(a,n)\max_{f=C(a,n)}|Y|_g^2,$$

where $C(a,n)$ is a positive constant that may vary from line to line. As the right-hand side of (A-12) is independent of ε, by letting ε tend to 0 on the left-hand side, we reach the desired conclusion. ◁

Since $|\nabla u_Y|_g = |JY|_g = |Y|_g$, one obtains the expected growth on u_Y by integrating the estimate on $|Y|_g$ established in the previous claim.

We next prove the lower bound on u_Y. First notice that since u_Y is proper and bounded below, u_Y is strictly positive outside a sufficiently large compact set of the form $\{f \leq c_0\}$. Recall from Remark 2.23 that the normalisation of f is determined by the soliton identities, which in this case yield $\Delta_g f - X\cdot f = -2f$. Using f as a barrier function together with (A-10), we compute the weighted Laplacian of the difference of the inverses of u_Y and f on the region where this difference makes sense. We have

$$\left(\Delta_\omega - \tfrac{1}{2}X\cdot\right)\left(\frac{1}{u_Y}-\frac{C}{f}\right) = \frac{1}{u_Y}-\frac{C}{f}+2\frac{|\nabla u_Y|^2}{u_Y^3}-2C\frac{|\nabla f|^2}{f^3}, \tag{A-13}$$

where C is a positive constant to be specified later. Assume that the function $u_Y^{-1}-Cf^{-1}$ attains its maximum at an interior point x_0 of a domain of the form $\{r_1 \le f \le r_2\}$ with r_1 sufficiently large so that both u_Y and f are strictly positive. At such a point x_0, one sees that $\nabla(u_Y^{-1}-Cf^{-1})(x_0)=0$, and from the maximum principle that

$$0 \ge \left(\Delta_\omega - \tfrac{1}{2}X\cdot\right)\left(\frac{1}{u_Y}-\frac{C}{f}\right)(x_0).$$

This information, together with (A-13), implies that at x_0,

$$\begin{aligned}
0 &\ge \frac{1}{u_Y}-\frac{C}{f}+2\frac{|\nabla u_Y|^2}{u_Y^3}-2C\frac{|\nabla f|^2}{f^3} = \frac{1}{u_Y}-\frac{C}{f}+2u_Y|\nabla u_Y^{-1}|^2-2Cf|\nabla f^{-1}|^2\\
&= \frac{1}{u_Y}-\frac{C}{f}+2u_YC^2|\nabla f^{-1}|^2-2Cf|\nabla f^{-1}|^2 = \left(\frac{1}{u_Y}-\frac{C}{f}\right)(1-2Cu_Yf|\nabla f^{-1}|^2)\\
&= \left(\frac{1}{u_Y}-\frac{C}{f}\right)\left(1-2C\frac{u_Y}{f^3}|\nabla f|^2\right).
\end{aligned}$$

Next, using the fact that $|\nabla f|^2$ grows quadratically by the soliton identities, we see from the upper bound on u_Y that on M,

$$2C\frac{u_Y}{f^3}|\nabla f|^2 \le \frac{2Cd}{f}$$

for some positive constant d uniform in r_1 and r_2. In particular, the term $(1-2Cu_Yf^{-3}|\nabla f|^2)$ is positive on $\{r_1\le f\le r_2\}$ as long as $2Cf^{-1}d$ is strictly less than 1, or equivalently, as long as $C<(2d)^{-1}r_1$.

In summary, for any heights $r_1<r_2$ and any constant C such that $C<(2d)^{-1}r_1$, we have that

$$\max_{r_1\le f\le r_2}\left(\frac{1}{u_Y}-\frac{C}{f}\right) \le \max\left\{0,\max_{f=r_1}\left(\frac{1}{u_Y}-\frac{C}{f}\right),\max_{f=r_2}\left(\frac{1}{u_Y}-\frac{C}{f}\right)\right\}.$$

Since u_Y (and f) tend to $+\infty$ as f approaches $+\infty$, one sees, by letting r_2 tend to $+\infty$, that

$$\max_{r_1\le f}\left(\frac{1}{u_Y}-\frac{C}{f}\right) \le \max\left\{0,\max_{f=r_1}\left(\frac{1}{u_Y}-\frac{C}{f}\right)\right\}. \tag{A-14}$$

We choose C and r_1 so that the right-hand side of (A-14) is nonpositive and so that $C<(2d)^{-1}r_1$. This is possible because by properness of u_Y, there exists some positive height r_1 such that $\min_{f=r_1}u_Y\ge 4d$. Then for $C:=(4d)^{-1}r_1$, we have that $C<(2d)^{-1}r_1$ and

$$\max_{f=r_1}\left(\frac{1}{u_Y}-\frac{C}{f}\right)\le 0,$$

as desired. This completes the proof of the proposition. □

References

[1] **D M Austin**, **P J Braam**, *Morse–Bott theory and equivariant cohomology*, from "The Floer memorial volume" (H Hofer, C H Taubes, A Weinstein, E Zehnder, editors), Progr. Math. 133, Birkhäuser, Basel (1995) 123–183 MR Zbl

[2] **W Ballmann**, *Lectures on Kähler manifolds*, Eur. Math. Soc., Zürich (2006) MR Zbl

[3] **W Barth**, **R Narasimhan** (editors), *Several complex variables*, *VI*: *Complex manifolds*, Encyclopaedia of Mathematical Sciences 69, Springer (1990) MR Zbl

[4] **F A Belgun**, *On the metric structure of non-Kähler complex surfaces*, Math. Ann. 317 (2000) 1–40 MR Zbl

[5] **N Berline**, **E Getzler**, **M Vergne**, *Heat kernels and Dirac operators*, corrected 1st edition, Grundl. Math. Wissen. 298, Springer (2004) MR Zbl

[6] **R J Berman**, **D W Nystrom**, *Complex optimal transport and the pluripotential theory of Kähler–Ricci solitons*, preprint (2014) arXiv 1401.8264

[7] **C Birkar**, **P Cascini**, **C D Hacon**, **J McKernan**, *Existence of minimal models for varieties of log general type*, J. Amer. Math. Soc. 23 (2010) 405–468 MR Zbl

[8] **C P Boyer**, **K Galicki**, *Sasakian geometry*, Oxford Univ. Press (2008) MR Zbl

[9] **E Brieskorn**, *Rationale Singularitäten komplexer Flächen*, Invent. Math. 4 (1967/68) 336–358 MR Zbl

[10] **R L Bryant**, *Gradient Kähler Ricci solitons*, from "Géométrie différentielle, physique mathématique, mathématiques et société, I" (O Hijazi, editor), Astérisque 321, Soc. Math. France, Paris (2008) 51–97 MR Zbl

[11] **H-D Cao**, **D Zhou**, *On complete gradient shrinking Ricci solitons*, J. Differential Geom. 85 (2010) 175–185 MR Zbl

[12] **C-W Chen**, **A Deruelle**, *Structure at infinity of expanding gradient Ricci soliton*, Asian J. Math. 19 (2015) 933–950 MR Zbl

[13] **X Chen**, **S Sun**, **G Tian**, *A note on Kähler–Ricci soliton*, Int. Math. Res. Not. 2009 (2009) 3328–3336 MR Zbl

[14] **B Chow**, **D Knopf**, *The Ricci flow: an introduction*, Mathematical Surveys and Monographs 110, Amer. Math. Soc., Providence, RI (2004) MR Zbl

[15] **B Chow**, **P Lu**, *Uniqueness of asymptotic cones of complete noncompact shrinking gradient Ricci solitons with Ricci curvature decay*, C. R. Math. Acad. Sci. Paris 353 (2015) 1007–1009 MR Zbl

[16] **B Chow**, **P Lu**, **B Yang**, *Lower bounds for the scalar curvatures of noncompact gradient Ricci solitons*, C. R. Math. Acad. Sci. Paris 349 (2011) 1265–1267 MR Zbl

[17] **C van Coevering**, *Examples of asymptotically conical Ricci-flat Kähler manifolds*, Math. Z. 267 (2011) 465–496 MR Zbl

[18] **T C Collins**, **G Székelyhidi**, *K–semistability for irregular Sasakian manifolds*, J. Differential Geom. 109 (2018) 81–109 MR Zbl

[19] **R J Conlon**, **A Deruelle**, *Expanding Kähler–Ricci solitons coming out of Kähler cones*, J. Differential Geom. 115 (2020) 303–365 MR Zbl

[20] **R J Conlon**, **H-J Hein**, *Asymptotically conical Calabi–Yau manifolds, I*, Duke Math. J. 162 (2013) 2855–2902 MR Zbl

[21] **R J Conlon**, **H-J Hein**, *Asymptotically conical Calabi–Yau metrics on quasi-projective varieties*, Geom. Funct. Anal. 25 (2015) 517–552 MR Zbl

[22] **X Dai**, **G Wei**, *Comparison geometry for Ricci curvature*, online book (2008) Available at `http://web.math.ucsb.edu/~dai/Ricci-book.pdf`

[23] **A S Dancer**, **M Y Wang**, *On Ricci solitons of cohomogeneity one*, Ann. Global Anal. Geom. 39 (2011) 259–292 MR Zbl

[24] **J-P Demailly**, *Complex analytic and differential geometry*, online book (2012) Available at `http://www-fourier.ujf-grenoble.fr/~demailly/manuscripts/agbook.pdf`

[25] **J-P Demailly**, **M Paun**, *Numerical characterization of the Kähler cone of a compact Kähler manifold*, Ann. of Math. 159 (2004) 1247–1274 MR Zbl

[26] **A Deruelle**, *Smoothing out positively curved metric cones by Ricci expanders*, Geom. Funct. Anal. 26 (2016) 188–249 MR Zbl

[27] **A Deruelle**, *Asymptotic estimates and compactness of expanding gradient Ricci solitons*, Ann. Sc. Norm. Super. Pisa Cl. Sci. 17 (2017) 485–530 MR Zbl

[28] **J J Duistermaat**, **G J Heckman**, *On the variation in the cohomology of the symplectic form of the reduced phase space*, Invent. Math. 69 (1982) 259–268 MR Zbl

[29] **F Fang**, **X-D Li**, **Z Zhang**, *Two generalizations of Cheeger–Gromoll splitting theorem via Bakry–Emery Ricci curvature*, Ann. Inst. Fourier (Grenoble) 59 (2009) 563–573 MR Zbl

[30] **M Feldman**, **T Ilmanen**, **D Knopf**, *Rotationally symmetric shrinking and expanding gradient Kähler–Ricci solitons*, J. Differential Geom. 65 (2003) 169–209 MR Zbl

[31] **T Frankel**, *Fixed points and torsion on Kähler manifolds*, Ann. of Math. 70 (1959) 1–8 MR Zbl

[32] **A Futaki**, *Kähler–Einstein metrics and integral invariants*, Lecture Notes in Math. 1314, Springer (1988) MR Zbl

[33] **A Futaki**, *The weighted Laplacians on real and complex metric measure spaces*, from "Geometry and analysis on manifolds" (T Ochiai, T Mabuchi, Y Maeda, J Noguchi, A Weinstein, editors), Progr. Math. 308, Springer (2015) 343–351 MR Zbl

[34] **A Futaki**, **M-T Wang**, *Constructing Kähler–Ricci solitons from Sasaki–Einstein manifolds*, Asian J. Math. 15 (2011) 33–52 MR Zbl

[35] **A M Gleason**, **R S Palais**, *On a class of transformation groups*, Amer. J. Math. 79 (1957) 631–648 MR Zbl

[36] **H Grauert**, *Über Modifikationen und exzeptionelle analytische Mengen*, Math. Ann. 146 (1962) 331–368 MR Zbl

[37] **W He**, **S Sun**, *Frankel conjecture and Sasaki geometry*, Adv. Math. 291 (2016) 912–960 MR Zbl

[38] **S Ishii**, *Introduction to singularities*, Springer (2014) MR Zbl

[39] **K Iwasawa**, *On some types of topological groups*, Ann. of Math. 50 (1949) 507–558 MR Zbl

[40] **S Kobayashi**, *Transformation groups in differential geometry*, Ergebnisse der Math. 70, Springer (1972) MR Zbl

[41] **J Kollár**, *Lectures on resolution of singularities*, Annals of Mathematics Studies 166, Princeton Univ. Press (2007) MR Zbl

[42] **B Kotschwar**, *Kählerity of shrinking gradient Ricci solitons asymptotic to Kähler cones*, J. Geom. Anal. 28 (2018) 2609–2623 MR Zbl

[43] **B Kotschwar**, **L Wang**, *Rigidity of asymptotically conical shrinking gradient Ricci solitons*, J. Differential Geom. 100 (2015) 55–108 MR Zbl

[44] **B Kotschwar**, **L Wang**, *Isometries of asymptotically conical shrinking Ricci solitons*, preprint (2019) arXiv 1901.00044

[45] **H B Laufer**, *Normal two-dimensional singularities*, Annals of Mathematics Studies 71, Princeton Univ. Press (1971) MR Zbl

[46] **R Lazarsfeld**, *Positivity in algebraic geometry, I: Classical setting: line bundles and linear series*, Ergebnisse der Math. (3) 48, Springer (2004) MR Zbl

[47] **C Li**, *On rotationally symmetric Kähler–Ricci solitons*, preprint (2010) arXiv 1004.4049

[48] **M T Lock**, **J A Viaclovsky**, *A smörgåsbord of scalar-flat Kähler ALE surfaces*, J. Reine Angew. Math. 746 (2019) 171–208 MR Zbl

[49] **D Martelli**, **J Sparks**, **S-T Yau**, *Sasaki–Einstein manifolds and volume minimisation*, Comm. Math. Phys. 280 (2008) 611–673 MR Zbl

[50] **Y Matsushima**, *Sur la structure du groupe d'homéomorphismes analytiques d'une certaine variété kählérienne*, Nagoya Math. J. 11 (1957) 145–150 MR Zbl

[51] **O Munteanu**, **J Wang**, *Holomorphic functions on Kähler–Ricci solitons*, J. Lond. Math. Soc. 89 (2014) 817–831 MR Zbl

[52] **O Munteanu**, **J Wang**, *Topology of Kähler Ricci solitons*, J. Differential Geom. 100 (2015) 109–128 MR Zbl

[53] **O Munteanu**, **J Wang**, *Conical structure for shrinking Ricci solitons*, J. Eur. Math. Soc. 19 (2017) 3377–3390 MR Zbl

[54] **O Munteanu**, **J Wang**, *Structure at infinity for shrinking Ricci solitons*, Ann. Sci. Éc. Norm. Supér. 52 (2019) 891–925 MR Zbl

[55] **T Ohsawa**, *Vanishing theorems on complete Kähler manifolds*, Publ. Res. Inst. Math. Sci. 20 (1984) 21–38 MR Zbl

[56] **P Orlik**, **P Wagreich**, *Isolated singularities of algebraic surfaces with* C^* *action*, Ann. of Math. 93 (1971) 205–228 MR Zbl

[57] **R S Palais**, *A global formulation of the Lie theory of transformation groups*, Mem. Amer. Math. Soc. 22, Amer. Math. Soc., Providence, RI (1957) MR Zbl

[58] **S Pigola**, **M Rimoldi**, **A G Setti**, *Remarks on noncompact gradient Ricci solitons*, Math. Z. 268 (2011) 777–790 MR Zbl

[59] **H Pinkham**, *Normal surface singularities with* C^* *action*, Math. Ann. 227 (1977) 183–193 MR Zbl

[60] **E Prato**, **S Wu**, *Duistermaat–Heckman measures in a noncompact setting*, Compositio Math. 94 (1994) 113–128 MR Zbl

[61] **H Rossi**, *Vector fields on analytic spaces*, Ann. of Math. 78 (1963) 455–467 MR Zbl

[62] **M Siepmann**, *Ricci flows of Ricci flat cones*, PhD thesis, ETH Zürich (2013) Available at `http://e-collection.library.ethz.ch/eserv/eth:7556/eth-7556-02.pdf`

[63] **G Tian**, **X Zhu**, *Uniqueness of Kähler–Ricci solitons*, Acta Math. 184 (2000) 271–305 MR Zbl

[64] **G Tian**, **X Zhu**, *A new holomorphic invariant and uniqueness of Kähler–Ricci solitons*, Comment. Math. Helv. 77 (2002) 297–325 MR Zbl

[65] **S Wu**, *An integration formula for the square of moment maps of circle actions*, Lett. Math. Phys. 29 (1993) 311–328 MR Zbl

[66] **S Wu**, *On the instanton complex of holomorphic Morse theory*, Comm. Anal. Geom. 11 (2003) 775–807 MR Zbl

[67] **W Wylie**, *Complete shrinking Ricci solitons have finite fundamental group*, Proc. Amer. Math. Soc. 136 (2008) 1803–1806 MR Zbl

[68] **B Yang**, *A characterization of noncompact Koiso-type solitons*, Internat. J. Math. 23 (2012) 1250054, 13 MR Zbl

[69] **Z-H Zhang**, *On the completeness of gradient Ricci solitons*, Proc. Amer. Math. Soc. 137 (2009) 2755–2759 MR Zbl

RJC: *Department of Mathematical Sciences, The University of Texas at Dallas*
Richardson, TX, United States

AD: *Université Paris-Saclay, Laboratoire de mathématiques d'Orsay*
Orsay, France

SS: *Institute for Advanced Study in Mathematics, Zhejiang University*
Hangzhou, China

Department of Mathematics, University of California, Berkeley
Berkeley, CA, United States

ronan.conlon@utdallas.edu, alix.deruelle@universite-paris-saclay.fr,
songsun1987@gmail.com

Proposed: Bruce Kleiner Received: 25 March 2021
Seconded: Tobias H Colding, Simon Donaldson Revised: 27 February 2022

msp

Geometry & Topology **28**:1 (2024) 353–392
DOI: 10.2140/gt.2024.28.353
Published: 27 February 2024

Embedding calculus and smooth structures

BEN KNUDSEN
ALEXANDER KUPERS

We study the dependence of the embedding calculus Taylor tower on the smooth structures of the source and target. We prove that embedding calculus does not distinguish exotic smooth structures in dimension 4, implying a negative answer to a question of Viro. In contrast, we show that embedding calculus does distinguish certain exotic spheres in higher dimensions. As a technical tool of independent interest, we prove an isotopy extension theorem for the limit of the embedding calculus tower, which we use to investigate several further examples.

58D10; 55P48, 57N35, 57R40

1 Introduction

We investigate the scope of a certain tool used to study the space $\mathrm{Emb}^s(N, M)$ of smooth embeddings from an n–manifold N into an m–manifold M. This investigation has consequences for spaces of embeddings themselves, as shown by the following result on knots and links, which answers a question of Viro [42, Section 6] in the negative and improves on a result of Arone and Szymik [3].

Theorem A *Let M and M' be smooth simply connected compact 4–manifolds. If M and M' are homeomorphic, then, for any $k \geq 0$,*

$$\mathrm{Emb}^s\Big(\bigsqcup_k S^1, M\Big) \simeq \mathrm{Emb}^s\Big(\bigsqcup_k S^1, M'\Big).$$

The tool in question is the embedding calculus of Goodwillie and Weiss [16; 46], which, at the coarsest level, provides a functorial comparison map

$$\mathrm{Emb}^s(N, M) \to T_\infty \mathrm{Emb}^s(N, M) = \operatorname*{holim}_k T_k \mathrm{Emb}^s(N, M),$$

whose target is assembled from the configuration spaces of N and M and maps among them (details are reviewed in Section 2). According to one of the main results of the subject (see Goodwillie and Klein [14] and [16]), this map is a weak equivalence in codimension at least 3; one says that *the Taylor tower converges* to the embedding space. In particular, the theorem applies to links in 4–manifolds as in Theorem A.

Little is known about convergence in low codimension. We begin to address this gap by proving that codimension-0 convergence largely fails in dimension 4.

Theorem B *Let M and N be smooth simply connected compact 4–manifolds. If M and N are homeomorphic, then $T_\infty \mathrm{Emb}^s(N, M) \neq \varnothing$. In particular, if M and N are not also diffeomorphic, then the map*

$$\mathrm{Emb}^s(N, M) \to T_\infty \mathrm{Emb}^s(N, M)$$

is not a weak equivalence.

In fact, we prove that there are homotopy invertible elements in $T_\infty \mathrm{Emb}^s(N, M)$, which one should think of as saying that N and M are diffeomorphic (or at least isotopy equivalent) in the eyes of embedding calculus.

Theorems A and B arise from a common source. Specifically, the data involved in the constructions of embedding calculus is a pair of presheaves, one for N and one for M. We show in Theorem 3.18 that these presheaves are largely insensitive to smooth structure in dimension 4, and the results follow; see Section 4.

The results above might lead one to suspect that embedding calculus is insensitive to smooth structure. The following contrasting result shows that the situation is not so simple (see Section 5.2 for further examples).

Theorem C *For any $n = 2^j$ with $j \geq 3$, there is an exotic n–sphere Σ such that $T_\infty \mathrm{Emb}^s(\Sigma, S^n) = \varnothing$. In particular, the map*

$$\mathrm{Emb}^s(\Sigma, S^n) \to T_\infty \mathrm{Emb}^s(\Sigma, S^n)$$

is a weak equivalence (both sides are empty).

Thus, embedding calculus distinguishes certain exotic spheres. Alternatively, one can interpret this as a convergence result in codimension 0. The crucial property distinguishing the exotic spheres of Theorem C from S^n is that they do not embed in $\mathbb{R}^{n+3}$.

To facilitate the further study of embedding calculus in the potential absence of convergence, we prove an isotopy extension theorem for $T_\infty \mathrm{Emb}^s(-,-)$; see Theorem 6.1. We close by demonstrating its utility with several applications.

Acknowledgments

Knudsen thanks John Francis for bringing his attention to the question of whether embedding calculus distinguishes exotic spheres. Kupers would like to thank Oscar Randal-Williams for helpful conversations, and in particular the suggestion that some exotic spheres do not admit embeddings into Euclidean space with low codimension. We would also like to thank the referees, as well as Manuel Krannich and Oscar Randal-Williams for corrections to an earlier version. Knudsen was supported by NSF grant DMS-1906174, the Natural Sciences and Engineering Research Council of Canada (NSERC) (funding reference numbers 512156 and 512250), as well as the Research Competitiveness Fund of the University of Toronto at Scarborough. Kupers was supported by NSF grant DMS-1803766.

2 Preliminaries

In this section, we gather what facts we need from the theory of embedding calculus, as well as some standard foundational material on topological manifolds. In our discussion of calculus, we adopt the perspective of [5], but see [6; 15; 41; 46] for other foundations.

2.1 Embedding calculus

Write $\mathcal{M}\mathrm{fld}^s$ for the simplicial category whose objects are smooth manifolds without boundary, of finite type and arbitrary dimension. The morphism space $\mathrm{Map}_{\mathcal{M}\mathrm{fld}^s}(N, M)$ has as n–simplices commuting diagrams

$$\begin{array}{ccc} \Delta^n \times N & \longrightarrow & \Delta^n \times M \\ & \searrow \quad \swarrow & \\ & \Delta^n & \end{array}$$

in which the top map is a neat smooth embedding of manifolds with corners. This category is symmetric monoidal under disjoint union.

Manifold calculus approximates simplicial presheaves on this category by extrapolating from their values on disjoint unions of disks of a fixed dimension. More formally, let $\mathcal{D}\mathrm{isk}^s_n \subset \mathcal{M}\mathrm{fld}^s$ be the full subcategory on those objects that are diffeomorphic to a disjoint union of finitely many copies of $\mathbb{R}^n$ with its standard smooth structure. Manifold calculus is the approximation of simplicial presheaves on $\mathcal{M}\mathrm{fld}^s$ by simplicial presheaves on $\mathcal{D}\mathrm{isk}^s_n$. *Embedding calculus* is the application of manifold calculus to the presheaf of embeddings into a smooth manifold M. Fixing n, we write $\mathbb{E}^s_M$ for the presheaf on $\mathcal{D}\mathrm{isk}^s_n$ obtained by restriction of the representable presheaf on $\mathcal{M}\mathrm{fld}^s$ determined by M; explicitly,

$$\mathbb{E}^s_M\Big(\bigsqcup_k \mathbb{R}^n\Big) := \mathrm{Emb}^s\Big(\bigsqcup_k \mathbb{R}^n, M\Big).$$

The reader is warned that our notation does not reflect the choice of n, which should always be clear from context.

Remark 2.1 Equivalently, writing $\mathbb{E}_n^s$ for the endomorphism operad of $\mathbb{R}^n$ with its standard smooth structure — equivalent to the framed little n–disks operad allowing translation, scaling, rotation and reflection of the little disks — the simplicial category $\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}_n^s)$ of simplicial presheaves on $\mathcal{D}\mathrm{isk}_n^s$ is equivalent to the simplicial category of right $\mathbb{E}_n^s$–modules [5, Section 6].

An embedding $N \hookrightarrow M$ determines a map $\mathbb{E}_N^s \to \mathbb{E}_M^s$ of presheaves. Since the category $\mathcal{D}\mathrm{isk}_n^s$ has a filtration by cardinality of path components, there results a canonical functorial cofiltration on mapping spaces between presheaves and localizing at the objectwise weak equivalences also on the derived mapping spaces. In the situation at hand, this cofiltration is called the *embedding calculus Taylor tower*.

Definition 2.2 (Boavida and Weiss) Let N and M be smooth manifolds of dimension n and m, respectively. The *embedding calculus Taylor tower* for smooth embeddings of N into M is the cofiltered derived mapping space of presheaves on truncations of the simplicial category $\mathcal{D}\mathrm{isk}_n^s$:

$$T_\bullet \mathrm{Emb}^s(N, M) := \mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}_n^s)}(\mathbb{E}_N^s, \mathbb{E}_M^s).$$

The cofiltered derived mapping space gives rise to a tower of comparison maps

$$\begin{array}{ccc} & & \vdots \\ & & \downarrow \\ & \overset{\eta_{k+1}}{\nearrow} & T_{k+1}\mathrm{Emb}^s(N, M) \\ & & \downarrow \\ \mathrm{Emb}^s(N, M) & \xrightarrow{\ \eta_k\ } & T_k\mathrm{Emb}^s(N, M) \\ & & \downarrow \\ & & \vdots \end{array}$$

We write $T_\infty \mathrm{Emb}^s(N, M)$ for the homotopy limit of the tower, which is to say the derived mapping space of presheaves on the untruncated simplicial category $\mathcal{D}\mathrm{isk}_n^s$. One can choose a model for the derived mapping space such that these constructions are functorial in $M, N \in \mathcal{M}\mathrm{fld}^s$, there are associative and unital composition maps, and both functoriality and composition are compatible with the above comparison maps. See [29, Section 3.3.1] for further discussion of this point.

The following is [16, Theorem 2.3], relying on excision estimates from Goodwillie and Klein [14].

Theorem 2.3 (Goodwillie, Klein and Weiss) *The map η_k is $(3-m+(k+1)(m-n-2))$–connected for $k > 0$. In particular, if $m-n \geq 3$, then η_∞ is a weak equivalence.*

In fact, we may replace n in the above result by the handle dimension $\mathrm{hdim}(N)$ of N. Recall that $\mathrm{hdim}(N) \leq h$ if N is the interior of a manifold which admits a handle decomposition with handles of index $\leq h$ only. For example, $\mathrm{hdim}(\mathbb{R}^n) = 0$.

If $M = N$, we write $T_\bullet\mathrm{Diff}(M) \subseteq T_\bullet\mathrm{Emb}^s(M, M)$ for the simplicial subset of homotopy invertible maps. This distinction may very well be unnecessary; however, even in cases where every self-embedding of M is a diffeomorphism, we do not know whether every path component of the limit of the Taylor tower is invertible.

Question 2.4 When are all elements of $\pi_0\,\mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}^s_m)}(\mathbb{E}^s_M, \mathbb{E}^s_M)$ invertible?

2.2 Calculus for manifolds with boundary

We close with a brief description of the modifications necessary to use embedding calculus in the setting of manifolds with boundary [5, Section 9]. Fixing a smooth manifold Z, we write $\mathcal{M}\mathrm{fld}^s_Z$ for the simplicial category of smooth manifolds with boundary identified with Z by a diffeomorphism, and morphism spaces given by smooth embeddings that are the identity on Z. In particular, $\mathcal{M}\mathrm{fld}^s = \mathcal{M}\mathrm{fld}^s_\varnothing$. Let $\mathcal{D}\mathrm{isk}^s_{n,Z} \subset \mathcal{M}\mathrm{fld}^s_Z$ be the full subcategory on those objects that are diffeomorphic relative to Z to a disjoint union of a collar $Z \times [0, 1)$ and finitely many copies of $\mathbb{R}^n$.

An object $P \in \mathcal{M}\mathrm{fld}^s_Z$ determines a representable presheaf on $\mathcal{M}\mathrm{fld}^s_Z$ and we denote its restriction to $\mathcal{D}\mathrm{isk}^s_{n,Z}$ by $\mathbb{E}^s_{P,\partial}$. As before, for an object $N \in \mathcal{M}\mathrm{fld}^s_Z$ of dimension n, we obtain an approximation

$$\mathrm{Emb}^s_\partial(N, P) \to T_\bullet\mathrm{Emb}^s_\partial(N, P) = \mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}^s_{n,Z})}(\mathbb{E}^s_{N,\partial}, \mathbb{E}^s_{P,\partial})$$

as a cofiltered derived mapping space of presheaves on $\mathcal{D}\mathrm{isk}^s_{n,Z}$. The conclusion of Theorem 2.3 holds for this approximation, though handle dimension needs to be replaced by handle dimension relative to Z.

2.3 A simplicial category of topological manifolds

Recall that a topological embedding $e\colon N \to M$ is *locally flat* if, for every $p \in N$, there exist open neighborhoods $p \in U$ and $e(U) \subseteq V$ and homeomorphisms $U \cong \mathbb{R}^n$ and $V \cong \mathbb{R}^m$ fitting into the commuting diagram

(1)
$$\begin{array}{ccc} N \supseteq U & \xrightarrow{e|_U} & V \subseteq M \\ \cong\downarrow & & \downarrow\cong \\ \mathbb{R}^n & \xrightarrow{j} & \mathbb{R}^m \end{array}$$

where $j : \mathbb{R}^n \to \mathbb{R}^m$ is the standard inclusion $(x_1, \dots, x_n) \mapsto (x_1, \dots, x_n, 0, \dots, 0)$.

The simplicial category $\mathcal{M}\mathrm{fld}^t$ has objects topological manifolds of finite type and arbitrary dimension, with the n–simplices of the mapping space $\mathrm{Map}_{\mathcal{M}\mathrm{fld}^t}(N, M)$ given by commuting diagrams

$$\begin{array}{ccc} \Delta^n \times N & \longrightarrow & \Delta^n \times M \\ & \searrow \quad \swarrow & \\ & \Delta^n & \end{array}$$

with the top map a locally flat embedding admitting charts as in (1) that commute with the projection to Δ^n. This definition is chosen so that the isotopy extension theorem holds.

As every smooth embedding is locally flat as a consequence of the tubular neighborhood theorem, forgetting the smooth structure defines a simplicial functor from $\mathcal{M}\mathrm{fld}^s$ to $\mathcal{M}\mathrm{fld}^t$.

2.4 Microbundles

Microbundles were defined by Milnor in [37] and play the role of vector bundles for topological manifolds.

Definition 2.5 A *retractive space* is a map $\pi\colon E \to B$ of topological spaces together with a section $\iota\colon B \to E$.

The spaces E and B are referred to as the *total space* and *base space*, and the maps π and ι as *projection* and *zero section*. Via the zero section, we identify B with its image in E, and we abusively refer to this image also as the zero section. We abusively refer to a retractive space simply by the letter E.

Definition 2.6 A *map* $F\colon E_1 \to E_2$ *of retractive spaces* is a continuous map $F\colon E_1 \to E_2$ such that the dashed filler exists in the commuting diagram

$$\begin{CD} B_1 @>{\iota_1}>> E_1 @>{\pi_1}>> B_1 \\ @V{f}VV @VV{F}V @VV{f}V \\ B_2 @>{\iota_2}>> E_2 @>{\pi_2}>> B_2 \end{CD}$$

Note that the map F determines the map $f = \pi_2 \circ F \circ \iota_1$. When we wish to emphasize the latter, we say that F is a map of retractive spaces *over* f, or *over* B in the case $f = \mathrm{id}_B$. Retractive spaces and morphisms between them form a category, Retr.

Definition 2.7 A *microbundle* is a retractive space E such that, for every $b \in B$, there is an open neighborhood $b \in U \subseteq E$ and a homeomorphism $U \cong \pi(U) \times \mathbb{R}^n$ such that the diagram

$$\begin{array}{ccccc} & & U & & \\ & \overset{\iota}{\nearrow} & & \overset{\pi}{\searrow} & \\ \pi(U) & & \downarrow \cong & & \pi(U) \\ & \searrow & & \nearrow & \\ & & \pi(U) \times \mathbb{R}^n & & \end{array}$$

commutes, where the bottom left map is induced by the inclusion of the origin and the bottom right is projection onto the first factor.

Example 2.8 The prototypical example of a microbundle is the tangent microbundle of a topological manifold — see Definition 2.13 below or [37, Example (3)].

The homeomorphisms which appear in the previous definition are called *microbundle charts*. Note that, by invariance of domain, the parameter n in Definition 2.7 is locally constant.

If E is a retractive space, so is any open neighborhood W of the zero section. The set of germs of maps $E_1 \to E_2$ of retractive spaces is the colimit

$$\operatorname*{colim}_{B_1 \subseteq U \subseteq E_1} \mathrm{Hom}_{\mathrm{Retr}}(U, E_2),$$

over the poset of open subsets U of E_1 containing B_1, which may be composed as follows:

$$\begin{array}{ccc}
\operatorname*{colim}_{B_1 \subseteq U \subseteq E_1} \mathrm{Hom}_{\mathrm{Retr}}(U, E_2) \times \operatorname*{colim}_{B_2 \subseteq V \subseteq E_2} \mathrm{Hom}_{\mathrm{Retr}}(V, E_3) & & (F_1, F_2) \\
\| \wr & & \big\downarrow \\
\operatorname*{colim}_{B_1 \subseteq U \subseteq E_1, B_2 \subseteq V \subseteq E_2} \mathrm{Hom}_{\mathrm{Retr}}(U, E_2) \times \mathrm{Hom}_{\mathrm{Retr}}(V, E_3) & & \\
\big\downarrow & & \\
\operatorname*{colim}_{B_1 \subseteq W \subseteq E_1} \mathrm{Hom}_{\mathrm{Retr}}(W, E_3) & & F_2 \circ F_1|_{F_1^{-1}(V)}
\end{array}$$

This composition is easily checked to be associative and unital.

Definition 2.9 A *map* $F \colon E_1 \to E_2$ *of microbundles* is a germ of a map of retractive spaces such that, for every $b \in B_1$, there are microbundle charts around b and $F(b)$ fitting into the commuting diagram

$$\begin{array}{ccc}
U_1 & \xrightarrow{\quad F|_{U_1} \quad} & U_2 \\
\cong \big\downarrow & & \big\downarrow \cong \\
\pi(U_1) \times \mathbb{R}^{n_1} & \xrightarrow{\ f|_{\pi(U_1)} \times j\ } & \pi(U_1) \times \mathbb{R}^{n_2}
\end{array}$$

where $j \colon \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ is the standard inclusion and $f \colon B_1 \to B_2$ the map on base spaces induced by F.

Note that maps of microbundles are fiberwise embeddings.

Remark 2.10 When E_1 and E_2 are microbundles of the same fixed dimension, this definition reduces to [37, Definition 6.3].

Example 2.11 The prototypical example of a map of microbundles is the topological derivative of a locally flat embedding — see Definition 2.14 below.

It is easy to check that maps of microbundles are closed under composition of germs of maps of retractive spaces, so we obtain a category Mic of numerable microbundles as a subcategory of the category Retr of retractive spaces.

A retractive space E with base B can be pulled back along a continuous map $f: A \to B$ to give a retractive space f^*E with base A; in the commutative diagram

$$\begin{array}{ccccc} A & \longrightarrow & f^*E & \longrightarrow & A \\ {\scriptstyle f}\downarrow & & \downarrow & & \downarrow{\scriptstyle f} \\ B & \xrightarrow{\iota} & E & \xrightarrow{\pi} & B \end{array}$$

the right-hand square is a pullback square, and the section $A \to f^*E$ is induced by the maps $\mathrm{id}: A \to A$ and $\iota \circ f: A \to E$. This exhibits a canonical map of retractive spaces $f^*E \to E$. If E is a microbundle, then f^*E is so as well, and the canonical map is a map of microbundles [37, Section 3]. Given a microbundle E with base B and a topological space X, we let $X \times E \to X \times B$ denote the pullback of E along the projection $X \times B \to B$.

Microbundles form a simplicial category $\mathcal{M}\mathrm{ic}$ via the declaration

$$\mathrm{Map}_{\mathcal{M}\mathrm{ic}}(E_1, E_2)_n := \mathrm{Hom}_{\mathrm{Mic}}(\Delta^n \times E_1, E_2).$$

Concretely, an n–simplex $F: \Delta^n \times E_1 \to E_2$ can be described as a germ near the zero section $\Delta^n \times B_1$ of a commutative diagram

$$\begin{array}{ccc} \Delta^n \times E_1 & \xrightarrow{(\pi_1, F)} & \Delta^n \times E_2 \\ \downarrow & & \downarrow \\ \Delta^n \times B_1 & \xrightarrow{(\pi_1, f)} & \Delta^n \times B_2 \\ & \searrow \quad \Delta^n \quad \swarrow & \end{array}$$

with the additional properties that

(i) (π_1, F) preserves the zero section, and

(ii) with respect to suitable microbundle charts, (π_1, F) is given by the germ of

$$(\mathrm{id}, f)|_{U_1} \times j : U_1 \times \mathbb{R}^{n_1} \to U_2 \times \mathbb{R}^{n_2}$$

with j the standard inclusion.

Using that every topological horn is a retract of the corresponding topological simplex, it is easy to see that these mapping objects are Kan complexes.

Microbundles adhere to a covering homotopy theorem analogous to the classical result for vector bundles and fiber bundles, which has the following consequence. In it, $\mathcal{T}\mathrm{op}$ denotes the simplicial category with objects topological spaces and morphism spaces the singular simplicial sets of mapping spaces.

Lemma 2.12 *The natural map* $\mathrm{Map}_{\mathcal{M}\mathrm{ic}}(E_1, E_2) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(B_1, B_2)$ *is a Kan fibration with fiber over* $f: B_1 \to B_2$ *canonically isomorphic to the simplicial subset of* $\mathrm{Map}_{\mathcal{M}\mathrm{ic}}(E_1, f^*E_2)$ *with underlying map* id_{B_1}.

Proof We check that the map $\mathrm{Map}_{\mathcal{M}\mathrm{ic}}(E_1, E_2) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(B_1, B_2)$ is a Kan fibration, as the identification of the fiber is straightforward. To check the lifting property in a commutative diagram

$$\begin{array}{ccc} \Lambda^n_k & \longrightarrow & \mathrm{Map}_{\mathcal{M}\mathrm{ic}}(E_1, E_2) \\ \downarrow & & \downarrow \\ \Delta^n & \longrightarrow & \mathrm{Map}_{\mathcal{T}\mathrm{op}}(B_1, B_2) \end{array}$$

we first, by gluing, represent the top map by a map of microbundles $F : \Lambda^n_k \times E_1 \to E_2$ (here, and throughout, we employ the same notation for a simplicial set and its geometric realization). We similarly represent the bottom map by an extension of the map f underlying F to a continuous map $g : \Delta^n \times B_1 \to B_2$.

Let us denote by $\widetilde{F}$, $\tilde{f}$ and $\tilde{g}$ the maps obtained from F, f and g using the homeomorphism of pairs

$$(\Delta^n, \Lambda^n_k) \cong (\Delta^{n-1} \times [0,1], \Delta^{n-1} \times \{0\}).$$

Under this identification, the lifting problem at hand is equivalent to extending a map of microbundles $\widetilde{F} : \Delta^{n-1} \times E_1 \to \tilde{f}^* E_2$ over $\Delta^{n-1} \times B_1$ to $\Delta^{n-1} \times [0,1] \times E_1 \to \tilde{g}^* E_2$ over $\Delta^{n-1} \times [0,1] \times B_1$. By the microbundle homotopy covering theorem [37, Theorem 3.1], there is an isomorphism of microbundles $\varphi : \tilde{g}^* E_2 \cong \tilde{f}^* E_2 \times [0,1]$ over $\Delta^{n-1} \times [0,1] \times B_1$. It is now evident that the desired extension exists, as we may form the product of $\widetilde{F}$ with $[0,1]$ and apply φ^{-1}. □

2.5 Topological tangency

We come now to the motivating example of a microbundle, the "tangent bundle" of a topological manifold [37, Lemma 2.1].

Definition 2.13 Let M be a topological manifold. The *topological tangent bundle* of M, denoted by $T^t M$, is the retractive space

$$M \xrightarrow{\Delta} M \times M \xrightarrow{\pi} M,$$

where π is the projection onto the first factor

To verify that $T^t M$ is a microbundle, it suffices, by locality, to assume $M = \mathbb{R}^m$, in which case we may appeal to the commuting diagram

$$\begin{array}{ccccc} & & \mathbb{R}^m \times \mathbb{R}^m & & \\ & \overset{\Delta}{\nearrow} & & \overset{\pi}{\searrow} & \\ \mathbb{R}^m & & \downarrow \cong & & \mathbb{R}^m \\ & \searrow & & \nearrow & \\ & & \mathbb{R}^m \times \mathbb{R}^m & & \end{array} \qquad \begin{array}{c} (x, y) \\ \downarrow \\ (x, y - x) \end{array}$$

A smooth embedding has a derivative, and likewise a locally flat embedding $\varphi : N \to M$ has a topological derivative.

Definition 2.14 If $\varphi\colon N \to M$ is a locally flat embedding, the *topological derivative* $T^t\varphi\colon T^t N \to T^t M$ of φ is the map of microbundles

$$\begin{array}{ccccc} N & \xrightarrow{\Delta} & N \times N & \xrightarrow{\pi} & N \\ \downarrow{\scriptstyle\varphi} & & \downarrow{\scriptstyle\varphi\times\varphi} & & \downarrow{\scriptstyle\varphi} \\ M & \xrightarrow{\Delta} & M \times M & \xrightarrow{\pi} & M \end{array}$$

To verify that $T^t\varphi$ is a map of microbundles, we may, by locality, assume that φ is the standard inclusion $\mathbb{R}^n \hookrightarrow \mathbb{R}^m$, in which case the bundle chart constructed above implies the claim. Thus, we obtain a simplicial functor $T^t\colon \mathcal{M}\mathrm{fld}^t \to \mathcal{M}\mathrm{ic}$

2.6 Comparing tangent bundles

We write $\mathcal{V}\mathrm{ec}$ for the simplicial category of numerable vector bundles and maps of vector bundles, which for us are always fiberwise linear injections. Specifically, given vector bundles $E_1 \to B_1$ and $E_2 \to B_2$, an n–simplex of $\mathrm{Map}_{\mathcal{V}\mathrm{ec}}(E_1, E_2)$ is a commuting diagram

$$\begin{array}{ccc} E_1 \times \Delta^n & \longrightarrow & E_2 \times \Delta^n \\ \downarrow & & \downarrow \\ B_1 \times \Delta^n & \longrightarrow & B_2 \times \Delta^n \\ & \searrow \quad \Delta^n \quad \swarrow & \end{array}$$

in which the top map is a fiberwise linear injection. As before, these mapping spaces are Kan complexes.

We record the following standard consequence of the covering homotopy theorem for vector bundles [19, Theorem 4.3], whose proof proceeds along the lines of Lemma 2.12.

Lemma 2.15 *The natural map* $\mathrm{Map}_{\mathcal{V}\mathrm{ec}}(E_1, E_2) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(B_1, B_2)$ *is a Kan fibration with fiber over* $f\colon B_1 \to B_2$ *canonically isomorphic to the simplicial subset of* $\mathrm{Map}_{\mathcal{V}\mathrm{ec}}(E_1, f^*E_2)$ *with underlying map* id_{B_1}.

Vector bundles are in particular microbundles, and assigning to a vector bundle its underlying microbundle extends to a simplicial functor $\mathcal{M}\mathrm{ic} \to \mathcal{V}\mathrm{ec}$.

We now have two ways of extracting a microbundle from a smooth manifold M: first, by considering its tangent bundle TM as a microbundle; second, by forgetting the smooth structure and considering $T^t M$. To compare these, we use the following construction:

Construction 2.16 Fix a Riemannian metric on the smooth manifold M. The $t = 1$ exponential map is defined on a neighborhood U of the zero section, and the assignment

$$\exp_M\colon TM \supseteq U \to T^t M, \quad (p, v) \mapsto (p, \exp(p, v))$$

defines a map of retractive spaces.

Proposition 2.17 (Milnor [37, Theorem 2.2]) *The map of Construction 2.16 defines an isomorphism of microbundles* $TM \xrightarrow{\sim} T^t M$.

3 Formally smooth manifolds

The first goal of this section is to factor the forgetful functor from smooth to topological manifolds as in the commuting diagram

$$\begin{array}{ccc} \mathcal{M}\mathrm{fld}^s & \longrightarrow & \mathcal{M}\mathrm{fld}^t \\ \simeq\uparrow & & \uparrow \\ \mathcal{M}\mathrm{fld}^r & \longrightarrow & \mathcal{M}\mathrm{fld}^f \end{array}$$

The simplicial category $\mathcal{M}\mathrm{fld}^r$ is a category of Riemannian manifolds under embeddings respecting the metric up to specified homotopy. It is introduced because Construction 2.16 requires a Riemannian metric. As a result of the homotopy equivalence between $O(n)$ and $\mathrm{GL}(n)$, the leftmost functor is an equivalence, and the role of $\mathcal{M}\mathrm{fld}^r$ is as a convenient proxy for $\mathcal{M}\mathrm{fld}^s$. The simplicial category $\mathcal{M}\mathrm{fld}^f$ is a category of *formally smooth* manifolds, which is to say manifolds equipped with vector bundle refinements of their topological tangent bundles.

The second goal of this section is to prove Theorem 3.18, which asserts that all information detectable by embedding calculus is contained in $\mathcal{M}\mathrm{fld}^f$.

3.1 Simplicial categories of Riemannian and formally smooth manifolds

In this section, we have in mind the model of the homotopy pullback of simplicial categories explained in Section 6.2.4, following [1].

We begin with the construction of $\mathcal{M}\mathrm{fld}^r$. We write $\mathcal{M}\mathrm{et}$ for the simplicial category whose objects are vector bundles endowed with Riemannian metrics and whose morphisms are fiberwise linear isometries, which are assembled into simplicial sets in the same manner as in $\mathcal{V}\mathrm{ec}$. As before, these mapping spaces are Kan complexes and we have the following consequence of local triviality:

Lemma 3.1 *The natural map* $\mathrm{Map}_{\mathcal{M}\mathrm{et}}(E_1, E_2) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(B_1, B_2)$ *is a Kan fibration with fiber over* $f : B_1 \to B_2$ *canonically isomorphic to the simplicial subset of* $\mathrm{Map}_{\mathcal{M}\mathrm{et}}(E_1, f^* E_2)$ *with underlying map* id_{B_1}.

There is a canonical simplicial forgetful functor from $\mathcal{M}\mathrm{et}$ to $\mathcal{V}\mathrm{ec}$.

Proposition 3.2 *The forgetful functor* $\mathcal{M}\mathrm{et} \to \mathcal{V}\mathrm{ec}$ *is essentially surjective and induces weak equivalences on mapping spaces.*

Proof The first claim follows from the fact that every numerable vector bundle admits a Riemannian metric. For the second claim, by Lemmas 2.15 and 3.1 it suffices to show that the maps induced on point–set fibers in the commuting diagram

$$\begin{array}{ccc} \mathrm{Map}_{\mathcal{M}\mathrm{et}}(E_1, E_2) & \longrightarrow & \mathrm{Map}_{\mathcal{V}\mathrm{ec}}(E_1, E_2) \\ \downarrow & & \downarrow \\ \mathrm{Map}_{\mathcal{T}\mathrm{op}}(B_1, B_2) & = & \mathrm{Map}_{\mathcal{T}\mathrm{op}}(B_1, B_2) \end{array}$$

are weak equivalences. By the same results, we may identify the left-hand (resp. right-hand) fiber over f with the singular simplicial set of the space of sections of the associated bundle of noncompact (resp. compact) Stiefel manifolds, whose fibers are general linear (resp. orthogonal) groups. The conclusion then follows as the inclusion of the orthogonal group into the general linear group is a homotopy equivalence. □

We use this to define $\mathcal{M}\mathrm{fld}^r$, which is a homotopy pullback as in Section 6.2.4.

Definition 3.3 The *simplicial category of Riemannian manifolds* is the homotopy pullback in the diagram

$$\begin{array}{ccc} \mathcal{M}\mathrm{fld}^r & \longrightarrow & \mathcal{M}\mathrm{fld}^s \\ \downarrow & & \downarrow T \\ \mathcal{M}\mathrm{et} & \longrightarrow & \mathcal{V}\mathrm{ec} \end{array}$$

of simplicial categories over $\mathcal{T}\mathrm{op}$.

Notation 3.4 We denote the morphism spaces in $\mathcal{M}\mathrm{fld}^r$ by $\mathrm{Emb}^r(-,-)$.

Thus, an object of $\mathcal{M}\mathrm{fld}^r$ is a smooth manifold with a choice of metric, and a morphism is a fiberwise isometry covering a smooth embedding, together with a fiberwise homotopy through linear injections to the derivative of the embedding.

As the following result illustrates, the forgetful functor exhibits $\mathcal{M}\mathrm{fld}^r$ as a proxy for $\mathcal{M}\mathrm{fld}^s$. This proxy is easier to map out of.

Proposition 3.5 *The forgetful functor $\mathcal{M}\mathrm{fld}^r \to \mathcal{M}\mathrm{fld}^s$ is essentially surjective, and induces weak equivalences on mapping spaces.*

Proof The first claim follows from the statement that every smooth manifold admits a Riemannian metric. The second claim follows from Propositions A.2 and 3.2 and Lemma 2.15. □

We continue with the construction of $\mathcal{M}\mathrm{fld}^f$, which is a homotopy pullback as in Section 6.2.4.

Definition 3.6 The *simplicial category of formally smooth manifolds* is the homotopy pullback in the diagram

$$\begin{array}{ccc} \mathcal{M}\mathrm{fld}^f & \longrightarrow & \mathcal{M}\mathrm{fld}^t \\ \downarrow & & \downarrow T^t \\ \mathcal{V}\mathrm{ec} & \longrightarrow & \mathcal{M}\mathrm{ic} \end{array}$$

of simplicial categories over $\mathcal{T}\mathrm{op}$.

Notation 3.7 We denote the morphism spaces in $\mathcal{M}\mathrm{fld}^f$ by $\mathrm{Emb}^f(-,-)$.

Thus, an object of $\mathcal{M}\mathrm{fld}^f$ is a topological manifold with a vector bundle refinement of its topological tangent bundle, and a morphism is a fiberwise linear injection covering a topological embedding, together with a fiberwise homotopy through embeddings to the topological derivative of the embedding.

It remains to construct the functor $\mathcal{M}\mathrm{fld}^r \to \mathcal{M}\mathrm{fld}^f$.

Construction 3.8 We obtain $\mathcal{M}\mathrm{fld}^r \to \mathcal{M}\mathrm{fld}^f$ as an instance of Construction A.4. The requisite data are the following:

(i) the simplicial functor $\mathcal{M}\mathrm{fld}^r \to \mathcal{M}\mathrm{et} \to \mathcal{V}\mathrm{ec}$ associating to a Riemannian manifold its tangent bundle,

(ii) the simplicial functor $\mathcal{M}\mathrm{fld}^r \to \mathcal{M}\mathrm{fld}^s \to \mathcal{M}\mathrm{fld}^t$ associating to a Riemannian manifold its underlying topological manifold,

(iii) the natural isomorphism indicated by the thick arrow between bottom-left and top-right compositions in the diagram

$$\begin{array}{ccc} \mathcal{M}\mathrm{fld}^r & \longrightarrow & \mathcal{M}\mathrm{fld}^s \\ \big\downarrow & \overset{\cong}{\nearrow\!\!\!\!\nearrow} & \big\downarrow \\ \mathcal{V}\mathrm{ec} & \longrightarrow & \mathcal{M}\mathrm{ic} \end{array}$$

arising from Construction 2.16.

Remark 3.9 Upon restricting to the respective full subcategories of manifolds of dimension different from 4, the functor of Construction 3.8 becomes an equivalence by smoothing theory [22, Essays IV and V].

3.2 Smooth embeddings of Euclidean spaces

In the next sections, we assemble results on various types of embeddings of Euclidean spaces, which will be used below in the proof of Theorem 3.18. We begin in the smooth context, where these results are standard, but we include proofs for the sake of completeness.

Fix a smooth m–manifold M and a natural number $0 < n \leq m$, as well as a point $p \in M$. We introduce four simplicial sets, the first three defined as pullbacks of diagrams of the form

$$\begin{array}{ccc} & & X \\ & & \big\downarrow \\ \{p\} & \longrightarrow & M \end{array}$$

(i) Taking $X = \mathrm{Map}_{\mathcal{V}\mathrm{ec}}(T\mathbb{R}^n, TM)$ mapping to M by evaluation at the origin followed by projection, we obtain $\mathrm{Map}_{\mathcal{V}\mathrm{ec},p}(T\mathbb{R}^n, TM)$.

(ii) Taking $X = \mathrm{Map}_{\mathcal{V}\mathrm{ec}}(T_0\mathbb{R}^n, TM)$ mapping to M in the same way, we obtain $\mathrm{Map}_{\mathcal{V}\mathrm{ec},p}(T_0\mathbb{R}^n, TM)$, otherwise known as the (noncompact) Stiefel manifold of n–planes in T_pM.

(iii) Fixing an open subset $0 \in U \subseteq \mathbb{R}^n$ and taking $X = \mathrm{Emb}^s(U, M)$ mapping to M by evaluation at the origin, we obtain $\mathrm{Emb}^s_p(U, M)$.

(iv) Finally, we write $G^s_p(n, M) := \mathrm{colim}_{0 \in U \subseteq \mathbb{R}^n} \mathrm{Emb}^s_p(U, M)$ for the simplicial set of germs of smooth embeddings (here U ranges over open subsets containing the origin).

Lemma 3.10 *All maps in the commuting diagram*

$$\begin{array}{ccc} \mathrm{Emb}^s_p(\mathbb{R}^n, M) & \longrightarrow & G^s_p(n, M) \\ \downarrow & & \downarrow \\ \mathrm{Map}_{\mathcal{V}\mathrm{ec},p}(T\mathbb{R}^n, TM) & \longrightarrow & \mathrm{Map}_{\mathcal{V}\mathrm{ec},p}(T_0\mathbb{R}^n, TM) \end{array}$$

are weak equivalences.

Proof For the top map, the restriction $\mathrm{Emb}^s_p(\mathbb{R}^n, M) \to \mathrm{Emb}^s_p(U, M)$ is a weak equivalence whenever U is an open ball centered at the origin, since the inclusion $U \subseteq \mathbb{R}^n$ is isotopic to a diffeomorphism relative to the origin. The claim now follows from the observation that the subposet of such open balls is final in the poset of all open neighborhoods of the origin, and both are filtered. For the bottom map, the claim is a consequence of the contractibility of $\mathbb{R}^n$ and the homotopy covering theorem. For the right-hand map, the claim may be tested on compact families of germs, so we may assume that $M = \mathbb{R}^m$. In this case, the Stiefel manifold includes canonically into $\mathrm{Emb}^s_p(\mathbb{R}^n, \mathbb{R}^m)$, and composing with the map to $G^s_p(n, \mathbb{R}^m)$ supplies a homotopy inverse. For the left-hand map, the claim follows by two-out-of-three. □

We will also have use for a mild generalization of the claim regarding the top map. Let $N = \bigsqcup_{i \in I} \mathbb{R}^{n_i}$ for some finite set I, and fix a collection $p_i \in M$ of points for each $i \in I$ such that $p_i \neq p_j$ if $i \neq j$. Let $\mathrm{Emb}^s_{p_I}(N, M) \subseteq \mathrm{Emb}^s(N, M)$ be the simplicial subset of embeddings sending the origin in $\mathbb{R}^{n_i}$ to p_i.

Lemma 3.11 *The canonical map* $\mathrm{Emb}^s_{p_I}(N, M) \to \prod_{i \in I} G^s_{p_i}(n_i, M)$ *is a weak equivalence.*

Proof The map in question factors through $G^s_{p_I}(n_I, M) := \mathrm{colim}_{U \subseteq N} \mathrm{Emb}^s_{p_I}(U, M)$, where U ranges over open subsets containing the origin in $\mathbb{R}^{n_i}$ for every $i \in I$. As in the previous argument, the subposet consisting of the disjoint unions of open balls around the respective origins is final, and both are filtered. Thus, since the inclusion of such an open set into N is isotopic to a diffeomorphism, the first map is a weak equivalence. On the other hand, the map

$$G^s_{p_I}(n_I, M) \to \prod_{i \in I} G^s_{p_i}(n_i, M)$$

is an isomorphism; indeed, injectivity is immediate, and surjectivity follows from the observation that any family of I–tuples of germs parametrized over a compact space (such as a simplex) can be represented by a family of I–tuples of embeddings whose images are pairwise disjoint at every point of the parameter space. □

We write $\mathrm{Conf}_I(M) := \{(p_i)_{i\in I} \mid p_i \neq p_j \text{ if } i \neq j\} \subseteq M^I$ for the *configuration space* of particles in M labeled by I.

Proposition 3.12 *Let M be a smooth manifold and $N = \bigsqcup_{i\in I} \mathbb{R}^{n_i}$. The diagram*

$$\begin{array}{ccc} \mathrm{Emb}^s(N, M) & \longrightarrow & \mathrm{Map}_{\mathcal{V}\mathrm{ec}}(TN, TM) \\ \downarrow & & \downarrow \\ \mathrm{Conf}_I(M) & \longrightarrow & M^I \end{array}$$

induced by evaluation at the respective origins is homotopy Cartesian.

Proof This square is the outer square in the commuting diagram

$$\begin{array}{ccccc} \mathrm{Emb}^s(N, M) & \longrightarrow & \prod_{i\in I} \mathrm{Emb}^s(\mathbb{R}^{n_i}, M) & \longrightarrow & \prod_{i\in I} \mathrm{Map}_{\mathcal{V}\mathrm{ec}}(T\mathbb{R}^{n_i}, TM) \\ \downarrow & & \downarrow & & \downarrow \\ \mathrm{Conf}_I(M) & \longrightarrow & M^I & = & M^I \end{array}$$

so it suffices to verify that each of the inner squares is homotopy Cartesian. The left two vertical maps are fibrations by the isotopy extension theorem [44, Chapter 6], and the right-hand vertical map is a product of fibrations, the i^{th} map being the composite of two fibrations

$$\mathrm{Map}_{\mathcal{V}\mathrm{ec}}(T\mathbb{R}^{n_i}, TM) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(\mathbb{R}^{n_i}, M) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(\{0\}, M).$$

Thus, it suffices to establish that the induced maps on fibers are weak equivalences.

For the right-hand square, the map on fibers is a product of weak equivalences by Lemma 3.10. For the left-hand square, the map on fibers is the top map in the commuting diagram

$$\begin{array}{ccc} \mathrm{Emb}^s_{p^I}(N, M) & \longrightarrow & \prod_{i\in I} \mathrm{Emb}^s_{p_i}(\mathbb{R}^{n_i}, M) \\ \downarrow & & \downarrow \\ \prod_{i\in I} G^s_{p_i}(n_i, M) & = & \prod_{i\in I} G^s_{p_i}(n_i, M) \end{array}$$

and the vertical maps are weak equivalences by Lemmas 3.10 and 3.11. □

3.3 Topological embeddings of Euclidean spaces

We turn now to the topological versions of these facts, our goal being a description of locally flat embeddings of Euclidean spaces in terms of microbundle maps and configuration spaces.

Taking M instead to be merely a topological manifold, we define the simplicial sets $\mathrm{Map}_{\mathcal{M}\mathrm{ic},p}(T\mathbb{R}^n, TM)$, $\mathrm{Map}_{\mathcal{M}\mathrm{ic},p}(T_0\mathbb{R}^n, TM)$, $\mathrm{Emb}^t_p(U, M)$ and $G^t_p(n, M)$ by replacing smooth embeddings with locally flat embeddings and vector bundles with microbundles in the definitions of the previous section.

Lemma 3.13 *All maps in the commuting diagram*

$$\begin{array}{ccc}\mathrm{Emb}^t_p(\mathbb{R}^n, M) & \longrightarrow & G^t_p(n, M) \\ \downarrow & & \downarrow \\ \mathrm{Map}_{\mathcal{M}\mathrm{ic},p}(T\mathbb{R}^n, TM) & \longrightarrow & \mathrm{Map}_{\mathcal{M}\mathrm{ic},p}(T_0\mathbb{R}^n, TM)\end{array}$$

are weak equivalences. In fact, the right-hand vertical map is an isomorphism.

Proof The claim regarding the right-hand map follows upon inspecting the definitions, and the same argument as in Lemma 3.10 suffices for the remaining three. □

As in the smooth case, extending our notation in the obvious way, we have the following generalization:

Lemma 3.14 *The canonical map* $\mathrm{Emb}^t_{p_I}(N, M) \to \prod_{i \in I} G^t_{p_i}(n_i, M)$ *is a weak equivalence.*

Given these inputs and isotopy extension for locally flat embeddings [11] (see [40, Theorem 6.17] for the variant with parameters), the topological analog of Proposition 3.12 follows by the same argument.

Proposition 3.15 *Let* M *be a topological manifold and* $N = \bigsqcup_{i \in I} \mathbb{R}^{n_i}$. *The diagram*

$$\begin{array}{ccc}\mathrm{Emb}^t(N, M) & \longrightarrow & \mathrm{Map}_{\mathcal{M}\mathrm{ic}}(T^t N, T^t M) \\ \downarrow & & \downarrow \\ \mathrm{Conf}_I(M) & \longrightarrow & M^I\end{array}$$

induced by evaluation at the respective origins is homotopy Cartesian.

3.4 Formally smooth embeddings of Euclidean spaces

The key calculation in the proof of Theorem 3.18 is a comparison between Riemannian embeddings and formally smooth embeddings. We start with a lemma concerning Riemannian embeddings:

Lemma 3.16 *Let* M *be a Riemannian manifold and* $N = \bigsqcup_{i \in I} \mathbb{R}^{n_i}$. *The diagram*

$$\begin{array}{ccc}\mathrm{Emb}^r(N, M) & \longrightarrow & \mathrm{Map}_{\mathcal{M}\mathrm{et}}(TN, TM) \\ \downarrow & & \downarrow \\ \mathrm{Conf}_I(M) & \longrightarrow & M^I\end{array}$$

induced by evaluation at the respective origins is homotopy Cartesian.

Proof The upper square of the commuting diagram

$$\begin{array}{ccc}
\mathrm{Emb}^r(N,M) & \longrightarrow & \mathrm{Map}_{\mathcal{M}\mathrm{et}}(TN,TM) \\
\downarrow & & \downarrow \\
\mathrm{Emb}^s(N,M) & \longrightarrow & \mathrm{Map}_{\mathcal{V}\mathrm{ec}}(TN,TM) \\
\Big\downarrow & & \downarrow \\
 & & \mathrm{Map}_{\mathcal{T}\mathrm{op}}(N,M) \\
\Big\downarrow & & \downarrow \\
\mathrm{Conf}_I(M) & \longrightarrow & M^I
\end{array}$$

is homotopy Cartesian by Propositions 3.2 and 3.5 (they imply that right and left vertical maps, respectively, are weak equivalences), and the bottom square is homotopy Cartesian by Proposition 3.12. □

We come now to the result of interest:

Proposition 3.17 *Let M be a Riemannian manifold and $N = \bigsqcup_{i\in I}\mathbb{R}^{n_i}$. Then the canonical map*

$$\mathrm{Emb}^r(N,M) \to \mathrm{Emb}^f(N,M)$$

is a weak equivalence.

Proof Consider the diagram

$$\begin{array}{ccccc}
\mathrm{Emb}^r(N,M) & \longrightarrow & \mathrm{Emb}^f(N,M) & \longrightarrow & \mathrm{Emb}^t(N,M) \\
\downarrow & & \downarrow & & \downarrow \\
\mathrm{Map}_{\mathcal{M}\mathrm{et}}(TN,TM) & \longrightarrow & \mathrm{Map}_{\mathcal{V}\mathrm{ec}}(TN,TM) & \longrightarrow & \mathrm{Map}_{\mathcal{M}\mathrm{ic}}(TM,TN)
\end{array}$$

The right square commutes, but the left square commutes only up to specified homotopy.

The maps from $\mathrm{Map}_{\mathcal{V}\mathrm{ec}}(TN,TM)$ and $\mathrm{Map}_{\mathcal{M}\mathrm{ic}}(TN,TM)$ to $\mathrm{Map}_{\mathcal{T}\mathrm{op}}(N,M)$ are fibrations by Lemma 2.15, so Proposition A.2 grants that the right-hand square is homotopy Cartesian. Therefore, since the lower left-hand map is a weak equivalence by Proposition 3.2, it suffices to show that the outer diagram is also homotopy Cartesian. By Proposition 3.15 and Lemma 3.16, the vertical homotopy fibers in the outer diagram are compatibly identified with the homotopy fiber of the inclusion $\mathrm{Conf}_I(M)\subseteq M^I$, and the claim follows. □

3.5 Consequences for embedding calculus

In order to state the main result, we extend our notation in the obvious way by writing $\mathcal{D}\mathrm{isk}_n^f$ and $\mathcal{D}\mathrm{isk}_n^r$ for the full subcategories on disjoint unions of finitely many copies of $\mathbb{R}^n$ in the appropriate categories of manifolds, and similarly for derived mapping spaces of simplicial presheaves on these categories.

Theorem 3.18 *Given Riemannian metrics on smooth manifolds M and N, there is a canonical weak equivalence*

$$T_\bullet \mathrm{Emb}^s(N, M) \simeq \mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}^f_n)}(\mathbb{E}^f_N, \mathbb{E}^f_M).$$

In particular, the embedding calculus Taylor tower depends only on M and N as formally smooth manifolds.

Remark 3.19 The choice of Riemannian metric on M and N is irrelevant; the space of Riemannian metrics on a smooth manifold is contractible, and our constructions are continuous in the Riemannian metric in the sense that a path of Riemannian metrics gives rise to a homotopy of zigzags of maps between the left- and right-hand sides.

Remark 3.20 Similar methods serve to establish a version of Theorem 3.18 for manifolds M and N with common boundary Z.

The theorem is an immediate consequence of the following result, which will follow easily from Proposition 3.17. Write $f\colon \mathcal{D}\mathrm{isk}^r_n \to \mathcal{D}\mathrm{isk}^s_n$ and $g\colon \mathcal{D}\mathrm{isk}^r_n \to \mathcal{D}\mathrm{isk}^f_n$ for the respective forgetful functors, and write $\Phi := \mathbb{L}g_! f^*$ for the composite of the (automatically derived) restriction and derived induction functors pertaining to these maps (a concrete model for the latter is available via a functorial cofibrant replacement, for example).

Proposition 3.21 *Fix $n \geq 0$.*

(i) *The functor $\Phi\colon \mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}^s_n) \to \mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}^f_n)$ is essentially surjective up to weak equivalence, and induces weak equivalences on derived mapping spaces.*

(ii) *For any Riemannian manifold M, there is a canonical weak equivalence $\Phi(\mathbb{E}^s_M) \simeq \mathbb{E}^f_M$.*

Proof By Proposition 3.17, the functors f and g are Dwyer–Kan equivalences and hence so are the induced maps on presheaf categories [25], implying the first claim. For the second, we observe the zigzag

$$\Phi(\mathbb{E}^s_M) = \mathbb{L}g_! f^* \mathbb{E}^s_M \xleftarrow{\simeq} \mathbb{L}g_! \mathbb{E}^r_M \xrightarrow{\simeq} \mathbb{L}g_! g^* \mathbb{E}^f_M \xrightarrow{\simeq} \mathbb{E}^f_M,$$

where the first two weak equivalences follow from Proposition 3.17. □

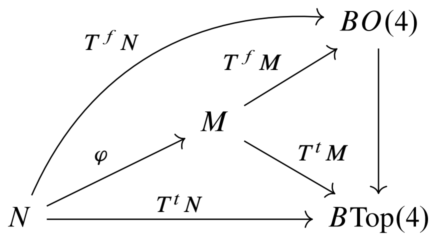
4 Embedding calculus in dimension 4

The goal of this section is to prove Theorems A and B. At the heart of the matter is the question of deciding when two 4–manifolds are formally diffeomorphic.

4.1 Formal diffeomorphisms of 4–manifolds

A homeomorphism $\varphi\colon N \to M$ between formally smooth 4–manifolds has an associated element $\mathrm{ks}(\varphi) \in H^3(N;\mathbb{Z}/2)$, called the *relative Kirby–Siebenmann invariant* (in higher dimensions it is sometimes called the *Casson–Sullivan invariant* [39]). A definition for smooth 4–manifolds is given in [12, Corollary 8.3D], and we define it now for formally smooth 4–manifolds.

By [23, Corollary 2], the topological tangent bundles of N and M have essentially unique lifts to $\mathbb{R}^4$–bundles with structure group $\mathrm{Top}(4)$, where we recall that $\mathrm{Top}(4)$ is the space of self-homeomorphisms of $\mathbb{R}^4$. Thus, we have the diagram



in which the right-hand and bottom triangles may be taken to commute strictly and the outer triangle to commute up to homotopy. The obstruction to the remaining 3–dimensional cell of the diagram commuting up to homotopy is the homotopy class of a map from N to $\mathrm{Top}(4)/O(4)$, which is an Eilenberg–Mac Lane space $K(\mathbb{Z}/2\mathbb{Z},3)$ through dimension 5 [12, Theorems 8.3B and 8.7A]. By definition, the resulting obstruction class in $H^3(N;\mathbb{Z}/2\mathbb{Z})$ is $\mathrm{ks}(\varphi)$. The following is immediate:

Proposition 4.1 *Suppose $\varphi\colon N \to M$ is a homeomorphism between formally smooth 4–manifolds. Then $\mathrm{ks}(\varphi) \in H^3(N;\mathbb{Z}/2)$ vanishes if and only if φ lifts to an isomorphism between N and M in $\mathcal{M}\mathrm{fld}^f$.*

Corollary 4.2 *Let N and M be smooth simply connected compact 4–manifolds. If N and M are homeomorphic, then N and M are isomorphic in $\mathcal{M}\mathrm{fld}^f$.*

Proof Choosing a homeomorphism φ, we have $\mathrm{ks}(\varphi) \in H^3(N;\mathbb{Z}/2\mathbb{Z}) = 0$ by Poincaré duality and the assumption that N is simply connected. □

Remark 4.3 Supposing N and M to be smooth, the *sum-stable smoothing theorem* in [12, Section 8.6] asserts that, if φ lifts to an isomorphism between N and M in $\mathcal{M}\mathrm{fld}^f$, then N and M are *stably diffeomorphic*: there exists $g \geq 0$ and a diffeomorphism

$$\tilde{\varphi}\colon N \#_g (S^2 \times S^2) \xrightarrow{\cong} M \#_g (S^2 \times S^2).$$

The converse is also true, as forming the connected sum with $S^2 \times S^2$ does not affect the value of the relative Kirby–Siebenmann invariant.

4.2 Proof of Theorems A and B

The proofs of these theorems are now a matter of stringing weak equivalences together.

Proof of Theorem A Assuming that N and M are smooth simply connected compact 4–manifolds which are homeomorphic, we have the equivalences

$$\begin{aligned}\mathrm{Emb}^s\Big(\bigsqcup_k S^1, N\Big) &\simeq T_\infty \mathrm{Emb}^s\Big(\bigsqcup_k S^1, N\Big) && \text{(Theorem 2.3)}\\ &\simeq \mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}_1^f)}(\mathbb{E}^f_{\bigsqcup_k S^1}, \mathbb{E}^f_N) && \text{(Theorem 3.18)}\\ &\simeq \mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}_1^f)}(\mathbb{E}^f_{\bigsqcup_k S^1}, \mathbb{E}^f_M) && \text{(Corollary 4.2)}\\ &\simeq T_\infty \mathrm{Emb}^s\Big(\bigsqcup_k S^1, M\Big) && \text{(Theorem 3.18)}\\ &\simeq \mathrm{Emb}^s\Big(\bigsqcup_k S^1, M\Big) && \text{(Theorem 2.3).}\end{aligned}$$ □

Proof of Theorem B Once more assuming that N and M are smooth simply connected compact 4–manifolds, we have

$$\begin{aligned}T_\infty \mathrm{Emb}^s(N, M) &\simeq \mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}_4^f)}(\mathbb{E}^f_N, \mathbb{E}^f_M) && \text{(Theorem 3.18)}\\ &\simeq \mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}_4^f)}(\mathbb{E}^f_N, \mathbb{E}^f_N) && \text{(Corollary 4.2),}\end{aligned}$$

and this last space is nonempty, as it contains the identity. On the other hand, any embedding of N into M is a diffeomorphism by compactness, so $\mathrm{Emb}^s(N, M)$ is nonempty if and only if N and M are diffeomorphic. □

Remark 4.4 Our proof of Theorem A implies that, under the same hypotheses, the finite stages $T_r\mathrm{Emb}^s\big(\bigsqcup_k S^1, N\big)$ and $T_r\mathrm{Emb}^s\big(\bigsqcup_k S^1, M\big)$ are also weakly equivalent. A related result appears in [3, Theorem A], where a study of the second stage of the Taylor tower is leveraged to show that, if N is n–dimensional, the $(2n-7)$–skeleton of $\mathrm{Emb}^s(S^1, N)$ does not depend on the smooth structure of N.

Remark 4.5 The element of $T_\infty\mathrm{Emb}^s(N, M)$ obtained in the course of the proof of Theorem B is homotopy-invertible.

4.3 Remarks on the study of smooth 4–manifolds

In this section, we discuss some expected consequences of our results for the study of smooth 4–manifolds. This discussion is informal and should be taken as motivation for further investigation.

One way to get invariants for smooth manifolds is from configuration space integrals. Pioneered by Kontsevich [24] and developed subsequently by many authors, this type of invariant is given schematically by a map of the form

$$H^*(\Gamma) \to H^*(\mathrm{Emb}^s(N, M)),$$

where Γ is a combinatorially defined cochain complex of graphs. We will remain vague about the coefficients and the precise flavor of graph complex in question (there are many options); suffice it to say that an element of the graph complex is typically interpreted as a set of instructions for combining differential forms on compactified configuration spaces.

Extrapolating from results in the literature, such as [38; 43], a positive answer to the following general question is expected:

Question 4.6 Do configuration space integrals factor through the limit of the embedding calculus Taylor tower?

$$\begin{array}{ccc} H^*(\Gamma) & \longrightarrow & H^*(\mathrm{Emb}^s(N,M)) \\ & \overset{\exists ?}{\searrow} & \uparrow \\ & & H^*(T_\infty \mathrm{Emb}^s(N,M)) \end{array}$$

If Question 4.6 has a positive answer, Theorem 3.18 implies that these invariants cannot distinguish exotic smooth structures on M by taking N to be homeomorphic but not diffeomorphic to M, unless they are already not formally diffeomorphic. For example, it would follow that this use of configuration space integrals can shed no light on the smooth Poincaré conjecture in dimension 4, or at least not directly.

A second use for configuration space integrals, accessed by setting $M = N$, is to study the classifying spaces of diffeomorphism groups. Again assuming a positive answer to Question 4.6, Theorem 3.18 implies that this approach is limited to detecting the algebraic topology of *formal* diffeomorphism groups; for example, the results of Watanabe [45] on the rational homotopy of $B\mathrm{Diff}_\partial(D^4)$ should be interpreted as results about the automorphisms of D^4 as a formally smooth manifold. This change in perspective has concrete consequences.

Proposition 4.7 *If Question 4.6 has a positive answer, then the natural map*

$$\mathrm{Top}(4)/O(4) \to \mathrm{Top}/O$$

is not a weak equivalence, even after rationalizing.

Proof By [45, Theorem 1.1], configuration space integrals produce many nontrivial classes of positive degree in $H^*(B\mathrm{Diff}_\partial(D^4);\mathbb{R})$, which our assumption implies are pulled back from $H^*(BT_\infty\mathrm{Diff}_\partial(D_4);\mathbb{R})$. A version of Theorem 3.18 with boundary implies that the map $\mathrm{Diff}_\partial(D^4) \to T_\infty\mathrm{Diff}_\partial(D_4)$ factors over the automorphisms of D^4 as a formally smooth manifold. By the Alexander trick, the latter are given by $\Omega^5\mathrm{Top}(4)/O(4)$, so $\mathrm{Top}(4)/O(4)$ is not rationally trivial. The claim then follows from the fact that Top/O is rationally trivial [22, Essay V]. □

A third use for configuration space integrals lies in distinguishing embeddings. As many open problems in the topology of smooth 4–manifolds are of this type, Theorem 3.18 likewise rules out the direct use of configuration space integrals in their solutions. For example, using configuration space integrals to distinguish isotopy classes of embeddings of S^3 into S^4 cannot negatively resolve the 4–dimensional smooth Schoenflies conjecture, as shown by the following result (here, the superscript $+$ indicates restriction to orientation-preserving embeddings):

Proposition 4.8 *The image of*

$$\mathrm{Emb}^{s,+}(S^3\times(-\epsilon,\epsilon),S^4)\to T_\infty\mathrm{Emb}^{s,+}(S^3\times(-\epsilon,\epsilon),S^4)$$

lies in a single path component.

Proof By Theorem 3.18, it suffices to show that $\mathrm{Emb}^{f,+}(S^3\times(-\epsilon,\epsilon),S^4)$ is path connected. Since the topological Schoenflies conjecture holds in dimension 4 [7], each locally flat embedding $S^3\times(-\epsilon,\epsilon)\hookrightarrow S^4$ extends to an orientation-preserving locally flat embedding $\mathbb{R}^4\hookrightarrow S^4$. This embedding can be lifted to one of formally smooth manifolds, since $\pi_4(\mathrm{Top}(4)/O(4))=0$ [12, Theorems 8.3B and 8.7A]. Thus, the restriction

$$\mathrm{Emb}^{f,+}(\mathbb{R}^4,S^4)\to\mathrm{Emb}^{f,+}(S^3\times(-\epsilon,\epsilon),S^4)$$

is surjective on path components. Finally,

$$\begin{aligned}\mathrm{Emb}^{f,+}(\mathbb{R}^4,S^4)&\simeq\mathrm{Emb}^{r,+}(\mathbb{R}^4,S^4) &&\text{(Proposition 3.17)}\\ &\simeq\mathrm{Emb}^{s,+}(\mathbb{R}^4,S^4) &&\text{(Proposition 3.5)}\\ &\simeq\mathrm{SO}(5) &&\text{(Lemma 3.10)},\end{aligned}$$

and the last space is path connected. □

Remark 4.9 Theorem 3.18 and the previous discussion suggests that it may be fruitful to study smooth 4–manifolds by

(a) studying formally smooth 4–manifolds, and, separately,

(b) studying the difference between smooth and formally smooth 4–manifolds.

The study of formally smooth 4–manifolds should be much like that of smooth manifolds in higher dimensions, since the Whitney trick is available under assumptions on fundamental groups [12]. In particular, it may be possible to obtain versions of the homological stability and stable homology results of Galatius and Randal-Williams in this setting (see [13] for a survey). If so, one can study the moduli space $\mathcal{M}^f(M)$ of formally smooth manifolds isomorphic to M using the methods of homotopy theory, just as one studies the moduli space $\mathcal{M}^s(M)$ of smooth manifolds diffeomorphic to M in higher dimensions.

Next, we wish to separate the "exotic smooth structures" from the "formally smooth structures" by defining a moduli space of "exotic" smooth manifolds formally isomorphic to M. Fixing a formally smooth manifold M, this moduli space is defined as the homotopy fiber

$$\mathcal{M}^{\mathrm{ex}}(M):=\mathrm{hofiber}[\mathcal{M}^s(M)\to\mathcal{M}^f(M)]$$

over the specified structure. As we argued above, configuration space integrals are likely blind to the topology of this moduli space.

5 Embedding calculus and exotic spheres

In this section we prove Theorem C, which asserts the existence of exotic n–spheres Σ for which $T_\infty\mathrm{Emb}^s(\Sigma,S^n)=\varnothing$.

5.1 Proof of Theorem C

Our proof uses the following convergence criterion:

Proposition 5.1 *Let N_1 and N_2 be nondiffeomorphic closed smooth n–manifolds and M a smooth m–manifold into which N_1 does not embed. If $m-n \geq 3$ and N_2 embeds in M, then $T_\infty \mathrm{Emb}^s(N_1, N_2) = \varnothing$. In particular, the map $\mathrm{Emb}^s(N_1, N_2) \to T_\infty \mathrm{Emb}^s(N_1, N_2)$ is a weak equivalence.*

Proof By Theorem 2.3 and the assumption on N_1, the target of the composition map

$$T_\infty \mathrm{Emb}^s(N_1, N_2) \times T_\infty \mathrm{Emb}^s(N_2, M) \to T_\infty \mathrm{Emb}^s(N_1, M) \simeq \mathrm{Emb}^s(N_1, M)$$

is empty, so the source must be empty as well. The assumption on N_2 says that the domain of the map $\mathrm{Emb}^s(N_2, M) \to T_\infty \mathrm{Emb}^s(N_2, M)$ is nonempty, so the right factor of the source is also nonempty. Thus the left factor is empty, as desired. □

The heavy lifting is handled by a collage of classical results; see also [35, page 408].

Theorem 5.2 (Hsiang, Levine, Szczarba and Mahowald) *If $n = 2^j$ with $j \geq 3$, then there is an exotic n–sphere Σ that does not embed in $\mathbb{R}^{n+3}$.*

Proof It suffices to show that there is an exotic n–sphere Σ that embeds in $\mathbb{R}^{2n-3}$ with nontrivial normal bundle. Indeed, our assumptions on n imply that $n < 2(n-3)-1$, so [18, Lemma 1.1] then guarantees that *every* embedding of Σ in $\mathbb{R}^{2n-3}$ has nontrivial normal bundle. Since every embedding of Σ in $\mathbb{R}^{n+3}$ has trivial normal bundle by [36, Corollary], there can be no such embedding, or else the composite

$$\Sigma \to \mathbb{R}^{n+3} \to \mathbb{R}^{2n-3}$$

has trivial normal bundle, a contradiction.

In order to find such a Σ, it suffices by [18, Theorem 1.2] to find a nonzero element $\alpha \in \pi_{n-1}(\mathrm{SO}(n-3))$ annihilated by the maps $i : \pi_{n-1}(\mathrm{SO}(n-3)) \to \pi_{n-1}(\mathrm{SO}) \cong \mathbb{Z}$ and $J : \pi_{n-1}(\mathrm{SO}(n-3)) \to \pi_{2n-4}(S^{n-3})$. When $n \equiv 0 \pmod 8$ we have $\pi_{n-1}(\mathrm{SO}(n-3)) \cong \mathbb{Z} \oplus \mathbb{Z}/2$ by [20, page 161], with the 2–torsion generated by the image $\partial(\nu)$ of a generator $\nu \in \pi_n(S^{n-3}) \cong \mathbb{Z}/24\mathbb{Z}$ under the connecting homomorphism

$$\partial\colon \pi_n(S^{n-3}) \to \pi_{n-1}\,\mathrm{SO}(n-3)$$

of the fibration sequence $\mathrm{SO}(n-3) \to \mathrm{SO}(n-2) \to S^{n-3}$ [20, Theorem 3(i)].

We now prove that $\alpha = \partial(\nu)$ is in the kernel of both i and J. According to [18, page 176], the composite

$$\pi_n(S^{n-3}) \xrightarrow{\partial} \pi_{n-1}(\mathrm{SO}(n-3)) \xrightarrow{J} \pi_{2n-4}(S^{n-3})$$

is the Whitehead product $[\iota, -]$, where $\iota \in \pi_{n-3}(S^{n-3})$ is a generator. Then $i(\alpha) \in \pi_{n-1}(\mathrm{SO}) \cong \mathbb{Z}$ is torsion, hence zero, while $J(\alpha) = [\iota, \nu] = 0$ by [34, page 249, (2)] because $n = 2^j$ with $j \geq 3$ (Theorem 1.1.2(b) of [32] proved there are no other cases). □

Proof of Theorem C Set $N_1 = \Sigma$ as in Theorem 5.2, $N_2 = S^n$ and $M = \mathbb{R}^{n+3}$ in Proposition 5.1. □

Given this result, several questions naturally arise.

Question 5.3 Given exotic n–spheres Σ and Σ', is $T_\infty\mathrm{Emb}^s(\Sigma,\Sigma')$ empty whenever Σ and Σ' are not diffeomorphic?

The argument for Theorem C proves something stronger:

Corollary 5.4 *For Σ as in Theorem C, the map*

$$\mathrm{Emb}^s(\Sigma, S^n) \to T_k\mathrm{Emb}^s(\Sigma, S^n)$$

is a weak equivalence for any $k \geq n-4$.

Proof By Theorem 2.3, the map $\mathrm{Emb}^s(\Sigma,\mathbb{R}^{n+3}) \to T_k\mathrm{Emb}^s(\Sigma,\mathbb{R}^{n+3})$ is a π_0–surjection when $k \geq n-4$, and similarly for S^n in place of Σ, so $T_k\mathrm{Emb}^s(\Sigma,\mathbb{R}^{n+3}) = \varnothing$ in this range, and the argument of Proposition 5.1 applies. □

Thus the $(n-4)^{\text{th}}$ stage of the embedding calculus Taylor tower can distinguish these exotic smooth structures. On the other hand, since the first stage is given by bundle maps between tangent bundles, the fact that exotic spheres have isomorphic tangent bundles shows that the first stage does not depend on the smooth structure of Σ. Thus, in the following question, k lies in the range $2 \leq k \leq n-4$.

Question 5.5 What is the smallest k such that $T_k\mathrm{Emb}^s(\Sigma, S^n) = \varnothing$?

The embedding calculus Taylor tower can be modeled geometrically in terms of stratified maps of bundles over compactified configuration spaces [6; 41]. Since the first stage of the tower is never empty in the case at hand, it follows that, in examples where $T_\infty\mathrm{Emb}^s(\Sigma, S^n) = \varnothing$, such a stratified map exists between compactified configuration spaces of $k-1$ points that does not extend to configurations of k points.

Question 5.6 Does the classification of exotic spheres admit an interpretation in terms of stratified obstruction theory applied to compactified configuration spaces?

5.2 Further examples

We indicate a few other exotic spheres for which the conclusion of Theorem C holds.

Example 5.7 The paper [2] studies the values of n and r for which the quotient of Θ_n, the group of oriented exotic spheres under connected sum, by the subgroup of oriented exotic spheres which embed in $\mathbb{R}^{n+r}$ with trivial normal bundle is nonzero. In particular, [2, Table 1] provides examples of exotic n–spheres in dimensions $n = 17, 18, 32, 33, 34, 37, 38$ which do not embed in $\mathbb{R}^{n+3}$.

Example 5.8 According to [30], the generators of Θ_n for $n = 8, 9, 10$ do not embed in $\mathbb{R}^{n+3}$.

In general, the homotopy-theoretic problem indicated by the proof of Theorem 5.2, which we believe to be of independent interest, remains open.

Question 5.9 Which elements of $\pi_{n-1}(\mathrm{SO}(n-3))$ lie in the common kernel of

$$i : \pi_{n-1}(\mathrm{SO}(n-3)) \to \pi_{n-1}(\mathrm{SO}) \quad \text{and} \quad J : \pi_{n-1}(\mathrm{SO}(n-3)) \to \pi_{2n-4}(S^{n-3})?$$

One can also vary the target in Theorem C.

Example 5.10 In [21, Theorem I] it is proven that an oriented exotic n–sphere Σ' embeds in $\mathbb{R}^{n+2}$ if and only if it represents an element of the subgroup $\mathrm{bP}_{n+1} \subset \Theta_n$ of oriented exotic n–spheres that bound a stably parallelizable $(n+1)$–manifold. In the proof of Theorem C, all we used about S^n is that it embeds in $\mathbb{R}^{n+3}$, so the same argument gives us that

$$T_\infty \mathrm{Emb}^s(\Sigma, \Sigma') = \varnothing$$

whenever Σ' represents an element of bP_{n+1} and Σ is as in Examples 5.7 and 5.8. (It is also true for Σ as in Theorem 5.2, but for even n the group bP_{n+1} is always trivial.)

6 Isotopy extension for embedding calculus

Fix manifolds M and N of equal dimension n, a compact smooth submanifold $P \subseteq N$ of codimension 0, and an embedding e of P in M. Even though P is not an object of $\mathcal{M}\mathrm{fld}^s$ we can still define the presheaf $\mathrm{Emb}^s(-, P)$, obtain a corresponding presheaf $\mathbb{E}^s_P$ on $\mathcal{D}\mathrm{isk}^s_n$, and define $T_\infty \mathrm{Emb}^s(P, M)$ to be the derived mapping space $\mathrm{Map}^h_{\mathcal{P}\mathrm{sh}(\mathcal{D}\mathrm{isk}^s_n)}(\mathbb{E}^s_P, \mathbb{E}^s_M)$ of presheaves on $\mathcal{D}\mathrm{isk}^s_n$. The goal of this section is to prove the following result:

Theorem 6.1 *Let M, N and P be as above. If $\mathrm{hdim}(P) \le \dim(M) - 3$ or $P = \bigsqcup_I D^n$ for some finite set I, then the diagram*

$$\begin{array}{ccc} T_\infty \mathrm{Emb}^s_\partial(N \setminus \mathring{P}, M \setminus \mathring{P}) & \longrightarrow & T_\infty \mathrm{Emb}^s(N, M) \\ \downarrow & & \downarrow \\ * & \longrightarrow & T_\infty \mathrm{Emb}^s(P, M) \end{array}$$

is homotopy Cartesian, where the bottom map is induced by the embedding e.

Removing the symbol T_∞ from the statement, one obtains the conclusion of the usual isotopy extension theorem [44, Chapter 6], an important tool in the study of spaces of embeddings and diffeomorphisms. Thus, Theorem 6.1 asserts that isotopy extension holds for limits of Taylor towers.

Remark 6.2 (i) We will see that the top horizontal map is the extension-by-identity map, as in Section 6.1.2.

(ii) In this theorem, two different incarnations of embedding calculus occur; the top left-hand corner uses the version for presheaves on $\mathcal{M}\mathrm{fld}^s_{\partial P}$, while the two right-hand corners use the version for presheaves on $\mathcal{M}\mathrm{fld}^s$.

(iii) Since P and $\mathring{P}$ are isotopy equivalent, the inclusion $\mathring{P} \to P$ induces a weak equivalence of presheaves $\mathrm{Emb}^s(-, \mathring{P}) \to \mathrm{Emb}^s(-, P)$, and thus a weak equivalence $T_\infty \mathrm{Emb}^s(P, M) \to T_\infty \mathrm{Emb}^s(\mathring{P}, M)$. Under the hypotheses of the theorem, the latter has the weak homotopy type of $\mathrm{Emb}^s(\mathring{P}, M)$ by Theorem 2.3.

(iv) A more technical hypothesis guaranteeing the conclusion of the theorem is that, for all $k \geq 0$, $\mathrm{Emb}^s\big(\mathring{P} \sqcup \bigsqcup_k D^n, M\big) \to T_\infty \mathrm{Emb}^s\big(\mathring{P} \sqcup \bigsqcup_k D^n, M\big)$ is a weak equivalence.

(v) Isotopy extension for embedding calculus generalizes to spaces of neat embeddings of manifolds with corners. Here the input is as follows: N and M are manifolds of equal dimension n with fixed embedding $\partial N \to \partial M$, and $P \subseteq N$ is a neatly embedded compact smooth submanifold of codimension 0 with corners whose boundary ∂P is the union of $\partial_0 P = \partial P \cap \partial N$ and a submanifold $\partial_1 P$, which meets at the subset of corners of P. Fixing a neat embedding $e \colon P \to N$ which is equal to the given embedding near $\partial_0 P$, we have the homotopy Cartesian square

$$\begin{array}{ccc} T_\infty \mathrm{Emb}^s_{\partial_1 P \cup \partial N \setminus \mathring{\partial}_0 P}(N \setminus \mathring{P}, M \setminus \mathring{P}) & \longrightarrow & T_\infty \mathrm{Emb}^s_{\partial N}(N, M) \\ \big\downarrow & & \big\downarrow \\ * & \longrightarrow & T_\infty \mathrm{Emb}^s_{\partial_0 P}(P, M) \end{array}$$

The argument is essentially the same as that given below, but with more involved notation.

6.1 Proof of Theorem 6.1

6.1.1 Complete Weiss covers We begin with a discussion of a well-known form of locality enjoyed by embedding calculus.

Definition 6.3 Let X be a topological space and $1 \leq k \leq \infty$. A collection of open subsets $\mathcal{U}$ of X is a *Weiss k–cover* if every finite subset of X with cardinality $\leq k$ is contained in some element of $\mathcal{U}$. A Weiss k–cover $\mathcal{U}$ is *complete* if it contains a Weiss k–cover of $\bigcap_{U \in \mathcal{U}_0} U$ for every finite subset $\mathcal{U}_0 \subseteq \mathcal{U}$.

The following result asserts that T_k has descent for complete Weiss k–covers. The intended application is to $k = \infty$ and $\mathbb{E}^s_{M,\partial}$.

Lemma 6.4 *Let N be a smooth manifold and $1 \leq k \leq \infty$. If F is a presheaf on $\mathcal{M}\mathrm{fld}_Z$ and $\mathcal{U}$ is a complete Weiss k–cover of N, each element of which contains ∂N, then the natural map*

$$T_k F(N) \to \operatorname*{holim}_{U \in \mathcal{U}} T_k F(U)$$

is a weak equivalence.

Proof Since derived mapping spaces convert homotopy colimits in the source to homotopy limits, it suffices to show that the natural map

$$\operatorname*{hocolim}_{U \in \mathcal{U}} \mathbb{E}^s_{U,\partial} \to \mathbb{E}^s_{N,\partial}$$

is a weak equivalence of presheaves on the full subcategory $\mathcal{D}\mathrm{isk}^s_{n,Z,\le k}$ whose objects are diffeomorphic to a disjoint union of $Z\times[0,1)$ and finitely many but at most k copies of $\mathbb{R}^n$. Since homotopy colimits of presheaves are computed pointwise, it suffices to check the corresponding claim for

$$\mathrm{Emb}^s_\partial\Big(Z\times[0,1)\sqcup\coprod_I\mathbb{R}^n,-\Big)$$

for every finite set I of cardinality $\le k$.

Assume first that $Z=\varnothing$. Given a configuration $\{p_i\}_{i\in I}\in\mathrm{Conf}_I(N)$ to serve as a basepoint, consider the commuting diagram

$$\begin{array}{ccc}
\prod_{i\in I}\mathrm{Emb}^s_{p_i}(\mathbb{R}^n,N) & \longrightarrow & \prod_{i\in I}\mathrm{Map}_{\mathcal{V}\mathrm{ec},p_i}(T\mathbb{R}^n,TN)\\
\downarrow & & \downarrow\\
\mathrm{Emb}^s\big(\coprod_I\mathbb{R}^n,N\big) & \longrightarrow & E\\
\downarrow & & \downarrow\\
\mathrm{Conf}_I(N) & = & \mathrm{Conf}_I(N)
\end{array}$$

where $E=\mathrm{Map}_{\mathcal{V}\mathrm{ec}}(T\mathbb{R}^n,TN)^I|_{\mathrm{Conf}_I(N)}$. As in the proof of Lemma 3.16, the vertical columns are fibration sequences and the top map is a weak equivalence, so the middle map is so. The same remarks apply after replacing N by U. The claim follows upon observing that the natural map

$$\operatorname*{hocolim}_{U\in\mathcal{U}}E|_U\to E$$

is a weak equivalence by [10, Proposition 4.6], since the collection $\{\mathrm{Conf}_I(U)\}_{U\in\mathcal{U}}$ is a complete cover of $\mathrm{Conf}_I(N)$ in the sense of [10, Definition 4.5].

In the general case, consider the commuting diagram

$$\begin{array}{ccc}
\operatorname*{hocolim}_{U\in\mathcal{U}}\mathrm{Emb}^s_\partial\big(Z\times[0,1)\sqcup\coprod_I\mathbb{R}^n,U\big) & \longrightarrow & \mathrm{Emb}^s_\partial\big(Z\times[0,1)\sqcup\coprod_I\mathbb{R}^n,N\big)\\
\downarrow & & \downarrow\\
\operatorname*{hocolim}_{U\in\mathcal{U}}\mathrm{Emb}^s\big(\coprod_I\mathbb{R}^n,\mathring{U}\big) & \longrightarrow & \mathrm{Emb}^s\big(\coprod_I\mathbb{R}^n,\mathring{N}\big)
\end{array}$$

where the vertical arrows are induced by restriction. Since the collection $\{\mathring{U}\}_{U\in\mathcal{U}}$ is a complete Weiss k–cover of $\mathring{N}$, the bottom arrow is a weak equivalence by the previous case. Since $\mathrm{Emb}^s_\partial(Z\times[0,1),N)$ is contractible and N is isotopy equivalent to its interior, isotopy extension implies that the right-hand map is an equivalence, and the same considerations applied to U show that the left-hand map is as well, implying the claim. □

Remark 6.5 The map $F\to T_kF$ can be described as homotopy sheafification with respect to Weiss k–covers [5, Theorem 1.2].

6.1.2 Extension-by-identity maps Suppose M, N and P are manifolds with a common boundary Z. Then we can form the manifolds $M \cup_\partial P$ and $N \cup_\partial P$, and construct an extension-by-identity map

$$\mathrm{Emb}^s_\partial(N, M) \to \mathrm{Emb}^s(N \cup_\partial P, M \cup_\partial P).$$

Lemma 6.6 *There is a dashed map making the following diagram commute up to preferred homotopy*:

$$\begin{array}{ccc} \mathrm{Emb}^s_\partial(N, M) & \longrightarrow & \mathrm{Emb}^s(N \cup_\partial P, M \cup_\partial P) \\ \downarrow & & \downarrow \\ T_\infty \mathrm{Emb}^s_\partial(N, M) & \dashrightarrow & T_\infty \mathrm{Emb}^s(N \cup_\partial P, M \cup_\partial P) \end{array}$$

Proof Consider the map

$$\mathrm{Emb}^s_\partial(-, M) \to \mathrm{Emb}^s(- \cup_\partial P, M \cup_{\partial P} P)$$

of presheaves on $\mathcal{D}\mathrm{isk}^s_{n,Z}$ induced by extension-by-identity, postcomposed with

$$\mathrm{Emb}^s(- \cup_\partial P, M \cup_\partial P) \to T_\infty \mathrm{Emb}^s(- \cup_\partial P, M \cup_\partial P).$$

As the target satisfies descent for complete Weiss ∞–covers by construction, and $T_\infty \mathrm{Emb}^s_\partial(-, M)$ is the homotopy sheafification of $\mathrm{Emb}^s_\partial(-, M)$ with respect to Weiss ∞–covers by Remark 6.5, this factors essentially uniquely over $T_\infty \mathrm{Emb}^s_\partial(-, M)$. Evaluating at N, we get the desired diagram. □

6.1.3 Proof of Theorem 6.1 We proceed by applying Lemma 6.4 with $k = \infty$ and $F = \mathbb{E}^s_{M,\partial}$ to a convenient cover. Write $\mathcal{D}_{P \subset N}$ for the collection of open subsets U of N that are disjoint unions of a finite number of open balls in $N \setminus P$ together with a collar neighborhood of P. In other words, U is diffeomorphic, relative to P, to the manifold

$$(P \cup \partial P[0, 1)) \sqcup \bigsqcup_I \mathbb{R}^n$$

for some finite set I.

The reader is invited to check that $\mathcal{D}_{P \subset N}$ is a complete Weiss cover of N. This cover also has the following pleasant property:

Lemma 6.7 *The poset $\mathcal{D}_{P \subset N}$ is contractible.*

Proof Let $\mathcal{C}_{P \subset N} \subseteq \mathcal{D}_{P \subset N}$ denote the full subposet spanned by the objects, so that the inclusion $P \hookrightarrow U$ is 0–connected, ie an object of $\mathcal{C}_{P \subset N}$ is simply a collar neighborhood of P. A retraction and right adjoint to the inclusion of this subcategory is obtained by sending U to the component of U containing P. The claim now follows upon noting that $\mathcal{C}_{P \subset N}$ is contractible, being cofiltered. □

Remark 6.8 By adapting [31, Section 5.5.2], something much stronger can be shown, namely that $\mathcal{D}_{P \subset N}$ is final in a sifted ∞–category.

We now prove the isotopy extension theorem.

Proof of Theorem 6.1 Suppose first that $\mathrm{hdim}(P) \leq \dim(M)-3$. Restricting to $U \in \mathcal{D}_{P\subset N}$ induces the commuting diagram

$$\begin{array}{ccccc}
T_\infty \mathrm{Emb}^s_\partial(N\setminus \mathring{P}, M\setminus \mathring{P}) & \xrightarrow{(1)} & \operatorname*{holim}_{U\in\mathcal{D}_{P\subset N}} T_\infty\mathrm{Emb}^s_\partial(U\setminus \mathring{P}, M\setminus \mathring{P}) & \xleftarrow{(4)} & \operatorname*{holim}_{U\in\mathcal{D}_{P\subset N}} \mathrm{Emb}^s_\partial(U\setminus \mathring{P}, M\setminus \mathring{P}) \\
\downarrow & & \downarrow & & \downarrow \\
T_\infty\mathrm{Emb}^s(N,M) & \xrightarrow{(2)} & \operatorname*{holim}_{U\in\mathcal{D}_{P\subset N}} T_\infty\mathrm{Emb}^s(U,M) & \xleftarrow{(5)} & \operatorname*{holim}_{U\in\mathcal{D}_{P\subset N}} \mathrm{Emb}^s(U,M) \\
\downarrow & & \downarrow & & \downarrow \\
T_\infty\mathrm{Emb}^s(P,M) & \xrightarrow{(3)} & \operatorname*{holim}_{U\in\mathcal{D}_{P\subset N}} T_\infty\mathrm{Emb}^s(P,M) & \xleftarrow{(6)} & \operatorname*{holim}_{U\in\mathcal{D}_{P\subset N}} \mathrm{Emb}^s(P,M)
\end{array}$$

where the top vertical maps are given by extension-by-identity and the bottom vertical maps by restriction to P.

For each $U \in \mathcal{D}_{P\subset N}$, the rightmost column is a fibration sequence by the usual isotopy extension theorem. Since all have $e\colon P \hookrightarrow M$ as a basepoint, it remains a fibration sequence after taking homotopy limits. The claim will follow upon verifying that each of the numbered arrows is a weak equivalence. For the maps (1) and (2) this follows from Lemma 6.4 applied with $k=\infty$ and $F=\mathbb{E}^s_{M\setminus\mathring{P},\partial}$ or $F=\mathbb{E}^s_M$, respectively; for (3) from Lemma 6.7; for (5) and (6) from Theorem 2.3 and our assumption on P; and for (4) from the Yoneda lemma.

The only modification in the case $P=\bigsqcup_I \mathbb{R}^n$ is for the sixth arrow, which is now an equivalence by the Yoneda lemma. □

6.2 Applications of isotopy extension

We now give some applications of Theorem 6.1.

6.2.1 Rephrasing Question 5.3 Let Σ and Σ' be exotic n–spheres. Fixing disks $D^n \subseteq \Sigma, \Sigma'$, we write $D_\Sigma := \Sigma\setminus \mathring{D}^n$ for the corresponding exotic disk with boundary identified with ∂D^n, and similarly for $D_{\Sigma'}$.

Corollary 6.9 *There is a fibration sequence*

$$T_\infty\mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'}) \to T_\infty\mathrm{Emb}^s(\Sigma,\Sigma') \to O(n+1)$$

with fiber taken over the identity.

Proof We apply Theorem 6.1 with $N=\Sigma$, $M=\Sigma'$ and $P=D^n$. The tangent bundle of an exotic sphere is isomorphic to that of the standard sphere (a well-known consequence of [9, Proposition 5.4(iv)]), so $\mathrm{Emb}^s(D^n,\Sigma')$ is weakly equivalent to the orthogonal frame bundle of TS^n, which is homeomorphic to $O(n+1)$. □

To connect to results about the groups Θ_n, we consider a version of Question 5.3 for oriented exotic n–spheres and orientation-preserving embeddings. This question is essentially equivalent: given two oriented exotic n–spheres Σ and Σ', then $T_\infty \mathrm{Emb}^s(\Sigma, \Sigma')$ contains an element which reverses orientation (this is well defined since T_∞ maps to T_1 via bundle maps) if and only if $T_\infty \mathrm{Emb}^{s,+}(\Sigma, \overline{\Sigma}') \neq \varnothing$, where $\overline{\Sigma}'$ denotes Σ' with opposite orientation. As before, we use a superscript $+$ to denote orientation-preserving embeddings.

Corollary 6.10 *Let Σ and Σ' be oriented exotic n–spheres. Then $T_\infty \mathrm{Emb}^{s,+}(\Sigma, \Sigma')$ is nonempty if and only if $T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'})$ is nonempty.*

Proof This follows directly from the oriented version of the fibration sequence in Corollary 6.9:

$$T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'}) \to T_\infty \mathrm{Emb}^{s,+}(\Sigma, \Sigma') \to \mathrm{SO}(n+1). \qquad \square$$

Let us define a relation on Θ_n by saying

$$[\Sigma] \sim_{T_\infty} [\Sigma'] \iff T_\infty \mathrm{Emb}^{s,+}(\Sigma, \Sigma') \neq \varnothing.$$

Lemma 6.11 *This is an equivalence relation, and is compatible with addition on Θ_n.*

Proof It is easy to see it is reflexive and transitive, so we prove it is symmetric. To do so, we claim that $T_\infty \mathrm{Emb}^{s,+}(\Sigma, \Sigma') \neq \varnothing$ if and only if $T_\infty \mathrm{Emb}^{s,+}(\Sigma \,\#\, \overline{\Sigma}', S^n) \neq \varnothing$. Using the previous corollary, the statement is equivalent to

$$T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'}) \neq \varnothing \iff T_\infty \mathrm{Emb}^s_\partial(D^n_{\Sigma \# \overline{\Sigma}'}, D^n) \neq \varnothing.$$

This follows from the fact that the operation of boundary connected sum with $D^n_{\overline{\Sigma}'}$, which is an instance of extension-by-identity, induces a map

$$T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'}) \to T_\infty \mathrm{Emb}^s_\partial(D^n_{\Sigma \# \overline{\Sigma}'}, D^n)$$

with homotopy inverse given by the boundary connected sum with $D^n_{\Sigma'}$. For symmetry we use that, by reversing orientations on both the domain and target, $T_\infty \mathrm{Emb}^{s,+}(\Sigma \,\#\, \overline{\Sigma}', S^n) \neq \varnothing$ if and only if $T_\infty \mathrm{Emb}^{s,+}(\overline{\Sigma} \,\#\, \Sigma', \overline{S^n}) \neq \varnothing$, and that S^n has an orientation-reversing self-diffeomorphism.

We now prove $\sim_{T_\infty}$ is compatible with the addition in Θ_n. By taking the boundary connected sum with $D^n_{\Sigma''}$ or $D^n_{\overline{\Sigma}''}$, we obtain that $T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'}) \neq \varnothing$ if and only if $T_\infty \mathrm{Emb}^s_\partial(D^n_{\Sigma \# \Sigma''}, D^n_{\Sigma' \# \Sigma''}) \neq \varnothing$, so

$$[\Sigma] \sim_{T_\infty} [\Sigma'] \iff [\Sigma] + [\Sigma''] \sim_{T_\infty} [\Sigma'] + [\Sigma'']. \qquad \square$$

Example 6.12 For Σ as in Theorem C, $T_\infty \mathrm{Emb}^s(S^n, \Sigma) = \varnothing$.

Example 6.13 The subset $\{[\Sigma] \in \Theta_n \mid [\Sigma] \sim_{T_\infty} [S^n]\}$ is a subgroup.

The results of [6] shed some light on the space $T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'})$. Their statement involves the operad $\mathbb{E}_n$ of little n–disks and its derived automorphisms.

Proposition 6.14 *There is a fibration sequence*

$$T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'}) \to X \to X'$$

with X an $\Omega^n\, O(n)$–torsor and X' an $\Omega^n \mathrm{Aut}^h(\mathbb{E}_n)$–torsor with preferred basepoint.

Proof According to [6, Theorem 1.1] (with modifications for manifolds with boundary as in [6, Section 6]), there is a homotopy Cartesian square

$$\begin{array}{ccc} T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n_{\Sigma'}) & \longrightarrow & Y \\ \downarrow & & \downarrow \\ \mathrm{Map}_{\mathcal{V}\mathrm{ec},\partial}(TD^n_\Sigma, TD^n_{\Sigma'}) & \longrightarrow & Y' \end{array}$$

where Y is contractible [6, Theorem 1.4] and Y' is a mapping space between certain "local configuration categories." We require only two pieces of information about Y': it is the space of compactly supported sections of a bundle over D^n_Σ, and the fibers are weakly equivalent to $\mathrm{Aut}^h(\mathbb{E}_n)$ by [6, Theorem 1.2]. These facts give the identification of the right-hand term, and the identification of the middle term follows from the aforementioned fact about tangent bundles of exotic spheres. □

The action of $O(n)$ on the little n–disks operad by rotation gives a map $O(n) \to \mathrm{Aut}^h(\mathbb{E}_n)$. We do not know much about its effect on homotopy groups. Nevertheless, using our results on exotic spheres, we can say the following:

Corollary 6.15 *The map $O(n) \to \mathrm{Aut}^h(\mathbb{E}_n)$ is not surjective on π_n when $n = 2^j$ with $j \geq 3$.*

Proof Let Σ be an exotic n–sphere as in Theorem 5.2. Looping the map $O(n) \to \mathrm{Aut}^h(\mathbb{E}_n)$, we obtain a map $\Omega^n O(n) \to \Omega^n \mathrm{Aut}^h(\mathbb{E}_n)$, and the torsor structures on the domain and target of $X \to X'$ are compatible with this. If the map $O(n) \to \mathrm{Aut}^h(\mathbb{E}_n)$ were surjective on π_n, Proposition 6.14 would imply that $X \to X'$ is surjective on path components, and hence $T_\infty \mathrm{Emb}^s_\partial(D^n_\Sigma, D^n) \neq \varnothing$. Corollary 6.10 then implies a contradiction of Theorem 5.2. □

Remark 6.16 The map in question *is* injective on π_n, at least when n is sufficiently large. Restricting to the $(n-1)$–sphere of binary operations in $\mathbb{E}_n$ and suspending produces the right-hand map in $O(n) \to \mathrm{Aut}^h(\mathbb{E}_n) \to \mathrm{Aut}^h_*(S^n)$ whose composite is the unstable J–homomorphism, which is injective on π_n for $n \geq 40$ [33].

6.2.2 Morlet's theorem for T_∞ Setting $\Sigma = \Sigma' = S^n$ we draw the following conclusion, with $\mathrm{Aut}^h(\mathbb{E}_n)/O(n)$ notation for the homotopy fiber of $BO(n) \to B\mathrm{Aut}^h(\mathbb{E}_n)$.

Corollary 6.17 *There are weak equivalences*

$$T_\infty \mathrm{Diff}_\partial(D^n) \simeq \Omega^{n+1}\mathrm{Aut}^h(\mathbb{E}_n)/O(n) \quad \textit{and} \quad T_\infty \mathrm{Diff}(S^n) \simeq O(n+1) \times \Omega^{n+1}\mathrm{Aut}^h(\mathbb{E}_n)/O(n).$$

Proof When $\Sigma = \Sigma' = S^n$, we have $D^n_\Sigma = D^n_{\Sigma'} = D^n$. In the fibration sequence

$$T_\infty \mathrm{Emb}^s_\partial(D^n, D^n) \to \Omega^n O(n) \to \Omega^n \mathrm{Aut}^h(\mathbb{E}_n)$$

from Proposition 6.14, the basepoint is provided by the constant map at the identity. So $T_\infty \mathrm{Emb}^s_\partial(D^n, D^n)$ is the fiber of a map of n–fold loop spaces over the unit, and hence it is grouplike. This implies that $T_\infty \mathrm{Diff}_\partial(D^n) = T_\infty \mathrm{Emb}^s_\partial(D^n, D^n)$, and the first claim follows. The second claim then follows from Corollary 6.9, using the splitting provided by the natural action of $O(n+1)$ on S^n. □

This result is to be compared to the classical theorem of Morlet, which asserts the same conclusion with T_∞ removed and $\mathrm{Aut}^h(\mathbb{E}_n)$ replaced by $\mathrm{Top}(n)$; see eg [8, Theorem 4.4(b); 22, Essay V]. Unlike Morlet's theorem, our results are valid even for $n = 4$.

Example 6.18 Since $\mathrm{Aut}(\mathbb{E}_2) \simeq O(2)$ [17, Theorem 8.5], we conclude that $\mathrm{Diff}(S^2) \to T_\infty \mathrm{Diff}(S^2)$ is a weak equivalence, furnishing another example of convergence in codimension 0. In fact, embedding calculus always converges for diffeomorphisms of surfaces, by [26, Theorem A].

6.2.3 Rephrasing the Weiss fibration sequence Consider a manifold M with $\partial M = S^{n-1}$ and disc $D^n \subset M$ such that $\partial M \cap D^n = D^{n-1} \subset \partial D^n$; that is, the disk meets the boundary of M in half its boundary. Then there is a fibration sequence which — informally speaking — describes $\mathrm{Diff}_\partial(M)$ as built from $\mathrm{Diff}_\partial(D^n)$ and a certain space of self-embeddings of M [28, Section 4; 47, Remark 2.1.2]. We will use Theorem 6.1 to reformulate this result.

Let $T_\infty \mathrm{Diff}^{\cong}_\partial(M) \subseteq T_\infty \mathrm{Diff}_\partial(M)$ denote the union of the path components lying in the image of $\mathrm{Diff}_\partial(M)$. The following result asserts that, with suitable assumptions on M, the homotopy fiber

$$M \mapsto \mathrm{hofiber}[B\mathrm{Diff}_\partial(M) \to BT_\infty \mathrm{Diff}^{\cong}_\partial(M)],$$

which we think of as the "error term" involved in applying embedding calculus to diffeomorphisms, is independent of M.

Corollary 6.19 *Let M be a 2–connected compact smooth manifold of dimension $n \geq 6$ with $\partial M = S^{n-1}$. The diagram*

$$\begin{array}{ccc} B\mathrm{Diff}_\partial(D^n) & \longrightarrow & B\mathrm{Diff}_\partial(M) \\ \downarrow & & \downarrow \\ BT_\infty \mathrm{Diff}^{\cong}_\partial(D^n) & \longrightarrow & BT_\infty \mathrm{Diff}^{\cong}_\partial(M) \end{array}$$

is homotopy Cartesian.

Proof Fix an embedded closed disk $D^{n-1} \subseteq \partial M$, and let $\mathrm{Emb}^s_{\partial/2}(M)$ denote the simplicial monoid of self-embeddings of M fixing D^{n-1} pointwise. There is the grouplike submonoid $\mathrm{Emb}^{s,\cong}_{\partial/2}(M) \subseteq \mathrm{Emb}^s_{\partial/2}(M)$ given by the union of the path components lying in the image of $\mathrm{Diff}_\partial(M)$. By naturality properties of embedding calculus (see [27, Sections 3 and 4] for a detailed proof of these), the diagram

$$\begin{array}{ccccc} B\mathrm{Diff}_\partial(D^n) & \longrightarrow & B\mathrm{Diff}_\partial(M) & \longrightarrow & B\mathrm{Emb}^{s,\cong}_{\partial/2}(M) \\ \downarrow & & \downarrow & & \downarrow \\ BT_\infty\mathrm{Diff}^{\cong}_\partial(D^n) & \longrightarrow & BT_\infty\mathrm{Diff}^{\cong}_\partial(M) & \longrightarrow & BT_\infty\mathrm{Emb}^{s,\cong}_{\partial/2}(M) \end{array}$$

commutes. In [28, Lemma 3.14] it is verified that M has handle dimension at most $n-3$ relative to D^{n-1}, so the right-hand vertical map is a weak equivalence (strictly speaking, to apply embedding calculus as discussed above we must remove the complement of D^{n-1} in S^{n-1}, which gives homotopy equivalent spaces). By isotopy extension (see [28, Theorem 4.17; 47, Remark 2.1.2]) the top row is a fibration sequence, and the bottom row is a fibration sequence by Theorem 6.1 (using the extension explained in Remark 6.2(v)). □

6.2.4 An example of convergence in handle codimension 2 We finish with an example of the convergence of the embedding calculus Taylor tower in handle codimension 2. For the sake of readability, we omit some details regarding boundary conditions; for example, strictly speaking, to apply embedding calculus as discussed above, one must remove parts of $S^2 = \partial D^3$ not in $\partial_0 D^1_{+,\epsilon}$.

Let $D^3 \subset \mathbb{R}^3$ be the closed unit disk, which contains the interval

$$D^1 = \{(x_1, 0, 0) \mid x_1 \in [-1, 1]\}$$

as a submanifold with boundary. We let

$$\mathbb{R}^3_+ := \{(x_1, x_2, x_3) \mid x_1 \geq 0\}$$

denote the half-plane and set $D^1_+ := D^1 \cap \mathbb{R}^3_+$. This is a manifold with boundary given by the union of the two points $\partial_0 D^1_+ = \{(0,0,0)\} = D^0$ and $\partial_1 D^1_+ := \{(1,0,0)\} = D^1_+ \cap S^2$.

The situation we will be interested in is obtained by "thickening" to codimension 0 the following simpler situation. By isotopy extension, there is a fibration sequence

$$\mathrm{Emb}^s_\partial(D^1_+, D^3) \to \mathrm{Emb}^s_{\partial_1}(D^1_+, D^3) \to \mathrm{Emb}^s(D^0, D^3),$$

where the fiber is taken over the inclusion. As the middle term is contractible, we obtain the weak equivalence $\mathrm{Emb}^s_\partial(D^1_+, D^3) \xrightarrow{\sim} \Omega\mathrm{Emb}^s(D^0, D^3) \simeq *$, a space-level version of the light bulb trick.

We now "thicken" all the submanifolds involved to codimension 0. Fixing a small $\epsilon > 0$, we replace D^0 by D^3_ϵ and D^1_+ by the union of D^3_ϵ with a closed $\frac{1}{2}\epsilon$–neighborhood of $D^1_{+,\epsilon}$ in D^3. We let C denote the closure of $D^1_{+,\epsilon} \setminus \mathring{D}^3_\epsilon$ in $D^1_{+,\epsilon}$, essentially a cylinder. Its boundary intersects the larger sphere

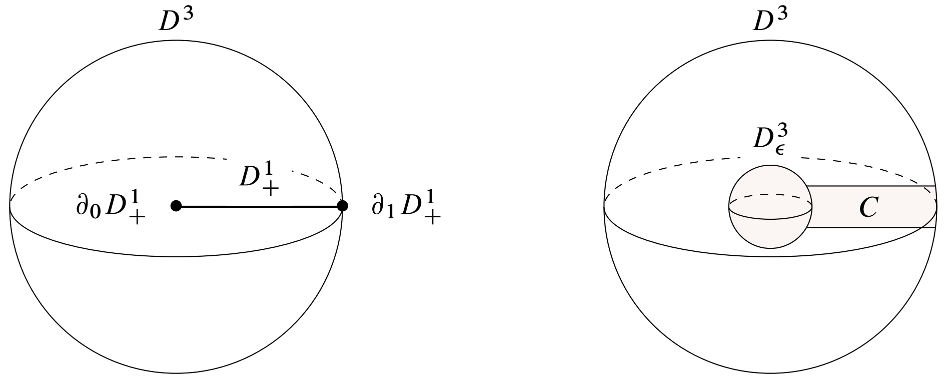


Figure 1: Left: the subspaces of D^3 involved in the earlier part of Section 6.2.4. Right: the subspaces of D^3 involved in the latter part of Section 6.2.4. The shaded region is $D^1_{+,\epsilon}$.

in $\partial_0 D^1_{+,\epsilon} := D^1_{+,\epsilon} \cap S^2$ and the smaller sphere in $\partial_1 D^1_{+,\epsilon} := D^1_{+,\epsilon} \cap S^2_\epsilon \cap \mathbb{R}^3_+$. As before, isotopy extension produces a fibration sequence with contractible middle term, whence the weak equivalence

$$\mathrm{Emb}^s_{\partial_0 \cup \partial_1}(C, D^3 \setminus \mathring{D}^3_\epsilon) \xrightarrow{\sim} \Omega \mathrm{Emb}^s(D^3_\epsilon, D^3) \simeq \Omega O(3).$$

We now show that embedding calculus captures this homotopy type; specifically, the left-hand vertical map is a weak equivalence in the commuting diagram

$$\begin{array}{ccccc} \mathrm{Emb}^s_{\partial_0 \cup \partial_1}(C, D^3 \setminus \mathring{D}^3_\epsilon) & \longrightarrow & \mathrm{Emb}^s_{\partial_1}(D^1_{+,\epsilon}, D^3) & \longrightarrow & \mathrm{Emb}^s(D^3_\epsilon, D^3) \\ \downarrow & & \downarrow & & \downarrow \\ T_\infty \mathrm{Emb}^s_{\partial_0 \cup \partial_1}(C, D^3 \setminus \mathring{D}^3_\epsilon) & \longrightarrow & T_\infty \mathrm{Emb}^s_{\partial_1}(D^1_{+,\epsilon}, D^3) & \longrightarrow & T_\infty \mathrm{Emb}^s(D^3_\epsilon, D^3) \end{array}$$

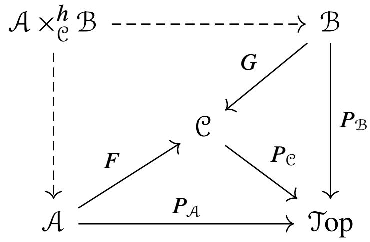
giving an example of convergence in codimension 2. Since D^3 has handle dimension 0, isotopy extension for embedding calculus — or, rather, the extension to neat embeddings of manifolds with corners — implies that the bottom row is also a fibration sequence, so it suffices to show that the middle and right-hand vertical maps are weak equivalences, both of which follow from the Yoneda lemma. For the latter map, we use that the inclusion of the interior D^3_ϵ induces a weak equivalence $T_\infty \mathrm{Emb}^s(D^3_\epsilon, D^3) \simeq T_\infty \mathrm{Emb}^s(\mathring{D}^3_\epsilon, D^3)$. For the former, we may similarly replace the source in $T_\infty \mathrm{Emb}^s(D^1_{+,\epsilon}, D^3)$ with an open collar on $\partial_1 D^1_{+,\epsilon}$.

Remark 6.20 These results generalize from dimension 3 to arbitrary dimension $n \geq 3$ by changing notation; this says that embedding calculus converges in codimension 2 for embeddings of $D^{n-3} \times C$ in $D^{n-3} \times (D^3 \setminus \mathring{D}^3_\epsilon)$.

Appendix Homotopy pullbacks of simplicial categories

In this appendix, we discuss a simplicial variant of a construction introduced in [1, Section 9] for topological categories.

Suppose given the solid commuting diagram of simplicial categories



where $\mathcal{T}\mathrm{op}$ denotes the simplicial category of topological spaces. Via the structure functors to $\mathcal{T}\mathrm{op}$, objects and morphisms in $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ have underlying spaces and maps.

Construction A.1 We define a simplicial category $\mathcal{A}\times^h_{\mathcal{C}}\mathcal{B}$ as follows:

(i) The objects of $\mathcal{A}\times^h_{\mathcal{C}}\mathcal{B}$ are triples (A, B, f), where $A \in \mathcal{A}$ and $B \in \mathcal{B}$ are objects with the same underlying space, and $f : F(A) \to G(B)$ is an isomorphism with underlying map the identity.

(ii) An n–simplex in the mapping space from (A_1, B_1, f_1) to (A_2, B_2, f_2) is a triple (φ, ψ, γ), where $\varphi \in \mathrm{Map}_{\mathcal{A}}(A_1, A_2)_n$ and $\psi \in \mathrm{Map}_{\mathcal{B}}(B_1, B_2)_n$ have the same underlying simplex in $\mathcal{T}\mathrm{op}$, and γ is a path $f_2 \circ F(\varphi) \Rightarrow G(\psi) \circ f_1$ in $(\mathrm{Map}_{\mathcal{C}}(F(A_1), G(B_2))^{\Delta^1})_n$ covering the constant path.

(iii) Composition is induced by composition in $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, and the diagonal of Δ^1.

The notation $\mathcal{A}\times^h_{\mathcal{C}}\mathcal{B}$ is justified by the following result, whose proof we defer to the end of this subsection and may be skipped on a first reading.

Proposition A.2 *Suppose that*

(i) *the simplicial sets* $\mathrm{Map}_{\mathcal{B}}(B_1, B_2)$ *and* $\mathrm{Map}_{\mathcal{C}}(F(A_1), G(B_2))$ *are Kan complexes, and*

(ii) *the structure maps* $\mathrm{Map}_{\mathcal{B}}(B_1, B_2) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(P_{\mathcal{B}}(B_1), P_{\mathcal{B}}(B_2))$ *and* $\mathrm{Map}_{\mathcal{C}}(F(A_1), G(B_2)) \to \mathrm{Map}_{\mathcal{T}\mathrm{op}}(P_{\mathcal{A}}(A_1), P_{\mathcal{B}}(B_2))$ *are Kan fibrations.*

The diagram

$$\begin{array}{ccc} \mathrm{Map}_{\mathcal{A}\times^h_{\mathcal{C}}\mathcal{B}}((A_1, B_1, f_1), (A_2, B_2, f_2)) & \longrightarrow & \mathrm{Map}_{\mathcal{B}}(B_1, B_2,) \\ \downarrow & & \downarrow \\ \mathrm{Map}_{\mathcal{A}}(A_1, A_2) & \longrightarrow & \mathrm{Map}_{\mathcal{C}}(F(A_1), G(B_2)) \end{array}$$

is homotopy Cartesian.

Note that the diagram in question commutes only up to specified homotopy.

Remark A.3 Proposition A.2 implies that $\mathcal{A}\times^h_{\mathcal{C}}\mathcal{B}$ is often the homotopy pullback of $\mathcal{A}$ and $\mathcal{B}$ over $\mathcal{C}$ in the Bergner model structure on simplicial categories [4]; specifically, we require the assumptions of the proposition to hold for all objects, and we require that $\mathrm{Ho}(P_{\mathcal{B}})$ and $\mathrm{Ho}(P_{\mathcal{C}})$ be isofibrations. Therefore, we think of $\mathcal{A} \times^h_{\mathcal{C}} \mathcal{B}$ as a (particularly convenient) model for the pullback of ∞–categories, whose homotopy theory is captured by the Bergner model structure.

Construction A.4 Suppose $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are as above. Let $\mathcal{D}$ be a simplicial category equipped with simplicial functors $H: \mathcal{D} \to \mathcal{A}$ and $K: \mathcal{D} \to \mathcal{B}$ over $\mathcal{T}\mathrm{op}$, together with the natural isomorphism χ in the diagram

$$\begin{array}{ccc} \mathcal{D} & \xrightarrow{K} & \mathcal{B} \\ {\scriptstyle H}\downarrow & \overset{\chi}{\nearrow} & \downarrow{\scriptstyle G} \\ \mathcal{A} & \xrightarrow[F]{} & \mathcal{C} \end{array}$$

We obtain a functor $\mathcal{D} \to \mathcal{A}_2 \times^h_{\mathcal{C}_2} \mathcal{B}_2$ as follows:

(i) The object $D \in \mathcal{D}$ is sent to the triple $(H(D), K(D), \chi_D)$.

(ii) The n–simplex $\sigma \in \mathrm{Map}_{\mathcal{D}}(D_1, D_2)$ is sent to the triple consisting of $H(\sigma)$, $K(\sigma)$ and the constant path at $\chi_{D_2} \circ H(\sigma) = K(\sigma) \circ \chi_{D_1}$.

To prove Proposition A.2, it will be convenient to put ourselves in a more general setting. Suppose we have the commutative diagram of simplicial sets

$$\begin{array}{ccccc} & & X & & \\ & & {\scriptstyle g}\downarrow & \overset{p_X}{\searrow} & \\ Z & \xrightarrow{h} & Y & & \\ & \underset{p_Z}{\searrow} & & \overset{p_Y}{\searrow} & \\ & & & & W \end{array}$$

Write P for the standard model of the homotopy pullback of X and Z over Y; explicitly, P is the limit of the diagram

$$\begin{array}{ccccccc} X & & & Y^{\Delta^1} & & & Z \\ & \overset{g}{\searrow} & \overset{\mathrm{ev}_0}{\swarrow} & & \overset{\mathrm{ev}_1}{\searrow} & \overset{h}{\swarrow} & \\ & & Y & & Y & & \end{array}$$

Finally, write P_0 for the pullback in the diagram

$$\begin{array}{ccc} P_0 & \xrightarrow{\iota} & P \\ \downarrow & & \downarrow{\scriptstyle q} \\ W & \longrightarrow & W^{\Delta^1} \end{array}$$

where the bottom arrow is the inclusion of the constant maps and q is the composition of the projection to Y^{Δ^1} with $(p_Y)^{\Delta^1}$. We think of P_0 as the subspace of the homotopy pullback lying over constant paths in W. In the example of interest, X, Y and Z are mapping spaces in the relevant simplicial categories, and W is the corresponding mapping space in $\mathcal{T}\mathrm{op}$.

The topological analog of the following result is asserted in [1, Section 9]. We include a proof for the sake of completeness.

Lemma A.5 *If p_Y and p_Z are fibrations, then ι is a weak equivalence.*

Proof Given the solid commuting diagram

$$\begin{array}{ccc} \partial\Delta^n & \longrightarrow & P_0 \\ \downarrow & \nearrow & \downarrow{\scriptstyle \iota} \\ \Delta^n & \longrightarrow & P \end{array}$$

we will produce the dashed arrow making the top triangle commute and the bottom triangle commute up to homotopy fixing $\partial\Delta^n$. First, using the assumption that p_Z is a fibration, we solve the lifting problem

$$\begin{array}{ccccc} \Delta^n \times \Delta^0 \sqcup_{\partial\Delta^n \times \Delta^0} \partial\Delta^n \times \Delta^1 & \longrightarrow & P & \longrightarrow & Z \\ \downarrow & & & \nearrow & \downarrow{\scriptstyle p_Z} \\ \Delta^n \times \Delta^1 & & \longrightarrow & & W \end{array}$$

where the bottom map is the adjunct of the composite $\Delta^n \to P \to Y^{\Delta^1} \to W^{\Delta^1}$, and the left-hand map is induced by the inclusion of the vertex 0. Composing with h and passing back through the adjunction, we obtain the top map in the commuting diagram

$$\begin{array}{ccccc} \Delta^n & & \longrightarrow & & Y^{\Delta^1} \\ \downarrow & & & & \downarrow{\scriptstyle \mathrm{ev}_0} \\ P & \longrightarrow & Y^{\Delta^1} & \xrightarrow{\mathrm{ev}_1} & Y \end{array}$$

There is an induced map $\Delta^n \times \Lambda^2_1 \to Y$, and we use the assumption that p_Y is a fibration to solve the lifting problem

$$\begin{array}{ccccc} \Delta^n \times \Lambda^2_1 \sqcup_{\partial\Delta^n \times \Lambda^2_1} \partial\Delta^n \times \Delta^2 & & \longrightarrow & & Y \\ \downarrow & & & \nearrow & \downarrow{\scriptstyle p_Y} \\ \Delta^n \times \Delta^2 & \longrightarrow & \Delta^n \times \Delta^1 & \longrightarrow & W \end{array}$$

Restricting to the third face of Δ^2, we obtain by adjunction the middle map in the commuting diagram

$$\begin{array}{ccccccc} & & & \Delta^n & & & \\ & \swarrow & & \downarrow & & \searrow & \\ X & & & Y^{\Delta^1} & & & Z \\ & {\scriptstyle g}\searrow & \swarrow{\scriptstyle \mathrm{ev}_0} & & {\scriptstyle \mathrm{ev}_1}\searrow & \swarrow{\scriptstyle h} & \\ & & Y & & Y & & \end{array}$$

where the left-hand map is the composite $\Delta^n \to P \to X$, and the right-hand map is the restriction of our earlier lift $\Delta^n \times \Delta^1 \to Z$ to the vertex 1. The resulting map $\Delta^n \to P$ factors through P_0 and restricts to the original map on $\partial\Delta^n$ by construction. Also by construction, the right-hand square of the above diagram comes equipped with a homotopy relative to $\partial\Delta^n$, which furnishes the desired homotopy. □

Proof of Proposition A.2 The first assumption guarantees that the standard model for the homotopy pullback has the correct weak equivalence type. The second assumption permits the invocation of Lemma A.5, which guarantees that the canonical map from $\mathrm{Map}_{\mathcal{A}\times^h_{\mathcal{C}}\mathcal{B}}((A_1, B_1, f_1), (A_2, B_2, f_2))$ to the standard model for the homotopy pullback is a weak equivalence. □

References

[1] **R Andrade**, *From manifolds to invariants of E_n–algebras*, preprint (2012) arXiv 1210.7909

[2] **P L Antonelli**, *On stable diffeomorphism of exotic spheres in the metastable range*, Canad. J. Math. 23 (1971) 579–587 MR Zbl

[3] **G Arone**, **M Szymik**, *Spaces of knotted circles and exotic smooth structures*, Canad. J. Math. 74 (2022) 1–23 MR Zbl

[4] **J E Bergner**, *A model category structure on the category of simplicial categories*, Trans. Amer. Math. Soc. 359 (2007) 2043–2058 MR Zbl

[5] **P Boavida de Brito**, **M Weiss**, *Manifold calculus and homotopy sheaves*, Homology Homotopy Appl. 15 (2013) 361–383 MR Zbl

[6] **P Boavida de Brito**, **M Weiss**, *Spaces of smooth embeddings and configuration categories*, J. Topol. 11 (2018) 65–143 MR Zbl

[7] **M Brown**, *A proof of the generalized Schoenflies theorem*, Bull. Amer. Math. Soc. 66 (1960) 74–76 MR Zbl

[8] **D Burghelea**, **R Lashof**, *The homotopy type of the space of diffeomorphisms, I*, Trans. Amer. Math. Soc. 196 (1974) 1–36 MR Zbl

[9] **D Burghelea**, **R Lashof**, *The homotopy type of the space of diffeomorphisms, II*, Trans. Amer. Math. Soc. 196 (1974) 37–50 MR Zbl

[10] **D Dugger**, **D C Isaksen**, *Topological hypercovers and $\mathbb{A}^1$–realizations*, Math. Z. 246 (2004) 667–689 MR Zbl

[11] **R D Edwards**, **R C Kirby**, *Deformations of spaces of imbeddings*, Ann. of Math. 93 (1971) 63–88 MR Zbl

[12] **M H Freedman**, **F Quinn**, *Topology of* 4*–manifolds*, Princeton Mathematical Series 39, Princeton Univ. Press (1990) MR Zbl

[13] **S Galatius**, **O Randal-Williams**, *Moduli spaces of manifolds*: *a user's guide*, from "Handbook of homotopy theory" (H Miller, editor), CRC, Boca Raton, FL (2020) 443–485 MR Zbl

[14] **T G Goodwillie**, **J R Klein**, *Multiple disjunction for spaces of smooth embeddings*, J. Topol. 8 (2015) 651–674 MR Zbl

[15] **T G Goodwillie**, **J R Klein**, **M S Weiss**, *A Haefliger style description of the embedding calculus tower*, Topology 42 (2003) 509–524 MR Zbl

[16] **T G Goodwillie**, **M Weiss**, *Embeddings from the point of view of immersion theory, II*, Geom. Topol. 3 (1999) 103–118 MR Zbl

[17] **G Horel**, *Profinite completion of operads and the Grothendieck–Teichmüller group*, Adv. Math. 321 (2017) 326–390 MR Zbl

[18] **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

[19] **D Husemoller**, *Fibre bundles*, 3rd edition, Graduate Texts in Math. 20, Springer (1994) MR Zbl

[20] **M A Kervaire**, *Some nonstable homotopy groups of Lie groups*, Illinois J. Math. 4 (1960) 161–169 MR Zbl

[21] **M A Kervaire**, *On higher dimensional knots*, from "Differential and combinatorial topology (a symposium in honor of Marston Morse)", Princeton Univ. Press (1965) 105–119 MR Zbl

[22] **R C Kirby**, **L C Siebenmann**, *Foundational essays on topological manifolds, smoothings, and triangulations*, Annals of Mathematics Studies 88, Princeton Univ. Press (1977) MR Zbl

[23] **J M Kister**, *Microbundles are fibre bundles*, Ann. of Math. 80 (1964) 190–199 MR Zbl

[24] **M Kontsevich**, *Feynman diagrams and low-dimensional topology*, from "First European Congress of Mathematics, II" (A Joseph, F Mignot, F Murat, B Prum, R Rentschler, editors), Progr. Math. 120, Birkhäuser, Basel (1994) 97–121 MR Zbl

[25] **A Körschgen**, *Dwyer–Kan equivalences induce equivalences on topologically enriched presheaves*, preprint (2017) arXiv 1704.07472

[26] **M Krannich**, **A Kupers**, *Embedding calculus for surfaces*, preprint (2021) arXiv 2101.07885 To appear in Algebr. Geom. Topol.

[27] **M Krannich**, **A Kupers**, *The $\mathcal{D}$isc–structure space*, preprint (2022) arXiv 2205.01755

[28] **A Kupers**, *Some finiteness results for groups of automorphisms of manifolds*, Geom. Topol. 23 (2019) 2277–2333 MR Zbl

[29] **A Kupers**, **O Randal-Williams**, *The cohomology of Torelli groups is algebraic*, Forum Math. Sigma 8 (2020) art. id. e64 MR Zbl

[30] **J Levine**, *A classification of differentiable knots*, Ann. of Math. 82 (1965) 15–50 MR Zbl

[31] **J Lurie**, *Higher algebra*, preprint (2017) Available at `https://url.msp.org/Lurie-HA`

[32] **M Mahowald**, *Some Whitehead products in S^n*, Topology 4 (1965) 17–26 MR Zbl

[33] **M Mahowald**, *The metastable homotopy of S^n*, Mem. Amer. Math. Soc. 72, Amer. Math. Soc., Providence, RI (1967) MR Zbl

[34] **M Mahowald**, *A new infinite family in ${}_2\pi_*{}^s$*, Topology 16 (1977) 249–256 MR Zbl

[35] **M Mahowald**, **R D Thompson**, *The EHP sequence and periodic homotopy*, from "Handbook of algebraic topology" (I M James, editor), North-Holland, Amsterdam (1995) 397–423 MR Zbl

[36] **W S Massey**, *On the normal bundle of a sphere imbedded in Euclidean space*, Proc. Amer. Math. Soc. 10 (1959) 959–964 MR Zbl

[37] **J Milnor**, *Microbundles, I*, Topology 3 (1964) 53–80 MR Zbl

[38] **N Prigge**, *On tautological classes of fibre bundles and self-embedding calculus*, PhD thesis, University of Cambridge (2020) Available at `https://www.repository.cam.ac.uk/handle/1810/317892`

[39] **A A Ranicki**, *On the Hauptvermutung*, from "The Hauptvermutung book", K–Monogr. Math. 1, Kluwer, Dordrecht (1996) 3–31 MR Zbl

[40] **L C Siebenmann**, *Deformation of homeomorphisms on stratified sets*, Comment. Math. Helv. 47 (1972) 123–163 MR Zbl

[41] **V Turchin**, *Context-free manifold calculus and the Fulton–MacPherson operad*, Algebr. Geom. Topol. 13 (2013) 1243–1271 MR Zbl

[42] **O Viro**, *Space of smooth* 1*–knots in a* 4*–manifold: is its algebraic topology sensitive to smooth structures?*, Arnold Math. J. 1 (2015) 83–89 MR Zbl

[43] **I Volić**, *Finite type knot invariants and the calculus of functors*, Compos. Math. 142 (2006) 222–250 MR Zbl

[44] **C T C Wall**, *Differential topology*, Cambridge Studies in Advanced Mathematics 156, Cambridge Univ. Press (2016) MR Zbl

[45] **T Watanabe**, *Some exotic nontrivial elements of the rational homotopy groups of* $\mathrm{Diff}(S^4)$, preprint (2018) arXiv 1812.02448

[46] **M Weiss**, *Embeddings from the point of view of immersion theory, I*, Geom. Topol. 3 (1999) 67–101 MR Zbl Correction in 15 (2011) 407–409

[47] **M S Weiss**, *Rational Pontryagin classes of Euclidean fiber bundles*, Geom. Topol. 25 (2021) 3351–3424 MR Zbl

Department of Mathematics, Northeastern University
Boston, MA, United States

Department of Mathematics, University of Toronto
Toronto, ON, Canada

b.knudsen@northeastern.edu, a.kupers@utoronto.ca

Proposed: Jesper Grodal
Seconded: Ulrike Tillmann, David Gabai

Received: 12 July 2021
Revised: 10 March 2022

Geometry & Topology **28**:1 (2024) 393–496
DOI: 10.2140/gt.2024.28.393
Published: 27 February 2024

Stable maps to Looijenga pairs

PIERRICK BOUSSEAU
ANDREA BRINI
MICHEL VAN GARREL

A log Calabi–Yau surface with maximal boundary, or Looijenga pair, is a pair (Y, D) with Y a smooth rational projective complex surface and $D = D_1 + \cdots + D_l \in |-K_Y|$ an anticanonical singular nodal curve. Under some natural conditions on the pair, we propose a series of correspondences relating five different classes of enumerative invariants attached to (Y, D):

(1) the log Gromov–Witten theory of the pair (Y, D),
(2) the Gromov–Witten theory of the total space of $\bigoplus_i \mathcal{O}_Y(-D_i)$,
(3) the open Gromov–Witten theory of special Lagrangians in a Calabi–Yau 3–fold determined by (Y, D),
(4) the Donaldson–Thomas theory of a symmetric quiver specified by (Y, D), and
(5) a class of BPS invariants considered in different contexts by Klemm and Pandharipande, Ionel and Parker, and Labastida, Mariño, Ooguri and Vafa.

We furthermore provide a complete closed-form solution to the calculation of all these invariants.

14J33, 14J81, 14N35, 16G20, 53D45

1 Introduction

1.1 Looijenga pairs

A log Calabi–Yau surface with maximal boundary, or Looijenga pair, is a pair $Y(D) := (Y, D)$ consisting of a smooth rational projective complex surface Y and an anticanonical singular nodal curve $D = D_1 + \cdots + D_l \in |-K_Y|$. A prototypical example of Looijenga pair is given by $(Y, D) = (\mathbb{P}^2, D_1 + D_2)$ for D_1 a line and D_2 a conic not tangent to D_1.

Looijenga pairs [79] were first systematically studied in relation with resolutions and deformations of elliptic surface singularities and with degenerations of K3 surfaces; see Friedman and Scattone [41]. More recently, Looijenga pairs have played an important role as two-dimensional examples for mirror symmetry; see Barrott [9], Bousseau [13], Gross, Hacking and Keel [53], Hacking and Keating [60], Mandel [81] and Yu [114; 115] and, for the theory of cluster varieties, Gross, Hacking and Keel [52], Mandel [82] and Zhou [117]. These new developments have had in return nontrivial applications to the classical geometry of Looijenga pairs; see Engel [38], Friedman [40] and Gross, Hacking and Keel [53; 54].

1.2 Summary of the main results

In this paper we develop a series of correspondences relating different enumerative invariants associated to a given Looijenga pair. We start off by giving a very succinct summary of the main objects we will consider, and the main statements we shall prove.

1.2.1 Geometries Let $(Y, D = D_1 + \cdots + D_l)$ be a Looijenga pair with $l \geq 2$. In this paper we will construct four different geometries out of (Y, D):

- the log Calabi–Yau surface geometry $Y(D)$;
- the local Calabi–Yau $(l+2)$–fold geometry $E_{Y(D)} := \mathrm{Tot}\big(\mathcal{O}_Y(-D_1) \oplus \cdots \oplus \mathcal{O}_Y(-D_l)\big)$;
- a noncompact Calabi–Yau threefold geometry canonically equipped with a disjoint union of $l-1$ Lagrangians,
$$Y^{\mathrm{op}}(D) := \big(\mathrm{Tot}\big(\mathcal{O}(-D_l) \to Y \setminus (D_1 \cup \cdots \cup D_{l-1})\big), L_1 \sqcup \cdots \sqcup L_{l-1}\big),$$
where L_i are fibred over real curves in D_i;
- for $l = 2$, a noncommutative geometry given by a symmetric quiver $\mathsf{Q}(Y(D))$ made from the combinatorial data of the divisors D_i and their intersections.

1.2.2 Enumerative theories Our main focus will be on the enumerative geometry of curves in these geometries. More precisely, to a Looijenga pair $Y(D)$ satisfying some natural positivity conditions, we shall associate several classes of a priori different enumerative invariants:

- **log GW** All genus log GW invariants of $Y(D)$, counting curves in the surface Y with maximal tangency conditions along the divisors D_i.
- **local GW** Genus-zero local GW invariants of the CY $(l+2)$–fold $E_{Y(D)}$.

- **open GW** All genus open GW invariants counting open Riemann surfaces in the CY3–fold $Y^{\mathrm{op}}(D)$ with $l-1$ boundary components mapping to $L_1 \sqcup \cdots \sqcup L_{l-1}$.
- **local BPS** Genus-zero local BPS invariants of $E_{Y(D)}$, in the form of Gopakumar–Vafa/Klemm–Pandharipande/Ionel–Parker (GV/KP/IP) BPS invariants.
- **open BPS** All genus open BPS invariants of $Y^{\mathrm{op}}(D)$, in the form of Labastida–Mariño–Ooguri–Vafa (LMOV) BPS invariants.
- **quiver DT** If $l=2$, Donaldson–Thomas (DT) invariants of $\mathsf{Q}(Y(D))$.

1.2.3 Correspondences Under some positivity conditions on (Y, D), we will prove that the invariants above essentially coincide with one another. In particular, we shall show

(i) an equality between log GW and local GW in genus zero (Theorem 1.4),

(ii) an equality between log GW and open GW in all genera (Theorem 1.5),

(iii) an equality between local BPS and open BPS in genus zero for all l,

(iv) an equality between local BPS and quiver DT for $l=2$, ie when the local geometry $E_{Y(D)}$ is CY_4 (Theorem 1.6).

The equality (i) establishes for log CY surfaces a version of a conjecture of van Garrel, Graber and Ruddat about log and local GW invariants [43], while (ii) and (iv) are new. Equality (iii) follows from (i)–(ii) after a BPS-type change of variables.

1.2.4 Integrality Furthermore, we shall prove that the enumerative invariants of Looijenga pairs considered in this paper obey strong integrality constraints (Theorem 1.7), reflecting the conjectured integrality of the open BPS and local BPS counts. This shows the existence of novel integral structures underlying the higher-genus log GW theory of $Y(D)$. Restricting to genus zero, we will obtain as a corollary an algebrogeometric proof of the conjectured integrality of the genus-zero Gopakumar–Vafa invariants of the CY $(l+2)$–fold $E_{Y(D)}$. In particular, for $l=2$, this proves for CY4 local surfaces an integrality conjecture of Klemm and Pandharipande [68, Conjecture 0].

1.2.5 Solutions Moreover, we will completely solve the enumerative counts for these geometries (Theorems 1.4 and 1.5), by finding explicit closed-form, nonrecursive expressions for the generating series of the invariants associated to our Looijenga pairs.

The rest of the introduction is organised as follows:

- Section 1.3 sets the stage by giving a self-contained account of the enumerative theories we shall consider.

- Section 1.4 illustrates the geometric picture underpinning the web of correspondences explored in the paper. We spell out the enumerative relations (i)–(iv) in the broadest generality where we believe them to hold, and describe in detail the geometric heuristics which led us to (i) in Section 1.4.1 (Conjecture 1.1), to (ii) in Section 1.4.2 (Conjectures 1.2 and 1.3), and to (iii)–(iv) in Section 1.4.3.
- Section 1.5 puts these ideas on a rigorous footing. We first place a natural positivity condition on the irreducible components D_i by requiring them to be all smooth and nef; depending on the context, we often supplement this with a mild condition of "quasi-tameness", whose rationale is justified in Sections 1.5.1 and 1.5.2. The statements of the proof of the correspondences, the integrality results, and the full solutions for our enumerative counts are spelled out in Theorems 1.4–1.7.
- Section 1.6 surveys the implications of our results for related work, with emphasis on the possible sheaf-theoretic interpretations of the BPS invariants we consider.

1.3 Enumerative problems

1.3.1 Higher-genus log Gromov–Witten invariants Log Gromov–Witten theory, which was developed by Abramovich and Chen [25; 1] and Gross and Siebert [58], provides a deformation-invariant way to count curves with prescribed tangency conditions along a normal crossings divisor, by virtual intersection theory on moduli spaces of stable log maps. For $Y(D)$ a Looijenga pair where D has $l \geq 2$ irreducible components, we consider rational curves in Y with given degree $d \in \mathrm{H}_2(Y,\mathbb{Z})$ that meet each component D_j in one point of maximal tangency $d \cdot D_j$ and pass through $l-1$ given points in Y. Counting such curves is an enumerative problem of expected dimension 0 and we denote by $N_{0,d}^{\log}(Y(D))$ the corresponding log Gromov–Witten invariants.

For $g \geq 0$, the expected dimension of the moduli space of genus g curves in Y with given degree $d \in \mathrm{H}_2(Y,\mathbb{Z})$ that meet each component D_j in one point of maximal tangency $d \cdot D_j$ and pass through $l-1$ given points in Y, is g. On the other hand, assigning to every stable log map $f: C \to Y(D)$ the vector space $\mathrm{H}^0(C,\omega_C)$ of sections of the dualising sheaf of the domain curve defines a rank g vector bundle over the moduli space, called the Hodge bundle, and we denote by λ_g its top Chern class. We define log Gromov–Witten invariants $N_{g,d}^{\log}(Y(D))$ by integration of $(-1)^g\lambda_g$ over the virtual fundamental class of the moduli space. For genus $g=0$, $N_{0,d}^{\log}(Y(D))$ recovers the naive count of rational curves but for $g>0$, the log Gromov–Witten invariants $N_{g,d}^{\log}(Y(D))$ no longer have an obvious interpretation in terms of naive enumeration of curves. Fixing the degree d and summing over all genera, we define generating series

$$\mathsf{N}_d^{\log}(Y(D))(\hbar) := \frac{1}{\left(2\sin\left(\frac{1}{2}\hbar\right)\right)^{l-2}} \sum_{g\geqslant 0} N_{g,d}^{\log}(Y(D))\hbar^{2g-2+l}. \tag{1-1}$$

The term $\left(2\sin\left(\frac{1}{2}\hbar\right)\right)^{2-l}$ is natural from the point of view of the q–refined scattering diagrams of Section 4. It is accounted for in the correspondence with the open invariants.

1.3.2 Local Gromov–Witten invariants To a Looijenga pair $Y(D) = (Y, D = D_1 + \dots + D_l)$, we associate the $(l+2)$–dimensional noncompact Calabi–Yau variety $E_{Y(D)} := \mathrm{Tot}\big(\bigoplus_{i=1}^{l}(\mathcal{O}_Y(-D_i))\big)$. We view Y in $E_{Y(D)}$ via the inclusion given by the zero section. We refer to $E_{Y(D)}$ as the local geometry attached to $Y(D)$. If each component D_i is nef, then for every $d \in \mathrm{H}_2(Y, \mathbb{Z})$ intersecting D_i generically, the moduli space of genus-zero stable maps to $E_{Y(D)}$ of degree d is compact: every stable map to $E_{Y(D)}$ of class d factors through the zero section Y. Thus, it makes sense to consider the local genus-zero Gromov–Witten invariants $N^{\mathrm{loc}}_{0,d}(Y(D))$, which define virtual counts of rational curves in $E_{Y(D)}$ passing through $l-1$ given points in Y.

1.3.3 Higher-genus open Gromov–Witten invariants Let X be a semiprojective toric Calabi–Yau 3–fold, ie a toric Calabi–Yau 3–fold which admits a presentation as the GIT quotient of a vector space by a torus action; see Hausel and Sturmfels [61]. We will be concerned with a class of Lagrangian submanifolds of X considered by Aganagic and Vafa [7], which we simply refer to as toric Lagrangians: symplectically, these are singular fibres of the Harvey–Lawson fibration associated to X. A toric Lagrangian is diffeomorphic to $\mathbb{R}^2 \times S^1$, and so its first homology group is isomorphic to $\mathbb{Z}$.

We fix $L = L_1 \cup \dots \cup L_s$ a disjoint union of toric Lagrangians L_i in X. In informal terms, the open Gromov–Witten theory of $(X, L = L_1 \cup \dots \cup L_s)$ should be a virtual count of maps to X from open Riemann surfaces of fixed genus, relative homology degree, and boundary winding data around $S^1 \hookrightarrow L$. A precise definition of such counts in the algebraic category has been given by Li, Liu, Liu and Zhou [77; 76] using relative Gromov–Witten theory and virtual localisation. These invariants depend on the choice of a framing f of L, which is a choice of integer f_i for each connected component L_i of L. Given partitions $\mu_1, \dots, \mu_s$ of lengths $\ell(\mu_1), \dots, \ell(\mu_s)$, we denote by $O_{g;\beta;(\mu_1,\dots,\mu_s)}(X, L, \mathsf{f})$ the invariants defined in [77; 76], which are informally open Gromov–Witten invariants counting connected genus g Riemann surfaces of class $\beta \in \mathrm{H}_2(X, L, \mathbb{Z})$ with, for every $1 \le i \le s$, $\ell(\mu_i)$ boundary components wrapping L_i with winding numbers given by the parts of μ_i. We package the open Gromov–Witten invariants $O_{g,\beta,\mu_1,\dots,\mu_s}(X, L, \mathsf{f})$ into formal generating functions

$$\mathsf{O}_{\beta;\vec{\mu}}(X, L, \mathsf{f})(\hbar) := \sum_{g \ge 0} \hbar^{2g-2+\ell(\vec{\mu})} O_{g;\beta;\vec{\mu}}(X, L, \mathsf{f}), \tag{1-2}$$

where $\ell(\vec{\mu}) = \sum_{i=1}^{s} \ell(\mu_i)$. We simply denote by $O_{g;\beta}(X, L, \mathsf{f})$ and $\mathsf{O}_{\beta}(X, L, \mathsf{f})(\hbar)$ the s–holed open Gromov–Witten invariants obtained when each partition μ_i consists of a single part (whose value is then determined by the class $\beta \in \mathrm{H}_2(X, L, \mathbb{Z})$).

1.3.4 Quiver DT invariants Let Q be a quiver with an ordered set Q_0 of n vertices $v_1, \dots v_n \in \mathsf{Q}_0$ and a set of oriented edges $\mathsf{Q}_1 = \{\alpha \colon v_i \to v_j\}$. We let $\mathbb{N}\mathsf{Q}_0$ be the free abelian semigroup generated by Q_0, and, for $\mathsf{d} = \sum d_i v_i$ and $\mathsf{e} = \sum e_i v_i \in \mathbb{N}\mathsf{Q}_0$, we write $E_{\mathsf{Q}}(\mathsf{d}, \mathsf{e})$ for the Euler form

$$E_{\mathsf{Q}}(\mathsf{d}, \mathsf{e}) := \sum_{i=1}^{n} d_i e_i - \sum_{\alpha \colon v_i \to v_j} d_i e_j. \tag{1-3}$$

Assume that Q is symmetric; that is, for every i and j, the number of oriented edges from v_i to v_j is equal to the number of oriented edges from v_j to v_i. The Euler form is then a symmetric bilinear form. The motivic DT invariants $\mathrm{DT}_{\mathsf{d};i}(\mathsf{Q})$ of Q are defined by the equality

$$\sum_{\mathsf{d}\in\mathbb{N}^n} \frac{(-q^{1/2})^{E_Q(\mathsf{d},\mathsf{d})}\mathsf{x}^{\mathsf{d}}}{\prod_{i=1}^n (q;q)_{d_i}} = \prod_{\mathsf{d}\neq 0}\prod_{i\in\mathbb{Z}}\prod_{k\geq 0}\big(1-(-1)^i\mathsf{x}^{\mathsf{d}}q^{-k-(i+1)/2}\big)^{-\mathrm{DT}_{\mathsf{d};i}(\mathsf{Q})}, \tag{1-4}$$

where $\mathsf{x}^{\mathsf{d}} = \prod_{i=1}^n x_i^{d_i}$; see Efimov [34], Kontsevich and Soibelman [70] and Reineke [107]. In other words, the motivic DT invariants are defined by taking the plethystic logarithm of the generating series of Poincaré rational functions of the stacks of representations of Q. The numerical DT invariants $\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}}(\mathsf{Q})$ are defined by

$$\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}}(\mathsf{Q}) := \sum_{i\in\mathbb{Z}} (-1)^i \mathrm{DT}_{\mathsf{d},i}(\mathsf{Q}). \tag{1-5}$$

According to Efimov [34], the numerical DT invariants $\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}}(\mathsf{Q})$ are nonnegative integers.

1.3.5 Open/closed BPS invariants Gromov–Witten invariants define virtual counts of curves and are in general rational numbers, but they are well-known to exhibit hidden integrality properties in terms of underlying BPS counts. The original physics definition, due to Gopakumar and Vafa [48; 47] in the classical context of closed Gromov–Witten invariants of Calabi–Yau 3–folds, predicted the form of these counts in terms of degeneracies of BPS particles in four/five dimensions arising from type IIA/M-theory as D2/M2–branes wrapping 2–cycles in the compactification. A longstanding effort has been made on multiple fronts to make the physics definition rigorous either using the associated cohomologies of sheaves (see Katz [64] and Maulik and Toda [90]), stable pairs (see Pandharipande and Thomas [101]), and direct symplectic methods (see Ionel and Parker [62]). In this paper, we will consider BPS invariants for genus-zero Gromov–Witten invariants of Calabi–Yau 4–folds and higher-genus open Gromov–Witten invariants of toric Calabi–Yau 3–folds. As an immediate corollary we obtain a new definition of all genus BPS invariants of Looijenga pairs (1-21).

Let $Y(D) = (Y, D = D_1 + D_2)$ be a 2–component Looijenga pair. The corresponding local geometry $E_{Y(D)}$ is a noncompact Calabi–Yau 4–fold. Following Greene, Morrison and Plesser [50, Appendix B] and Klemm and Pandharipande in [68, Section 1.1], we define BPS invariants $\mathrm{KP}_d(E_{Y(D)})$ in terms of the local genus-zero Gromov–Witten invariants $N^{\mathrm{loc}}_{0,d}(Y(D))$ by the formula

$$\mathrm{KP}_d(E_{Y(D)}) = \sum_{k|d} \frac{\mu(k)}{k^2} N^{\mathrm{loc}}_{d/k}(Y(D)). \tag{1-6}$$

Let X be a toric Calabi–Yau 3–fold, $L = L_1 \cup \cdots \cup L_s$ a disjoint union of toric Lagrangian branes and f a choice of framing. Following Labastida and Mariño [73], Labastida, Mariño and Vafa [74], Mariño and Vafa [88] and Ooguri and Vafa [100], we define the Labastida–Mariño–Ooguri–Vafa (LMOV)

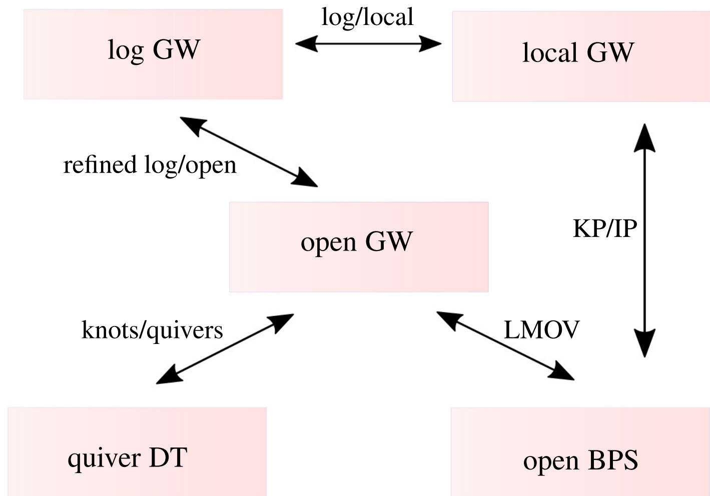


Figure 1: Enumerative invariants of $Y(D)$ and their mutual relations.

generating function of BPS invariants $\Omega_d(X, L, f)(q) \in \mathbb{Q}(q^{1/2})$ in terms of the s–holed higher-genus open Gromov–Witten generating series $\mathsf{O}_\beta(X, L, \mathsf{f})(\hbar)$ by the formula

$$\Omega_\beta(X, L, \mathsf{f})(q) = [1]_q^2 \left(\prod_{i=1}^{s} \frac{w_i}{[w_i]_q} \right) \sum_{k|\beta} \frac{\mu(k)}{k} \mathsf{O}_{\beta/k}(X, L, \mathsf{f})(-\mathrm{i}k \log q), \tag{1-7}$$

where $w_1, \dots, w_s$ are the winding numbers around the Lagrangians $L_1, \dots, L_s$ of the boundary components of an s–holed Riemann surface with relative homology class β, and where $[n]_q := q^{n/2} - q^{-n/2}$ are the q–integers, defined for all integers n.

1.4 The web of correspondences: geometric motivation

The enumerative theories of the previous section have superficially distant flavours, but they will turn out to be in close and often surprising relation to each other (Figure 1). We start by explaining the general geometric motivation behind the web of relations below, deferring rigorous statements for the case of Looijenga pairs to Section 1.5.

1.4.1 From log to local invariants Let $(Y, D = D_1 + \cdots + D_l)$ be a log smooth pair of maximal boundary; unless specified at this stage we do not restrict to Y being a surface, and neither do we impose the condition that (Y, D) be log Calabi–Yau, nor any positivity conditions on D_j. We will say that a curve class $d \in \mathrm{H}_2(Y, \mathbb{Z})$ is *D–convex* if $d \cdot D_i > 0$ for all i, and for every decomposition $d = [C_1] + \cdots + [C_m] \in \mathrm{H}_2(Y, \mathbb{Z})$, with each C_j an effective curve, we have $C_j \cdot D_i \geq 0$ for all i and j.

We begin by introducing some intermediate geometries built from $Y(D)$: for $\mathsf{m} = 1, \dots, l+1$, let

$$Y^{(\mathsf{m})} := \mathrm{Tot}\Big(\bigoplus_{k \geq \mathsf{m}} \mathcal{O}_Y(-D_k) \Big), \tag{1-8}$$

and $D^{(\mathsf{m})}$ be the preimage $\pi^{-1}\big(\bigcup_{k<\mathsf{m}} D_k\big)$ by the projection $\pi\colon Y^{(\mathsf{m})} \to Y$. Note that, by definition, $Y^{(1)}(D^{(1)}) = E_{Y(D)}$ and $Y^{(l+1)}(D^{(l+1)}) = Y(D)$: the geometries $Y^{(\mathsf{m})}(D^{(\mathsf{m})})$ for $1 < \mathsf{m} \le l$ consist of intermediate setups where a log condition is imposed on $\{D_k\}_{k<\mathsf{m}}$, and a local one on $\{D_k\}_{k\ge\mathsf{m}}$. For d a $D^{(\mathsf{m})}$–convex curve class, we denote by $N^{\log}_{0,d}(Y^{(\mathsf{m})}(D^{(\mathsf{m})}))$ a genus-zero maximal tangency log GW invariant of class d of $Y^{(\mathsf{m})}(D^{(\mathsf{m})})$ with a choice of point and ψ–class insertions; see Section 4.1. $D^{(\mathsf{m})}$–convexity ensures that this is well-defined, despite $Y^{(\mathsf{m})}(D^{(\mathsf{m})})$ not being proper for $\mathsf{m} \le l$.

Assume first that $l = 1$, ie that D is a smooth divisor. In van Garrel, Graber and Ruddat [43], the genus-zero local Gromov–Witten invariants of $E_{Y(D)}$ were related to the genus-zero maximal tangency Gromov–Witten theory of (Y, D) by the *stationary log/local correspondence*,

$$N^{\mathrm{loc}}_{0,d}(Y(D)) = \frac{(-1)^{d\cdot D-1}}{d\cdot D} N^{\log}_{0,d}(Y(D)). \tag{1-9}$$

The argument of [43] is geometric, and it gives a stronger statement at the level of virtual fundamental classes: $E_{Y(D)}$ is degenerated to $Y \times \mathbb{A}^1$ glued along $D \times \mathbb{A}^1$ to a line bundle over the projective bundle $\mathbb{P}(\mathcal{O}_D \oplus \mathcal{O}_D(-D))$. This degeneration moves genus-zero stable maps in $E_{Y(D)}$ to genus-zero stable maps splitting along both components of the central fibre: the degeneration formula then states that $N^{\mathrm{loc}}_{0,d}(Y(D))$ equals the weighted sum over splitting type of the product of invariants associated to each component, and a careful analysis shows that only one term is nonzero, leading to (1-9). In [43, Conjecture 6.4], a conjectural cycle-level log-local correspondence was also proposed for simple normal crossing pairs: we propose here a slight variation of its restriction to stationary invariants and anticanonical D in the following conjecture.

Conjecture 1.1 (the stationary log/local correspondence for maximal log CY pairs) *Suppose that $(Y, D = D_1 + \cdots + D_l)$ is a log smooth log Calabi–Yau pair of maximal boundary, d a D–convex curve class, and $1 \le \mathsf{n} < \mathsf{m} \le l+1$. Then*

$$N^{\log}_{0,d}(Y^{(\mathsf{m})}(D^{(\mathsf{m})})) = \left(\prod_{i=\mathsf{n}}^{\mathsf{m}-1} (-1)^{d\cdot D_i+1} d\cdot D_i\right) N^{\log}_{0,d}(Y^{(\mathsf{n})}(D^{(\mathsf{n})})). \tag{1-10}$$

In particular, when $(\mathsf{n},\mathsf{m}) = (1, l+1)$,

$$N^{\log}_{0,d}(Y(D)) = \left(\prod_{i=1}^{l} (-1)^{d\cdot D_i+1} d\cdot D_i\right) N^{\mathrm{loc}}_{0,d}(Y(D)). \tag{1-11}$$

When all D_j are nef and $(\mathsf{n},\mathsf{m}) = (1, l+1)$, this gives the numerical version of [43, Conjecture 6.4] for point insertions and anticanonical D. When $\mathsf{m} - \mathsf{n} = 1$, (1-10) is an extension of the main result of [43] to the noncompact case.

The extent to which the argument of [43] generalises to the case of simple normal crossings pairs of Conjecture 1.1 is a somewhat thorny issue. In particular, the cycle-level conjecture of [43, Conjecture 6.4] is known to fail in the nonstationary sector for general l, as recently observed in a non-log Calabi–Yau example by Nabijou and Ranganathan [96]. At the same time, there is a nontrivial body of evidence that a generalisation of the stationary sector equality (1-9) (ie with descendent point insertions only) might hold for simple normal crossings log Calabi–Yau pairs $Y(D)$; see Bousseau, Brini and van Garrel [18] for a proof for toric orbifold pairs. It is therefore an open question to find the exact boundaries of validity of the stationary log-local correspondence, and in this paper we chart a conceptual pathway to delineate them for the (special, but central) case of log Calabi–Yau pairs of Conjecture 1.1, as follows.

At a geometric level, the degeneration of [43] can be generalised to a birational modification of one where the generic fibre is $E_{Y(D)}$, and the special fibre is obtained by gluing, for each $j = 1, \dots, l$, $Y \times (\mathbb{A}^1)^l$ along $D_j \times (\mathbb{A}^1)^l$ to a rank l vector bundle over $\mathbb{P}(\mathcal{O}_{D_j} \oplus \mathcal{O}_{D_j}(-D_j))$. After an (explicit) birational modification this gives a log smooth family: we describe the details of the degeneration for the case of surfaces in Section 5.1. When $l > 1$, instead of the degeneration formula the decomposition formula [2] applies, expressing $N^{\mathrm{loc}}_{0,d}(Y(D))$ as a weighted sum of terms, indexed by tropical curves $h\colon \Gamma \to \Delta$, where Δ is the dual intersection complex of the central fibre:

$$N^{\mathrm{loc}}_{0,d}(Y(D)) = \sum_{h\colon \Gamma \to \Delta} \frac{m_h}{|\mathrm{Aut}(h)|} N^{\mathrm{loc},h}_{0,d}(Y(D)). \tag{1-12}$$

The geometric picture above, and the ensuing decomposition formula (1-12), provides a rather general and geometrically motivated blueprint to measure the deviation, or lack thereof, of the local invariants from their expected relation to maximal tangency log invariants in (1-11). As a proof-of-concept step, and as we shall describe in detail in Section 5.1, in this paper we show how this framework bears fruit in the context of Looijenga pairs:[1] here correction terms indexed by nonmaximal tangency tropical curves turn out, remarkably, to *all* individually vanish, whilst the maximal tangency tropical contribution exactly returns the right-hand side of (1-11).

1.4.2 From log to open invariants Let $Y(D)$ be a log Calabi–Yau surface. By (1-8), the complement $Y^{(l)} \setminus D^{(l)}$ is isomorphic to the total space of $\mathcal{O}(-D_l) \to Y \setminus (D_1 \cup \dots \cup D_{l-1})$; since D is anticanonical, this is a noncompact Calabi–Yau threefold. We propose that the log invariants $N^{\log}_{0,d}(Y(D))$ can be precisely related to open Gromov–Witten invariants of $Y^{(l)} \setminus D^{(l)}$ with boundary in fixed disjoint Lagrangians L_k, with $k < l$, near the divisor $D^{(l)}$. These Lagrangians should have a specific structure as described in [8, Section 7], namely they should be fibred over Lagrangians L'_k in $\pi^{-1}(D_k)$ with fibres Lagrangians in the normal bundle $(N_{\pi^{-1}(D_k)/Y^{(m)}})|_{L'_k}$. Writing $L := \bigcup_{k<l} L_k$ and $Y^{\mathrm{op}}(D) := (Y^{(l)} \setminus D^{(l)}, L)$, there is a natural isomorphism $\iota\colon \mathrm{H}_2^{\mathrm{rel}}(Y^{\mathrm{op}}(D), \mathbb{Z}) \to \mathrm{H}_2(Y, \mathbb{Z})$ induced by the embedding $Y^{(l)} \setminus D^{(l)} \hookrightarrow Y^{(l)}$ and the identification of winding degrees along L_k with contact orders along D_k; see Proposition 6.6 for details.

[1] It is an intriguing question, and one well beyond the scope of this paper, to test how this philosophy generalises to log Calabi–Yau varieties of any dimension, and to revisit the non-log Calabi–Yau, nonstationary negative result of [96] in this light.

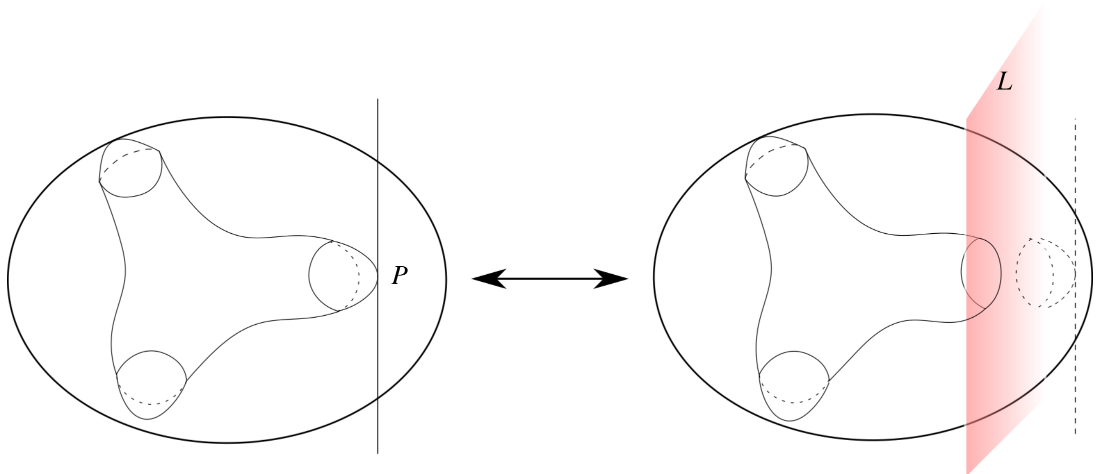


Figure 2: Exchanging log and open conditions.

Suppose now that there is a well-posed definition[2] of genus-zero open GW counts $O_{0;d}(Y^{\mathrm{op}}(D))$ as in Georgieva [45] and Solomon and Tukachinsky [109]. In such a scenario, we expect a close relationship between these and the log invariant $N_{0,d}^{\log}(Y(D))$.

Conjecture 1.2 (log-open correspondence for surfaces) *Let $Y(D)$ be a log Calabi–Yau surface with maximal boundary and d a D–convex curve class. Then*

$$O_{0;\iota^{-1}(d)}(Y^{\mathrm{op}}(D)) = \left(\prod_{k=1}^{l} \frac{(-1)^{d\cdot D_k - 1}}{d \cdot D_k}\right) N_{0,d}^{\log}(Y(D)). \tag{1-13}$$

There is an intuitive symplectic heuristics behind Conjecture 1.2: removing a tubular neighbourhood of $D^{(l)}$ turns pseudoholomorphic log curves in $Y^{(l)}$ with prescribed tangencies along $D^{(l)}$ into pseudoholomorphic open Riemann surfaces with boundaries in L, with winding numbers determined by the tangencies; see Figure 2. The relative factor $\prod_{k<l}(-1)^{d\cdot D_k-1}(d\cdot D_k)^{-1}$ at the level of GW counts in Conjecture 1.2 can be understood by looking at the simplest example where $Y=\mathbb{P}^1\times\mathbb{A}^1$, $D_1=\{0\}\times\mathbb{A}^1$ and $D_2=\{\infty\}\times\mathbb{A}^1$, where $0,\infty\in\mathbb{P}^1$. For the curve class d times the class of $\mathbb{P}^1$ we have $N_{0,d}^{\log}(Y(D))=1$, as there exists a unique degree d cover of $\mathbb{P}^1$ fully ramified over two points, and the order d automorphism group of this cover is killed by the point condition. By Conjecture 1.1, and in particular (1-10) with $\mathsf{m}=2$, we deduce that $N_{0,d}^{\log}(Y^{(2)}(D^{(2)}))=(-1)^{d-1}/d$; on the other hand, the open geometry $Y^{\mathrm{op}}(D)$ is $\mathbb{C}^3$ with a singular Harvey–Lawson Lagrangian L of framing zero (see Construction 6.4): the degree d multicovers of the unique embedded disk [65, Theorem 7.2] contribute $O_{0;\iota^{-1}(d)}(Y^{\mathrm{op}}(D))=1/d^2$, from which the relative factor in (1-13) is recovered.

[2]An example of this situation (see Construction 6.4) is when, up to deformation, both Y and the divisors D_k $(k<l)$ are toric, implying that $Y^{\mathrm{op}}(D)$ is a toric Calabi–Yau threefold geometry equipped with framed toric Lagrangians L_k: in this case the open GW invariants were introduced in Section 1.3.3.

Much as in Conjecture 1.1, the invariants in Conjecture 1.2 live in different dimensions: (1-13) relates log invariants of the log CY surface $Y(D)$ to open invariants of special Lagrangians in a Calabi–Yau threefold. Note that combining Conjectures 1.1 and 1.2 further gives a surprising conjectural relation

$$O_{0;\iota^{-1}(d)}(Y^{\mathrm{op}}(D)) = N^{\mathrm{loc}}_{0,d}(Y(D)), \tag{1-14}$$

which equates the GW invariants of the CY3 open geometry $Y^{\mathrm{op}}(D)$ with the local GW invariants of the CY $(l+2)$ variety $E_{Y(D)}$.[3]

We also expect a precise uplift of this picture to higher-genus invariants. For a single irreducible divisor, an all-genus version of the log-local correspondence of [43] was described in [19, Theorems 1.1–1.2]. Its generalisation to a log-open correspondence in higher genus for a completely general pair is likely to take an unwieldy form, but we expect it to be particularly simple for a maximal boundary log Calabi–Yau surface. Indeed, in the degeneration to the normal cone along D_l, only multiple covers of a $\mathbb{P}^1$–fibre in $\mathbb{P}(\mathcal{O}_{D_l} \oplus \mathcal{O}_{D_l}(-D_l))$ will contribute. The resulting combination of the multiplicity $d \cdot D_l$ in the degeneration formula with the higher-genus multiple cover contribution

$$\frac{(-1)^{d\cdot D_l+1}}{(d\cdot D_l)[d\cdot D_l]_q}$$

leads us to predict a precise, and tantalisingly simple q–analogue of Conjecture 1.2.

Conjecture 1.3 (the all-genus log-open correspondence for surfaces) *Let $Y(D)$ be a log CY surface with maximal boundary and d a D–convex curve class. With notation as in Sections 1.3.1 and 1.3.3, we have*

$$\mathsf{O}_{\iota^{-1}(d)}(Y^{\mathrm{op}}(D))(-\mathrm{i}\log q) = [1]_q^{l-2}\,\frac{(-1)^{d\cdot D_l+1}}{[d\cdot D_l]_q}\prod_{k=1}^{l-1}\frac{(-1)^{d\cdot D_k+1}}{d\cdot D_k}\,\mathsf{N}^{\log}_d(Y(D))(-\mathrm{i}\log q). \tag{1-15}$$

The factor $[1]_q^{l-2}$ corresponds to the relative normalisation of the higher-genus generating functions in (1-1) and (1-2). The allusive hints of this section will be put on a rigorous footing in Section 1.5.2.

1.4.3 Quivers and BPS invariants Given $Y(D = D_1 + D_2)$ a 2–component Looijenga pair, the virtual count of curves in the noncompact Calabi–Yau 4–fold $E_{Y(D)}$ (see Klemm and Pandharipande [68]) is expected to be expressible in terms of sheaf counting (see Cao, Maulik and Toda [23; 24]). More precisely, it is expected that the BPS invariants of $E_{Y(D)}$ are extracted from a DT_4 virtual fundamental

[3]The relation (1-14) is in tune with physics expectations from type IIA string theory compactification on $\mathbb{R}^{1,1}\times X$, where X is a Calabi–Yau fourfold: the low energy effective theory is a $\mathcal{N}=(2,2)$ QFT, whose effective holomorphic superpotential is computed by the genus-zero Gromov–Witten invariants of X. Now precisely the same type of holomorphic F–terms can be engineered by considering D4–branes wrapping special Lagrangians on a Calabi–Yau 3–fold [100]: the superpotential here is a generating function of holomorphic disk counts with boundary on the Lagrangian three-cycle. It was suggested by Mayr [91] (see also [4]) that there exist cases where 2d superpotentials can be engineered in both ways, resulting in an identity between local genus-zero invariants of CY 4–folds and disk invariants of CY 3–folds: the equality in (1-14) asserts just that.

class associated to the moduli space of one-dimensional coherent sheaves on $E_{Y(D)}$. As coherent sheaves are often very closely related to modules over quivers, it might be tempting to ask if curve counting in $E_{Y(D)}$ (and, via the arguments of the previous section, the log/open GW theory of $Y(D)$) can be described in terms of some quiver DT theory.

This is more than a suggestive speculation. Consider for example $Y = \mathbb{P}^2$ and $D = D_1 + D_2$ the union of a line D_1 and a conic D_2, so that $E_{Y(D)}$ is the total space of $\mathcal{O}_{\mathbb{P}^2}(-1) \oplus \mathcal{O}_{\mathbb{P}^2}(-2)$. Let $\mathcal{M}_{\mathbb{P}^1}^{\mathrm{Higgs}}(d,n)$ be the moduli space of rank-d, degree-n $\mathcal{O}(1)$–twisted Higgs bundles $\mathcal{O}_{\mathbb{P}^1}^{\oplus d} \to \mathcal{O}_{\mathbb{P}^1}^{\oplus d} \otimes \mathcal{O}_{\mathbb{P}^1}(1)$ on $\mathbb{P}^1$. The total space of $\mathcal{O}_{\mathbb{P}^1}(1)$ is the complement of a point in $\mathbb{P}^2$, and as $\mathbb{P}^1$ has normal bundle $\mathcal{O}(1)$ in $\mathbb{P}^2$, $\mathcal{M}_{\mathbb{P}^1}^{\mathrm{Higgs}}(d,n)$ sits as an open part of the moduli space of one-dimensional coherent sheaves on $E_{Y(D)}$. At the same time, as $\mathcal{O}(1)$ has two sections on $\mathbb{P}^1$, $\mathcal{M}_{\mathbb{P}^1}^{\mathrm{Higgs}}(d,n)$ is isomorphic to the moduli space of representations of the quiver with one vertex and two loops. Strikingly, we remark here that this is reflected into a completely unexpected identity for the corresponding invariants: the Klemm–Pandharipande BPS invariants of $E_{Y(D)}$ computed in [68, Section 3.2] simultaneously coincide (up to sign) with the DT invariants of the 2–loop quiver computed in Reineke [107, Theorem 4.2], as well as with the top Betti numbers[4] $\mathfrak{B}_d^{\mathrm{Higgs}}(\mathbb{P}^1) := \dim \mathrm{H}^{\mathrm{top}}(\mathcal{M}_{\mathbb{P}^1}^{\mathrm{Higgs}}(d,n), \mathbb{Q})$ of the moduli spaces of $\mathcal{O}(1)$–twisted Higgs bundles on the line considered in Rayan [106, Section 5]:

$$
\begin{aligned}
\text{(1-16)} \qquad |\mathrm{KP}_d(\mathcal{O}_{\mathbb{P}^2}(-1) \oplus \mathcal{O}_{\mathbb{P}^2}(-2))| &= \mathfrak{B}_d^{\mathrm{Higgs}}(\mathbb{P}^1) = \mathrm{DT}_d^{\mathrm{num}}(\text{2–loop quiver}) \\
&= (1, 1, 1, 2, 5, 13, 35, 100, 300, 925, 2915, 9386, \dots)_d.
\end{aligned}
$$

From a sheafy point of view, this raises the question how the definition of Calabi–Yau 4–fold invariants from the moduli space of coherent sheaves [23; 24] interacts with the quiver description, and whether such a startling coincidence is an isolated example — or not.

An upshot of Conjectures 1.1 and 1.2 is a surprising Gromov–Witten-theoretic take on this question: for $l = 2$ and when $Y^{\mathrm{op}}(D)$ is an open geometry given by toric Lagrangians in a toric CY3, the quiver can be reconstructed systematically from the geometry of $Y(D)$ via a version of the "branes–quivers" correspondence introduced in Ekholm, Kucharski and Longhi [36; 35], Kucharski, Reineke, Stošić and Sułkowski [72] and Panfil, Stošić and Sułkowski [102]. According to the open GW/quiver dictionary of [35], the quiver nodes are identified with basic (in the sense of [36; 35]) embedded holomorphic disks with boundary on L, edges and self-edges correspond to linking and self-linking numbers of the latter, and the DT invariants of the quiver return (up to signs) the genus-zero LMOV count of holomorphic disks obtained as "boundstates" of the basic ones [36, Section 4].

Now, by the $q \to 1$ limit of (1-7), the genus-zero LMOV and GW invariants of $Y^{\mathrm{op}}(D)$ are related to each other by the *same* BPS change of variables relating KP invariants and local GW invariants of $E_{Y(D)}$ in (1-6). Then a direct consequence of the conjectural open = local GW equality (1-14) is that the KP invariants of the local CY4–fold $E_{Y(D)}$ coincide with the LMOV invariants of the open CY3 geometry $Y^{\mathrm{op}}(D)$ —

[4] The degree-independence of these Betti numbers, at least for $(d,n) = 1$, is explained in [106, Section 5].

which by the branes-quivers correspondence above are in turn DT invariants of a symmetric quiver! In particular, for the example above of $Y=\mathbb{P}^2$ and $D=D_1+D_2$ the union of a line and a conic, we shall find the open geometry $Y^{\mathrm{op}}(D)$ to be three-dimensional affine space with a single toric Lagrangian at framing one (see Construction 6.4) — and as expected, in this case the quiver construction in [102, Section 5.1] returns exactly the quiver with one vertex and two loops we had found in (1-16). In general, this connection leads to some nontrivial implications for the Gopakumar–Vafa/Donaldson–Thomas theory of CY4 local surfaces from log Gromov–Witten theory, which we describe precisely in Sections 1.6.1 and 1.6.3.

1.5 The web of correspondences: results

In order to state our results, we introduce some notions of positivity for Looijenga pairs. A Looijenga pair $Y(D)=(Y,D=D_1+\cdots+D_l)$ is *nef* if each irreducible component D_i of D is smooth and nef: note that the condition that the components D_i are smooth implies in particular that $l\geq 2$, and nefness entails that a generic stable map to Y is D–convex, which implies that the corresponding local Gromov–Witten invariants are well-defined.

A nef Looijenga pair $Y(D)$ is *tame* if either $l>2$ or $D_i^2>0$ for all i, and *quasi-tame* if the associated local geometry $E_{Y(D)}$ is deformation equivalent to the local geometry $E_{Y'(D')}$ associated to a tame Looijenga pair $Y'(D')$: we explain the relevance of these two properties in Section 1.5.1. As we will show in Section 2, there are 18 smooth deformation types of nef Looijenga pairs in total, 11 of which are tame and 15 are quasi-tame. In particular, a nef Looijenga pair $Y(D)$ is uniquely determined by Y and the self-intersection numbers D_i^2, and we sometimes use the notation $Y(D_1^2,\dots,D_l^2)$ for $Y(D)$; see Table 1. We state our results in a slightly discursive form below, including pointers to their precise versions in the main body of the text.

1.5.1 The stationary log-local correspondence

Our first result establishes the stationary log/local correspondence of Conjecture 1.1 in the form given by (1-11).

Theorem 1.4 (Theorem 5.1, Lemma 3.1, Theorem 3.3, Theorem 3.5, Proposition 3.6) *For every nef Looijenga pair $Y(D)$, the genus-zero log invariants $N^{\log}_{0,d}(Y(D))$ and the genus-zero local invariants $N^{\mathrm{loc}}_{0,d}(Y(D))$ are related by*

$$N^{\mathrm{loc}}_{0,d}(Y(D))=\left(\prod_{j=1}^{l}\frac{(-1)^{d\cdot D_j-1}}{d\cdot D_j}\right)N^{\log}_{0,d}(Y(D)). \tag{1-17}$$

Moreover, we provide a closed-form solution to the calculation of both sets of invariants in (1-17).

As explained in Section 1.4.1, the key idea to prove Theorem 1.4 is by a degeneration argument, illustrated in Section 5.1 for $l=2$: we follow the general strategy of [43] to deduce log-local relations from a degeneration to the normal cone, and we solve in our case of interest the difficulties of the normal-crossings situation through a detailed study of the tropical curves contributing in the decomposition formula of Abramovich, Chen, Gross and Siebert [2] for log Gromov–Witten invariants. For $l > 2$, and more generally when $Y(D)$ is tame, it turns out to be more convenient to structure the proof so that an uplift to the all-genus story, absent in other approaches, is immediate. The notion of tameness is first shown to be synonymous of finite scattering, and for tame pairs we compute closed-form solutions for the log Gromov–Witten invariants using tropical geometry, more precisely two-dimensional scattering diagrams; see Gross [51], Gross, Hacking and Keel [53], Gross, Pandharipande and Siebert [56] and Mandel [84]. The statement of the theorem for tame cases follows by subsequently comparing with a closed-form solution of the local theory via Givental-style mirror theorems: the proof follows from a general statement valid for local invariants of toric Fano varieties in any dimension twisted by a sum of concave line bundles (Lemma 3.1), and the notion of tameness is shown to coincide here with the vanishing of quantum corrections to the mirror map. For non-quasi-tame cases, we use a blowup formula which allows to restrict to the case of highest Picard number; the proof of the equality (1-17) in this case, in Theorem 3.3, requires a highly nontrivial mirror map calculation.

1.5.2 The all-genus log-open correspondence A notable property of the scattering approach to Theorem 1.4 for $l > 2$ (and, in general, for tame Looijenga pairs) is that it can be bootstrapped to obtain all-genus results for the log invariants through the q–deformed version of the two-dimensional scattering diagrams of Gross [51], Gross, Hacking and Keel [53], Gross, Pandharipande and Siebert [56] and Mandel [84] and the general connection between higher-genus log invariants of surfaces with λ_g–insertion and q–refined tropical geometry studied in Bousseau [12; 14]. This is key to establishing the following version of Conjectures 1.2 and 1.3.

Theorem 1.5 (Theorems 4.5, 4.9, 4.10 and 6.7) *For every quasi-tame Looijenga pair $Y(D)$ distinct from $\mathrm{dP}_3(0,0,0)$, there exists a triple $Y^{\mathrm{op}}(D) = (X, L, \mathsf{f})$, geometrically related to $Y(D)$ by Construction 6.4, where X is a semiprojective toric Calabi–Yau 3–fold, $L = L_1 \cup \cdots \cup L_{l-1}$ is a disjoint union of $l-1$ toric Lagrangians in X, f is a framing for L, and there exists an isomorphism $\iota\colon \mathrm{H}_2(X, L, \mathbb{Z}) \xrightarrow{\sim} \mathrm{H}_2(Y, \mathbb{Z})$ such that*

$$O_{0;\iota^{-1}(d)}(Y^{\mathrm{op}}(D)) = N^{\mathrm{loc}}_{0,d}(Y(D)) = \prod_{i=1}^{l} \frac{(-1)^{d\cdot D_i+1}}{d\cdot D_i} N^{\log}_{0,d}(Y(D)). \tag{1-18}$$

Furthermore, if $Y(D)$ is tame,

$$\mathsf{O}_{\iota^{-1}(d)}(Y^{\mathrm{op}}(D))(-\mathrm{i}\log q) = [1]_q^{l-2} \frac{(-1)^{d\cdot D_l+1}}{[d\cdot D_l]_q} \prod_{i=1}^{l-1} \frac{(-1)^{d\cdot D_i+1}}{d\cdot D_i} \mathsf{N}^{\log}_{d}(Y(D))(-\mathrm{i}\log q). \tag{1-19}$$

Moreover, we provide a closed-form solution to the calculation of the invariants in (1-18)–(1-19).

The open geometry $Y^{\mathrm{op}}(D)$ is constructed following the ideas of Section 1.4.2; see Section 6.2 for full details. Key to the proof of Theorem 1.5 is the fact that quasi-tame Looijenga pairs can always be deformed to pairs for which the both surface Y and the divisors D_i with $i < l$ are toric: as we shall explain in Section 6.2, the corresponding open geometry $Y^{\mathrm{op}}(D)$ is given by suitable Aganagic–Vafa (singular Harvey–Lawson) Lagrangian branes in a toric Calabi–Yau threefold, whose open Gromov–Witten theory can be compactly encoded through the topological vertex.[5] Conjecture 1.3 then predicts a completely unexpected relation between the q–scattering and topological vertex formalisms, which Theorem 1.5 establishes for tame pairs. The combinatorics underlying the resulting comparison of invariants is in general extremely nontrivial: for $l = 2$, it can be shown to be equivalent to Jackson's q–analogue of the Pfaff–Saalschütz summation for the ${}_3\phi_2$ generalised q–hypergeometric function.

We furthermore conjecture that the higher-genus log-open correspondence of Theorem 1.5 extends to all quasi-tame pairs. The scattering diagrams become substantially more complicated in the nontame cases, and (1-18) translates into an intricate novel set of q–binomial identities: see Conjecture B.3 for explicit examples.[6] The log-local correspondence of Theorem 1.4 establishes their limit for $q \to 1$.

1.5.3 BPS invariants and quiver DT invariants As anticipated in Section 1.4.3, the log/open correspondence of Theorem 1.5 can be leveraged to produce a novel correspondence between log/local Gromov–Witten invariants and quiver DT theory.

Theorem 1.6 (Theorem 7.3) *Let $Y(D) = (Y, D_1 + D_2)$ be a 2–component quasi-tame Looijenga pair. Then there exists a symmetric quiver $\mathsf{Q}(Y(D))$ with $\chi(Y) - 1$ vertices and a lattice isomorphism $\kappa\colon \mathbb{Z}(\mathsf{Q}(Y(D)))_0 \xrightarrow{\sim} \mathrm{H}_2(Y, \mathbb{Z})$ such that*

$$\mathrm{DT}_d^{\mathrm{num}}(\mathsf{Q}(Y(D))) = \Big|\mathrm{KP}_{\kappa(d)}(E_{Y(D)}) + \sum_i \alpha_i \delta_{d,v_i}\Big|, \tag{1-20}$$

with $\alpha_i \in \{-1, 0, 1\}$. In particular, $\mathrm{KP}_d(E_{Y(D)}) \in \mathbb{Z}$.

A symplectic proof of the integrality of genus-zero BPS invariants for projective Calabi–Yau 4–folds, although likely adaptable to the noncompact setting, was given by Ionel–Parker in [62]. In Theorem 1.6, the integrality for the local Calabi–Yau 4–folds $E_{Y(D)}$ follows from the identification of the BPS invariants with DT invariants of a symmetric quiver.[7] We construct the symmetric quiver $\mathsf{Q}(Y(D))$ by combining the log-open correspondence given by Theorem 1.5 with a correspondence previously established by Panfil and Sułkowski [103] between toric Calabi–Yau 3–folds with "strip geometries" and symmetric quivers; see also Kimura, Panfil, Sugimoto and Sułkowski [67].

Theorem 1.5 associates to a Looijenga pair $Y(D)$ satisfying Property O the toric Calabi–Yau 3–fold geometry $Y^{\mathrm{op}}(D)$. Denote by $\Omega_d(Y(D))(q) := \Omega_{\iota^{-1}(d)}(Y^{\mathrm{op}}(D))(q)$ the open BPS invariants defined

[5] A conceptual explanation for the exclusion of $\mathrm{dP}_3(0,0,0)$ from the statement of Theorem 1.5 is given by the notion of Property O, which we introduce in Definition 6.3.

[6] After the first version of this paper appeared on the arXiv, we received a combinatorial proof of Conjecture B.3 from C Krattenthaler [71].

[7] The equality modulo the integral shift by $\sum_i \alpha_i \delta_{d,v_i}$ in (1-20) can be traded to an actual equality of absolute values at the price of considering a larger disconnected quiver $\widetilde{\mathsf{Q}}$, and a corresponding epimorphism $\tilde{\kappa}\colon \mathbb{Z}(\widetilde{\mathsf{Q}}(Y(D)))_0 \to \mathrm{H}_2(Y, \mathbb{Z})$; see [103].

in (1-7). In general, for any Looijenga pair we can define

$$\Omega_d(Y(D))(q) := [1]_q^2 \left(\prod_{i=1}^{l} \frac{1}{[d \cdot D_i]_q} \right) \sum_{k|d} \frac{(-1)^{d/k \cdot D + l} \mu(k)}{[k]_q^{2-l} k^{2-l}} \mathsf{N}^{\log}_{d/k}(Y(D))(-\mathrm{i}k \log q). \tag{1-21}$$

When $Y(D)$ is tame and satisfies Property O, the equivalence of the definitions (1-7) and (1-21) is a rephrasing of the log-open correspondence of Theorem 1.5 at the level of BPS invariants.

A priori, $\Omega_d(Y(D))(q) \in \mathbb{Q}(q^{1/2})$. By a direct arithmetic argument, we prove the following integrality result, which in particular establishes the existence of an integral BPS structure underlying the higher-genus log Gromov–Witten theory of $Y(D)$.

Theorem 1.7 (Theorem 8.1) *Let $Y(D)$ be a quasi-tame Looijenga pair. Then*

$$\Omega_d(Y(D))(q) \in q^{-\frac{1}{2}\mathrm{g}_{Y(D)}(d)} \mathbb{Z}[q]$$

for an integral quadratic polynomial $\mathrm{g}_{Y(D)}(d)$.

1.5.4 Orbifolds In the present paper, we mainly focus on the study of the finitely many deformation families of nef Looijenga pairs (Y, D) with Y smooth. Nevertheless, most of our techniques and results should extend to the more general setting where we allow Y to have orbifold singularities at the intersection of the divisors: the log Gromov–Witten theory is then well-defined since $Y(D)$ is log smooth, and the local Gromov–Witten theory makes sense by viewing Y and $E_{Y(D)}$ as smooth Deligne–Mumford stacks. There are infinitely many examples of nef/tame/quasi-tame Looijenga pairs in the orbifold sense. Deferring a treatment of more general examples to our companion note [17], we content ourselves here to show in Section 9 that the log-local, log-open and Gromov–Witten/quiver correspondences still hold for the infinite family of examples obtained by taking $Y = \mathbb{P}_{(1,1,n)}$, the weighted projective plane with weights $(1, 1, n)$, and $D = D_1 + D_2$ with D_1 a line passing through the orbifold point and D_2 a smooth member of the linear system given by the sum of the two other toric divisors.

1.6 The web of correspondences: implications

The results of the previous section subsume and were motivated by several disconnected strands of development in the study of the enumerative invariants in Sections 1.3.1–1.3.5. We briefly describe here how they relate to and impact ongoing progress in some allied contexts.

1.6.1 BPS structures in log/local GW theory The relation of log GW invariants to BPS invariants in Theorem 1.7 echoes very similar[8] statements relating log GW theory to DT and LMOV invariants in Bousseau [15; 14], and in particular it partly demystifies the interpretation of log GW partition functions

[8]A nontrivial difference is that here the log Gromov–Witten invariants are *not* interpreted as BPS invariants themselves, unlike in [15; 14], but rather are related to them via (1-21).

as related to some putative open curve counting theory on a Calabi–Yau 3–fold in [14, Section 9] by realising the open BPS count in terms of actual, explicit special Lagrangians in a toric Calabi–Yau threefold. Aside from its conceptual appeal, its power is revealed by some of its immediate consequences: the Klemm–Pandharipande conjectural integrality [68, Conjecture 0] for local CY4 surfaces follows as a zero-effort corollary of the log-open correspondence of Theorem 1.5 by constructing the associated quiver in Theorem 1.6, identifying the KP invariants of the local surface with its DT invariants, and applying Efimov's theorem [34].

We note that this chain of connections opens the way to a proof of the Calabi–Yau 4–fold Gromov–Witten/Donaldson–Thomas correspondence [23; 24], which is an open conjecture even for the simplest local surfaces. The analysis of the underlying integrality of the q–scattering calculation in Theorem 1.7 furthermore gives, in the limit $q \to 1$, an algebrogeometric version of symplectic results of Ionel and Parker [62] for Calabi–Yau vector bundles on toric surfaces; and away from this limit, it provides a refined integrality statement whose enumerative salience for the local theory is hitherto unknown, and worthy of further study: see Section 1.6.3.

1.6.2 The general log-open correspondence for surfaces Throughout the heuristic description of the motivation for Conjectures 1.1–1.3, we have been mindful not to impose any nefness condition on the divisors D_i: the only request we made was for the genus-zero obstruction theory of the local theory to be encoded by a genuine obstruction bundle over the untwisted moduli space. This was taken into account by the condition of D–convexity for the stable maps: restricting to D–convex maps widens the horizon of the log/local correspondence of [43] to a vast spectrum of cases which were not accounted for in previous studies of the correspondence. And indeed, in the broadest generality where the open invariants can be defined in the algebraic category, the methods proposed here extend straightforwardly to treat the cases when one or more of the irreducible components D_i have negative self-intersection: Conjectures 1.1–1.3 hold with flying colours in these cases as well, with all $l > 2$ anticanonical pairs satisfying Property O that remarkably enjoy the same salient properties of the *tame* nef pairs, such as finite scattering, closed-form resummation of the topological vertex, and triviality of the mirror map; their detailed study will appear in work-in-progress of Brini, van Garrel and Schüler.

The discussion of Section 1.4.2 also opens the door to pushing the log/open correspondence beyond the maximal contact setting: it is tempting to see how the maximal tangency condition could be removed from Conjecture 1.2, with the splitting of contact orders amongst multiple points on the same divisor being translated to windings of multiple boundary disks ending on the same Lagrangian. The multicovering factor of (1-19) would then be naturally given by a product of individual contact orders/disk windings — an expectation that the reader can verify to be fulfilled in the basic example presented there of $(Y, D) = (\mathbb{P}^1 \times \mathbb{A}^1, \mathbb{A}^1 \cup \mathbb{A}^1)$. More generally, the link to the topological vertex and open GW invariants of arbitrary topology raises a fascinating question how much the topological vertex knows of the log theory of the surface — and how it can be effectively used in the construction of (quantum) SYZ mirrors.

1.6.3 Relation to the Cao–Maulik–Toda conjecture Another direction towards a geometric understanding of the integrality of KP invariants is provided by sheaf-counting theories for Calabi–Yau 4–folds, which were originally introduced by Borisov and Joyce [11] (see also Cao and Leung [22]) and have recently been given an algebraic construction by Oh and Thomas [98]. More precisely, Cao, Maulik and Toda have conjectured in [24] (resp. in [23]) explicit relations between genus-zero KP invariants and stable pair invariants (resp. counts of one-dimensional coherent sheaves) on Calabi–Yau 4–folds. Recently, Cao, Kool and Monavari [21] have checked the conjecture of [24] for low-degree classes on local toric surfaces; their proof hinges on the solution of the Gromov–Witten/Klemm–Pandharipande side given by Theorems 1.4 and 1.6 in this paper.

The results of Theorems 1.6 and 1.7 also raise a host of new questions. First and foremost, it would be extremely interesting to find for local toric surfaces a direct connection between the symmetric quivers appearing in Theorem 7.3 and the moduli spaces of coherent sheaves appearing in the conjectures of [24; 23]. Furthermore, since for $l = 2$ we have $\mathrm{KP}_d(E_{Y(D)}) = \Omega_d(Y(D))$, a fascinating direction would be to find an interpretation of the q–refined invariants $\Omega_d(Y(D))(q)$ in terms of the Calabi–Yau 4–fold $E_{Y(D)}$. A natural suggestion is that $\Omega_d(Y(D))(q)$ should take the form of some appropriately refined Donaldson–Thomas invariants of $E_{Y(D)}$. As the topic of refined DT theory of Calabi–Yau 4–folds is still in its infancy, we leave the question open for now.

Acknowledgements

Discussions with T Bridgeland, M Gross, A Klemm, M Kool, P Kucharski, P Longhi, C Manolache, N Nabijou, D Ranganathan and P Sulkowski are gratefully acknowledged. We thank A Afandi, Y Schüler, and especially R Thomas for very helpful comments on previous versions of the manuscript, and C Krattenthaler for providing us with a copy of [71], where Conjecture B.3 is proved for $q \neq 1$.

This project has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement 746554 (van Garrel), the Engineering and Physical Sciences Research Council under grant agreement EP/S003657/2 (Brini) and by Dr Max Rössler, the Walter Haefner Foundation, the ETH Zürich Foundation, and the NSF grant DMS-2302117 (Bousseau).

2 Nef Looijenga pairs

We start off by establishing some general facts about the classical geometry of nef log Calabi–Yau (CY) surfaces. We first proceed to classify them in the smooth case, recall some basics of their birational geometry and the construction of toric models, and describe the structure of their pseudoeffective cone in preparation for the study of curve counts in them. We then end by introducing the notions of (quasi)tameness.

2.1 Classification

We start by giving the following definition.

Definition 2.1 An *l–component log CY surface with maximal boundary*, or *l–component Looijenga pair*, is a pair $Y(D) := (Y, D = D_1 + \cdots + D_l)$ consisting of a smooth rational projective surface Y and a singular nodal anticanonical divisor D that admits a decomposition $D = D_1 + \cdots + D_l$. We say that an l–component log CY surface is *nef* if $l \geq 2$ and each D_i is a smooth, irreducible and nef rational curve.

Examples of log CY surfaces arise when Y is a projective toric surface and D is the complement of the maximal torus orbit in Y; we call these pairs *toric*. By definition, if $Y(D)$ is nef, Y is a weak Fano surface together with a choice of distribution of the anticanonical degree amongst components D_i preserving the condition that $D_i \cdot C \geq 0$ for any effective curve C and all $i = 1, \dots, l$ with $l \geq 2$. We classify these by recalling some results of di Rocco [33]; see also [27, Section 2] and [28; 29].

Let $\mathrm{dP_r}$ be the surface obtained from blowing up $r \geq 1$ general points in $\mathbb{P}^2$. The Picard group of $\mathrm{dP_r}$ is generated by the hyperplane class H and the classes E_i of the exceptional divisors. The anticanonical class is $-K_{\mathrm{dP_r}} = 3H - \sum_{i=1}^r E_i$. Recall that a *line class* on $\mathrm{dP_r}$ is $l \in \mathrm{Pic}(\mathrm{dP_r})$ such that $l^2 = -1$ and $-K_{\mathrm{dP_r}} \cdot l = 1$; for $r \leq 5$ and up to permutation of the E_i, they are given by E_i, $H - E_1 - E_2$ or $2H - \sum_{i=1}^5 E_i$. Furthermore, for $n \geq 0$, denote by $\mathbb{F}_\mathrm{n}$ the n^{th} Hirzebruch surface. Its Picard group is of rank 2 generated by the sections C_{-n} (resp. C_n), with self-intersections $-n$ (resp. n), and by the fibre class f, subject to the relation that $C_{-n} + nf = C_n$. Note that $\mathbb{F}_0 \simeq \mathbb{P}^1 \times \mathbb{P}^1$ and $\mathbb{F}_1 \simeq \mathrm{dP}_1$ is the blowup of $\mathbb{P}^2$ in one point.

Lemma 2.1 [33] *Assume that $1 \leq r \leq 5$ and let $D \in \mathrm{Pic}(\mathrm{dP_r})$. Then D is nef if and only if*

(i) *for $r = 1$, $D \cdot l \geq 0$ for all line classes l and $D \cdot (H - E_1) \geq 0$,*

(ii) *for $5 \geq r \geq 2$, $D \cdot l \geq 0$ for all line classes l.*

2.1.1 $l = 2$ Let's start by setting $l = 2$. With the sole exception of $\mathrm{dP}_4(H, 2H - E_1 - E_2 - E_3 - E_4)$ versus $\mathrm{dP}_4(H - E_1, 2H - E_2 - E_3 - E_4)$, the next proposition shows that up to deformation and permutation of the factors, and assuming that D_1 and D_2 are nef, $Y(D)$ is determined by Y and the self-intersections $\{D_1^2, D_2^2\}$. We will consequently employ the shorthand notation $Y(D) \leftrightarrow Y(D_1^2, D_2^2)$ to indicate this, making precise which one is meant in the case of $\mathrm{dP}_4(0, 1)$.

Proposition 2.2 *Let $Y(D = D_1 + D_2)$ be a 2–component nef log CY surface. Then up to deformation and interchange of D_1 and D_2, $Y(D_1, D_2)$ is one of the following, abbreviated by $Y(D_1^2, D_2^2)$ except in cases* (4) *and* (5):

(1) $\mathbb{P}^2(1, 4)$,

(2) $\mathrm{dP_r}(1, 4 - r)$ *for* $1 \leq r \leq 3$,

(3) $\mathrm{dP_r}(0, 5-r)$ *for* $1 \le r \le 3$,
(4) $\mathrm{dP}_4(H, 2H - E_1 - E_2 - E_3 - E_4)$,
(5) $\mathrm{dP}_4(H - E_1, 2H - E_2 - E_3 - E_4)$,
(6) $\mathrm{dP}_5(0,0)$,
(7) $\mathbb{F}_0(0,4)$,
(8) $\mathbb{F}_0(2,2)$.

Proof A minimal model of Y is given by $\mathbb{P}^2$, $\mathbb{F}_0$ or $\mathbb{F}_n$ for $n \ge 2$. By assumption $-K_Y = D_1 + D_2$ is nef, ruling out $\mathbb{F}_n$ for $n > 2$. If $Y = \mathbb{F}_0$, then the stipulated decompositions of $-K_{\mathbb{F}_0}$ are immediate. If $\mathbb{F}_2$ is a minimal model of Y, then $Y = \mathbb{F}_2$. In this case, the only possible decomposition of $-K_{\mathbb{F}_2}$ into nef divisors is as $D_1 = C_{-2} + 2f = C_2$ and $D_2 = C_2$. The resulting pair $\mathbb{F}_2(2,2)$ is deformation equivalent to $\mathbb{F}_0(2,2)$; see the proof of Proposition 2.6.

Assume now that $\mathbb{P}^2$ is a minimal model of Y. If $Y = \mathbb{P}^2$, we are done. Otherwise, up to deformation, we may assume that $Y = \mathrm{dP}_r$. Since $-K_Y$ is nef, $r \le 9$. As D_1 and D_2 are nef, they are of the form $dH - \sum_{i=1}^r a_i E_i$ for $d \ge 1$ and $a_i \ge 0$. Applying Lemma 2.1, we find that the only nef decompositions are as follows:

- either $D_1 = H$, $D_2 = 2H - \sum_{i=1}^r E_i$ for $r \le 4$,
- or $D_1 = H - E_j$, $D_2 = 2H - \sum_{i \ne j}^r E_i$ for $r \le 5$.

They are all basepoint-free by [33] (see [27, Lemma 2.7]) and hence a general member will be smooth by Bertini. □

2.1.2 $l = 3$ Next, we classify the surfaces with $l = 3$ nef components. The shorthand notation $Y(D_1^2, D_2^2, D_3^2)$ is employed as in the previous section.

Proposition 2.3 *Let $Y(D = D_1 + D_2 + D_3)$ be a 3–component log CY surface with Y smooth and D_1, D_2 and D_3 nef. Then up to deformation and permutation of D_1, D_2 and D_3, $Y(D_1^2, D_2^2, D_3^2)$ is one of the following*:

(1) $\mathbb{P}^2(1,1,1)$,
(2) $\mathrm{dP}_1(1,1,0)$,
(3) $\mathrm{dP}_2(1,0,0)$,
(4) $\mathrm{dP}_3(0,0,0)$,
(5) $\mathbb{F}_0(2,0,0)$.

Proof A minimal model of Y is given by $\mathbb{P}^2$, $\mathbb{F}_0$ or $\mathbb{F}_n$ for $n \ge 2$. By assumption $-K_Y = D_1 + D_2 + D_3$ is nef, ruling out $\mathbb{F}_n$ for $n \ge 2$. For $\mathbb{P}^2$, the only possibility is to choose D_1, D_2, D_3 in class H. For $\mathbb{F}_0$, it is to choose $D_1 = H_1 + H_2$ the diagonal and $D_2 = H_1$, $D_3 = H_2$. Necessarily, all other surfaces are given by iterated blowups of the minimal models, keeping the divisors nef, leading to the list. As in the previous proposition, they are all basepoint-free and thus a general member will be smooth. □

2.1.3 $\boldsymbol{l \geq 4}$ For $l = 4$, a minimal model for Y is $\mathbb{F}_0$, for which the only possibility is given by D being its toric boundary. There are no other cases preserving nefness of the divisors. For 5 components or more, there are no surfaces keeping each divisor nef.

2.2 Toric models

We consider two basic operations on log CY surfaces $Y(D)$.

- Let $\widetilde{Y}$ be the blowup of Y at a node of D and let $\widetilde{D}$ be the preimage of D in $\widetilde{Y}$. Then the log CY surface $(\widetilde{Y}, \widetilde{D})$ is said to be a *corner blowup* of $Y(D)$.
- Let $\widetilde{Y}$ be the blowup of Y at a smooth point of D. Let $\widetilde{D}$ be the strict transform of D in $\widetilde{Y}$. Then the log CY surface $(\widetilde{Y}, \widetilde{D})$ is said to be an *interior blowup* of $Y(D)$.

A corner blowup does not change the complement $Y \setminus D$, whereas an interior blowup does; accordingly corner blowups do not change log Gromov–Witten invariants [3].

Definition 2.2 Let $\pi\colon Y(D) \longrightarrow \overline{Y}(\overline{D})$ be a sequence of interior blowups between log CY surfaces such that $\overline{Y}(\overline{D})$ is toric. Then π is said to be a *toric model* of $Y(D)$.

We will describe toric models by giving the fan of $(\overline{Y}, \overline{D})$ with *focus–focus singularities* on its rays. A focus–focus singularity on the ray corresponding to a toric divisor F encodes blowing up F at a smooth point. Each focus–focus singularity produces a wall and interactions of them create a scattering diagram Scatt($Y(D)$), as we discuss in Section 4.2.

Proposition 2.4 [53, Proposition 1.3] *Let $Y(D)$ be a log CY surface. Then there exist log CY surfaces $\widetilde{Y}(\widetilde{D})$ and $\overline{Y}(\overline{D})$, with the latter toric, and a diagram*

(2-1)
$$\begin{array}{ccc} & \widetilde{Y}(\widetilde{D}) & \\ {\scriptstyle\varphi}\swarrow & & \searrow{\scriptstyle\pi} \\ Y(D) & & \overline{Y}(\overline{D}) \end{array}$$

such that φ is a sequence of corner blowups and π is a toric model.

The diagrams as in (2-1) are far from unique, and they are related by cluster mutations [52]. Because of the invariance of log Gromov–Witten invariants by corner blowups, we can calculate the log Gromov–Witten invariants of $Y(D)$ on the scattering diagram Scatt($Y(D)$) associated to the toric model π.

2.3 The effective cone of curves

Given $Y(D)$ a nef log CY surface and $d \in A_1(Y)$, it will be convenient for the discussion in the foregoing sections to determine numerical conditions for d to be an element of the pseudoeffective cone. If $Y = \mathbb{F}_n$, $\mathrm{NE}(Y)$ is just the monoid generated by C_{-n} and f, so let us assume that $Y = \mathrm{dP_r}$. We will write a curve class d as $d_0\big(H - \sum_{i=1}^r E_i\big) + \sum_{i=1}^r d_i E_i$. If $\rho(Y) \geq 2$, then the extremal rays of the effective cone $\mathrm{NE}(Y)$ of Y are generated by extremal classes D with $D^2 \leq 0$, and in the case of del Pezzo surfaces these are the line and fibre classes described above. Using the classification [27, Examples 2.3 and 2.11], up to permutation of the E_i and E_j, we find the following lists of generators of extremal rays of $\mathrm{NE}(Y)$:

- If $r = 1$,

$$E_1, \quad H - E_1. \tag{2-2}$$

- If $2 \leq r \leq 4$,

$$E_i, \quad H - E_i, \quad H - E_i - E_j \text{ for } i \neq j. \tag{2-3}$$

- If $r = 5$,

$$E_i, \quad H - E_i, \quad H - E_i - E_j \text{ for } i \neq j, \quad 2H - \sum_{i=1}^{5} E_i. \tag{2-4}$$

Note that the effective cone is closed since it is generated by finitely many elements. The following proposition can be specialised to the del Pezzo surfaces dP_r for $r \leq 5$ by setting the corresponding d_i to 0 and removing the superfluous equations such as the last one, which only holds for $r = 5$.

Proposition 2.5 *A class $d = d_0\big(H - \sum_{i=1}^5 E_i\big) + \sum_{i=1}^5 d_i E_i$ of dP_5 is effective if and only if*

$$d_0 \geq 0, \quad d_i \geq 0, \quad d_i + d_j + d_k \geq d_0, \quad d_i + d_j + d_k + d_l \geq 2d_0, \quad 2d_i + \sum_{j \neq i} d_j \geq 3d_0, \tag{2-5}$$

where the i, j, k, l are always pairwise distinct.

The statement follows from the explicit description of the effective cone as generated by extremal rays. A direct calculation using the Polymake package in Macaulay2 computes the halfspaces defining the cone, yielding the above inequalities for the effective curves.

2.4 Tame and quasi-tame Looijenga pairs

The computation of curve-counting invariants of nef Looijenga pairs is strongly affected by the number l of smooth irreducible components of D and the positivity of D_i for $i = 1, \ldots, l$. We spell this out with the following definition, whose significance will be worked out in Sections 3.1 and 4.2.

Let $Y(D)$ be a nef Looijenga pair and let

$$E_{Y(D)} := \mathrm{Tot}\Bigl(\bigoplus_{i=1}^{l} \mathcal{O}_Y(-D_i)\Bigr) \tag{2-6}$$

be the total space of the direct sum of the dual line bundles to D_i for $i = 1, \dots, l$.

Definition 2.3 We call a nef log CY surface $(Y, D = D_1 + \cdots + D_l)$ *tame* if either $l > 2$ or $D_i^2 > 0$ for all i. A nef log CY surface $Y(D)$ is *quasi-tame* if $E_{Y(D)}$ is deformation equivalent to $E_{Y'(D')}$, with $Y'(D')$ tame.

We will use the abbreviated notation $E_{Y(D_1^2, D_2^2)}$ for the local Calabi–Yau fourfold $E_{Y(D_1+D_2)}$ associated by (2-6) to a 2–component log CY surface $Y(D_1^2, D_2^2)$ in the classification of Proposition 2.2. Quasitame pairs are classified by the following proposition.

Proposition 2.6 *The following varieties are deformation-equivalent*:

1. $E_{\mathbb{F}_0(0,4)}$, $E_{\mathbb{F}_0(2,2)}$ *and* $E_{\mathbb{F}_2(2,2)}$;
2. $E_{\mathrm{dP}_r(1,4-r)}$ *and* $E_{\mathrm{dP}_r(0,5-r)}$, *where* $1 \le r \le 4$.

Proof For the first part of the proposition, denote by H_1 and H_2 the two generators of the Picard group of $\mathbb{F}_0$ corresponding to the pullbacks of a point in $\mathbb{P}^1$ along $\mathrm{proj}_{1,2}\colon \mathbb{F}_0 \to \mathbb{P}^1$. The Euler sequence on $\mathbb{P}^1$, pulled back to $\mathbb{F}_0$ along proj_1 and tensored by $\mathcal{O}(-H_2)$, yields

$$0 \to \mathcal{O}(-2H_1 - H_2) \to \mathcal{O}(-H_1 - H_2) \oplus \mathcal{O}(-H_1 - H_2) \to \mathcal{O}(-H_2) \to 0. \tag{2-7}$$

This determines a family with general fibre the total space of $\mathcal{O}(-H_1 - H_2) \oplus \mathcal{O}(-H_1 - H_2)$ and special fibre the total space of $\mathcal{O}(-H_2) \oplus \mathcal{O}(-2H_1 - H_2)$, hence a deformation between $E_{\mathbb{F}_0(0,4)}$ and $E_{\mathbb{F}_0(2,2)}$.

Next, consider again the Euler sequence over $\mathbb{P}^1$,

$$0 \to \mathcal{O}(-2) \to \mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathcal{O} \to 0, \tag{2-8}$$

and the associated deformation of the total space of $\mathcal{O}(-1) \oplus \mathcal{O}(-1)$ into the total space of $\mathcal{O} \oplus \mathcal{O}(-2)$. Taking the projectivisation of this family yields a deformation between $\mathbb{F}_0$ and $\mathbb{F}_2$. In this deformation, $-H_1 - H_2$ specialises to $-C_2$. Taking twice the associated line bundles yields the deformation between $E_{\mathbb{F}_0(2,2)}$ and $E_{\mathbb{F}_2(2,2)}$.

To prove the second part, assume first that $r = 1$. We start with the relative (dual) Euler sequence for the fibration $\mathrm{dP}_1 \to \mathbb{P}^1$ with distinct sections with image H and E_1

$$0 \to \mathcal{O} \to \mathcal{O}(H) \oplus \mathcal{O}(E_1) \to \mathcal{O}(H + E_1) \to 0. \tag{2-9}$$

We tensor it with $\mathcal{O}(-2H)$ to obtain

$$0 \to \mathcal{O}(-2H) \to \mathcal{O}(-H) \oplus \mathcal{O}(-2H + E_1) \to \mathcal{O}(-H + E_1) \to 0. \tag{2-10}$$

This determines a family with general fibre the total space of $\mathcal{O}(-H) \oplus \mathcal{O}(-2H + E_1)$ and special fibre the total space of $\mathcal{O}(-2H) \oplus \mathcal{O}(-H + E_1)$, hence a deformation between $E_{\mathrm{dP}_1(1,3)}$ and $E_{\mathrm{dP}_1(0,4)}$.

$Y(D)$	l	K_Y^2	D_1	D_2	D_3	D_4	tame	quasi-tame
$\mathbb{P}^2(1,4)$	2	9	H	$2H$	–	–	✓	✓
$\mathbb{F}_0(2,2)$	2	8	H_1+H_2	H_1+H_2	–	–	✓	✓
$\mathbb{F}_0(0,4)$	2	8	H_1	H_1+2H_2	–	–	✗	✓
$\mathrm{dP}_1(1,3)$	2	8	H	$2H-E_1$	–	–	✓	✓
$\mathrm{dP}_1(0,4)$	2	8	$H-E_1$	$2H$	–	–	✗	✓
$\mathrm{dP}_2(1,2)$	2	7	H	$2H-E_1-E_2$	–	–	✓	✓
$\mathrm{dP}_2(0,3)$	2	7	$H-E_1$	$2H-E_2$	–	–	✗	✓
$\mathrm{dP}_3(1,1)$	2	6	H	$2H-E_1-E_2-E_3$	–	–	✓	✓
$\mathrm{dP}_3(0,2)$	2	6	$H-E_1$	$2H-E_2-E_3$	–	–	✗	✓
$\mathrm{dP}_4(1,0)$	2	5	H	$2H-E_1-E_2-E_3-E_4$	–	–	✗	✗
$\mathrm{dP}_4(0,1)$	2	5	$H-E_1$	$2H-E_2-E_3-E_4$	–	–	✗	✗
$\mathrm{dP}_5(0,0)$	2	4	$H-E_1$	$2H-E_2-E_3-E_4-E_5$	–	–	✗	✗
$\mathbb{P}^2(1,1,1)$	3	9	H	H	H	–	✓	✓
$\mathbb{F}_0(2,0,0)$	3	8	H_1+H_2	H_1	H_2	–	✓	✓
$\mathrm{dP}_1(1,1,0)$	3	8	H	H	$H-E_1$	–	✓	✓
$\mathrm{dP}_2(1,0,0)$	3	7	H	$H-E_1$	$H-E_2$	–	✓	✓
$\mathrm{dP}_3(0,0,0)$	3	6	$H-E_1$	$H-E_2$	$H-E_3$	–	✓	✓
$\mathbb{F}_0(0,0,0,0)$	4	8	H_1	H_2	H_1	H_2	✓	✓

Table 1: Classification of smooth nef Looijenga pairs.

Dually, we have

$$0 \to \mathrm{H}^0(\mathcal{O}(H-E_1)) \to \mathrm{H}^0(\mathcal{O}(H)) \oplus \mathrm{H}^0(\mathcal{O}(2H-E_1)) \to \mathrm{H}^0(\mathcal{O}(2H)), \tag{2-11}$$

and a section of $\mathcal{O}(2H-E_1)$ in the general fibre gives a section of $\mathcal{O}(2H)$ in the special fibre. Hence we have a divisor $\mathcal{D}$ in the family in class $2H-E_1$ for the general fibre and of class $2H$ for the special fibre. Blowing up a general point of $\mathcal{D}$ in the family gives a deformation between $E_{\mathrm{dP}_2(1,2)}$ and $E_{\mathrm{dP}_2(0,3)}$. Iterating the process, we obtain the desired deformations. □

We summarise the discussion of this section in Table 1. There are 18 smooth deformation types of nef Looijenga pairs in total, 11 of which are tame and 15 of which are quasi-tame. The three non-quasi-tame cases occur when Y is a del Pezzo surface of degree 5 or less.

3 Local Gromov–Witten theory

3.1 1–Pointed local Gromov–Witten invariants

In this section, we provide general formulas for the Gromov–Witten invariants with point insertions of toric Fano varieties in any dimension twisted by a sum of concave line bundles. For the remainder of this section, let Y be an n–dimensional smooth projective variety of Picard rank r, let $D = D_1 + \cdots + D_l \in A_{n-1}(Y)$ with $D \in |-K_Y|$ and each D_i smooth and irreducible, and let d be a D–convex curve class.

Let $E_{Y(D)} := \mathrm{Tot}\big(\bigoplus_{i=1}^{l}(\mathcal{O}_Y(-D_i))\big)$ be as in (2-6) and let $\pi_Y\colon E_{Y(D)} \to Y$ be the natural projection. Since d is D–convex, the moduli space $\overline{\mathrm{M}}_{0,m}(E_{Y(D)}, d)$ of genus-zero m–marked stable maps $[f\colon C \to E_{Y(D)}]$ with $f_*([C]) = d \in \mathrm{H}_2(Y,\mathbb{Z})$ is scheme-theoretically the moduli stack $\overline{\mathrm{M}}_{0,m}(Y,d)$ of stable maps to the base Y, as every stable map to the total space factors through the zero section $Y \hookrightarrow E_{Y(D)}$. In particular, $\overline{\mathrm{M}}_{0,m}(E_{Y(D)}, d)$ is proper. Consider the universal curve $\pi\colon \mathcal{C} \to \overline{\mathrm{M}}_{0,m}(Y,d)$, and denote by $f\colon \mathcal{C} \to Y$ the universal stable map. Then $\mathrm{H}^0(C, f^*\mathcal{O}_Y(-D_i)) = 0$ and we have obstruction bundles $\mathrm{Ob}_{D_j} := R^1\pi_* f^*\mathcal{O}_Y(-D_j)$, of rank $d \cdot D_j - 1$ with fibre $\mathrm{H}^1(C, f^*\mathcal{O}_Y(-D_j))$ over a stable map $[f\colon C \to Y]$. The virtual fundamental class on $\overline{\mathrm{M}}_{0,m}(E_{Y(D)}, d)$ is defined by intersecting the virtual fundamental class on $\overline{\mathrm{M}}_{0,m}(Y,d)$ with the top Chern class of $\bigoplus_j \mathrm{Ob}_{D_j}$:

$$[\overline{\mathrm{M}}_{0,m}(E_{Y(D)}, d)]^{\mathrm{vir}} := c_{\mathrm{top}}(\mathrm{Ob}_{D_1}) \cap \cdots \cap c_{\mathrm{top}}(\mathrm{Ob}_{D_l}) \cap [\overline{\mathrm{M}}_{0,m}(Y,d)]^{\mathrm{vir}} \in \mathrm{H}_{m+l-1}(\overline{\mathrm{M}}_{0,m}(Y,d),\mathbb{Q}). \tag{3-1}$$

There are tautological classes $\psi_i := c_1(L_i)$, where L_i is the i^{th} tautological line bundle on $\overline{\mathrm{M}}_{0,m}(Y,d)$ whose fibre at $[f\colon (C, x_1, \dots, x_m) \to Y]$ is the cotangent line of C at x_i, and we denote by ev_i the evaluation maps at the i^{th} marked point. For an effective D–convex curve class $d \in \mathrm{H}_2(Y,\mathbb{Z})$, genus-zero local Gromov–Witten invariants of $E_{Y(D)}$ with point insertions on the base are defined as

$$N_{0,d}^{\mathrm{loc}}(Y(D)) := \int_{[\overline{\mathrm{M}}_{0,l-1}(E_{Y(D)},d)]^{\mathrm{vir}}} \prod_{j=1}^{l-1} \mathrm{ev}_j^*(\pi_Y^*[\mathrm{pt}_Y]), \tag{3-2}$$

$$N_{0,d}^{\mathrm{loc},\psi}(Y(D)) := \int_{[\overline{\mathrm{M}}_{0,1}(E_{Y(D)},d)]^{\mathrm{vir}}} \mathrm{ev}_1^*(\pi_Y^*[\mathrm{pt}_Y]) \cup \psi_1^{l-2}, \tag{3-3}$$

which we think of as the virtual counts of curves through $l-1$ points (resp. 1–point with a ψ–condition) on the zero section of the vector bundle $E_{Y(D)}$.

Since D is anticanonical, $E_{Y(D)}$ is a noncompact Calabi–Yau $(n+l)$–fold. The case $n + l = 3$ has been the main focus in the study of local mirror symmetry, and as such it has been abundantly studied in the literature [26]. It turns out that the lesser studied situation when $n + l > 3$ has a host of simplifications, often leading to closed-form expressions for (3-2)–(3-3). We start by fixing some notation which will be of further use throughout this section. Let $T \simeq (\mathbb{C}^\star)^l \circlearrowright E_{Y(D)}$ be the fibrewise torus action and denote by $\lambda_i \in \mathrm{H}(BT)$, with $i = 1, \dots, l$, its equivariant parameters. Let $\{\phi_\alpha\}_\alpha$ be a graded $\mathbb{C}$–basis for the nonequivariant cohomology of the image of the zero section $Y \hookrightarrow E_{Y(D)}$ with $\deg\phi_\alpha \le \deg\phi_{\alpha+1}$; in particular, $\phi_1 = \mathbf{1}_{\mathrm{H}(Y)}$. Its elements have canonical lifts $\phi_\alpha \to \varphi_\alpha$ to T–equivariant cohomology forming a $\mathbb{C}(\lambda_1, \dots, \lambda_l)$ basis for $\mathrm{H}_T(E_{Y(D)})$. The latter is furthermore endowed with a perfect pairing

$$\eta_{E_{Y(D)}}(\varphi_\alpha, \varphi_\beta) := \int_Y \frac{\phi_\alpha \cup \phi_\beta}{\bigcup_i \mathrm{e}_T(\mathcal{O}_Y(-D_i))}, \tag{3-4}$$

with e_T denoting the T–equivariant Euler class. In what follows, we will indicate by $\eta^{-1}_{E_{Y(D)}}(\varphi_\alpha, \varphi_\beta)$ the inverse of the Gram matrix (3-4).

Let now $\tau \in \mathrm{H}_T(E_{Y(D)})$. The J–function of $E_{Y(D)}$ is the formal power series

$$J_{\mathrm{big}}^{E_{Y(D)}}(\tau, z) := z + \tau + \sum_{d \in \mathrm{NE}(Y)} \sum_{n \in \mathbb{Z}^+} \sum_{\alpha,\beta} \frac{\eta^{-1}_{E_{Y(D)}}(\varphi_\alpha, \varphi_\beta)}{n!} \left\langle \tau, \dots, \tau, \frac{\varphi_\alpha}{z-\psi} \right\rangle^{E_{Y(D)}}_{0,n+1,d} \varphi_\beta, \tag{3-5}$$

where we employed the usual correlator notation for GW invariants,

$$\langle \tau_1 \psi_1^{k_1}, \dots, \tau_n \psi_n^{k_n} \rangle^{E_{Y(D)}}_{0,n,d} := \int_{[\overline{\mathrm{M}}_{0,m}(E_{Y(D)},d)]^{\mathrm{vir}}} \prod_i \mathrm{ev}_i^*(\tau_i) \psi_i^{k_i}. \tag{3-6}$$

Restriction to $t = \sum_{i=1}^{r+1} t_i \varphi_i$ and use of the divisor axiom gives the small J–function

$$J_{\mathrm{small}}^{E_{Y(D)}}(t, z) := z \mathrm{e}^{\sum t_i \varphi_i / z} \left(1 + \sum_{d \in \mathrm{NE}(Y)} \sum_{\alpha,\beta} \eta^{-1}(\varphi_\alpha, \varphi_\beta)_{E_{Y(D)}} \mathrm{e}^{t(d)} \left\langle \frac{\varphi_\alpha}{z(z-\psi_1)} \right\rangle^{E_{Y(D)}}_{0,1,d} \varphi_\beta \right). \tag{3-7}$$

Lemma 3.1 *Suppose that Y is a toric Fano variety and either $n+l=4$ and D_i is ample for all i, or $n+l>4$ and D_i is nef for all i. Let $\mathcal{T} := \{T_i \in A_{\dim Y-1}(Y)\}_{i=1}^{n+r}$ be the collection of its prime toric divisors, and $\bigsqcup_{i=1}^m S_i = \{1, \dots, n+r\}$ a length-m partition of $n+r$ such that $D_i := \sum_{j \in S_i} T_j$. For an effective curve class $d \in \mathrm{NE}(Y)$, write $\mathsf{d}_i := d \cdot D_i$ and $\mathsf{t}_i := d \cdot T_i$ for its intersection multiplicities with the nef divisors D_i and the toric divisors T_i, respectively. Then*

$$N_{0,d}^{\mathrm{loc},\psi}(Y(D_1 + \dots + D_l)) = \frac{(-1)^{\sum_{i=1}^l (\mathsf{d}_i - 1)}}{\prod_{i=1}^l \mathsf{d}_i} \prod_{i=1}^l \binom{\mathsf{d}_i}{\{\mathsf{t}_j\}_{j \in S_i}}, \tag{3-8}$$

where

$$\binom{k}{\{i_j\}_{j=1}^m} = \frac{k!}{\prod_{j=1}^m i_j!}$$

is the multinomial coefficient.

Proof By (3-3) and (3-7), we have

$$N_{0,d}^{\mathrm{loc},\psi}(Y(D)) = \sum_\beta \eta(\varphi_{\overline{\alpha}}, \varphi_\beta)_{E_{Y(D)}} [z^{1-l} \mathrm{e}^{t(d) + \sum t_i \varphi_i / z} \varphi^\beta] J_{\mathrm{small}}^{E_{Y(D)}}(t, z), \tag{3-9}$$

where $\overline{\alpha}$ is defined by $\varphi_{\overline{\alpha}} = [\mathrm{pt}]$. From (3-4), we have $\eta(\varphi_{\overline{\alpha}}, \varphi_\beta)_{E_{Y(D)}} = \delta_{\overline{\alpha}1} \prod_{i=1}^l \lambda_i^{-1}$, hence

$$N_{0,d}^{\mathrm{loc},\psi}(Y(D)) = \frac{1}{\prod_{i=1}^l \lambda_i} [z^{1-l} \mathrm{e}^{t(d) + \sum t_i \varphi_i / z} \mathbf{1}_{H_T(Y)}] J_{\mathrm{small}}^{E_{Y(D)}}(t, z). \tag{3-10}$$

The right-hand side can be computed by Givental-style toric mirror theorems. Let $\theta_a := T_a^\vee \in \mathrm{H}^2(Y)$ be the Poincaré dual class of the a^{th} toric divisor of Y, let $\kappa_i := c_1(\mathcal{O}(-D_i))$ be the T–equivariant Chern class of D_i, and let y_i with $i = 1, \dots, r+1$ be variables in a formal disk around the origin. Writing

$(x)_n := \Gamma(x+n)/\Gamma(x)$ for the Pochhammer symbol of (x,n) with $n \in \mathbb{Z}$, the I–function of $E_{Y(D)}$ is the $\mathrm{H}_T(E_{Y(D)})$–valued Laurent series

$$I^{E_{Y(D)}}(y,z) := z + \prod_i y_i^{\varphi_i/z} \sum_{0 \neq d \in \mathrm{NE}(Y)} \prod_i y_i^{d_i} z^{1-l} \frac{\prod_i \kappa_i (\kappa_i/z+1)_{\mathsf{d}_i - 1}}{\prod_a (\theta_a/z+1)_{\mathsf{t}_a}}, \tag{3-11}$$

and its mirror map is their formal $\mathcal{O}(z^0)$ coefficient,

$$\tilde{t}^i_{E_{Y(D)}}(y) := [z^0 \varphi_i] I^{E_{Y(D)}}(y,z). \tag{3-12}$$

Then [46; 31; 30]

$$J^{E_{Y(D)}}_{\text{small}}(\tilde{t}_{E_{Y(D)}(D)}(y), z) = I^{E_{Y(D)}}(y,z). \tag{3-13}$$

Inspecting (3-11) shows that if either $n+l > 4$, or $n+l = 4$ and D_i is ample, the mirror map does not receive quantum corrections:

$$\tilde{t}^i_{E_{Y(D)}}(y) = \log y_i. \tag{3-14}$$

Therefore, under the assumptions of the lemma,

$$\begin{aligned} N^{\mathrm{loc},\psi}_{0,d}(Y(D)) &= \frac{1}{\prod_{i=1}^l \lambda_i} \Big[z^{1-l} \prod_i y_i^{d_i} \prod_i y_i^{\varphi_i/z} \mathbf{1}_{H_T(Y(D))} \Big] I^{E_{Y(D)}}(y,z) \\ &= \frac{1}{\prod_{i=1}^l \lambda_i} \Big[z^{1-l} \prod_i y_i^{d_i} \Big] I^{E_{Y(D)}}(y,z) \Big|_{\varphi_\alpha \to 0}. \end{aligned} \tag{3-15}$$

The claim then follows by substituting $\theta_a|_{\varphi_\alpha \to 0} = 0$ and $\kappa_i|_{\varphi_\alpha \to 0} = -\lambda_i$ into (3-11). □

3.1.1 Quasitame Looijenga pairs Let us now go back to the case of log CY surfaces and specialise the discussion in the previous section to $Y(D)$ a tame Looijenga pair. The key observation in the proof of Lemma 3.1 was that no contributions to the $\mathcal{O}(z^0)$ Laurent coefficient of the I–function could possibly come from any stable maps in any degrees, which is automatic for $n+l > 4$, and requires that $\mathsf{d}_i > 0$ when $n+l = 4$. We can in fact partly relax the condition that D_i is ample by just requiring by *fiat* that no curves with $\mathsf{d}_i = d \cdot D_i = 0$ contribute to the mirror map. A direct calculation from (3-11) shows that in the case of nef log CY surfaces with Y a Fano surface, this relaxed assumption coincides with $Y(D)$ being tame as in Definition 2.1. Since, by Proposition 2.2, Y is toric for all tame cases, Lemma 3.1 computes (3-3) for all of them.

Example 3.1 Let $Y(D) = \mathbb{P}^2(1,4)$. Then Lemma 3.1 gives for the degree-d local invariants of the projective plane

$$N^{\mathrm{loc},\psi}_{0,d}(Y(D)) = N^{\mathrm{loc}}_{0,d}(Y(D)) = \frac{(-1)^d}{2d^2} \binom{2d}{d}. \tag{3-16}$$

This recovers a direct localisation calculation by Klemm and Pandharipande in [68, Proposition 2].

Example 3.2 Let $Y(D)=\mathbb{F}_0(2,2)$ and write $d=d_1H_1+d_2H_2$. Lemma 3.1 yields

$$N^{\mathrm{loc},\psi}_{0,d}(Y(D))=N^{\mathrm{loc}}_{0,d}(Y(D))=\frac{1}{(d_1+d_2)^2}\binom{d_1+d_2}{d_1}^2 \tag{3-17}$$

as in [68, Proposition 3].

Moreover, if $Y(D)$ is a quasi-tame Looijenga pair, the Calabi–Yau vector bundle $E_{Y(D)}$ is deformation equivalent to $E_{Y'(D')}$ for some tame Looijenga pair by definition. It therefore carries the same local Gromov–Witten theory, and the calculation of $N^{\mathrm{loc},\psi}_{0,d}(Y(D))=N^{\mathrm{loc},\psi}_{0,d}(Y'(D'))$ from Lemma 3.1 extends immediately to these cases as well.

3.1.2 Nonquasi-tame Looijenga pairs Lemma 3.1 cannot be immediately extended to non-quasi-tame pairs $Y(D)$, as Y is not toric and $E_{Y(D)}$ does not deform to $E_{Y'(D')}$ for tame $Y'(D')$. We will proceed by exhibiting a closed-form solution for the case of lowest anticanonical degree $Y(D)=\mathrm{dP}_5(0,0)$. This recovers all other cases with $l=2$ by blowing down, as per the following.

Proposition 3.2 (blowup formula for local GW invariants) *Let $Y(D)$ be an l–component log CY surface. Let $\pi\colon Y'(D')\to Y(D)$ be the l–component log CY surface obtained by an interior blowup at a general point of D with exceptional divisor E. Let d be a curve class of $Y(D)$ and let $d':=\pi^*d$. Then*

$$N^{\mathrm{loc}}_{0,d}(Y(D))=N^{\mathrm{loc}}_{d'}(Y'(D'))\quad\text{and}\quad N^{\mathrm{loc},\psi}_{0,d}(Y(D))=N^{\mathrm{loc},\psi}_{d'}(Y'(D')). \tag{3-18}$$

Proof By [87, Proposition 5.14],

$$\overline{\pi}_*[\overline{\mathrm{M}}_{0,m}(Y',d')]^{\mathrm{vir}}=[\overline{\mathrm{M}}_{0,m}(Y,d)]^{\mathrm{vir}}, \tag{3-19}$$

where $\overline{\pi}$ is the morphism between the moduli spaces induced by π. Since $E\cdot d'=0$,

$$\overline{\pi}_*[\overline{\mathrm{M}}_{0,m}(E_{Y'(D')},d')]^{\mathrm{vir}}=[\overline{\mathrm{M}}_{0,m}(E_{Y(D)},d)]^{\mathrm{vir}}. \tag{3-20}$$

□

Theorem 3.3 *With notation as in Proposition 2.5, we have*

$$\begin{aligned} N^{\mathrm{loc}}_{0,d}(\mathrm{dP}_5(0,0))= &\sum_{j_1,\dots,j_4=0}^{\infty}\Bigg[\frac{(-1)^{d_1+d_2+d_3+d_4+d_5}(d_1+d_2+d_3+d_4+d_5-3d_0+j_1+j_2-1)!}{j_1!\,j_2!\,j_3!\,j_4!\,(d_1+d_2+d_4-2d_0+j_1)!\,(-d_1+d_0-j_1-j_2)!\,(-d_3+d_5+j_4)!}\\ &\times\frac{(d_1+d_4-d_0+j_1+j_3-1)!\,(d_1+d_5-d_0+j_2+j_4-1)!\,(d_4+d_5-d_0+j_3+j_4-1)!}{(d_1+d_3+d_5-2d_0+j_2)!\,(-d_4+d_0-j_1-j_3)!\,(-d_2+d_4+j_3)!\,(-d_5+d_0-j_2-j_4)!}\\ &\times\frac{1}{(d_2+d_3-d_0-j_3-j_4)!\,((d_1+d_4+d_5-2d_0+j_1+j_2+j_3+j_4-1)!)^2}\Bigg]. \end{aligned} \tag{3-21}$$

Sketch of the proof The strategy of the proof runs by deforming dP_5 to the blowup of $\mathbb{F}_0$ at four toric points, which is only weak Fano but allows to work torically along the lines of Lemma 3.1 at the price of extending the pseudoeffective cone by four generators of self-intersection -2. These contribute nontrivially to the mirror map, alongside curves with zero intersections with the boundary divisors $D_1 = H - E_1$ and $D_2 = 2H - \sum_{i\neq 1} E_i$. However, the mirror map turns out to be algebraic, and furthermore it remarkably has a closed-form rational inverse, leading to the final result (3-21). Full details are given in Appendix A. □

Remark 3.4 The final expression (3-21) is significantly more involved than (3-8), to which it reduces when blowing down to the quasi-tame del Pezzo cases dP_k for $k \le 3$ by setting $d_i = d_0$ for all $i \ge k+1$ using (3-18), since then only the summand with $j_i = 0$ for all i survives. It is also noteworthy that, while the summands in (3-21) are not symmetric under permutation of the degrees $\{d_2, d_3, d_4, d_5\}$, the final sum is highly nonobviously warranted to be S_4–invariant since the left-hand side is,[9] and we verified this explicitly in low degrees. The BPS invariants arising from (3-21) should also be integers, and we checked this is indeed the case for a large sample of nonprimitive classes with multicovers of order up to 11.

3.2 Multipointed local GW invariants

The primary multipoint invariants (3-2) of nef Looijenga pairs with $l > 2$ can be reconstructed from the descendent single-insertion invariants (3-3). We shall show how this arises by combining the associativity of the quantum product with the vanishing of quantum corrections for particular classes.

3.2.1 $\boldsymbol{l = 3}$ It suffices to compute the invariants for the case of maximal Picard rank, $dP_3(0,0,0)$, from which the other $l = 3$ cases can be recovered by blowing down.

Theorem 3.5 *With notation as in Proposition 2.5, we have*

$$N^{\mathrm{loc}}_{0,d}(dP_3(0,0,0)) = (d_0^2 - d_1d_0 - d_2d_0 - d_3d_0 + d_1d_2 + d_1d_3 + d_2d_3)N^{\mathrm{loc},\psi}_{0,d}(dP_3(0,0,0)), \tag{3-22}$$

$$N^{\mathrm{loc},\psi}_{0,d}(dP_3(0,0,0)) = \frac{(-1)^{d_1+d_2+d_3+1}(d_1-1)!\,(d_2-1)!\,(d_3-1)!}{(d_1+d_2-d_0)!\,(d_1+d_3-d_0)!\,(d_2+d_3-d_0)!\,(d_0-d_1)!\,(d_0-d_2)!\,(d_0-d_3)!}. \tag{3-23}$$

Proof In the notation of the proof of Lemma 3.1, for $i, j = 2, \dots, 5$ the components of the small J–function of $E_{dP_3(0,0,0)}$ satisfy the quantum differential equations

$$z\nabla_{\varphi_i}\nabla_{\varphi_j} J^{E_{dP_3(0,0,0)}}_{\mathrm{small}}(t,z) = \nabla_{\varphi_i \star_t \varphi_j} J^{E_{dP_3(0,0,0)}}_{\mathrm{small}}(t,z), \tag{3-24}$$

where $\alpha \star_t \beta$ denotes the small quantum cohomology product, and the cohomology classes

$$\{\varphi_1 = \mathbf{1}_{H_T(E_{dP_3(0,0,0)})}, \dots, \varphi_5\}$$

[9] There is an obvious S_5 symmetry under permutation of the exceptional classes E_i in Y, which is reduced to an S_4 symmetry in the degrees (d_2, d_3, d_4, d_5) in $E_{Y(D)}$ by the splitting $D_1 = H - E_1$, $D_2 = 2H - E_2 - E_3 - E_4 - E_5$.

are denoted as in the proof of Lemma 3.1. We take $\{\varphi_i\}_{i=2}^5$ to be the basis elements of $\mathrm{H}_T(E_{\mathrm{dP}_3(0,0,0)})$ given by lifts to T–equivariant cohomology of the integral Kähler classes dual to $\{C_i \in \mathrm{H}_2(\mathrm{dP}_3,\mathbb{Z})\}_i$ with $C_{i+1}=E_i$ for $i=1,2,3$, and $C_5=H-E_1-E_2-E_3$, and an effective curve will be written $d=d_0C_5+\sum_{i=1}^3 d_iC_{i+2}$. From the proof of Lemma 3.1, the small J–function in the tame setting equates the I–function,

$$(3\text{-}25)\quad J_{\mathrm{small}}^{E_{\mathrm{dP}_3(0,0,0)}}(t,z)= \sum_{d_i>0} \mathrm{e}^{\sum_{i=0}^3 t_{i+2}d_i}\Bigg[\frac{(-1)^{d_1+d_2+d_3}(\phi_2-\lambda_1)(\phi_3-\lambda_2)}{z^2\left(\frac{z+\phi_2+\phi_3-\phi_5}{z}\right)_{-d_0+d_1+d_2}\left(\frac{z+\phi_2+\phi_4-\phi_5}{z}\right)_{-d_0+d_1+d_3}} \frac{(\phi_4-\lambda_3)\left(\frac{z-\lambda_1+\phi_2}{z}\right)_{d_1-1}\left(\frac{z-\lambda_2+\phi_3}{z}\right)_{d_2-1}\left(\frac{z-\lambda_3+\phi_4}{z}\right)_{d_3-1}}{\left(\frac{z+\phi_3+\phi_4-\phi_5}{z}\right)_{-d_0+d_2+d_3}\left(\frac{z-\phi_2+\phi_5}{z}\right)_{d_0-d_1}\left(\frac{z-\phi_3+\phi_5}{z}\right)_{d_0-d_2}\left(\frac{z-\phi_4+\phi_5}{z}\right)_{d_0-d_3}}\Bigg].$$

By (3-7), the small quantum product can be computed from the $\mathcal{O}(z^{-1})$ formal Taylor coefficient of (3-25) as

$$(3\text{-}26)\qquad \varphi_i\star_t\varphi_j=\sum_\alpha \varphi_\alpha[z^{-1}\varphi_\alpha]\partial^2_{t_it_j}J_{\mathrm{small}}^{E_{\mathrm{dP}_3(0,0,0)}}(t,z).$$

Inspection of (3-25) shows that the right-hand side receives quantum corrections of the form $1/n^2$ from curves with either $d_i=\delta_{ij}n$ or $d_i=(1-\delta_{ij})n$ and $j\neq 0$, $n\in\mathbb{N}^+$, with vanishing contributions in all other degrees. This implies that

$$(3\text{-}27)\qquad (\partial^2_{t_5}-\partial_{t_2}\partial_{t_5}-\partial_{t_3}\partial_{t_5}-\partial_{t_4}\partial_{t_5}+\partial_{t_2}\partial_{t_3}+\partial_{t_3}\partial_{t_4}+\partial_{t_2}\partial_{t_4})[z^{-1}]J_{\mathrm{small}}^{E_{\mathrm{dP}_3(0,0,0)}}(t,z)=0,$$

which amounts to

$$(3\text{-}28)\qquad \varphi_5\star_t\varphi_5-\sum_{j=2}^4\varphi_5\star_t\varphi_i+\sum_{j>i=2}^5\varphi_i\star_t\varphi_j=\varphi_5\cup\varphi_5-\sum_{j=2}^4\varphi_5\cup\varphi_i+\sum_{j>i=2}^4\varphi_i\cup\varphi_j.$$

It is immediate to verify that the right-hand side is the Poincaré dual of the point class. Therefore, from (3-24),

$$(3\text{-}29)\quad N^{\mathrm{loc}}_{0,d}(\mathrm{dP}_3(0,0,0))=(d_0^2-d_1d_0-d_2d_0-d_3d_0+d_1d_2+d_1d_3+d_2d_3)N^{\mathrm{loc},\psi}_{0,d}(\mathrm{dP}_3(0,0,0)),$$

and the second equation in the statement follows by Lemma 3.1. □

3.2.2 $l=4$ In this case D is the toric boundary, and the invariants $N^{\mathrm{loc}}_{0,d}(\mathbb{F}_0(0,0,0,0))$ were computed in [18] by a strategy similar to that of Theorem 3.5. The final result is the following proposition.

Proposition 3.6 [18, Theorem 3.1, Corollary 6.4]

$$(3\text{-}30)\qquad N^{\mathrm{loc}}_{0,d}(\mathbb{F}_0(0,0,0,0))=d_1^2d_2^2N^{\mathrm{loc},\psi}_{0,d}(\mathbb{F}_0(0,0,0,0))=1.$$

$Y(D)$	$N^{\mathrm{loc},\psi}_{0,d}$	$N^{\mathrm{loc}}_{0,d}/N^{\mathrm{loc},\psi}_{0,d}$
$\mathbb{P}^2(1,4)$	$\frac{1}{2d^2}\binom{2d}{d}$	1
$\mathbb{F}_0(2,2)$ $\mathbb{F}_0(0,4)$	$\left(\frac{1}{d_1+d_2}\binom{d_1+d_2}{d_1}\right)^2$	1
$\mathrm{dP}_1(1,1)$ $\mathrm{dP}_2(0,4)$	$\frac{(-1)^{d_1}}{d_0(d_1+d_0)}\binom{d_0}{d_1}\binom{d_1+d_0}{d_0}$	1
$\mathrm{dP}_2(1,2)$ $\mathrm{dP}_2(0,3)$	$\frac{(-1)^{d_0+d_1+d_2}}{d_0(d_1+d_2)}\binom{d_0}{d_1}\binom{d_0}{d_2}\binom{d_1+d_2}{d_0}$	1
$\mathrm{dP}_3(1,1)$ $\mathrm{dP}_3(0,2)$	$\frac{(-1)^{d_1+d_2+d_3}(d_0-1)!(d_1+d_2+d_3-d_0-1)!}{(d_0-d_1)!(d_0-d_2)!(d_0-d_3)!(d_1+d_2-d_0)!(d_1+d_3-d_0)!(d_2+d_3-d_0)!}$	1
$\mathrm{dP}_4(1,0)$ $\mathrm{dP}_4(0,1)$	(3-21)$\big\vert_{d_5\to d_0}$	1
$\mathrm{dP}_5(0,0)$	(3-21)	1
$\mathbb{P}^2(1,1,1)$	$\frac{(-1)^{d+1}}{d^3}$	d^2
$\mathbb{F}_0(2,0,0)$	$-\frac{1}{d_1d_2(d_1+d_2)}\binom{d_1+d_2}{d_2}$	d_1d_2
$\mathrm{dP}_1(1,1,0)$	$\frac{(-1)^{d_1+1}}{d_0^2d_1}\binom{d_0}{d_1}$	d_1d_0
$\mathrm{dP}_2(1,0,0)$	$\frac{(-1)^{d_0+d_1+d_2+1}}{d_0d_1d_2}\binom{d_0}{d_1}\binom{d_1}{d_0-d_2}$	d_1d_2
$\mathrm{dP}_3(0,0,0)$	$\frac{(-1)^{d_1+d_2+d_3+1}}{d_1d_2d_3}\binom{d_1}{d_0-d_2}\binom{d_2}{d_0-d_3}\binom{d_3}{d_0-d_1}$	$d_0^2-(d_1+d_2+d_3)d_0$ $+d_1d_2+d_1d_3+d_2d_3$
$\mathbb{F}_0(0,0,0,0)$	1	$d_1^2d_2^2$

Table 2: Local Gromov–Witten invariants of nef Looijenga pairs.

This concludes the calculation of local invariants with point insertions for nef Looijenga pairs. We collate the results in Table 2.

4 Log Gromov–Witten theory

4.1 Log Gromov–Witten invariants of maximal tangency

Let $Y(D)$ be an l–component log CY surface with maximal boundary. We endow Y with the divisorial log structure coming from D. The log structure is used to impose tangency conditions along the components D_j of D. In this paper we will be looking at genus g stable maps into Y of class $d \in \mathrm{H}_2(Y,\mathbb{Z})$

that meet each component D_j in one point of maximal tangency $d \cdot D_j$. The appropriate moduli space $\overline{\mathrm{M}}^{\log}_{g,m}(Y(D), d)$ of maximally tangent basic stable log maps was constructed in all generality in [58; 25; 1].

There are tautological classes $\psi_i := c_1(L_i)$ for L_i the i^{th} tautological line bundle on $\overline{\mathrm{M}}^{\log}_{g,m}(Y(D), d)$ whose fibre at $[f : (C, x_1, \dots, x_m) \to Y]$ is the cotangent line of C at x_i. Let ev_i be the evaluation map at the i^{th} marked point, and for $\pi : \mathcal{C} \to \overline{\mathrm{M}}^{\log}_{g,m}(Y(D), d)$ the universal curve with relative dualising sheaf ω_π, denote by $\mathbb{E} := \pi_* \omega_\pi$ the Hodge bundle, which is a rank g vector bundle on $\overline{\mathrm{M}}^{\log}_{g,m}(Y(D), d)$. The g^{th} *lambda class* is its top Chern class $\lambda_g := c_g(\mathbb{E})$.

We will be concerned with the virtual log GW count of genus g curves in Y of degree d meeting D_j in one point of maximal tangency $d \cdot D_j$, passing through $l-1$ general points of Y and with insertion λ_g,

$$N^{\log}_{g,d}(Y(D)) := \int_{[\overline{\mathrm{M}}^{\log}_{g,l-1}(Y(D),d)]^{\mathrm{vir}}} (-1)^g \lambda_g \prod_{j=1}^{l-1} \mathrm{ev}_j^*([\mathrm{pt}]). \tag{4-1}$$

Furthermore, we will denote by $N^{\log,\psi}_{0,d}(Y(D))$ the genus-zero log GW invariants of maximal tangency passing through one general point of Y with psi class to the power $l-2$,

$$N^{\log,\psi}_{0,d}(Y(D)) := \int_{[\overline{\mathrm{M}}^{\log}_{0,1}(Y(D),d)]^{\mathrm{vir}}} \mathrm{ev}_1^*([\mathrm{pt}]) \cup \psi_1^{l-2}. \tag{4-2}$$

It will be useful in the following to define all-genus generating functions for the logarithmic invariants of $Y(D)$ at fixed degree,

$$\mathsf{N}^{\log}_d(Y(D))(\hbar) := \frac{1}{\left(2\sin\left(\frac{1}{2}\hbar\right)\right)^{l-2}} \sum_{g \geqslant 0} N^{\log}_{g,d}(Y(D)) \hbar^{2g-2+l}. \tag{4-3}$$

In the setting of Proposition 2.4, it follows from [3] that $N^{\log}_{g,d}(Y(D))$ (resp. $N^{\log,\psi}_{0,d}(Y(D))$) equals the log GW invariant of $(\widetilde{Y}, \widetilde{D})$, of class $\varphi^* d$, with maximal tangency along each of the strict transforms of D_i, not meeting the other boundary components and meeting $l-1$ general points of $\widetilde{Y}$ (resp. one point with psi class to the power of $l-2$). The above numbers are deformation invariant in log smooth families [85].

4.2 Scattering diagrams

Our main tool for the calculation of (4-1)–(4-2) will be their associated quantum scattering diagrams and quantum broken lines [83; 14; 13; 32; 16]. In the classical limit, in dimension 2 this is treated in [56; 53; 51] and in full generality in [57; 55]. The quantum scattering diagram consists of an affine integral manifold B and a collection of walls $\mathfrak{d}$ with wall-crossing functions $f_{\mathfrak{d}}$. The latter are functions on open subsets of the mirror.

Let $\pi : (\widetilde{Y}, \widetilde{D}) \longrightarrow (\overline{Y}, \overline{D})$ be a toric model as in Proposition 2.4 with s interior blowups. Up to deformation, we may assume that the blowup points are disjoint. Note that $s = \chi_{\mathrm{top}}(Y \setminus D) = \chi_{\mathrm{top}}(\widetilde{Y} \setminus \widetilde{D})$ is an invariant of the interior. We construct an affine integral manifold B from π as follows. First, we start with the fan of $(\overline{Y}, \overline{D})$. Then, for every interior blowup, we add a focus–focus singularity in the direction of the corresponding ray. In practice, we introduce cuts connecting the singularities to infinity and we use charts to identify the complements of the cuts with an open subset of $\mathbb{R}^2$.

Let $\delta_1, \dots, \delta_s$ denote the focus–focus singularities and $B(\mathbb{Z})$ be the set of integral points of $B \setminus \{\delta_1, \dots, \delta_s\}$. In the limit where the singularities are sent to infinity, $B(\mathbb{Z})$ can be identified with the integral points of the fan of $(\overline{Y}, \overline{D})$. The singularity δ_j corresponds to an interior blowup on a toric divisor $D(\delta_j)$ of $(\overline{Y}, \overline{D})$ with exceptional divisor $\mathcal{E}_j$. Viewing the ray of the fan of $(\overline{Y}, \overline{D})$ corresponding to $D(\delta_j)$ as going from $(0,0)$ to infinity, denote by ρ_j its primitive direction.

Each δ_j creates a quantum wall $\mathfrak{d}_j$ propagating into the direction $-\rho_j$ and decorated with the wall-crossing function $f_{\mathfrak{d}_j} := 1 + t_j z^{\rho_j}$, where $t_j = t^{[\mathcal{E}_j]}$ is a formal variable keeping track of the exceptional divisor and z^{ρ_j} is the tangent monomial $x^a y^b$ if $\rho_j = (a,b)$. Note that the wall also propagates into the ρ_j direction (decorated with $1 + t_j z^{-\rho_j}$), but that part of the scattering diagram is not relevant to us.

When two walls meet, this creates scattering: up to perturbation, we may assume that at most two walls $\mathfrak{d}_j$ and $\mathfrak{d}_k$ come together at one point, which in the following is taken to be the origin for simplicity. We refer to [56] for the general case and only describe the explicit result in the two cases relevant to us:

- **Simple scattering** $(\det(\rho_j, \rho_k) = \pm 1)$ The scattering algorithm draws an additional quantum wall $\mathfrak{d}$ in the direction $-\rho_j - \rho_k$ decorated with the function $1 + t_j t_k z^{\rho_j + \rho_k}$.
- **Infinite scattering** $(\det(\rho_j, \rho_k) = \pm 2)$ The algorithm creates a central quantum wall $\mathfrak{d}$ in the direction $-\rho_j - \rho_k$ decorated with the function

$$\prod_{\ell = -\frac{1}{2}(\operatorname{ind}(\rho_j+\rho_k)-1)}^{\frac{1}{2}(\operatorname{ind}(\rho_j+\rho_k)-1)} (1 - q^{-\frac{1}{2}+\ell} t_j t_k z^{\rho_j+\rho_k})^{-1} (1 - q^{\frac{1}{2}+\ell} t_j t_k z^{\rho_j+\rho_k})^{-1}, \tag{4-4}$$

where $\operatorname{ind}(\rho_j + \rho_k)$ is the index of $\rho_j + \rho_k$. We then add quantum walls $\mathfrak{d}_1, \dots, \mathfrak{d}_n, \dots$ in the directions $-(n+1)\rho_j - n\rho_k$ decorated with functions

$$1 + t_j^{n+1} t_k^n z^{(n+1)\rho_j + n\rho_k}, \tag{4-5}$$

for $n \geq 0$, as well as quantum walls ${}_1\mathfrak{d}, \dots, {}_n\mathfrak{d}, \dots$ in the directions $-n\rho_j - (n+1)\rho_k$ decorated with functions

$$1 + t_j^n t_k^{n+1} z^{n\rho_j + (n+1)\rho_k} \tag{4-6}$$

for $n \geq 0$.

The classical scattering algorithm is recovered in the classical limit $q^{\frac{1}{2}} = 1$. Only the central quantum wall in the case $\det(\rho_j, \rho_k) = \pm 2$ is different from its classical version, for which the wall-crossing function specialises to $(1 - t_j t_k z^{\rho_j+\rho_k})^{-2\operatorname{ind}(\rho_j+\rho_k)}$.

If $\mathfrak{u}$ and $\mathfrak{u}'$ are adjacent chambers of B separated by the quantum wall $\mathfrak{d}$ decorated with $f_{\mathfrak{d}}$, we can define a *quantum wall-crossing transformation* $\theta_{\mathfrak{d}}$ from $\mathfrak{u}$ to $\mathfrak{u}'$ as follows. Denote by $n_{\mathfrak{d}/\mathfrak{u}}$ the primitive orthogonal vector pointing from $\mathfrak{d}$ into $\mathfrak{u}$. Let m be such that $\langle n_{\mathfrak{d}/\mathfrak{u}}, m\rangle \geq 0$. For a polynomial a in the variables t_j, consider an expression az^m, which we think of as a function on $\mathfrak{u}$. Then, writing

$$f_{\mathfrak{d}} = \sum_{r\geq 0} c_r z^{r\rho_{\mathfrak{d}}}, \tag{4-7}$$

where $-\rho_{\mathfrak{d}}$ is the primitive direction of $\mathfrak{d}$,

$$\theta_{\mathfrak{d}} \colon az^m \mapsto az^m \prod_{\ell=-\frac{1}{2}(\langle n_{\mathfrak{d}/\mathfrak{u}}, m\rangle -1)}^{\frac{1}{2}(\langle n_{\mathfrak{d}/\mathfrak{u}}, m\rangle -1)} \left(\sum_{r\geq 0} c_r q^{r\ell} z^{r\rho_{\mathfrak{d}}}\right). \tag{4-8}$$

Note that in the classical limit $q^{\frac{1}{2}} = 1$, we recover the formula for the classical wall-crossing transformation, which is $\theta_{\mathfrak{d}}^{\mathrm{cl}} \colon az^m \mapsto f_{\mathfrak{d}}^{\langle n_{\mathfrak{d}/\mathfrak{u}}, m\rangle} az^m$. Writing $\theta_{\mathfrak{d}}(az^m) = \sum_i a_i z^{m_i}$, any summand $a_i z^{m_i}$ is called *a result of quantum transport of az^m* from $\mathfrak{u}$ to $\mathfrak{u}'$.

The final object we will need is the algebra of quantum broken lines associated to the scattering diagram, which we describe in the generality needed here; see [55] for full details in the classical limit. Let $B_0 := B \setminus \{\delta_1, \dots, \delta_s, \mathfrak{d}_j \cap \mathfrak{d}_k \mid \text{for all } j,k\}$. Let z^m be an *asymptotic monomial*, in our case this means that $m = (a,b) \neq (0,0)$, and let $p \in B$. Then a *quantum broken line β* with asymptotic monomial z^m and endpoint p consists of

(1) a directed piecewise straight path in B_0 of rational slopes, coming from infinity in the direction $-m$, bending only at quantum walls and ending at p;

(2) a labelling of the initial ray by L_1 and the successive line segments in order by $L_2, \dots, L_s$, where p is the endpoint of L_s;

(3) if $L_i \cap L_{i+1} \in \mathfrak{d}_i$, then, iteratively defined from 1 to s, the assignment of a monomial $a_i z^{m_i}$, where
   - $a_1 z^{m_1} = z^m$,
   - $a_{i+1} z^{m_{i+1}}$ is a result of the quantum transport of $a_i z^{m_i}$ across $\mathfrak{d}_i$,
   - L_i is directed in the direction $-m_i$.

Note that if $n_{\mathfrak{d}_i/L_i}$ is the primitive orthogonal vector to $\mathfrak{d}_i$ pointing into the half-plane containing L_i, then, as L_i is directed in the direction $-m_i$, we have $\langle n_{\mathfrak{d}_i/L_i}, m_i\rangle \geq 0$, and so the quantum transport of $a_i z^{m_i}$ across $\mathfrak{d}_i$ is indeed well-defined. We call $a_{\mathrm{end}} z^{m_{\mathrm{end}}} = a_s z^{m_s}$ the *end monomial* of β and a_{end} the *end coefficient* of β.

If z^m is an asymptotic monomial, the *theta function* ϑ_m is the sum of the end monomials of all broken lines with asymptotic monomial z^m and ending at p. Note that a priori ϑ_m depends on p, but it is one of the main results of [55] that it is constant in chambers and transforms from chamber to chamber according to the wall-crossing transformations.

We first describe the classical algebra of theta functions, ie we set $q^{\frac{1}{2}} = 1$. For A an element in the algebra of theta functions, we denote by $\langle A, \vartheta_m \rangle$ the coefficient of ϑ_m in A; note that $\langle A, \vartheta_m \rangle$ is a polynomial in the t_j. Then the identity component $\langle \vartheta_{m^1} \cdot \vartheta_{m^2}, \vartheta_0 \rangle$ is given as the sum of products of end coefficients $a^1_{\mathrm{end}} a^2_{\mathrm{end}}$ over all broken lines β_1 with asymptotic monomial z^{m^1} and β_2 with asymptotic monomial z^{m^2} such that $m^1_{\mathrm{end}} = -m^2_{\mathrm{end}}$. The identity component $\langle \vartheta_{m^1} \cdot \vartheta_{m^2} \cdot \vartheta_{m^3}, \vartheta_0 \rangle$ is given as the sum of products of end coefficients $a^1_{\mathrm{end}} a^2_{\mathrm{end}} a^3_{\mathrm{end}}$ over all broken lines β_1, β_2, β_3, with asymptotic monomials z^{m^1}, z^{m^2}, z^{m^3} and such that $m^1_{\mathrm{end}} + m^2_{\mathrm{end}} + m^3_{\mathrm{end}} = 0$.

For $(Y(D = D_1 + \cdots + D_l))$, consider the scattering diagram associated to a toric model π coming from a diagram as in Proposition 2.4,

(4-9)
$$\begin{array}{ccc} & \widetilde{Y}(\widetilde{D}) & \\ \swarrow^{\varphi} & & \searrow^{\pi} \\ Y(D) & & \overline{Y}(\overline{D}) \end{array}$$

Then the proper transform and pushforward of D_j is a toric divisor in $\overline{Y}$ corresponding to a ray in B. Up to reordering the indices, we assume that the ray corresponding to D_j is directed by ρ_j.

Proposition 4.1 [84] *Let $Y(D)$ be an l–component log Calabi–Yau surface of maximal boundary. Let $d \in \mathrm{H}_2(Y, \mathbb{Z})$ be an effective curve class and write $e_j := d \cdot \varphi_* \mathcal{E}_j$ for $j = 1, \ldots, s$, where φ is as in* (4-9).

- *Assume that $l = 2$. Set $m^1 = (d \cdot D_1)\rho_1$ and $m^2 = (d \cdot D_2)\rho_2$. Then $N^{\log}_{0,d}(Y(D))$ is the coefficient of $\prod_{j=1}^s t_j^{e_j}$ in $\langle \vartheta_{m^1} \cdot \vartheta_{m^2}, \vartheta_0 \rangle$.*
- *Assume that $l = 3$. Set $m^1 = (d \cdot D_1)\rho_1$, $m^2 = (d \cdot D_2)\rho_2$ and $m^3 = (d \cdot D_3)\rho_3$. Then $N^{\log,\psi}_{0,d}(Y(D))$ is the coefficient of $\prod_{j=1}^s t_j^{e_j}$ in $\langle \vartheta_{m^1} \cdot \vartheta_{m^2} \cdot \vartheta_{m^3}, \vartheta_0 \rangle$.*

We return to the algebra of quantum theta functions. For every m^1, m^2 and $p \in B_0$, denote by C_{m^1,m^2} the polynomial in the variables t_j with coefficients in $\mathbb{Z}[q^{\pm\frac{1}{2}}]$ given as the sum of products of end coefficients $a^1_{\mathrm{end}} a^2_{\mathrm{end}}$ over all quantum broken lines β_1 with asymptotic monomial z^{m^1} and β_2 with asymptotic monomial z^{m^2}, with common endpoint p and such that $m^1_{\mathrm{end}} = -m^2_{\mathrm{end}}$. The polynomial C_{m^1,m^2} is independent of the choice of $p \in B_0$.

Proposition 4.2 *Let $Y(D)$ be an l–component log Calabi–Yau surface of maximal boundary. Let $d \in \mathrm{H}_2(Y, \mathbb{Z})$ be an effective curve class and write $e_j := d \cdot \varphi_* \mathcal{E}_j$ for $j = 1, \ldots, s$, where φ is as in* (4-9).

- *Assume that $l = 2$. Set $m^1 = (d \cdot D_1)\rho_1$ and $m^2 = (d \cdot D_2)\rho_2$. Then after the change of variables $q = e^{i\hbar}$, the series*

(4-10)
$$\mathsf{N}^{\log}_d(Y(D))(\hbar) = \sum_{g \geqslant 0} N^{\log}_{g,d}(Y(D))\hbar^{2g}$$

is the $\hbar$–expansion of the q–polynomial which is the coefficient of $\prod_{j=1}^s t_j^{e_j}$ in C_{m^1,m^2}.

- *Assume that $l=3$. Set $m^1=(d\cdot D_1)\rho_1$, $m^2=(d\cdot D_2)\rho_2$ and $m^3=(d\cdot D_3)\rho_3$. Then after the change of variables $q=e^{i\hbar}$, the series*

$$\mathsf{N}_d^{\log}(Y(D))(\hbar)=\frac{1}{2\sin\left(\frac{1}{2}\hbar\right)}\sum_{g\geqslant 0}N_{g,d}^{\log}(Y(D))\hbar^{2g+1} \tag{4-11}$$

is the $\hbar$–expansion of the q–polynomial obtained as the sum over all quantum broken lines β_1 with asymptotic monomial z^{m_1}, β_2 with asymptotic monomial z^{m_2}, and β_3 with asymptotic monomial z^{m_3}, with common endpoint and such that $m^1_{\mathrm{end}}+m^2_{\mathrm{end}}+m^3_{\mathrm{end}}=0$, of

$$\frac{[|\det(m^1_{\mathrm{end}},m^2_{\mathrm{end}})|]_q}{[1]_q}a^1_{\mathrm{end}}a^2_{\mathrm{end}}a^3_{\mathrm{end}}. \tag{4-12}$$

Here $a^i_{\mathrm{end}}z^{m^i_{\mathrm{end}}}$ are the end monomials of the broken lines β_i and the q–integers $[\,\cdot\,]_q$ are defined in (4-18) *below.*

Proof We only give a sketch of the proof as it is an adaptation of the proof of the Frobenius structure conjecture of [84], which in the setting relevant to us is stated in Proposition 4.1 above.

Recall first the geometric argument of the proof of [84]. The starting point is to consider the degeneration of [56] of $Y(D)$ to a toric situation: using toric transversality in the cluster setting, the curves do not fall into the codimension-one strata of D and one may apply the degeneration formula, expressing $N_{0,d}^{\log,\psi}(Y(D))$ in terms of log GW invariants of the central fibre, which can in turn be computed via the toric tropical correspondence theorem [97; 85]. In the scattering diagram, the tropical curves correspond to constellations of broken lines, and the product of the end coefficients equals the product of the multiplicity of the tropical curve with the terms coming from the degeneration formula.

To see how this is modified to obtain higher-genus invariants, the study of the degeneration of [56] is done using the techniques introduced in [14], and then the result follows from the toric tropical correspondence theorem for higher-genus log Gromov–Witten invariants with λ_g–insertion proven in [12]. The toric transversality of the log maps in the degeneration is a consequence of the vanishing result of [12, Lemma 8].

To further encompass two-pointed insertions, one can see that tropically, by [85], a ψ–class corresponds to a marked 3–valent vertex with multiplicity 1. In the case of 2–pointed invariants, one can carry out the same degeneration as above, therefore leading to the same tropical curves in the fan of the central fibre. The one difference is that previously one 3–valent vertex corresponded to a point with a ψ–class, and hence carried multiplicity 1, whereas in the case of 2–pointed invariants this vertex is no longer marked and carries its Block–Göttsche [10] multiplicity

$$\frac{[|\det(m^1_{\mathrm{end}},m^2_{\mathrm{end}})|]_q}{[1]_q},$$

and instead there are two marked 2–valent vertices elsewhere (which do not carry any multiplicity). □

4.2.1 Binomials and q–binomial coefficients In our applications of Propositions 4.1 and 4.2, we will mostly consider (quantum) broken lines bending along (quantum) walls $\mathfrak{f}_{\mathfrak{d}}$ decorated by a function of the form

$$f_{\mathfrak{d}} = 1 + t z^{\rho_{\mathfrak{d}}}, \tag{4-13}$$

where $-\rho_{\mathfrak{d}}$ is the primitive direction of $\mathfrak{d}$. By the binomial theorem, we have

$$f_{\mathfrak{d}}^{\langle n,m\rangle} = (1 + t z^{\rho_{\mathfrak{d}}})^{\langle n,m\rangle} = \sum_{k=0}^{\langle n,m\rangle} \binom{\langle n,m\rangle}{k} t^k z^{k\rho_{\mathfrak{d}}}. \tag{4-14}$$

Therefore, each application of transport across such a wall will produce a binomial coefficient, and so our genus-zero log Gromov–Witten invariants will be product of binomial coefficients. By the q–binomial theorem, we have

$$\prod_{\ell=-\frac{1}{2}(\langle n,m\rangle-1)}^{\frac{1}{2}(\langle n,m\rangle-1)} \left(1 + t q^{\ell} z^{\rho_{\mathfrak{d}}}\right) = \sum_{k=0}^{\langle n,m\rangle} \begin{bmatrix} \langle n,m\rangle \\ k \end{bmatrix}_q t^k z^{k\rho_{\mathfrak{d}}}, \tag{4-15}$$

where the q–binomial coefficients

$$\begin{bmatrix} N \\ k \end{bmatrix}_q := \frac{[N]_q!}{[k]_q![(N-k)]_q!} \tag{4-16}$$

are defined in terms of the q–factorials

$$[n]_q! := \prod_{j=1}^{n} [j]_q, \tag{4-17}$$

where the q–integers are

$$[n]_q := q^{n/2} - q^{-n/2}. \tag{4-18}$$

It follows that the formulas for the higher-genus log Gromov–Witten invariants $\mathsf{N}_d^{\log}(Y(D))(\hbar)$ will be obtained by replacing binomial coefficients by q–binomial coefficients in the formulas for the genus-zero invariant $N_{0,d}^{\log}(Y(D))$.

4.3 Log Gromov–Witten invariants under interior blowup

Proposition 4.3 (blowup formula for log GW invariants) *Let $Y(D)$ be an l–component log CY surface with maximal boundary. Let $\pi\colon Y'(D') \to Y(D)$ be the l–component log CY surface with maximal boundary obtained by an interior blowup at a general point of D with exceptional divisor E. Let d be a curve class of $Y(D)$ and let $d' := \pi^* d$. Then*

$$N_{g,d}^{\log}(Y(D)) = N_{g,d'}^{\log}(Y'(D')), \tag{4-19}$$

$$N_{0,d}^{\log,\psi}(Y(D)) = N_{0,d'}^{\log,\psi}(Y'(D')). \tag{4-20}$$

Proof Let D_j be the irreducible component of D containing the point that we blow up. We consider the degeneration of $Y(D)$ to the normal cone of D_j: the fibre over any point of $\mathbb{A}^1 - \{0\}$ is $Y(D)$ and the special fibre over $\{0\}$ has two irreducible components, which are isomorphic to $Y(D)$ and a $\mathbb{P}^1$–bundle $\mathbb{P}_j$ over D_j, and are glued together along a copy of D_j. Let $\mathcal{D}_j$ be the closure of $D_j \times (\mathbb{A}^1 - \{0\})$ in the total space of the degeneration. After blowing up a section of $\mathcal{D}_j \to \mathbb{A}^1$, we obtain a family with fibre $Y'(D')$ over any point of $\mathbb{A}^1 - \{0\}$, and special fibre over $\{0\}$ given by the union of two irreducible components, which are isomorphic to $Y(D)$ and to the blowup $\widetilde{\mathbb{P}}_j$ of $\mathbb{P}_j$ at one point. We compare the invariants $N_{g,d}^{\log}$ and $N_{0,d}^{\log,\psi}$ of $Y(D)$ and $Y'(D')$ using this degeneration. Following the general strategy of [14, Section 5], using in particular the vanishing result of [12, Lemma 8] to guarantee toric transversality of the log maps in the degeneration, we obtain that the invariants of $Y(D)$ and $Y'(D')$ only differ by a multiplicative factor coming from multiple covers of a fibre of $\widetilde{\mathbb{P}}_j \to D_j$. By deformation invariance, we can assume that this fibre is a smooth $\mathbb{P}^1$–fibre, with trivial normal bundle in $\widetilde{\mathbb{P}}_j$. Therefore, the correction factor is an integral over a moduli space of stable log maps to $\mathbb{P}^1$ with extra insertion of the class $e(\mathrm{H}^1(C, \mathcal{O}_C)) = (-1)^g \lambda_g$. Because our genus g invariants already contain an insertion of λ_g and $\lambda_g^2 = 0$ for $g > 0$ by Mumford's relation [95], the correction factor only receives contributions from genus zero. The genus-zero corrections involves degree $d \cdot D_j$ stable log maps to $(\mathbb{P}^1, \{0\} \cup \{\infty\})$, fully ramified over 0 and ∞. The corresponding moduli space is a point with an automorphism group of order $d \cdot D_j$ and so contributes $1/(d \cdot D_j)$. Because of the extra $(d \cdot D_j)$ multiplicity factor in the degeneration formula, the total multiplicative correction factor is 1. □

As a consequence of Proposition 4.3, if we calculate $N_{g,d}^{\log}(Y(D))$ and $N_{0,d}^{\log,\psi}(Y(D))$ for all g and d, then we will know the invariants for all interior blowdowns of $Y(D)$. Therefore it is enough to calculate the invariant for the cases of highest Picard rank in Propositions 2.2 and 2.3. In the following section, we calculate the higher-genus log invariants $\mathsf{N}_d^{\log}(Y(D))(\hbar)$ for all tame Looijenga pairs: using Proposition 4.3, it is enough to consider the pairs $\mathrm{dP}_3(1,1)$, $\mathrm{dP}_3(0,0,0)$ and $\mathbb{F}_0(0,0,0,0)$, which are treated in Theorems 4.5, 4.9 and 4.10.

For nontame pairs, the genus-zero invariants can be obtained by combining the log-local correspondence of Theorem 5.1 and (3-21) in Theorem 3.3 giving the local invariants. For quasi-tame pairs we furthermore make the following general conjecture for the higher-genus invariants $\mathsf{N}_d^{\log}(Y(D))(\hbar)$.

Conjecture 4.4 *Let $Y(D)$ and $Y'(D')$ be nef 2–component log CY surfaces with maximal boundary such that the corresponding local geometries $E_{Y(D)}$ and $E_{Y'(D')}$ are deformation equivalent. Then, under suitable identification of d, we have*

$$\left(\prod_{j=1}^{l=2} [d \cdot D'_j]_q\right) \mathsf{N}_d^{\log}(Y(D))(\hbar) = \left(\prod_{j=1}^{l=2} [d \cdot D_j]_q\right) \mathsf{N}_d^{\log}(Y'(D'))(\hbar). \tag{4-21}$$

Conjecture 4.4 holds in the genus-zero, ie $q^{\frac{1}{2}} = 1$, limit, as a corollary of the log-local correspondence given by Theorem 5.1 and of the deformation invariance of local Gromov–Witten invariants. In higher genus, Conjecture 4.4 translates to conjectural, new nontrivial q–binomial identities: see eg Conjecture B.3 for the cases of $\mathrm{dP}_1(0,4)$ and $\mathbb{F}_0(0,4)$.

4.4 Toric models: $l = 2$

Extending [15, Section 5] we find toric models for all $l = 2$ nef log Calabi–Yau surfaces except for $\mathbb{F}_0(2,2)$, which we leave to the reader as an exercise. For each toric model, we draw the corresponding fans with focus–focus singularities. By [40, Lemma 2.10], a log Calabi–Yau surface with maximal boundary $(\overline{Y}, \overline{D})$ is toric if the sequence of self-intersection numbers of irreducible components of $\overline{D}$ is realised as the sequence of self-intersection numbers of toric divisors on a toric surface. Once we have the toric models, we calculate the part of the scattering diagram relevant to us, and by Proposition 4.1 the relevant structural coefficients for the multiplication of theta functions yield the maximal tangency log Gromov–Witten invariants.

4.4.1 Tame pairs: simple scattering By Proposition 4.3, it suffices to consider the case $Y(D) = \mathrm{dP}_3(1,1)$. Start with $\mathbb{P}^2(1,4)$. The anticanonical decomposition of D is given by D_1 a line and D_2 a smooth conic not tangent to D_1. For notational convenience, in what follows we will identify D_1 and D_2 (resp. F_1 and F_2) with their strict transforms (resp. pushforwards) under blowups (resp. blowdowns).

Denote by pt one of the intersection points of D_1 and D_2 and by L the line tangent to D_2 at pt. We blow up pt, leading to the exceptional divisor F_1. We further blow up the intersection of F_1 with D_2 and write F_2 for the exceptional divisor. Denote the resulting log Calabi–Yau surface with maximal boundary by $(\widetilde{\mathbb{P}^2(1,4)}, \widetilde{D})$, where $\widetilde{D}$ is the strict transform of D.

The toric model $(\overline{\mathbb{P}^2(1,4)}, \overline{D})$ is given by blowing down the strict transform of L, so that $\overline{\mathbb{P}^2(1,4)} = \mathbb{F}_2$ and $\overline{D} = D_1 \cup F_1 \cup F_2 \cup D_2$, with F_1 the (-2)–curve of $\mathbb{F}_2$, D_2 a section of self-intersection 2, and D_1 and F_2 linearly equivalent to fibre classes. Labelling the toric boundary divisors with their self-intersections, we obtain the diagram at the left of Figure 3.

To obtain the toric model for $\mathrm{dP}_3(1,1)$, we need to blow up a nontoric point on F_2 (thus reproducing L), and three nontoric points on D_2. Tropically, this amounts to introducing a focus–focus singularity on the ray of F_2 and three on the ray of D_2 as in Figure 3 to the right. Walls emanate out of these focus–focus singularities. While they propagate into two directions, for our calculations only one direction matters (the other ray being close to infinity and thus noninteracting). We perturb the focus–focus singularities on D_2 horizontally.

The cone of curves is generated by $H - E_i - E_j$ for $1 \le i < j \le r$ and the E_i. In particular, any curve class $d \in \mathrm{H}_2(\mathrm{dP}_3, \mathbb{Z})$ can be written as $d = d_0(H - E_1 - E_2 - E_3) + d_1 E_1 + d_2 E_2 + d_3 E_3$.

Theorem 4.5 *Putting $q = \mathrm{e}^{\mathrm{i}\hbar}$, we have*

$$\mathsf{N}_d^{\log}(\mathrm{dP}_3(1,1))(\hbar) = \begin{bmatrix} d_3 \\ d_0 - d_1 \end{bmatrix}_q \begin{bmatrix} d_3 \\ d_0 - d_2 \end{bmatrix}_q \begin{bmatrix} d_0 \\ d_3 \end{bmatrix}_q \begin{bmatrix} d_1 + d_2 + d_3 - d_0 \\ d_3 \end{bmatrix}_q . \tag{4-22}$$

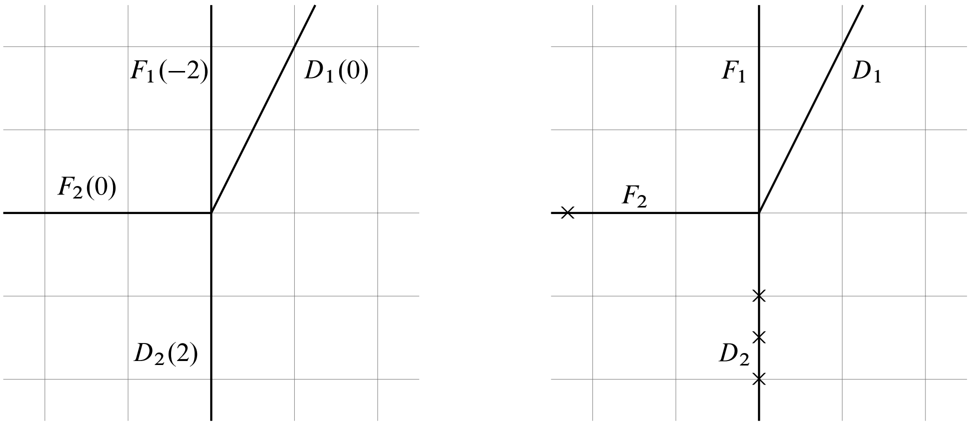


Figure 3: Left: the toric model of $\mathbb{P}^2(1,4)$. Right: the toric model of $\mathrm{dP}_3(1,1)$.

Proof Write $t = z^{[L]}$ and let $t_i = z^{[E_i]}$. Since $D_1 = H$ and $D_2 = 2H - E_1 - E_2 - E_3$, we have the following intersection multiplicities:

$$d \cdot D_1 = d_0, \quad d \cdot D_2 = d_1 + d_2 + d_3 - d_0 \quad \text{and} \quad d \cdot E_i = d_0 - d_i.$$

All of the scattering is simple. The initial wall-crossing functions are drawn in Figure 4, and all successive functions are easily obtained. We have two broken lines, one coming from the D_1–direction with attaching monomial $(xy^2)^{d \cdot D_1}$ and one coming from the D_2–direction with attaching monomial $(y^{-1})^{d \cdot D_2}$. Provided we choose our endpoint p to be sufficiently far into the x–direction, Figure 4 contains all the relevant walls. We start from the broken line coming from the D_2 direction and summarise the wall-crossing functions attached to the walls it meets:

$$① \ 1 + tx^{-1}, \qquad ② \ 1 + tt_3x^{-1}y^{-1}, \qquad ③ \ 1 + tt_2x^{-1}y^{-1},$$
$$④ \ 1 + tt_1x^{-1}y^{-1}, \qquad ⑤ \ 1 + t^2t_1t_2t_3x^{-2}y^{-3}.$$

Crossing these walls leads to $y^{d_0 - d_1 - d_2 - d_3}$ mapping to

$$\begin{bmatrix} d_1 + d_2 + d_3 - d_0 \\ k \end{bmatrix}_q \begin{bmatrix} d_1 + d_2 + d_3 - d_0 - k \\ k_1 \end{bmatrix}_q \begin{bmatrix} d_1 + d_2 + d_3 - d_0 - k \\ k_2 \end{bmatrix}_q$$
$$\cdot \begin{bmatrix} d_1 + d_2 + d_3 - d_0 - k \\ k_3 \end{bmatrix}_q \begin{bmatrix} 2d_1 + 2d_2 + 2d_3 - 2d_0 - 3k - k_1 - k_2 - k_3 \\ k_4 \end{bmatrix}_q$$
$$\cdot t^{k + k_1 + k_2 + k_3 + 2k_4} t_3^{k_1 + k_4} t_2^{k_2 + k_4} t_1^{k_3 + k_4} x^{-k - k_1 - k_2 - k_3 - 2k_4} y^{d_0 - d_1 - d_2 - d_3 - k_1 - k_2 - k_3 - 3k_4}.$$

The intersection multiplicities with the divisors impose the following conditions:

$$k + k_1 + k_2 + k_3 + 2k_4 = d_0, \quad k_1 + k_4 = d_0 - d_3, \quad k_2 + k_4 = d_0 - d_2, \quad k_3 + k_4 = d_0 - d_1. \tag{4-23}$$

Choose as indeterminate k. For the coefficient to be nonzero, $0 \le k \le d_1 + d_2 + d_3 - d_0$. Then

$$\begin{aligned} k_4 &= k + 2d_0 - d_1 - d_2 - d_3, & k_1 &= d_1 + d_2 - d_0 - k, \\ k_2 &= d_1 + d_3 - d_0 - k, & k_3 &= d_2 + d_3 - d_0 - k. \end{aligned} \tag{4-24}$$

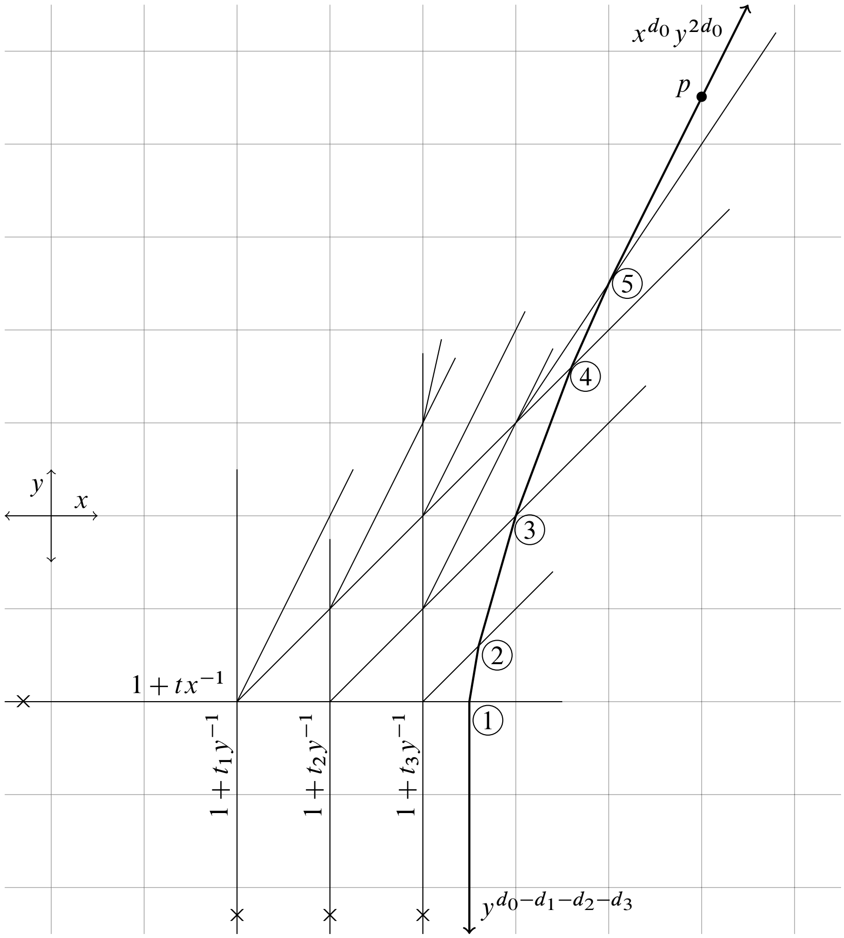


Figure 4: Scatt $\mathrm{dP}_3(1,1)$.

Hence the sum of the coefficients of the broken lines is

$$\sum_{k=0}^{d_1+d_2+d_3-d_0}\left(\begin{bmatrix} d_1+d_2+d_3-d_0 \\ k \end{bmatrix}_q \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_3 \end{bmatrix}_q \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_2 \end{bmatrix}_q \cdot \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_1 \end{bmatrix}_q \begin{bmatrix} 2d_1+2d_2+2d_3-2d_0-3k-k_1-k_2-k_3 \\ k_4 \end{bmatrix}_q\right)$$

$$= \sum_{k=0}^{k(d_0,d_1,d_2,d_3)}\left(\begin{bmatrix} d_1+d_2+d_3-d_0 \\ k \end{bmatrix}_q \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_3 \end{bmatrix}_q \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_2 \end{bmatrix}_q \cdot \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_1 \end{bmatrix}_q \begin{bmatrix} d_0 \\ k+2d_0-d_1-d_2-d_3 \end{bmatrix}_q\right),$$

where $k(d_0,d_1,d_2,d_3) := \min\{d_0, d_1+d_2-d_0, d_1+d_3-d_0, d_2+d_3-d_0\}$.

Therefore, we obtain

$$\text{(4-25)}\quad \mathsf{N}_d^{\log}(\mathrm{dP}_3(1,1))(\hbar) = \sum_{k\geqslant 0}\Bigg(\begin{bmatrix} d_1+d_2+d_3-d_0 \\ k \end{bmatrix}_q \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_3 \end{bmatrix}_q \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_2 \end{bmatrix}_q \cdot \begin{bmatrix} d_1+d_2+d_3-d_0-k \\ d_1 \end{bmatrix}_q \begin{bmatrix} d_0 \\ k+2d_0-d_1-d_2-d_3 \end{bmatrix}_q\Bigg).$$

Writing the q–binomial coefficients in terms of q–factorials, and changing the indexing variable

$$k \mapsto k - d_0 + \tfrac{1}{2}(d_1+d_2+d_3),$$

we have

$$\text{(4-26)}\quad \frac{[d_0]_q![d_1+d_2+d_3-d_0]_q!}{[d_1]_q![d_2]_q![d_3]_q!} \cdot \sum_k \frac{\big[\frac{1}{2}(d_1+d_2+d_3)-k\big]_q!}{\Big(\big[\frac{1}{2}(d_1+d_2-d_3)-k\big]_q!\big[\frac{1}{2}(d_1+d_3-d_2)-k\big]_q!\big[\frac{1}{2}(d_2+d_3-d_1)-k\big]_q! \cdot \big[k-d_0+\frac{1}{2}(d_1+d_2+d_3)\big]_q!\big[k+d_0-\frac{1}{2}(d_1+d_2+d_3)\big]_q!\Big)}.$$

We re-sum this explicitly using the q–Pfaff–Saalschütz identity[10] in the form given in [116, Equation (1q)]:

$$\text{(4-27)}\quad \sum_k \frac{[a+b+c-k]_q!}{[a-k]_q![b-k]_q![c-k]_q![k-m]_q![k+m]_q!} = \begin{bmatrix} a+b \\ a+m \end{bmatrix}_q \begin{bmatrix} a+c \\ c+m \end{bmatrix}_q \begin{bmatrix} b+c \\ b+m \end{bmatrix}_q.$$

Therefore, specialising (4-27) to $a+b=d_1$, $b+c=d_3$, $a+c=d_2$, $a+m=d_0-d_3$, $b+m=d_0-d_2$ and $c+m=d_0-d_1$, we have

$$\text{(4-28)}\quad \mathsf{N}_d^{\log}(\mathrm{dP}_3(1,1))(\hbar) = \frac{[d_0]_q![d_1+d_2+d_3-d_0]_q!}{[d_1]_q![d_2]_q![d_3]_q!}\begin{bmatrix} d_1 \\ d_0-d_3 \end{bmatrix}_q \begin{bmatrix} d_2 \\ d_0-d_1 \end{bmatrix}_q \begin{bmatrix} d_3 \\ d_0-d_2 \end{bmatrix}_q,$$

which after elementary simplifications gives (4-22). □

Remark 4.6 It follows from the above proof that Theorem 4.5 is in fact equivalent to the q–Pfaff–Saalschütz identity. In genus zero, Theorem 5.1 applied to $\mathrm{dP}_3(1,1)$ gives a geometric proof of Theorem 4.5. Thus, we obtain a new geometric, albeit quite indirect, proof of the classical ($q=1$) Pfaff–Saalschütz identity.

4.4.2 Nontame pairs: infinite scattering Figure 19 gives the toric model of $\mathbb{F}_0(0,4)$. For the other nontame pairs, let $1 \le r \le 5$. Then $\mathrm{dP_r}(0,5-r)$ is obtained from $\mathbb{P}^2(1,4)$ by blowing up the first point on the line D_1 and the remaining $r-1$ points on the conic D_2. Hence we obtain the toric model of $\mathrm{dP_r}(0,5-r)$ by adding 1 focus–focus singularity on the ray D_1 and $r-1$ focus–focus singularities on the ray D_2, as in Figure 5. The singularities on the ray of D_2 can be perturbed horizontally.

[10]Unlike [44; 116], we are using q–factorials and q–binomial coefficients symmetric under $q \mapsto q^{-1}$. This explains the absence in the above expression of the power $q^{n^2-k^2}$, which is present in [116, Equation (1q)].

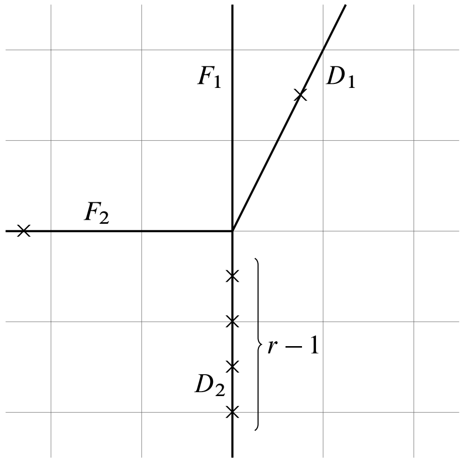


Figure 5: $\mathrm{dP_r}(0, 5-r)$.

Write a curve class $d \in \mathrm{H}_2(\mathrm{dP}_3(0,2), \mathbb{Z})$ as $d = d_0(H - E_1 - E_2 - E_3) + d_1 E_1 + d_2 E_2 + d_3 E_3$. As $D_1 = H - E_1$ and $D_2 = 2H - E_2 - E_3$, we get that $d \cdot D_1 = d_1$ and $d \cdot D_2 = d_2 + d_3$. As $E_{\mathrm{dP}_3(0,2)}$ is deformation equivalent to $E_{\mathrm{dP}_3(1,1)}$ by Proposition 2.6, Conjecture 4.4 and Theorem 4.5 give the following conjecture.

Conjecture 4.7 *The generating function* $\mathsf{N}_d^{\log}(\mathrm{dP}_3(0,2))(\hbar)$ *equals*

$$\frac{[d_1]_q [d_2+d_3]_q}{[d_0]_q [d_1+d_2+d_3-d_0]_q} \begin{bmatrix} d_3 \\ d_0 - d_1 \end{bmatrix}_q \begin{bmatrix} d_3 \\ d_0 - d_2 \end{bmatrix}_q \begin{bmatrix} d_0 \\ d_3 \end{bmatrix}_q \begin{bmatrix} d_1+d_2+d_3-d_0 \\ d_3 \end{bmatrix}_q, \tag{4-29}$$

where $q = \mathrm{e}^{\mathrm{i}\hbar}$.

Theorem 5.1 in Section 5.1 implies that Conjecture 4.7 holds in the classical limit $q^{\frac{1}{2}} = 1$. Direct scattering computation for $\mathrm{dP}_r(0, 5-r)$ with $r > 1$ are particularly daunting owing to the presence of infinite scattering, and in particular the final formulas take the shape of somewhat intricate multiple q–sums, which Conjecture 4.7 predicts should take a remarkably simple q–binomial form. We exemplify this for the blowdown geometries $\mathrm{dP}_1(0,4)$ and $\mathbb{F}_0(0,4)$ in Section B. For these cases, the specialisation of Conjecture 4.7 reduces to nontrivial, and apparently novel, conjectural q–binomial identities; see eg Conjecture B.3.

4.5 Toric models: $l = 3$

For $l = 3$, recall from (4-2) (resp. (4-1)) that $N_{0,d}^{\log,\psi}(Y(D))$ (resp. $N_{0,d}^{\log}(Y(D))$) is the genus-zero log Gromov–Witten invariant of maximal tangency passing through one point with psi-class (resp. passing through 2 points). By Proposition 4.3, it is enough to treat $\mathrm{dP}_3(0,0,0)$ as the other cases are obtained from it by interior blowdowns. We leave the description of the other toric models as an exercise to the reader.

Via Proposition 4.1, the invariant $N_{0,d}^{\log,\psi}(Y(D))$ is calculated from the scattering diagram as a structural coefficient of the product of three theta functions. For each constellation of three broken lines, the union of these corresponds to a tropical curve in the degeneration encoded by the scattering diagram. It is counted with multiplicity given by the product of the coefficients of the final monomials of the broken lines. Using Proposition 4.2, one can compute the generating series $\mathsf{N}_d^{\log}(Y(D))(\hbar)$ of higher-genus 2–point log Gromov–Witten invariants. The relevant tropical curves are identical to those entering the computation of $N_{0,d}^{\log,\psi}(Y(D))$. The difference is in the weighting of the tropical curves. For the ψ class, the trivalent vertex at the endpoint of the broken lines carry weight 1. For the 2–point invariants, we consider quantum broken lines, and the trivalent vertex is counted with Block–Göttsche multiplicity.

Let $Y(D) = \mathrm{dP}_3(0,0,0)$. We take D_1 in class $H - E_3$, D_2 in class $H - E_2$, D_3 in class $H - E_1$ and $d = d_0(H - E_1 - E_2 - E_3) + d_1 E_1 + d_2 E_2 + d_3 E_3$. Then $d \cdot D_1 = d_3$, $d \cdot D_2 = d_2$, $d \cdot D_3 = d_1$, $d \cdot E_1 = d_0 - d_1$, $d \cdot E_2 = d_0 - d_2$ and $d \cdot E_3 = d_0 - d_3$. The calculations of Figure 6 give for the broken line ① the contribution

$$\binom{d_1}{d_0 - d_2} t_2^{d_0-d_2} x^{-d_1} y^{d_2-d_0}, \tag{4-30}$$

for the broken line ② the contribution

$$\binom{d_2}{d_0 - d_3} t_3^{d_0-d_3} x^{d_0-d_3} y^{d_0-d_2-d_3}, \tag{4-31}$$

and from the broken line ③ the contribution

$$\binom{d_3}{d_0 - d_1} t_1^{d_0-d_1} x^{d_3+d_1-d_0} y^{d_3}. \tag{4-32}$$

Taken together, we obtain the following result.

Theorem 4.8 *We have*

$$N_{0,d}^{\log,\psi}(\mathrm{dP}_3(0,0,0)) = \binom{d_1}{d_0 - d_2}\binom{d_2}{d_0 - d_3}\binom{d_3}{d_0 - d_1}. \tag{4-33}$$

For the 2–point invariant, the tropical multiplicity at p is

$$\left|\det\begin{pmatrix} d_1 & d_1 + d_3 - d_0 \\ -d_2 + d_0 & d_3 \end{pmatrix}\right| = |d_1 d_3 + d_1 d_2 + d_2 d_3 - d_0 d_2 - d_0 d_1 - d_0 d_3 + d_0^2|. \tag{4-34}$$

For the invariant to be nonzero, the curve class needs to lie in the effective cone determined (see Proposition 2.5) by

$$d_0 \geq 0, \quad d_i \geq 0, \quad d_1 + d_2 + d_3 \geq d_0. \tag{4-35}$$

Also, for the binomial coefficients to be nonzero, the curve class needs to satisfy the equations

$$0 \leq d_0 - d_2 \leq d_1, \quad 0 \leq d_0 - d_3 \leq d_2, \quad 0 \leq d_0 - d_1 \leq d_3. \tag{4-36}$$

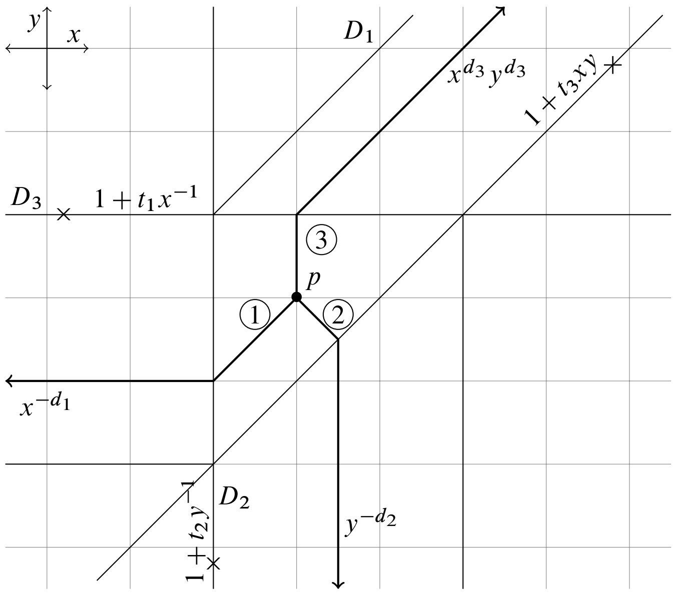


Figure 6: $\mathrm{dP}_3(0,0,0)$.

These inequalities determine a cone. Using the *Polyhedra* package of Macaulay2, in the basis

$$(H-E_1-E_2-E_3, E_1, E_2, E_3)$$

we find extremal rays generated by

$$(1,1,1,0),\quad (1,1,0,1),\quad (1,0,1,1),\quad (2,1,1,1). \tag{4-37}$$

Using this as a new basis, we find that the quadratic form in (4-34) is given by

$$xy+xz+yz+w(x+y+z)+w^2, \tag{4-38}$$

which is always positive in the cone. Therefore, we have proven the following result.

Theorem 4.9 *The generating function* $\mathsf{N}_d^{\log}(\mathrm{dP}_3(0,0,0))(\hbar)$ *equals*

$$\frac{[d_0^2-d_1(d_0-d_2)-d_2(d_0-d_3)-d_3(d_0-d_1)]_q}{[1]_q}\begin{bmatrix} d_1 \\ d_0-d_2\end{bmatrix}_q\begin{bmatrix} d_2 \\ d_0-d_3\end{bmatrix}_q\begin{bmatrix} d_3 \\ d_0-d_1\end{bmatrix}_q, \tag{4-39}$$

where $q=\mathrm{e}^{\mathrm{i}\hbar}$.

4.6 Toric models: $l=4$

There is only one 4–component log Calabi–Yau surface with maximal boundary, namely the toric surface $\mathbb{F}_0(0,0,0,0)$. For $d=d_1H_1+d_2H_2$, through tropical correspondence [92; 97; 85; 86], we calculated in [18] that

$$N_{0,d}^{\log,\psi}(\mathbb{F}_0(0,0,0,0))=1 \quad\text{and}\quad N_{0,d}^{\log}(\mathbb{F}_0(0,0,0,0))=d_1^2d_2^2. \tag{4-40}$$

To obtain the higher-genus invariant, we replace the tropical multiplicities by the Block–Göttsche multiplicities [10]. Applying [12] we obtain the following result.

Theorem 4.10 *We have*

$$\mathsf{N}^{\log}_{d}(\mathbb{F}_0(0,0,0,0))(\hbar) = \frac{[d_1 d_2]_q^2}{[1]_q^2}. \tag{4-41}$$

5 Log-local correspondence

In this section, we prove the following log-local correspondence theorem.

Theorem 5.1 *For every nef Looijenga pair $Y(D)$, the genus-zero log invariants $N^{\log}_{0,d}(Y(D))$ and the genus-zero local invariants $N^{\text{loc}}_{0,d}(Y(D))$ are related by*

$$N^{\text{loc}}_{0,d}(Y(D)) = \left(\prod_{j=1}^{l} \frac{(-1)^{d\cdot D_j - 1}}{(d \cdot D_j)} \right) N^{\log}_{0,d}(Y(D)). \tag{5-1}$$

The proof will be divided into two parts. In Section 5.1, we prove the result for $l = 2$ by a degeneration to the normal cone argument. In Section 5.2, we prove the result for $l = 3$ and $l = 4$ by direct comparison of the local results of Section 3 with the log results of Section 4.

5.1 Log-local for 2 components

For convenience in the following proof, we state separately the case $l = 2$ of Theorem 5.1.

Theorem 5.2 *For every 2–component nef Looijenga pair $Y(D)$, we have*

$$N^{\text{loc}}_{0,d}(Y(D)) = \left(\prod_{j=1}^{l} \frac{(-1)^{d\cdot D_j - 1}}{(d \cdot D_j)} \right) N^{\log}_{0,d}(Y(D)). \tag{5-2}$$

The proof of Theorem 5.2 takes the remainder of Section 5.1, and is a degeneration argument in log Gromov–Witten theory.

5.1.1 Construction of the degeneration We first construct the relevant degeneration for a general l–component nef Looijenga pair $Y(D) = (Y, D_1 + \cdots + D_l)$.

Let $\overline{\nu}_{\overline{\mathcal{Y}}} \colon \overline{\mathcal{Y}} \to \mathbb{A}^1$ be the degeneration of Y to the normal cone of D, obtained by blowing up $D \times \{0\}$ in $Y \times \mathbb{A}^1$. Irreducible components of the special fibre $\overline{\mathcal{Y}}_0 := \overline{\nu}_{\overline{\mathcal{Y}}}^{-1}(0)$ are Y and, for every $1 \le j \le l$, $\overline{\mathbb{P}}_j := \mathbb{P}(\mathcal{O} \oplus N_{D_j|Y})$, where $N_{D_j|Y}$ is the normal bundle to D_j in Y. For every double point $p \in D_j \cap D_{j'}$ of D, a local description of $\overline{\mathcal{Y}}_0$ is given by Figure 7, left. In particular, we have a point p^{∂} in $\overline{\mathcal{Y}}_0$ where

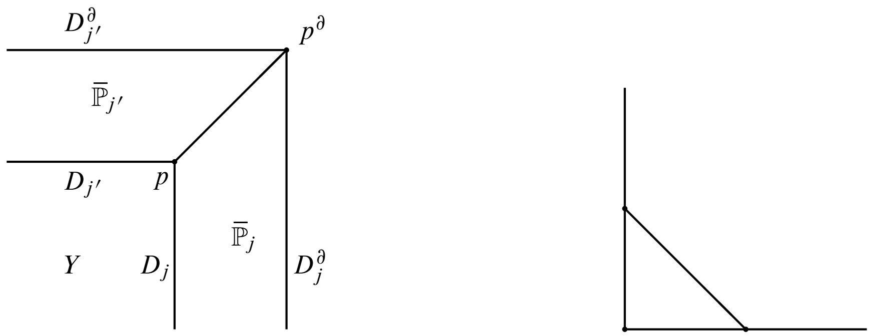


Figure 7: Left: local description of $\overline{\mathcal{Y}}_0$. Right: toric polyhedral decomposition of $\mathbb{R}^2_{\geq 0}$ describing locally $\overline{\mathcal{Y}}_0$ (fan picture).

the total space $\overline{\mathcal{Y}}$ is singular. This can be seen as follows. Locally near a double point $p \in D_{j'} \cap D_j$, the degeneration to the normal cone admits a toric description, whose fan is given by the closure of the cone over the polyhedral decomposition of $\mathbb{R}^2_{\geq 0}$ in Figure 7, right. The point p^{∂} corresponds to the 3–dimensional cone obtained by taking the closure of the cone over the unbounded region of $\mathbb{R}^2_{\geq 0}$ in Figure 7, right. This cone is generated by four rays, so is not simplicial and so p^{∂} is a singular point. More precisely, p^{∂} is an ordinary double point in $\overline{\mathcal{Y}}$. Every singular point of $\overline{\mathcal{Y}}$ is of the form p^{∂} for p a double point of D.

We resolve the singularities of $\overline{\mathcal{Y}}$ by blowing up the ordinary double points p^{∂}, and we obtain a new degeneration $\nu_{\mathcal{Y}}\colon \mathcal{Y} \to \mathbb{A}^1$. The total space $\mathcal{Y}$ is now smooth and the special fibre $\mathcal{Y}_0 := \nu_{\mathcal{Y}}^{-1}(0)$ is a normal crossings divisor on $\mathcal{Y}$. We view $\mathcal{Y}$ as a log scheme for the divisorial log structure defined by $\mathcal{Y}_0 \subset \mathcal{Y}$. Viewing $\mathbb{A}^1$ as a log scheme for the divisorial log structure defined by $\{0\} \subset \mathbb{A}^1$, the morphism $\nu_{\mathcal{Y}}\colon \mathcal{Y} \to \mathbb{A}^1$ can naturally be viewed as a log smooth log morphism.

Irreducible components of $\mathcal{Y}_0$ consist of Y, the strict transform $\mathbb{P}_j$ of the $\overline{\mathbb{P}}_j$ for every $1 \leq j \leq l$, and for every double point p of D the exceptional divisor $\mathbb{S}_p \simeq \mathbb{P}^1 \times \mathbb{P}^1$ created by the blowup of p^{∂}. Locally near a double point $p \in D_j \cap D_{j'}$, irreducible components of $\mathcal{Y}_0$ are glued together as in Figure 8, left. Locally near p, the total space $\mathcal{Y}$ admits a toric description whose fan is the closure of the cone over the polyhedral decomposition of $\mathbb{R}^2_{\geq 0}$ given in Figure 8, right. We remark that the log structure that we consider on $\mathcal{Y}$ is only partially compatible with this local toric description: one needs to remove from the toric boundary the horizontal toric divisors in order to obtain the divisorial log structure defined by the special fibre.

For every $1 \leq j \leq l$, let $\mathcal{D}_j$ be the closure in $\mathcal{Y}$ of the divisor $D_j \times (\mathbb{A}^1 - \{0\}) \subset Y \times (\mathbb{A}^1 - \{0\})$. We have

$$\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{\mathcal{Y}_0} = \mathcal{O}_{\mathcal{Y}_0}\Big(-\Big(D^{\partial}_j \cup \bigcup_{p \in D_j} D^{\partial}_{j,p}\Big)\Big), \tag{5-3}$$

where the union is taken over the double points p of D contained in D_j.

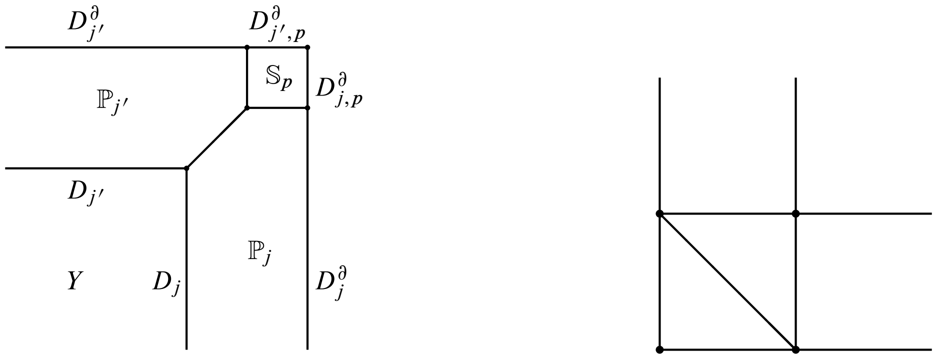


Figure 8: Left: local description of $\mathcal{Y}_0$. Right: toric polyhedral decomposition of $\mathbb{R}^2_{\geq 0}$ describing locally $\mathcal{Y}_0$ (fan picture).

We define $\mathcal{V} := \mathrm{Tot}\big(\bigoplus_{j=1}^l \mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)\big)$ and denote by $\pi_{\mathcal{V}} : \mathcal{V} \to \mathcal{Y}$ and $\nu_{\mathcal{V}} : \mathcal{V} \to \mathbb{A}^1$ the natural projections. We also denote by $\mathcal{V}_0 := \nu_{\mathcal{V}}^{-1}(0)$ the special fibre and by $\pi_{\mathcal{V}_0} : \mathcal{V}_0 \to \mathcal{Y}_0$ the restriction of $\pi_{\mathcal{V}}$ to the special fibre.

The irreducible components of $\mathcal{V}_0$ are

$$\mathcal{V}_{0,Y} := \mathrm{Tot}(\mathcal{O}_Y^{\oplus l}), \tag{5-4}$$

$$\mathcal{V}_{0,j} := \mathrm{Tot}(\mathcal{O}_{\mathbb{P}_j}(-D_j^\partial) \oplus \mathcal{O}_{\mathbb{P}_j}^{\oplus (l-1)}) \quad \text{for every } 1 \leq j \leq l, \tag{5-5}$$

and, for every double point $p \in D_j \cap D_{j'}$ of D,

$$\mathcal{V}_{0,p} := \mathrm{Tot}(\mathcal{O}_{\mathbb{S}_p}(-D^\partial_{j,p}) \oplus \mathcal{O}_{\mathbb{S}_p}(-D^\partial_{j',p}) \oplus \mathcal{O}_{\mathbb{S}_p}^{\oplus(l-2)}). \tag{5-6}$$

We view $\mathcal{V}$ as a log scheme for the divisorial log structure defined by $\mathcal{V}_0 \subset \mathcal{V}$, and then $\nu_{\mathcal{V}} : \mathcal{V} \to \mathbb{A}^1$ is naturally a log smooth log morphism. We remark that the log structure on $\mathcal{V}$ is the pullback of the log structure on $\mathcal{Y}$, ie the log morphism $\pi_{\mathcal{V}} : \mathcal{V} \to \mathcal{Y}$ is strict. In particular, $\mathcal{V}$ and $\mathcal{Y}$ have identical tropicalisations.

For every $1 \leq j \leq l$, we consider the projectivisation $\mathbb{P}(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j) \oplus \mathcal{O}_{\mathcal{Y}})$ of $\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)$ and the corresponding fibrewise compactification

$$\boldsymbol{P} := \mathbb{P}(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_1) \oplus \mathcal{O}_{\mathcal{Y}}) \times_{\mathcal{Y}} \cdots \times_{\mathcal{Y}} \mathbb{P}(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_l) \oplus \mathcal{O}_{\mathcal{Y}}) \tag{5-7}$$

of $\mathcal{V}$. We denote by $\pi_{\boldsymbol{P}} : \boldsymbol{P} \to \mathcal{Y}$ and $\nu_{\boldsymbol{P}} : \boldsymbol{P} \to \mathbb{A}^1$ the natural projections. We also denote by $\boldsymbol{P}_0 := \nu_{\boldsymbol{P}}^{-1}(0)$ the special fibre and by $\pi_{\boldsymbol{P}_0} : \boldsymbol{P}_0 \to \mathcal{Y}_0$ the restriction of $\pi_{\boldsymbol{P}}$ to the special fibre. We denote by $\boldsymbol{P}_{0,Y}$, $\boldsymbol{P}_{0,j}$ and $\boldsymbol{P}_{0,p}$ the irreducible components of $\boldsymbol{P}_0$ obtained by compactification of $\mathcal{V}_{0,Y}$, $\mathcal{V}_{0,j}$ and $\mathcal{V}_{0,p}$.

We view $\boldsymbol{P}$ as a log scheme for the divisorial log structure defined by $\boldsymbol{P}_0 \subset \boldsymbol{P}$, and then $\nu_{\boldsymbol{P}} : \boldsymbol{P} \to \mathbb{A}^1$ is naturally a log smooth log morphism. We remark that the log structure on $\boldsymbol{P}$ is the pullback of the log structure on $\mathcal{Y}$, ie the log morphism $\pi_{\boldsymbol{P}} : \boldsymbol{P} \to \mathcal{Y}$ is strict. In particular, $\boldsymbol{P}$ and $\mathcal{Y}$ have identical tropicalisations.

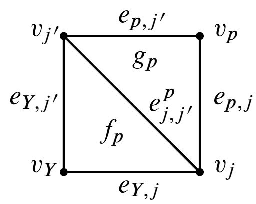


Figure 9: Local description of Δ.

Let Δ be the polyhedral complex obtained by taking the fibre over 1 of the tropicalisation of $\nu_{\boldsymbol{P}} : \boldsymbol{P} \to \mathbb{A}^1$. Combinatorially, Δ is the dual intersection complex of the special fibre $\mathcal{V}_0$; see Figure 9. Vertices of Δ consist of

- v_Y corresponding to the irreducible component $\boldsymbol{P}_{0,\mathcal{Y}}$,
- v_j corresponding to the irreducible component $\boldsymbol{P}_{0,j}$ for every $1 \le j \le l$,
- v_p corresponding to the irreducible component $\boldsymbol{P}_{0,p}$ for every double point p of D.

Edges of Δ consist of

- $e_{Y,j}$ connecting v_Y and v_j for every $1 \le j \le l$, corresponding to the divisor $\boldsymbol{P}_{0,Y} \cap \boldsymbol{P}_{0,j}$,
- $e^p_{j,j'}$ connecting v_j and $v_{j'}$ for every double point $p \in D_j \cap D_{j'}$ of D, corresponding to the component of the divisor $\boldsymbol{P}_{0,j} \cap \boldsymbol{P}_{0,j'}$ containing p,
- $e_{p,j}$ connecting v_p and v_j for every double point $p \in D_j \cap D_{j'}$ of D, corresponding to the divisor $\boldsymbol{P}_{0,j} \cap \boldsymbol{P}_{0,p}$, and $e_{p,j'}$ connecting v_p and $v_{j'}$, corresponding to the divisor $\boldsymbol{P}_{0,j'} \cap \boldsymbol{P}_{0,p}$.

Faces of Δ consist of

- a triangle f_p of sides $e_{Y,j}$, $e_{Y,j'}$, $e^p_{j,j'}$ for every double point $p \in D_j \cap D_{j'}$ of D, corresponding to the triple intersection $\boldsymbol{P}_{0,Y} \cap \boldsymbol{P}_{0,j} \cap \boldsymbol{P}_{0,j'}$,
- a triangle g_p of sides $e^p_{j,j'}$, $e_{p,j}$, $e_{p,j'}$ for every double point $p \in D_j \cap D_{j'}$ of D, corresponding to the triple intersection $\boldsymbol{P}_{0,p} \cap \boldsymbol{P}_{0,j} \cap \boldsymbol{P}_{0,j'}$.

As we are assuming that the components of D form a cycle, the boundary $\partial\Delta$ of Δ can be described as

$$\partial\Delta = \bigcup_{1 \le j \le l} (\partial\Delta)_j, \tag{5-8}$$

where for every $1 \le j \le l$,

$$(\partial\Delta)_j := \bigcup_{p \in D_j \cap D_{j'}} e_{p,j}. \tag{5-9}$$

We view $\boldsymbol{P}_0$ as a log scheme by restriction of the log structure on $\boldsymbol{P}$. We denote by $\nu_{\boldsymbol{P}_0} : \boldsymbol{P}_0 \to \mathrm{pt}_{\mathbb{N}}$ the corresponding log smooth log morphism to the standard log point. We view the curve class d as a class on $\boldsymbol{P}_0$ via the embedding $\mathcal{Y}_0 \to \boldsymbol{P}_0$ induced by the zero section of $\mathcal{V}_0$. Let $\overline{\mathrm{M}}_{0,m}(Y^{\mathrm{loc}}(D), d)$ be the moduli space of genus-zero class d stable log maps to $\nu_{\boldsymbol{P}_0}\colon \boldsymbol{P}_0 \to \mathrm{pt}_{\mathbb{N}}$ with m marked points with contact order 0 with $\boldsymbol{P}_{0,Y}$. Let $[\overline{\mathrm{M}}_{0,m}(\boldsymbol{P}_0, d)]^{\mathrm{vir}}$ be the corresponding virtual fundamental class, of dimension $l-1+m$. Using the nefness of the divisors D_j, the condition $d \cdot D_j > 0$ for every $1 \le j \le l$, and the deformation invariance of log Gromov–Witten invariants, we have

$$N^{\mathrm{loc}}_{0,d}(Y(D)) = \int_{[\overline{\mathrm{M}}_{0,l-1}(\boldsymbol{P}_0,d)]^{\mathrm{vir}}} \prod_{k=1}^{l-1} \mathrm{ev}_k^*(\pi^*_{\boldsymbol{P}_0}[\mathrm{pt}_Y]), \tag{5-10}$$

where ev_k is the evaluation at the k^{th} interior marked point and $[\mathrm{pt}_Y]$ is the class of a point on $Y \subset \mathcal{Y}_0$.

5.1.2 Degeneration formula According to the decomposition formula of Abramovich, Chen, Gross and Siebert [2], we have

$$[\overline{\mathrm{M}}_{0,l-1}(\boldsymbol{P}_0, d)]^{\mathrm{vir}} = \sum_{h\colon \Gamma \to \Delta} \frac{m_h}{|\mathrm{Aut}(h)|} [\overline{\mathrm{M}}^h_{0,l-1}(\boldsymbol{P}_0, d)]^{\mathrm{vir}}. \tag{5-11}$$

The sum is over the genus-zero rigid decorated parametrised tropical curves $h\colon \Gamma \to \Delta$, where Γ has $l-1$ unbounded edges, all contracted by h to v_Y, and the sum of classes attached to the vertices of Γ is d. The moduli space $\overline{\mathrm{M}}^h_{0,l-1}(\boldsymbol{P}_0, d)$ parametrises genus-zero class d stable log maps to $\nu_{\boldsymbol{P}_0}\colon \boldsymbol{P}_0 \to \mathrm{pt}_{\mathbb{N}}$ marked by h.

Therefore, we have

$$N^{\mathrm{loc}}_{0,d}(Y(D)) = \sum_{h\colon \Gamma \to \Delta} \frac{m_h}{|\mathrm{Aut}(h)|} N^{\mathrm{loc},h}_{0,d}(Y(D)), \tag{5-12}$$

where

$$N^{\mathrm{loc},h}_{0,d}(Y(D)) := \int_{[\overline{\mathrm{M}}^h_{0,l-1}(\boldsymbol{P}_0,d)]^{\mathrm{vir}}} \prod_{k=1}^{l-1} \mathrm{ev}_k^*(\pi^*_{\boldsymbol{P}_0}[\mathrm{pt}_Y]), \tag{5-13}$$

and ev_k is the evaluation at the k^{th} marked point. Thus, for every $h\colon \Gamma \to \Delta$, we have to compute $N^{\mathrm{loc},h}_{d}(Y(D))$.

Let Δ^h be a polyhedral complex obtained by refining the polyhedral decomposition of Δ and containing the $h(\Gamma)$ in its one-skeleton, ie such that, for every vertex V of Γ, $h(V)$ is a vertex of Δ^h, and for every edge E of Γ, $h(E)$ is an edge of Δ^h. We denote by $\mathcal{Y}^h_0$, $\mathcal{V}^h_0$ and $\boldsymbol{P}^h_0$ the corresponding log modifications of $\mathcal{Y}_0$, $\mathcal{V}_0$ and $\boldsymbol{P}_0$. Let $\overline{\mathrm{M}}^h_{0,l-1}(\boldsymbol{P}^h_0, d)$ the moduli space of stable log maps to $\boldsymbol{P}^h_0$ marked by h. By the invariance of log Gromov–Witten invariants under log modification [3], we have

$$N^{\mathrm{loc},h}_{0,d}(Y(D)) := \int_{[\overline{\mathrm{M}}^h_{0,l-1}(\boldsymbol{P}^h_0,d)]^{\mathrm{vir}}} \prod_{k=1}^{l-1} \mathrm{ev}_k^*(\pi^*_{\boldsymbol{P}^h_0}[\mathrm{pt}_Y]). \tag{5-14}$$

For every vertex V of Γ, let $\boldsymbol{P}_V$ be the irreducible component of $\boldsymbol{P}^h_0$ corresponding to the vertex $h(V)$ of Δ^h. We view $\boldsymbol{P}_V$ as a log scheme for the divisorial log structure defined by the divisor $\partial \boldsymbol{P}_V$, which is the union of intersection divisors with the other irreducible components of $\boldsymbol{P}_0^h$. Similarly, we define the component Y_V of $\mathcal{Y}_0^h$ and ∂Y_V, so that $\boldsymbol{P}_V$ is a $(\mathbb{P}^1)^l$–bundle over Y_V.

If $h(V) \in \Delta^h - \partial\Delta^h$, then $\boldsymbol{P}_V$ is the trivial $(\mathbb{P}^1)^l$–bundle over Y_V. If furthermore, $h(V) \notin \bigcup_{j=1}^l e_{Y,j}$, then $(Y_V, \partial Y_V)$ is a toric variety with its toric boundary.

If $h(V) \in e_{p,j} - v_p$ for some $p \in D_j \cap D_{j'}$, let $D_{j,V}^{\partial}$ be the irreducible component of $\mathcal{D}_j \cap \mathcal{Y}_0^h$ contained in Y_V. Then $\boldsymbol{P}_V$ is the fibrewise product over Y_V of the $\mathbb{P}^1$–bundle

$$\mathbb{P}\big(\mathcal{O}_{Y_V}(-D_{j,V}^{\partial}) \oplus \mathcal{O}_{Y_V}\big) \tag{5-15}$$

with the trivial $(\mathbb{P}^1)^{l-1}$–bundle. Moreover, $(Y_V, \partial Y_V \cup D_{j,V}^{\partial})$ is a toric variety with its toric boundary.

If $h(V) = v_p$ for some $p \in D_j \cap D_{j'}$, then, still denoting by $\mathbb{S}_p$, $D_{j,p}^{\partial}$ and $D_{j,p}^{\partial}$ the strict transforms in $\mathcal{Y}_0^h$ of $\mathbb{S}_p$, $D_{j,p}^{\partial}$, and $D_{j,p}^{\partial}$, $\boldsymbol{P}_V$ is the fibrewise product over $Y_V = \mathbb{S}_p$ of the $\mathbb{P}^1$–bundle

$$\mathbb{P}\big(\mathcal{O}_{\mathbb{S}_p}(-D_{j,p}^{\partial}) \oplus \mathcal{O}_{\mathbb{S}_p}\big), \tag{5-16}$$

of the $\mathbb{P}^1$–bundle

$$\mathbb{P}\big(\mathcal{O}_{\mathbb{S}_p}(-D_{j',p}^{\partial}) \oplus \mathcal{O}_{\mathbb{S}_p}\big), \tag{5-17}$$

and of the trivial $(\mathbb{P}^1)^{l-2}$–bundle. Moreover, $(Y_V, \partial Y_V \cup D_{j,p}^{\partial} \cup D_{j',p}^{\partial})$ is a toric variety with its toric boundary.

For every vertex V of Γ, let M_V be the moduli space of genus-zero stable log maps to $\boldsymbol{P}_{0,V}^h$, with class given by the class decoration of V, and contact orders specified by the local behaviour of h around V. Our goal is to compute the invariant $N_d^{\mathrm{loc},h}(Y(D))$ in terms of the virtual classes $[M_V]^{\mathrm{vir}}$. We are in a particularly favourable situation: we consider curves of genus zero and the dual intersection complex Δ^h has dimension 2. In such case, the degeneration formula in log Gromov–Witten theory has a particular simple form, as described in Section 6.5.2 of [105]; see also Section 4 of [104] for the corresponding discussion in the language of exploded manifolds.

We choose a flow on Γ such that unbounded edges are incoming and such that every vertex has at most one outgoing edge. Such flow exists as Γ has genus zero and then there is exactly one vertex, which we denote by V_0, without outgoing incident edge, and which we call the *sink* of the flow. All vertices distinct from V_0 have exactly one outgoing edge. In fact, for every vertex V of Γ, we can find such flow with sink $V_0 = V$.

For every edge E of Γ, we denote by $\boldsymbol{P}_E$ the stratum of $\boldsymbol{P}_0^h$ dual to E. The stratum $\boldsymbol{P}_E$ is a divisor if E is bounded and is the irreducible component $\mathbb{P}_V$ if E is unbounded and incident to the vertex V. For every E, $\mathbb{P}_E$ is a $(\mathbb{P}^1)^l$–bundle over a stratum Y_E of $\mathcal{Y}_0^h$, and we denote by $\pi_E \colon \boldsymbol{P}_E \to Y_E$ the corresponding projection.

For every edge E incident to a vertex V, we have the evaluation map

$$\mathrm{ev}_{V,E} \colon M_V \to \boldsymbol{P}_E. \tag{5-18}$$

For every vertex V distinct from V_0, let $\mathcal{E}_{\mathrm{in}}(V)$ be the set of incoming incident edges to V, and let E_V be the outgoing incident edge to V. The virtual class $[M_V]^{\mathrm{virt}}$ defines a map

$$\eta_V\colon \prod_{E\in\mathcal{E}_{\mathrm{in}}(V)} \mathrm{H}^*(\boldsymbol{P}_E)\to \mathrm{H}^*(\boldsymbol{P}_{E_V}) \tag{5-19}$$

by

$$\eta_V\biggl(\prod_{E\in\mathcal{E}_{\mathrm{in}}(V_0)} \alpha_E\biggr) := (\mathrm{ev}_{V,E_V})_*\biggl(\biggl(\prod_{E\in\mathcal{E}_{\mathrm{in}}(V)} \mathrm{ev}^*_{V,E}\,\alpha_E\biggr)\cap [M_V]^{\mathrm{virt}}\biggr). \tag{5-20}$$

Note that if $\mathcal{E}_{\mathrm{in}}(V)$ is empty, then η_V is a map of the form

$$\eta_V\colon \mathbb{Q}\to \mathrm{H}^*(\boldsymbol{P}_{E_V}). \tag{5-21}$$

Denote by $\mathcal{E}_{\mathrm{in}}(V_0)$ the set of incoming incident edges to V_0. The virtual class $[M_{V_0}]^{\mathrm{virt}}$ defines a map

$$\eta_{V_0}\colon \prod_{E\in\mathcal{E}_{\mathrm{in}}(V_0)} \mathrm{H}^*(\boldsymbol{P}_E)\to \mathbb{Q} \tag{5-22}$$

by

$$\eta_{V_0}\biggl(\prod_{E\in\mathcal{E}_{\mathrm{in}}(V_0)} \alpha_E\biggr) := \int_{[M_{V_0}]^{\mathrm{virt}}} \prod_{E\in\mathcal{E}_{\mathrm{in}}(V_0)} \mathrm{ev}^*_{V_0,E}\,\alpha_E. \tag{5-23}$$

Denote by $\mathcal{E}_\infty(\Gamma)$ the set of unbounded edges of Γ. Composing the maps η_V and η_{V_0}, we obtain a map

$$\eta_h\colon \prod_{E\in\mathcal{E}_\infty(\Gamma)} \mathrm{H}^*(\boldsymbol{P}_E)\to \mathbb{Q}. \tag{5-24}$$

For every edge E of Γ, let $\mathrm{pt}_E\in \mathrm{H}^2(Y_E)$ be the class of a point on Y_E. We consider the class $\pi_E^*\mathrm{pt}_E\in \mathrm{H}^2(\boldsymbol{P}_E)$. The degeneration formula is then

$$N_d^{\mathrm{loc},h}(Y(D)) = \eta_h\biggl(\prod_{E\in\mathcal{E}_\infty(\Gamma)} \pi_E^*\mathrm{pt}_E\biggr). \tag{5-25}$$

We define a rigid genus-zero parametrised tropical curve $\bar{h}\colon \bar{\Gamma}\to\Delta$ as follows. Let $\bar{\Gamma}$ be the star-shaped graph consisting of vertices V_j for $0\le j\le l$, and edges E_j connecting V_0 and V_j for $1\le j\le l$. We assign the length $1/(d\cdot D_j)$ to the edge E_j. Let $\bar{h}\colon\bar{\Gamma}\to\Delta$ be the piecewise linear map such that $\bar{h}(V_0)=v_Y$ and $\bar{h}(V_j)=v_j$ for $1\le j\le l$. In particular, we have $\bar{h}(E_j)=e_{Y,j}$ for $1\le j\le j$. As $e_{Y,j}$ has integral length 1, we deduce that E_j has weight $d\cdot D_j$. Finally, curve classes decoration of the vertices are given by: $d_{V_0}=d$ and d_{V_j} is equal to $(d\cdot D_j)$ times the class of a $\mathbb{P}^1$–fibre of $\mathbb{P}_j$ for $1\le j\le l$.

Lemma 5.3 *We have*

$$N_{0,d}^{\mathrm{loc},\bar{h}}(Y(D)) = \biggl(\prod_{j=1}^{l} \frac{(-1)^{d\cdot D_j-1}}{(d\cdot D_j)^2}\biggr) N_{0,d}^{\log}(Y(D)). \tag{5-26}$$

Proof We choose the flow on Γ with sink V_0. Applying the degeneration formula gives immediately the result, using the fact that the normal bundle in $\boldsymbol{P}_0$ to a $\mathbb{P}^1$–fibre of $\mathbb{P}_j$ is $\mathcal{O}(-1) \oplus \mathcal{O}^{\oplus(l+1)}$ and so the corresponding multicover contribution is

$$\frac{(-1)^{d \cdot D_j - 1}}{(d \cdot D_j)^2}, \tag{5-27}$$

by [20, Proof of Theorem 5.1]. □

Theorem 5.4 *Assume $l = 2$. Let $h\colon \Gamma \to \Delta$ be a rigid decorated parametrised tropical curve as above with $N_{0,d}^{\mathrm{loc},h}(Y(D)) \neq 0$. Then $h = \overline{h}$.*

Theorem 5.4 follows from a judicious analysis of the possible topologies of contributing tropical curves, which we perform in Appendix C. Theorem 5.2 then follows from the combination of Theorem 5.4, Lemma 5.3, and the decomposition formula using that $|\mathrm{Aut}(\overline{h})| = 1$ and $m_{\overline{h}} = \prod_{j=1}^{l} (d \cdot D_j)$.

5.2 The log-local correspondence for 3 and 4 components

We end the proof of Theorem 5.1 for $l = 3$ and $l = 4$. For $l = 3$, it is enough to treat the case of $\mathrm{dP}_3(0,0,0)$, as all the other 3–component cases are obtained from it by blowup, and the result is preserved under blowup by combination of Propositions 3.2 and 4.3. The result for $\mathrm{dP}_3(0,0,0)$ follows by comparing the local result given by Theorem 3.5 with the log result given by Theorem 4.9.

For $l = 4$, the result follows by comparing the local result given by (3-30) and the log result given by (4-41).

6 Open Gromov–Witten theory

In this section we relate the quantised scattering calculations of Section 4 to the higher-genus open Gromov–Witten theory of Aganagic–Vafa A–branes. We first give in Section 6.1 an overview of the framework of [76] to cast open toric Gromov–Witten theory within the realm of formal relative invariants, and recall the topological vertex formalism of Aganagic, Klemm, Mariño and Vafa. Our treatment throughout this section, while self-contained, will keep the level of detail to the necessary minimum, and we refer the reader to [76; 39] for further details. The reader who is familiar with this material may wish to skip to Section 6.2, where the stable log counts of Section 4 are related to open Gromov–Witten theory, with the main statement condensed in Theorem 6.7, and proved in Section 6.3.

In the following, for a partition $\lambda \vdash d$ of $d \in \mathbb{N}$ we write $|\lambda| = d$ for the order of λ, $\ell_\lambda = r$ for the cardinality of the partitioning set, $\kappa_\lambda := \sum_{i=1}^{\ell_\lambda} \lambda_i(\lambda_i - 2i + 1)$ for its second Casimir invariant, and let $m_j(\lambda) := \#\{\lambda_i \mid \lambda_i = j\}_{i=1}^{\ell_\lambda}$ and $z_\lambda := \prod_j m_j(\lambda)!\, j^{m_j(\lambda)}$. We furthermore denote by $\mathcal{P}$ the set of partitions, and $\mathcal{P}^d$ the set of partitions of order d. We will extensively need, particularly in the proof of Theorem 6.7, some classical results on principally specialised shifted symmetric functions, for which notation and necessary basic results are collected in Appendix D.

6.1 Toric special Lagrangians

Let X be a smooth complex toric threefold with $K_X \simeq \mathcal{O}_X$. If the affinisation morphism to $\mathrm{Spec}(\Gamma(X, \mathcal{O}_X))$ is projective, X can be realised as a symplectic quotient $\mathbb{C}^{r+3} /\!\!/ G$, where $G \simeq U(1)^r$ acts on the affine coordinates $\{z_i\}_{i=1}^{r+3}$ of $\mathbb{C}^{r+3} = \mathrm{Spec}\mathbb{C}[z_1, \dots, z_{r+3}]$ by

$$(t_1, \dots, t_r) \cdot (z_1, \dots, z_{r+3}) = \left(\prod_{i=1}^{r} t_i^{w_1^{(i)}} \cdot z_1, \dots, \prod_{i=1}^{r} t_i^{w_{r+3}^{(i)}} \cdot z_{r+3} \right),$$

where $w_j^{(i)} \in \mathbb{Z}$ for $i = 1, \dots, r$ and $j = 1, \dots, r+3$ are the weights of the G–action [61]. This is a Hamiltonian action with respect to the canonical Kähler form on $\mathbb{C}^{r+3}$,

$$\omega := \frac{\mathrm{i}}{2} \sum_{i=1}^{r+3} \mathrm{d}z_i \wedge \mathrm{d}\bar{z}_i, \tag{6-1}$$

with moment map

$$\tilde{\mu}(z_1, \dots, z_{r+3}) = \left(\sum_{i=1}^{r+3} w_i^{(1)} |z_i|^2, \dots, \sum_{i=1}^{r+3} w_i^{(r)} |z_i|^2 \right).$$

If $(t_1, \dots, t_k) \in \mathrm{H}^{1,1}(X; \mathbb{R}) \simeq (\mathfrak{u}(1)^r)^\star$ is a Kähler class, then X is the geometric quotient

$$X = \tilde{\mu}^{-1}(t_1, \dots, t_r)/G, \tag{6-2}$$

with symplectic structure given by the Marsden–Weinstein reduction ω_t of (6-1) onto the quotient (6-2), where $[\omega_t] = (t_1, \dots, t_r) \in \mathrm{H}^{1,1}(X; \mathbb{R})$.

We will be concerned with a class of special Lagrangian submanifolds $L = L_{\hat{w},c}$ of (X, ω_t) constructed by Aganagic and Vafa [7], which are invariant under the natural Hamiltonian torus action on X. They are defined by

$$\sum_{i=1}^{r+3} \hat{w}_i^1 |z_i|^2 = c, \quad \sum_{i=1}^{r+3} \hat{w}_i^2 |z_i|^2 = 0, \quad \sum_{i=1}^{r+3} \arg z_i = 0, \tag{6-3}$$

with $\hat{w}_i^a \in \mathbb{Z}$, $\sum_{i=1}^{r+3} \hat{w}_i^a = 0$ and $c \in \mathbb{R}$. These Lagrangians have the topology of $\mathbb{R}^2 \times S^1$, and they intersect a unique torus fixed curve C_L along an S^1: we say that L is an *inner* (resp. *outer*) brane if $C_L \simeq \mathbb{P}^1$ (resp. $\mathbb{C}$). Throughout the foregoing discussion we will assume that L is always an outer brane.

Let $T \simeq (\mathbb{C}^\star)^2$ be the algebraic subtorus of $(\mathbb{C}^\star)^3 \subset X$ acting trivially on K_X, and $T_{\mathbb{R}} \simeq U(1)^2$ be its maximal compact subgroup. Then by construction any toric Lagrangian L is preserved by $T_{\mathbb{R}}$, which acts on $\mathbb{C} \times S^1$ by scaling $(\lambda_1, \lambda_2) \cdot (w, \theta) \to (\lambda_1 w, \lambda_2 \theta)$. Writing $\mu_T : X \to \mathbb{R}^2 \simeq (\mathfrak{u}(1)^2)^*$ for the moment map of the $T_{\mathbb{R}}$–action, the union of the 1–dimensional $(\mathbb{C}^\star)^3$ orbit closures of X is mapped by μ_T to a planar trivalent metric graph Γ_X whose sets of vertices $(\Gamma_X)_0$, compact edges $(\Gamma_X)_1^{\mathrm{cp}}$ and noncompact edges $(\Gamma_X)_1^{\mathrm{nc}}$ correspond to T–fixed points, T–invariant proper curves, and T–invariant affine lines in X respectively. Since the moment map is an integral quadratic form, the tangent directions of the edges have rational slopes in $\mathbb{R}^2$: we can explicitly keep track of this information by regarding Γ_X as a topological graph[11] decorated by the assignment to each vertex $v \in (\Gamma_X)_0$ of primitive integral lattice vectors $\mathfrak{p}_v^e \in \mathbb{Z}^2$, representing the directions of the edges e emanating from $v \in (\Gamma_X)_0$. The graph Γ_X is determined bijectively by the weights $w_j^{(i)}$, and knowing it suffices to reconstruct X.

Remark 6.1 Let $\Sigma(X)$ be the fan of X. As $K_X \simeq \mathcal{O}_X$, $\Sigma(X)$ can be described as a cone in $\mathbb{R}^3$ over a polyhedral decomposition of an integral polygon P in $\mathbb{R}^2 \times \{1\} \subset \mathbb{R}^3$. The graph Γ_X can be obtained as the dual graph of the polyhedral decomposition of P taking orientations to be outgoing at every vertex. Conversely, one can recover (the $\mathrm{SL}(2,\mathbb{Z})$–equivalence class of) $P \subset \mathbb{R}^2$ and its decomposition as the dual polygon of Γ_X, and then $\Sigma(X)$ as the cone in $\mathbb{R}^3$ over $P \subset \mathbb{R}^2 \times \{1\}$.

If L is a toric outer Lagrangian, its image under μ_T is a point $\mu_T(L)$ lying on the noncompact edge $\mu_T(C)$ representing the curve it is incident to. Write $e_L := \mu_T(C)$, v for its adjacent vertex, and e'_L for the first edge met by moving clockwise from e_L with respect to the orientation determined by the plane containing Γ_X.

Definition 6.1 A *framing* of L is the choice of an integral vector f such that $\mathsf{p}_v^{e_L} \wedge \mathsf{p}_v^{e'_L} = \mathsf{p}_v^{e_L} \wedge \mathsf{f}$; equivalently, $\mathsf{f} = \mathsf{p}_v^{e'_L} - f\,\mathsf{p}_v^{e_L}$ for some $f \in \mathbb{Z}$. We say that L is canonically framed if $f = 0$, ie $\mathsf{f} = \mathsf{p}_v^{e'_L}$.

Remark 6.2 By construction, since $\mathsf{f} \wedge \mathsf{p}_v^{e_L} > 0$, a framing at an outer vertex is always pointing in the clockwise direction.

Definition 6.2 We call (X, L, f) a *toric Lagrangian triple* if

- X is a semiprojective toric CY3 variety,
- $L = \bigsqcup_i L_{\widehat{w}_i, c_i}$ is a disjoint union of Aganagic–Vafa special Lagrangian submanifolds of X, and
- f is the datum of a framing choice for each connected component of L.

We will write $\Gamma_{(X,L,\mathsf{f})}$ for the graph obtained from Γ_X by the extra decoration of an integral vector incident to the edge e_L representing the toric outer Lagrangian L at framing f; see Figure 10.

Example 6.1 Let $w^{(1)} = (1,1,-1,-1)$, $\widehat{w}^{(2)} = (1,0,-1,0)$ and $\widehat{w}^{(3)} = (0,1,-1,0)$. For any $t \neq 0$, the corresponding toric variety X is the resolved conifold $\mathrm{Tot}\big(\mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)\big)$, with $\int_{0_*\mathbb{P}^1} \omega_t = t$. The compact edge e_5 corresponds to the $\mathbb{P}^1$ given by the zero section of X. The edges e_i for $i = 1,2,3,4$ correspond to the T–invariant $\mathbb{A}^1$–fibres above the points $[1:0]$ and $[0:1]$ of the $\mathbb{P}^1$ base. The weights $\widehat{w}^{(i)}$ furthermore determine a toric Lagrangian, whose image in the toric graph lies in e_1, and is depicted in Figure 10 at framing $\mathsf{f} = \mathsf{p}_{v_1}^{e_5} - \mathsf{p}_{v_1}^{e_1}$.

[11] In doing so we forget the metric information about Γ_X which stems from a choice of a Kähler structure on X: this is inconsequential for the definition of the invariants in the next section. We thus make a slight abuse of notation, by indicating the decorated topological graph obtained by forgetting the information about the lengths of the edges by the same symbol Γ_X.

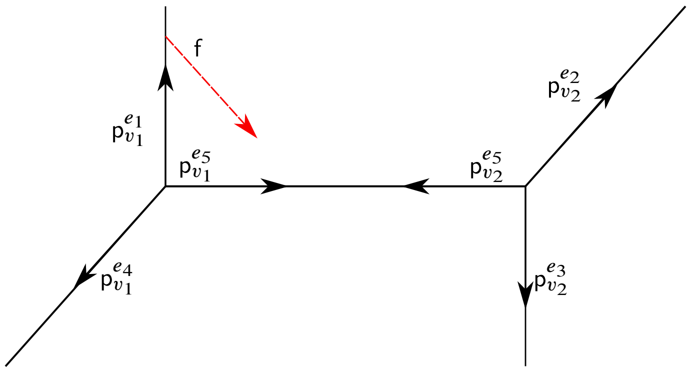


Figure 10: The toric Calabi–Yau graph $\Gamma_{(X,L,\mathsf{f})}$ of the resolved conifold with an outer Lagrangian at framing $\mathsf{f}=\mathsf{f}_{\mathrm{can}}-\mathsf{p}_{v_1}^{e_1}$, ie $f=1$.

6.1.1 Open Gromov–Witten invariants In informal terms, the open Gromov–Witten theory of $(X, L = L_1 \cup \cdots \cup L_s)$ for toric Lagrangians L_i with $i=1,\dots,s$ is a virtual count of maps to X from open Riemann surfaces of fixed genus, relative homology degree, and boundary winding data around $S^1 \hookrightarrow L$. This raises two orders of problems when trying to define these counts in the algebraic category, as the boundary conditions for the curve counts are imposed in odd real dimension, and the target geometry is noncompact. A strategy to address both issues simultaneously for framed outer toric Lagrangians, and which we will follow for the purposes for the paper, was put forward by Li, Liu, Liu and Zhou [76], which we briefly review below. The main idea in [76] is to replace the toric Lagrangian triple (X, L, f) by a formal relative Calabi–Yau pair $(\hat{X}, \hat{D})$, where $\hat{X}$ is obtained as the formal neighbourhood along a partial compactification, specified by L and the framing f, of the toric 1–skeleton of X, and $\hat{D} = \hat{D}_1 + \cdots + \hat{D}_s$ is a formal divisor[12] in the partial compactification $\hat{X}$ with $K_{\hat{X}} + \hat{D} = 0$, the aim being to trade the theory of open stable maps with prescribed windings along the boundary circles on L by a theory of relative stable maps with prescribed ramification profile above torus fixed points in $\hat{X}$, as previously suggested in [77]. The resulting moduli space $\overline{\mathcal{M}}^{\mathrm{rel}}_{g;\beta;\mu_1,\dots,\mu_s}(\hat{X}, \hat{D})$ of degree β stable maps from $\ell(\mu_1)+\cdots+\ell(\mu_s)$–pointed, arithmetic genus g nodal curves with ramification profile μ_i above $\hat{D}_i$ at the punctures is a formal Deligne–Mumford stack carrying a perfect obstruction theory $[\mathcal{T}^1 \to \mathcal{T}^2]$ of virtual dimension $\ell(\mu_1)+\cdots+\ell(\mu_s)$. While the moduli space is not itself proper, it inherits a $T \simeq (\mathbb{C}^\star)^2$ action from $\hat{X}$ with compact fixed loci, and open Gromov–Witten invariants

$$O_{g;\beta;(\mu_1,\dots,\mu_s)}(X, L, \mathsf{f}) := \frac{1}{|\mathrm{Aut}(\vec{\mu})|} \int_{[\overline{\mathcal{M}}^{\mathrm{rel}}_{g;\beta;\mu_1,\dots,\mu_s}(\hat{X},\hat{D})]^{\mathrm{vir},T}} \frac{\mathrm{e}_T(\mathcal{T}^{1,m})}{\mathrm{e}_T(\mathcal{T}^{2,m})}, \tag{6-4}$$

where the $\mathcal{T}^{i,m}$ for $i=1,2$ denote the moving parts of the obstruction theory, and are defined in a standard manner by T–virtual localisation [49]. It is a central result of [76] that the Calabi–Yau condition on T entails that $O_{g,\beta,(\mu_1,\dots,\mu_s)}(X, L, \mathsf{f})$ are nonequivariantly well-defined rational numbers: the invariants

[12] See [76, Section 5] for the details of the relevant construction.

however do depend on the framings f_i specified to construct the formal relative Calabi–Yau $(\widehat{X}, \widehat{D})$, in keeping with expectations from large N duality [100].

It will be helpful to package the open Gromov–Witten invariants $O_{g,\beta,\mu_1,\dots,\mu_s}(X, L, \mathsf{f})$ into formal generating functions. Let $\mathsf{x}^{(i)} = (\mathsf{x}_1^{(i)}, \mathsf{x}_2^{(i)}, \dots)$ for $i = 1, \dots, s$ be formal variables and for a partition μ define $\mathsf{x}_\mu^{(i)} := \prod_{j=1}^{\ell(\mu)} \mathsf{x}_{\mu_j}^{(i)}$. We further abbreviate $\vec{\mathsf{x}}_{\vec{\mu}} = (\mathsf{x}_{\mu_1}^{(1)}, \dots, \mathsf{x}_{\mu_s}^{(s)})$, $\vec{\mu} = (\mu_1, \dots, \mu_s)$, $|\vec{\mu}| = \sum_i |\mu_i|$ and $\ell(\vec{\mu}) = \sum_i \ell(\mu_i)$, and define the connected generating functions

$$
\begin{aligned}
\mathsf{O}_{\beta;\vec{\mu}}(X, L, \mathsf{f})(\hbar) &:= \sum_g \hbar^{2g-2+\ell(\vec{\mu})} O_{g;\beta;\vec{\mu}}(X, L, \mathsf{f}), \\
\mathcal{O}_{\vec{\mu}}(X, L, \mathsf{f})(Q, \hbar) &:= \sum_\beta \mathsf{O}_{\beta,\vec{\mu}}(X, L, \mathsf{f})(\hbar) Q^\beta, \\
\mathfrak{O}(X, L, \mathsf{f})(Q, \hbar, \mathsf{x}) &:= \sum_{\vec{\mu} \in (\mathcal{P})^s} \mathcal{O}_{\vec{\mu}}(X, L, \mathsf{f})(Q, \hbar) \vec{\mathsf{x}}_{\vec{\mu}},
\end{aligned} \tag{6-5}
$$

as well as generating functions of disconnected invariants in the winding number and representation bases

$$
\begin{aligned}
\mathfrak{Z}(X, L, \mathsf{f})(Q, \hbar, \mathsf{x}) &:= \exp\left(\mathfrak{O}(X, L, \mathsf{f})(Q, \hbar, \mathsf{x})\right) \\
&=: \sum_{\vec{\mu} \in (\mathcal{P})^s} \mathcal{Z}_{\vec{\mu}}(X, L, \mathsf{f})(Q, \hbar) \vec{\mathsf{x}}_{\vec{\mu}} \\
&=: \sum_{\vec{\mu} \in (\mathcal{P})^s} \sum_{\vec{\nu} \in (\mathcal{P})^s} \prod_{i=1}^s \frac{\chi_{\nu_i}(\mu_i)}{z_{\mu_i}} \mathcal{W}_{\vec{\nu}}(X, L, \mathsf{f})(Q, \hbar) \vec{\mathsf{x}}_{\vec{\mu}}.
\end{aligned} \tag{6-6}
$$

Here $\chi_\nu(\mu)$ denotes the irreducible character of $S_{|\nu|}$ evaluated on the conjugacy class labelled by μ. When $\mathsf{x} = 0$, (6-6) reduces to the ordinary generating function of disconnected Gromov–Witten invariants of X.

6.1.2 The topological vertex The invariants (6-6) can be computed algorithmically to all genera using the topological vertex of Aganagic, Klemm, Mariño and Vafa [6]. We can succinctly condense this into the following three statements:

(1) Let $X = \mathbb{C}^3$, $L = \bigcup_{i=1}^3 L_i$, and $L_i = L_{\widehat{w}^{(i)},c}$ with $\widehat{w}_j^{(i)} = \delta_{i,j} - \delta_{i,j+1 \bmod 3}$, $i = 1, \dots, 3$ be the outer Lagrangians of $\mathbb{C}^3$ as in Figure 11, and fix framing vectors f_i for each of them. Then

$$
\mathcal{W}_{\vec{\mu}}(\mathbb{C}^3, L, \mathsf{f}) = \prod_{i=1}^3 q^{f_i \kappa(\mu_i)/2} (-1)^{f_i |\mu_i|} \mathcal{W}_{\vec{\mu}}(\mathbb{C}^3, L, \mathsf{f}_{\mathrm{can}}), \tag{6-7}
$$

where $q = \mathrm{e}^{\mathrm{i}\hbar}$.

(2) Let $(X^{(1)}, L^{(1)}, \mathsf{f}^{(1)})$ and $(X^{(2)}, L^{(2)}, \mathsf{f}^{(2)})$ be smooth toric Calabi–Yau 3–folds with framed outer toric Lagrangians $L^{(i)} = \bigcup_{j=1}^{s_i} L_j^{(i)}$. Suppose that there exist noncompact edges $\widetilde{e}_i \in (\Gamma_{(X^{(i)}, L^{(i)}, \mathsf{f}^{(i)})})_1^{\mathrm{nc}}$

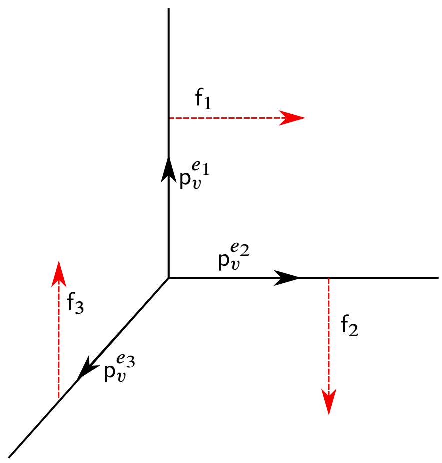


Figure 11: The framed vertex $(\mathbb{C}^3, L_1 \cup L_2 \cup L_3)$, depicted with framings $\mathsf{f}_1 = \mathsf{p}_v^{e_2}$, $\mathsf{f}_2 = \mathsf{p}_v^{e_2} + \mathsf{p}_v^{e_3}$ and $\mathsf{f}_3 = \mathsf{p}_v^{e_1}$.

emanating from vertices $\widetilde{v}_i \in (\Gamma_{(X^{(i)}, L^{(i)}, \mathsf{f}^{(i)})})_0$ such that $\mu_T(L^{(i)}_{s_i}) \cap \widetilde{e}_i \neq \varnothing$, and that moreover, $\mathsf{p}_{\widetilde{v}_1}^{\widetilde{e}_1} = -\mathsf{p}_{\widetilde{v}_2}^{\widetilde{e}_2}$ and $\mathsf{f}^{(1)}_{s_1} = \mathsf{f}^{(2)}_{s_2}$; see Figure 12. We can construct a planar trivalent graph $\Gamma_{X_1 \cup_{e_{12}} X_2}$ decorated with triples of primitive integer vectors at every vertex by considering the disconnected union of $\Gamma_{X^{(1)}}$ and $\Gamma_{X^{(2)}}$, deleting $\widetilde{e}_1$ and $\widetilde{e}_2$, and adding a compact edge e_{12} connecting $\widetilde{v}_1$ to $\widetilde{v}_2$. A toric Calabi–Yau 3–graph reconstructs uniquely a smooth toric CY3 with a T action isomorphic to the T–equivariant formal neighbourhood of the configuration of rational curves specified by the edges, and we call X the threefold determined by the glueing procedure such that $\Gamma_X = \Gamma_{X_1 \cup_{e_{12}} X_2}$. In the same vein, the collection of framed Lagrangians $L^{(i)}$ on X_i determine framed outer Lagrangians $L = \bigcup_{i=1}^{s_1+s_2-2} L_i$ on X: we have canonical projection maps $\pi_i : \Gamma_X \to \Gamma_{X^{(i)}}$, and we place an outer Lagrangian brane at framing f_j on all noncompact edges e such that $\pi_i(e) \cap \mu_T(L^{(i)}_j) \neq \varnothing$ for some j. Write

$$\vec{\mu} = (\mu^{(1)}_1, \dots, \mu^{(1)}_{s_1-1}, \mu^{(2)}_1, \dots, \mu^{(2)}_{s_2-1}),$$
$$\vec{\mu}^{(1)}_{12} = (\mu^{(1)}_1, \dots, \mu^{(1)}_{s_1-1}, \nu_{12}) \quad \text{and} \quad \vec{\mu_{12}}^{(2)} = (\mu^{(2)}_1, \dots, \mu^{(2)}_{s_2-1}, \nu^T_{12}).$$

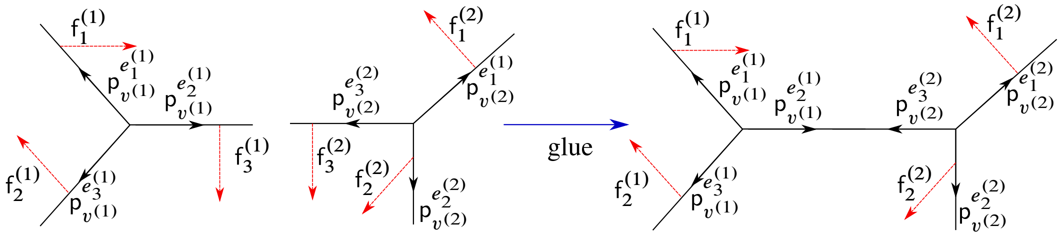


Figure 12: The glueing procedure for the topological vertex. In the notation of the text, we have $s_1 = s_2 = 3$, $\widetilde{e}_1 = e^{(1)}_2$, $\widetilde{e}_2 = e^{(2)}_3$, $\widetilde{e}'_1 = e^{(1)}_3$ and $\widetilde{e}'_2 = e^{(2)}_1$.

Then the following glueing formula holds:

$$\text{(6-8)}\quad \mathcal{W}_{\vec{\mu}}(X,L,\mathsf{f})(Q,\hbar) = \sum_{\nu_{12}\in\mathcal{P}} (-Q_{\beta_{12}})^{|\nu_{12}|} q^{f_{12}\kappa(\nu_{12})/2}(-1)^{f_{12}|\nu_{12}|}\mathcal{W}_{\vec{\mu}^{(1)}_{12}}(X^{(1)},L^{(1)},\mathsf{f}^{(1)})(Q,\hbar) \cdot \mathcal{W}_{\vec{\mu}^{(2)}_{12}}(X^{(2)},L^{(2)},\mathsf{f}^{(2)})(Q,\hbar).$$

Here $f_{12} = \det(\mathsf{p}_{\widetilde{e}'_1}, \mathsf{p}_{\widetilde{e}'_2})$, where $\widetilde{e}'_i \in (\Gamma_{(X,L,\mathsf{f})})_1$ is the first edge met when moving counterclockwise from $\widetilde{e}_i$, and $Q_{\beta_{12}}$ is the exponentiated Kähler parameter associated to the homology class $\beta_{12} = [\mu_T^{-1}(e_{12})] \in \mathrm{H}_2(X,\mathbb{Z})$. The glueing formula (6-8) originally proposed by [6] is derived in [76] as a consequence of Li's degeneration formula for relative Gromov–Witten theory [75].

(3) The glueing formula (6-8) allows us to recursively compute open Gromov–Witten invariants of any toric Lagrangian triple (X,L,f) starting from those of the framed vertex, ie affine 3–space with framed toric Lagrangians incident to each coordinate line. The framing transformation (6-7) further reduces the problem to the knowledge of the open Gromov–Witten invariants of $(\mathbb{C}^3, L = L_1 \cup L_2 \cup L_3, \mathsf{f}^{\mathrm{can}})$ in canonical framing $\mathsf{f}_i = \mathsf{f}_i^{\mathrm{can}} := \mathsf{p}_{i+1} \bmod 3$. This is given by

$$\text{(6-9)}\qquad \mathcal{W}_{\mu_1,\mu_2,\mu_3}(\mathbb{C}^3, L, \mathsf{f}^{\mathrm{can}})(\hbar) = q^{\kappa(\mu_1)/2} \sum_{\delta\in\mathcal{P}} s_{\mu_1^t/\delta}(q^{\rho+\mu_3}) s_{\mu_2/\delta}(q^{\rho+\mu_3^t}) s_{\mu_3}(q^{\rho}),$$

where the shifted skew Schur function $s_{\alpha/\beta}(q^{\rho+\gamma})$ is defined in (D-12). The formula (6-9) follows from an explicit evaluation of formal relative Gromov–Witten invariants in terms of descendent triple Hodge integrals. It was first proved in [78; 76] when $\mu_3 = \varnothing$, and the general case was established in [89].

An immediate consequence of (6-8) and (6-9) is that if $i\colon (X,L,\mathsf{f}) \hookrightarrow (X',L',\mathsf{f}')$ is an embedding of toric Lagrangian triples corresponding to an embedding of graphs $i_\#\colon \Gamma_{(X,L,\mathsf{f})} \hookrightarrow \Gamma_{(X',L',\mathsf{f}')}$, where $\Gamma_{(X',L',\mathsf{f}')}$ is obtained from $\Gamma_{(X,L,\mathsf{f})}$ by addition of a single vertex v_2 and glueing along a compact edge e_{12} to a vertex $v_1 \in (\Gamma_{(X,L,\mathsf{f})})_0$ by the above procedure, then

$$\text{(6-10)}\qquad \mathcal{W}_{\vec{\mu}}(X,L,\mathsf{f})(Q,\hbar) = \mathcal{W}_{\vec{\mu}}(X',L',\mathsf{f}')(Q,\hbar)|_{Q_{\beta_{12}}=0}.$$

6.2 The higher-genus log-open principle

In this section we associate certain toric Lagrangian triples to the geometry of Looijenga pair, under an additional condition given by the following definition.

Definition 6.3 Let $Y(D = D_1 + \cdots + D_l)$ be a nef Looijenga. We say that it satisfies *Property O* if $E_{Y(D)}$ deforms to $E_{Y'(D')}$ for a Looijenga pair $Y'(D' = D'_1 + \cdots + D'_l)$ such that

- Y' is a toric surface,
- D'_i is a prime toric divisor for all $i = 1, \dots, l-1$, and
- any nontrivial effective curve in Y' is D'_l–convex.

Example 6.2 Denote by $Y'(D'_1, D'_2)$ the toric surface whose fan is given by Figure 14, with $D'_1 = H - E_3$ and the class of D'_2 corresponding to the sum of the other rays. $Y'(D'_1, D'_2)$ is obtained from $\mathbb{P}^2(1,4) = \mathbb{P}^2(D_1, D_2)$ by blowing up a smooth point on D_1 and two infinitesimally close points on D_2. Moving the latter two apart (while staying on D_2) determines a deformation to $\mathrm{dP}_3(0,2)$. Given nefness of D'_2, it follows that $\mathrm{dP}_3(0,2)$ satisfies Property O. By Proposition 2.6, $\mathrm{dP}_3(1,1)$ also satisfies Property O. The property holds after blowing down (-1)–curves, including for $\mathbb{F}_0(0,4)$. Applying Proposition 2.6 it thus also holds for $\mathbb{F}_0(2,2)$ and $\mathbb{F}_2(2,2)$.

Example 6.3 Consider now $\mathrm{dP}_4(D_1, D_2)$ with $D_1^2 = 0$. Deforming $\mathrm{dP}_4(D_1, D_2)$ to a smooth toric surface with D_1 a toric divisor leads to the fan of Figure 14 with an additional ray in the lower half-plane. Up to deformation, there are two ways of doing so: by adding a ray either between $H - E_1 - E_2$ and E_2, or between E_2 and $E_1 - E_2$. Either way, this creates a curve C with $C \cdot (-K - D_1) < 0$ and therefore $\mathrm{dP}_4(0,1)$ does not satisfy Property O. The same argument applies to $\mathrm{dP}_5(0,0)$.

Example 6.4 When $l > 2$, Property O is always satisfied for all surfaces except for $\mathrm{dP}_3(0,0,0)$. The only way of deforming dP_3 to a toric surface with D_1 and D_2 toric is to take the fan of Figure 15 and add a ray in the lower-left quadrant. But this creates a curve C with $C \cdot (-K - D_1 - D_2) < 0$, and hence $\mathrm{dP}_3(0,0,0)$ does not satisfy Property O.

From Table 1, Property O coincides with quasi-tameness of $Y(D)$, with the sole exception of $\mathrm{dP}_3(0,0,0)$.

We make some informal comments about the geometric transition from stable log maps to open maps, which inform the construction of the open geometries below. This discussion is motivated by [8, Section 7] and in particular a natural generalisation of [8, Conjecture 7.3]. That description applied to our setting makes clear the structure of the toric Lagrangians. Denote by $(Y, D = D_1 + \cdots + D_l)$ a possibly noncompact log Calabi–Yau variety. For a maximally tangent stable log map to (Y, D), the expectation is that maximal tangency d_j with D_j can be replaced by an open boundary condition of winding number d_j with a special Lagrangian L_j near D_j. The special Lagrangian needs to have the property that it bounds a holomorphic disk $\mathcal{D}$ in the normal bundle to D_j; see [8, Section 7]. This property dictates how to compactify $Y \setminus D_j$: in a toric limit, $\mathcal{D}$ is simply the disk used to compactify the edge the framing lies on. If d is a D_j–convex curve class, then we can alternatively remove the maximal tangency condition by twisting the geometry by $\mathcal{O}_Y(-D_j)$. D_j–convexity then guarantees that no maps move into the fibre direction. To obtain the Calabi–Yau threefold geometry from a surface, we adopt the convention of twisting by the last divisor D_l.

In the toric limits of Construction 6.4, the choice of framing corresponds to a choice of compactification. If an outer edge e has framing f, then (see [76, Section 3.2]) the normal bundle of the compactification C of e is $\mathcal{O}(f) \oplus \mathcal{O}(-1-f)$. In our setting, one line bundle is the normal bundle $\mathcal{O}_C(C^2)$ of the curve C in the surface and the other is the normal bundle $\mathcal{O}_Y(-D_l)|_C$ of the curve in the fibre direction. In Construction 6.4, it follows from our conventions that if the framing points to the interior of the polytope, then the normal bundle of C in the surface is $\mathcal{O}(f)$, and if the framing points to the outside of the polytope, then the normal bundle of C in the surface is $\mathcal{O}(-1-f)$. In particular, from a Looijenga pair $Y(D_1,\dots,D_l)$ satisfying Property O, we construct a dual Aganagic–Vafa open Gromov–Witten geometry via the following construction.

Construction 6.4 Let $Y(D_1,\dots,D_l)$ be a Looijenga pair satisfying Property O for $Y'(D'_1,\dots,D'_l)$. Denote by $\Delta_{Y'}$ the polytope of Y' polarised by $-K_{Y'}$. We assume that $\Delta_{Y'\setminus\bigcup_{j\neq l} D'_j}$ is 2–dimensional or, equivalently, that D'_l is not toric, implying $l<4$. Denote by e_j the edge of $\Delta_{Y'}$ corresponding to D'_j for $1\le j\le l-1$ and denote by $e_l,\dots,e_{l+r}$ the remaining edges. Up to reordering, we may assume that the e_i are oriented clockwise. We construct a toric Lagrangian triple $Y^{\mathrm{op}}(D):=(X,L,\mathsf{f})$ as follows. In $\Delta_{Y'}$ remove the edge e_1 and replace it by a framing f_1 on e_{l+r} parallel to e_1. By Definition 6.1 and Remark 6.2, there is a unique way to do so, and f_1 points into the interior of $\Delta_{Y'}$. Denote the resulting graph by Δ^1. If $l=2$, add outer edges to Δ^1 so that each vertex satisfies the balancing condition and denote the resulting toric Calabi–Yau graph by Γ. If $l=3$, in Δ^1 remove the edge e_2 and replace it by a framing on e_3 parallel to e_2. Denote the resulting graph by Δ^2. Add edges to Δ^2 so that each vertex satisfies the balancing condition, and denote the resulting toric Calabi–Yau graph by Γ.

The graph Γ in Construction 6.4 gives the discriminant locus of the SYZ fibration of the toric Calabi–Yau threefold $X=\mathrm{Tot}(K_{Y'\setminus\bigcup_{j\neq l} D'_j})$. The base of the fibration is an $\mathbb{R}$–bundle over the polyhedron $\Delta_{Y'\setminus\bigcup_{j\neq l} D'_j}$. The framings determine toric special Lagrangians L_j, and the added outer edges correspond to the toric fibres of $\mathcal{O}(-D'_l)$. As is readily seen from the fan, f_1 (resp. $-1-f_2$) is the degree of the normal bundle of the divisor in Y' corresponding to e_{l+r} (resp. to e_3). The framing keeps track of the compactification of $Y'\setminus\bigcup_{j\neq l} D'_j$.

Remark 6.3 Tangency with more than one point can be incorporated by having parallel framings on different outer edges.

Remark 6.4 If $\Delta_{Y'\setminus D'_1\cup D'_2}$ is not 2–dimensional, then we blow up Y in a smooth point of D such that the resulting $\widetilde{Y}(\widetilde{D})$ satisfies Property O. We construct $\widetilde{Y}^{\mathrm{op}}(\widetilde{D})$, and recover the open invariants of $Y(D)$ by considering the curve classes that do not meet the exceptional divisor. In particular, for $l>3$ we stipulate that Construction 6.4 can be extended through suitable flopping of $(-1,-1)$–curves in the toric Calabi–Yau 3–fold geometry. We leave a precise formulation to future work, and develop the sole example relevant to our paper to illustrate this.

Example 6.5 We adapt the construction to the only nef Looijenga pair with 4 boundary components $\mathbb{F}_0(H_1,H_2,H'_1,H'_2)$. Since $\Delta_{\mathbb{F}_0\setminus(H_1\cup H_2\cup H'_1)}$ is 1–dimensional, we start by blowing up a smooth point of H'_1. In a toric deformation, we obtain $\mathrm{dP}_2(0,0,-1,0)$ with divisors $D'_1=H_1$, $D'_2=H_2$, $D'_3=H'_1-E$ and $D'_4=H'_2$. We assume that the corresponding $e_1,\dots,e_5$ are ordered clockwise in Δ_{dP_2}. Start with the graph Δ^2 from Construction 6.4. Balance the vertices and flop the inner edge. On the inner edge, add

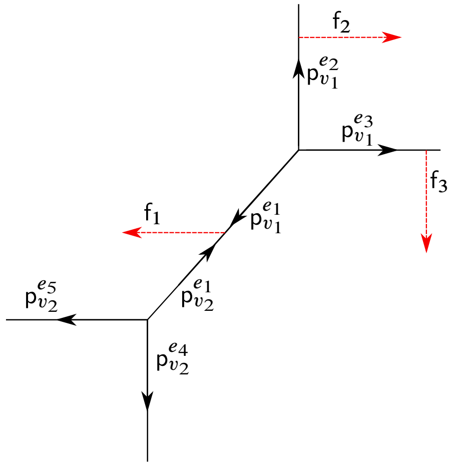


Figure 13: The toric CY3 graph of $\mathrm{dP}_2^{\mathrm{op}}(0,0,-1,0)$.

a framing parallel to e_3. The result is the graph of Figure 13, with the notational shift $\mathsf{f}_2 \leftrightarrow D'_1$, $\mathsf{f}_3 \leftrightarrow D'_2$, $\mathsf{f}_1 \leftrightarrow D'_3$. To obtain the graph for $\mathbb{F}_0(0,0,0,0)$, we remove the two outer edges that have no framing. The result is Figure 16.

Lemma 6.5 *Let $Y(D_1,\dots,D_l)$ and $Y'(D'_1,\dots,D'_l)$ be as in Construction 6.4. Then $\mathrm{H}_2(Y,\mathbb{Z}) = \mathrm{H}_2(Y',\mathbb{Z})$ is generated by the divisors corresponding to $e_3,\dots,e_{l+r}$.*

Proof In the fan of Y', define an ordering of the 2–dimensional cones by letting σ_i be the cone corresponding to $e_i \cap e_{i+1}$ when $1 \le i < l+r$ and σ_{l+r} be the cone corresponding to $e_{l+r} \cap e_1$. Define cones $\tau_i := \sigma_i \cap \bigcap_{j \in J_i} \sigma_j$, where J_i is the set of $j > i$ such that $\sigma_i \cap \sigma_j$ is 1–dimensional. Then $\tau_1 = \{0\}$, τ_i is the ray corresponding to e_{i+1} for $2 \le i < l+r$ and $\tau_{l+r} = \sigma_{l+r}$. By [42, Section 5.2, Theorem], these cones generate $\mathrm{H}_*(Y',\mathbb{Z})$; hence the divisors corresponding to $e_3,\dots,e_{l+r}$ generate $\mathrm{H}_2(Y',\mathbb{Z})$. □

Note that $\mathrm{H}_2^{\mathrm{rel}}(Y^{\mathrm{op}}(D),\mathbb{Z})$ is generated by the curve classes $[e]$ corresponding to inner edges e and by the relative disk classes $[\mathcal{D}_e]$ corresponding to outer edges e with framings. By the corresponding short exact sequence, the latter can be identified with $[S^1] \in \mathrm{H}_1(S^1,\mathbb{Z})$, where $L \supset S^1 = \partial\mathcal{D}_e$ and the degrees in the $[S^1]$ keep track of the winding numbers. By construction, the e thus described are edges of $\Delta_{Y'}$.

Definition 6.5 Let $Y(D_1,\dots,D_l)$ and $Y'(D'_1,\dots,D'_l)$ be as in Construction 6.4. Define

$$\iota\colon \mathrm{H}_2^{\mathrm{rel}}(Y^{\mathrm{op}}(D),\mathbb{Z}) \to \mathrm{H}_2(Y,\mathbb{Z}) \tag{6-11}$$

by sending $[e]$ to the divisor corresponding to e in Y.

Proposition 6.6 *The morphism ι is an isomorphism.*

Proof This is a direct consequence of Lemma 6.5. □

Example 6.6 We continue with Example 6.5. Following Figure 13, denote by e_i the edge with framing f_i for $i = 1, 2, 3$. Generalising Definition 6.5, we define $\iota\colon \mathrm{H}_2^{\mathrm{rel}}(\mathrm{dP}_2^{\mathrm{op}}(0,0,-1,0),\mathbb{Z}) \xrightarrow{\sim} \mathrm{H}_2(\mathrm{dP}_2,\mathbb{Z})$ by

$$\iota[e_1] = [D_2'] = [H_2], \quad \iota[e_2] = [D_4' - E] = [H_2 - E], \quad \iota[e_3] = [D_3'] = [H_1 - E], \tag{6-12}$$

which yields an isomorphism.

Theorem 6.7 (the higher-genus log-open principle) *Suppose $Y(D)$ satisfies Property O. Then*

$$O_{0;\iota^{-1}(d)}(Y^{\mathrm{op}}(D)) = N_{0,d}^{\mathrm{loc}}(Y(D)) = \prod_{i=1}^{l} \frac{(-1)^{d\cdot D_i+1}}{d\cdot D_i} N_{0,d}^{\log}(Y(D)). \tag{6-13}$$

Moreover, if $Y(D)$ is tame,

$$\mathsf{O}_{\iota^{-1}(d)}(Y^{\mathrm{op}}(D))(-\mathrm{i}\log q) = [1]_q^{l-2} \frac{(-1)^{d\cdot D_l+1}}{[d\cdot D_l]_q} \prod_{i=1}^{l-1} \frac{(-1)^{d\cdot D_i+1}}{d\cdot D_i} \mathsf{N}_d^{\log}(Y(D))(-\mathrm{i}\log q). \tag{6-14}$$

Remark 6.8 As is evident from Example 6.2, $Y^{\mathrm{op}}(D)$ depends on the toric model and hence is not unique. However, it can be checked directly for the examples of Table 1 that if $(X^{(1)}, L^{(1)}, \mathsf{f}^{(1)})$ and $(X^{(2)}, L^{(2)}, \mathsf{f}^{(2)})$ correspond to two such choices, then there exists a $\varpi\colon \mathrm{H}_2^{\mathrm{rel}}(X^{(1)}, L^{(1)}, \mathbb{Z}) \xrightarrow{\sim} \mathrm{H}_2^{\mathrm{rel}}(X^{(2)}, L^{(2)}, \mathbb{Z})$ such that $\iota^{(1)} = \iota^{(2)} \circ \varpi$.

6.3 Proof of Theorem 6.7

In order to work our way to a general $Y(D)$ satisfying Property O, we first show that if $\pi : Y'(D') \to Y(D)$ is an interior blowup, Construction 6.4 implies that the higher-genus open GW invariants $\mathsf{O}_{\iota^{-1}(d)}(Y^{\mathrm{op}}(D))$ satisfy the same blowup formula (4-19) of the log invariants on the right-hand side of (6-14).

Proposition 6.9 (blowup formula for open GW invariants) *Let $\pi : \widetilde{Y}(\widetilde{D}) \to Y(D)$ be an interior blowup of Looijenga pairs with both $\widetilde{Y}(\widetilde{D})$ and $Y(D)$ satisfying Property O, and denote by π_{op}^* the monomorphism of abelian groups defined by*

$$\begin{array}{ccc} \mathrm{H}_2(Y(D),\mathbb{Z}) & \xhookrightarrow{\ \pi^*\ } & \mathrm{H}_2(\widetilde{Y}(\widetilde{D}),\mathbb{Z}) \\ \big\downarrow{\scriptstyle \iota^{-1}} & & \big\downarrow{\scriptstyle \iota^{-1}} \\ \mathrm{H}_2^{\mathrm{rel}}(Y^{\mathrm{op}}(D),\mathbb{Z}) & \xhookrightarrow{\ \pi_{\mathrm{op}}^*\ } & \mathrm{H}_2^{\mathrm{rel}}(\widetilde{Y}^{\mathrm{op}}(\widetilde{D}),\mathbb{Z}) \end{array} \tag{6-15}$$

Then $\mathsf{O}_j(Y^{\mathrm{op}}(D)) = \mathsf{O}_{\pi_{\mathrm{op}}^* j}(\widetilde{Y}^{\mathrm{op}}(\widetilde{D}))$ for all $j \in \mathrm{H}_2^{\mathrm{rel}}(Y^{\mathrm{op}}(D),\mathbb{Z})$.

Sketch of the proof We provide an overview here and leave the details to the reader. The claim is proved by noting that Construction 6.4 implies the following: if $Y(D)$ is obtained from $\widetilde{Y}(\widetilde{D})$ by contraction of a (-1)–curve, then $Y^{\mathrm{op}}(D)$ is an open embedding into a flop of $\widetilde{Y}^{\mathrm{op}}(\widetilde{D})$ along a $(-1,-1)$–curve. The resulting nontrivial equality of open Gromov–Witten invariants under restriction on the image of π_{op}^* is then a combination of the invariance of open Gromov–Witten invariants under "forgetting an edge" in (6-10) and the flop invariance of the topological vertex [69]. □

By the previous proposition it then suffices to prove Theorem 6.7 for the pairs $Y(D)$ of highest Picard number for each value of $l = 2, 3, 4$, as all other pairs are recovered from these by blowing-down. We show this from a direct use of the topological vertex to determine the left-hand side of (6-14). The reader is referred to Appendix D for notation and basic results for shifted power sums $p_\alpha(q^{\rho+\gamma})$ and shifted skew Schur functions $s_{\alpha/\beta}(q^{\rho+\gamma})$ in the principal stable specialisation. The notation $\{\alpha, \beta\}_Q$ indicates the symmetric pairing on $\mathcal{P}$ of (D-15).

6.3.1 $l = 2$: holomorphic disks The classification of Propositions 2.2 and 2.3, the deformation equivalences in Proposition 2.6 and the definition of Property O in Definition 6.3 together imply that if $Y(D = D_1 + \cdots + D_l)$ and $Y(D' = D'_1 + \cdots + D'_l)$ are l–component Looijenga pairs both satisfying Property O, then there is a toric model for both with resulting $Y^{\mathrm{op}}(D) = Y^{\mathrm{op}}(D')$: in other words a model $Y^{\mathrm{op}}(D)$ for the open geometry only depends on Y and the number of irreducible components of D. Since 2–component log CY surfaces with maximal boundary come in pairs $Y(D)$ and $Y(D')$ from Table 1, throughout this section we will simplify notation and write $\Upsilon(Y) := Y^{\mathrm{op}}(D) = Y^{\mathrm{op}}(D')$ for the toric Lagrangian triple they share.

By Proposition 6.9, it suffices to consider the case of highest Picard rank $Y = \mathrm{dP}_3$. If $Y(D)$ is either $\mathrm{dP}_3(1,1)$ or $\mathrm{dP}_3(0,2)$, a toric model for Y is given by the toric surface Y' described by the fan of Figure 14, and in particular $D' = H - E_3$ is a toric divisor. Therefore $Y(D)$ satisfies Property O and, by Remark 6.1, $\Upsilon(\mathrm{dP}_3)$ is described by the toric CY3 graph of Figure 14. With conventions as in Figure 14, let $C_1 = \mu_T^{-1}(e_2)$, $C_2 = \mu_T^{-1}(e_5)$ and $C_3 = \mu_T^{-1}(e_7)$, and for a relative 2–homology class $j \in \mathrm{H}_2(\Upsilon(\mathrm{dP}_3), \mathbb{Z})$, write $j = j_0[S^1] + \sum_{i=1}^3 j_i[C_i]$.

We will compute generating functions of higher-genus 1–holed open Gromov–Witten invariants of $\Upsilon(\mathrm{dP}_3)$ in class j, using the theory of the topological vertex. For simplicity, we'll employ the shorthand notation $\mathrm{O}_{j_1,j_2,j_3;j_0}(\Upsilon(\mathrm{dP}_3))$ (resp. $\mathcal{O}_{j_0}(\Upsilon(\mathrm{dP}_3))$) to denote the generating function $\mathrm{O}_{\beta;\mu}(\Upsilon(\mathrm{dP}_3))$ (resp. $\mathcal{O}_\beta(\Upsilon(\mathrm{dP}_3))$) in (6-6) with $\beta = \sum_{i=1}^3 j_i[C_i]$ and $\mu = (j_0)$ a 1–row partition of length j_0. From

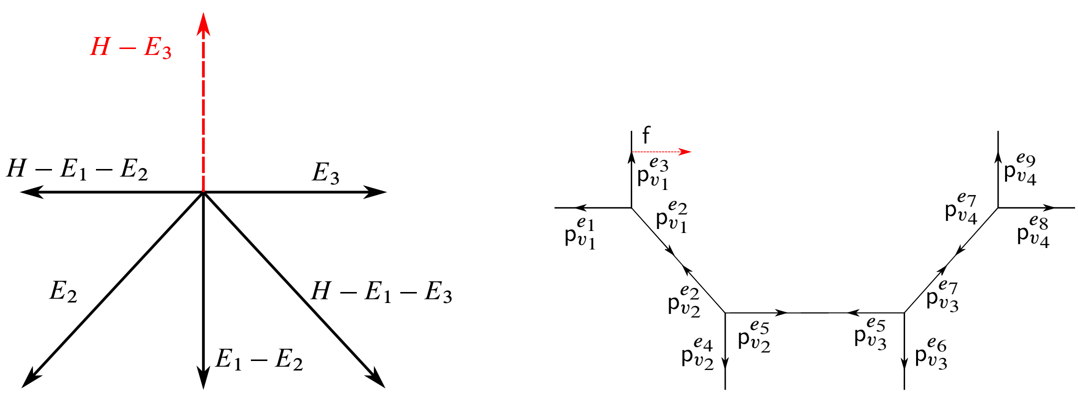


Figure 14: $\Upsilon(\mathrm{dP}_3) = \mathrm{dP}_3^{\mathrm{op}}(0,2) = \mathrm{dP}_3^{\mathrm{op}}(1,1)$ from the blowup of the plane at three nongeneric toric points.

(6-5) and (6-6), we have

$$(6\text{-}16)\quad \mathcal{O}_{j_0}(\Upsilon(\mathrm{dP}_3)) = \frac{\mathcal{Z}_{(j_0)}(\Upsilon(\mathrm{dP}_3))}{\mathcal{Z}_{\varnothing}(\Upsilon(\mathrm{dP}_3))} = \sum_{\nu\in\mathcal{P}} \frac{\chi_\nu((j_0))}{z_{(j_0)}} \frac{\mathcal{W}_\nu(\Upsilon(\mathrm{dP}_3))}{\mathcal{W}_\varnothing(\Upsilon(\mathrm{dP}_3))} = \sum_{s=0}^{j_0-1} \frac{(-1)^s}{j_0} \frac{\mathcal{W}_{(j_0-s,1^s)}(\Upsilon(\mathrm{dP}_3))}{\mathcal{W}_\varnothing(\Upsilon(\mathrm{dP}_3))},$$

where we have used the Murnaghan–Nakayama rule [112, Corollary 7.17.5]

$$(6\text{-}17)\qquad \chi_\nu((j_0)) = \begin{cases} (-1)^s & \text{if } \nu = (j_0-s, 1^s), \\ 0 & \text{else.} \end{cases}$$

The framing f in Figure 14 is shifted by one unit $f=-1$ from the canonical choice $\mathsf{f}_{\mathrm{can}} = \mathsf{p}_{v_1}^{e_2}$. From (6-7), (6-8) and (6-9), we then have, for any $\alpha\in\mathcal{P}$, that

$$\begin{aligned}(6\text{-}18)\quad & \mathcal{W}_\alpha(\Upsilon(\mathrm{dP}_3))(Q,\hbar) \\ &= (-1)^{|\alpha|} q^{-\frac{1}{2}\kappa(\alpha)} \\ &\quad\cdot \sum_{\lambda,\mu,\nu,\delta,\epsilon\in\mathcal{P}} s_{\lambda^t}(-Q_1 q^{\rho+\alpha}) s_\alpha(q^\rho) s_{\lambda/\delta}(q^\rho) s_{\mu/\delta}(q^\rho) Q_2^{|\mu|} s_{\mu/\epsilon}(q^\rho) s_{\nu/\epsilon}(q^\rho) s_{\nu^t}(-Q_3 q^\rho) \\ &= \frac{(-1)^{|\alpha|} s_{\alpha^t}(q^\rho)\{\alpha,\varnothing\}_{Q_1}\{\alpha,\varnothing\}_{Q_1Q_2}\{\varnothing,\varnothing\}_{Q_3}\{\varnothing,\varnothing\}_{Q_2Q_3}}{\{\varnothing,\varnothing\}_{Q_2}\{\alpha,\varnothing\}_{Q_1Q_2Q_3}},\end{aligned}$$

where we have used (D-13) and, repeatedly, (D-15) to express the sums over partitions in terms of Cauchy products. Then, specialising to $\alpha=(j_0-s,1^s)$ a hook partition with j_0 boxes and $s+1$ rows, and using (D-11) and (D-18), we have

$$\begin{aligned}(6\text{-}19)\quad \frac{\mathcal{W}(\Upsilon(\mathrm{dP}_3))_{(j_0-s,1^s)}}{\mathcal{W}(\Upsilon(\mathrm{dP}_3))_\varnothing} &= \frac{(-1)^{j_0} s_{(s+1,1^{j_0-s-1})}(q^\rho)\{(j_0-s,1^s),\varnothing\}_{Q_1}\{(j_0-s,1^s),\varnothing\}_{Q_1Q_2}}{\{(j_0-s,1^s),\varnothing\}_{Q_1Q_2Q_3}} \\ &= \frac{(-1)^{j_0} q^{-\frac{1}{2}\binom{j_0}{2}+\frac{1}{2}j_0 s}}{[j_0]_q[j_0-s-1]_q![s]_q!} \frac{\prod_{k=0}^{j_0-1}(1-q^k Q_1 q^{-s})\prod_{l=0}^{j_0-1}(1-q^l Q_1Q_2 q^{-s})}{\prod_{m=0}^{j_0-1}(1-q^m Q_1Q_2Q_3 q^{-s})}.\end{aligned}$$

Replacing this into (6-16) we get

$$\begin{aligned}(6\text{-}20)\quad & \mathcal{O}_{j_0}(\Upsilon(\mathrm{dP}_3))(Q,\hbar) \\ &= \sum_{s=0}^{j_0-1} \frac{(-1)^s}{j_0} \frac{\mathcal{W}_{(j_0-s,1^s)}(\Upsilon(\mathrm{dP}_3))(Q,\hbar)}{\mathcal{W}_\varnothing(\Upsilon(\mathrm{dP}_3))(Q,\hbar)} \\ &= \frac{(-1)^{j_0} q^{-\frac{1}{2}\binom{j_0}{2}}}{j_0[j_0]_q!} \sum_{j_1,j_2,j_3=0}^{\infty} \Bigg(q^{\frac{1}{2}j_1(j_0-1)} \begin{bmatrix} j_0 \\ j_1-j_2 \end{bmatrix}_q \begin{bmatrix} j_0 \\ j_2-j_3 \end{bmatrix}_q \begin{bmatrix} j_0+j_3-1 \\ j_3 \end{bmatrix}_q \\ &\qquad\qquad \cdot (-1)^{j_1+j_3} Q_1^{j_1} Q_2^{j_2} Q_3^{j_3} \sum_{s=0}^{j_0-1} \begin{bmatrix} j_0-1 \\ s \end{bmatrix}_q (-q^{-j_1})^s q^{\frac{1}{2}j_0 s} \Bigg) \\ &= \frac{(-1)^{j_0}}{j_0[j_0]_q!} \sum_{j_1,j_2,j_3}^{\infty} \begin{bmatrix} j_0 \\ j_1-j_2 \end{bmatrix}_q \begin{bmatrix} j_0 \\ j_2-j_3 \end{bmatrix}_q \begin{bmatrix} j_0+j_3-1 \\ j_3 \end{bmatrix}_q (-1)^{j_1+j_3} Q_1^{j_1} Q_2^{j_2} Q_3^{j_3} \frac{[j_1-1]_q!}{[j_1-j_0]_q!},\end{aligned}$$

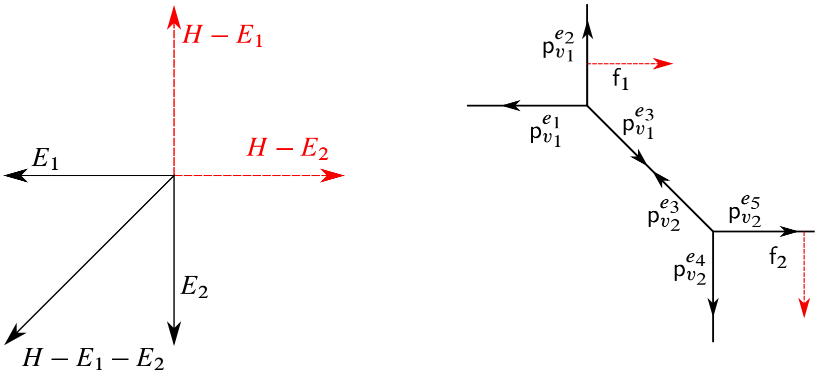


Figure 15: $\mathrm{dP}_2^{\mathrm{op}}(1,0,0)$ from dP_2 and $D_1 = H - E_1$, $D_2 = H - E_2$.

where the q–binomial theorem has been used to expand the products in (6-19) and to perform the summation over s in the last line. Isolating the $\mathcal{O}(Q_1^{j_1} Q_2^{j_2} Q_2^{j_3})$ coefficient yields

$$\mathsf{O}_{j_1,j_2,j_3;j_0}(\Upsilon(\mathrm{dP}_3))(\hbar) = \frac{(-1)^{j_1+j_0+j_3}[j_0]_q}{j_0[j_1]_q[j_0+j_3]_q} \begin{bmatrix} j_0 \\ j_1-j_2 \end{bmatrix}_q \begin{bmatrix} j_0 \\ j_2-j_3 \end{bmatrix}_q \begin{bmatrix} j_0+j_3 \\ j_3 \end{bmatrix}_q \begin{bmatrix} j_1 \\ j_0 \end{bmatrix}_q. \tag{6-21}$$

From Figure 14, the lattice isomorphism $\iota\colon \mathrm{H}_2^{\mathrm{rel}}(\Upsilon(\mathrm{dP}_3),\mathbb{Z}) \to \mathrm{H}_2(\mathrm{dP}_3,\mathbb{Z})$ in this case reads

$$\iota[S^1] = [H-E_1-E_2], \quad \iota[C_1] = [E_2], \quad \iota[C_2] = [E_1-E_2], \quad \iota[C_3] = [H-E_1-E_3], \tag{6-22}$$

and the change of variables relating the curve degrees (d_0,d_1,d_2,d_3) in $\mathrm{H}_2(\mathrm{dP}_3,\mathbb{Z})$ and the relative homology variables $(j_0; j_1, j_2, j_3)$ in $\mathrm{H}_2^{\mathrm{rel}}(\Upsilon(\mathrm{dP}_3),\mathbb{Z})$ is therefore

$$d_0 \to j_0 + j_3, \quad d_1 \to j_2, \quad d_2 \to j_1 - j_2 + j_3, \quad d_3 \to j_0. \tag{6-23}$$

Combining the change of variables (6-23) and the log result of (4-22) in Theorem 4.5 returns (6-21), establishing (6-14) for $Y(D) = \mathrm{dP}_3(1,1)$. Furthermore, taking the genus-zero limit $q \to 1$ and using Theorem 5.1, Lemma 3.1 and Proposition 2.6 implies (6-13), completing the proof of Theorem 6.7 for $Y(D) = \mathrm{dP}_3(D_1^2, D_2^2)$. Use of Propositions 6.9 and 4.3 then concludes the proof of Theorem 6.7 for any $Y(D)$ satisfying Property O with $l = 2$.

6.3.2 $l = 3$: holomorphic annuli The 3–component Looijenga pair of highest Picard rank satisfying Property O is $Y(D) = \mathrm{dP}_2(1,0,0)$. Taking $D_1 = H - E_1$, $D_2 = H - E_2$ we have that Y, D_1 and D_2 are toric, and $\mathrm{dP}_2^{\mathrm{op}}(1,0,0)$ is described by the toric CY3 graph on the left in Figure 15. Write $C = \mu_T^{-1}(e_3)$, and for a relative 2–homology class $j \in \mathrm{H}_2^{\mathrm{rel}}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0),\mathbb{Z})$ write $j = j_1[\mathcal{D}_1] + j_2[\mathcal{D}_2] + j_C[C]$, where $[\mathcal{D}_i]$ are integral generators of the first homology of the outer Lagrangians incident to edges adjacent to the vertices v_i for $i = 1, 2$ in Figure 15. As in the previous section, we will write $\mathsf{O}_{j_C;j_1,j_2}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))$ (resp. $\mathcal{O}_{j_1,j_2}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))$) for the generating function $\mathsf{O}_{\beta,\vec{\mu}}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))$ (resp. $\mathcal{O}_{\vec{\mu}}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))$), with $\beta = j_C[C]$ and $\vec{\mu} = ((j_1), (j_2))$ a pair of 1–row partitions of length (j_1, j_2). From (6-5), (6-6) and (6-17), we have

$$\begin{aligned}(6\text{-}24)\quad \mathcal{O}_{j_1,j_2}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))(Q,\hbar) &= \frac{\mathcal{Z}_{j_1,j_2}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))}{Z_{\varnothing,\varnothing}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))} - \frac{\mathcal{Z}_{(j_1),\varnothing}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))\mathcal{Z}_{\varnothing,(j_2)}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))}{\mathcal{Z}_{\varnothing,\varnothing}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))^2} \\ &= \sum_{s_1=0}^{j_1-1}\sum_{s_2=0}^{j_2-1} \frac{(-1)^{s_1+s_2}}{j_1 j_2}\Bigg[\frac{\mathcal{W}_{(j_1-s_1,1^{s_1}),(j_2-s_2,1^{s_2})}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))}{\mathcal{W}_{\varnothing,\varnothing}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))} \\ &\qquad - \frac{\mathcal{W}_{(j_1-s_1,1^{s_1}),\varnothing}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))\mathcal{W}_{\varnothing,(j_2-s_2,1^{s_2})}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))}{\mathcal{W}_{\varnothing,\varnothing}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))^2}\Bigg].\end{aligned}$$

The framings f_1 and f_2 in Figure 15 are, respectively, shifted by one unit $f = -1$ from the canonical choice $\mathsf{f}_{\mathrm{can}} = \mathsf{p}_{v_1}^{e_3}$, and equal to the canonical framing $\mathsf{f}_2 = \mathsf{p}_{v_2}^{e_4}$. Then (6-7), (6-8) and (6-9) give

$$\begin{aligned}(6\text{-}25)\qquad \mathcal{W}_{\alpha\beta}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))(Q,\hbar) &= (-1)^{|\alpha|}q^{-\kappa(\alpha)/2}\sum_{\mu,\delta,\in\mathcal{P}} s_{\mu^t}(-q^{\rho+\alpha}Q)s_\alpha(q^\rho)s_{\mu/\delta}(q^\rho)s_{\beta/\delta}(q^\rho) \\ &= (-1)^{|\alpha|}s_{\alpha^t}(q^\rho)\{\alpha,\varnothing\}_Q\sum_{\delta\in\mathcal{P}} s_{\beta^t/\delta^t}(-q^{-\rho})s_{\delta^t}(-q^{\rho+\alpha}Q) \\ &= (-1)^{|\alpha|}s_{\alpha^t}(q^\rho)\{\alpha,\varnothing\}_Q s_{\beta^t}(-q^{-\rho},-q^{\rho+\alpha}Q),\end{aligned}$$

where we have used (D-14), (D-7) and (D-15) to perform the summations over partitions. Then, restricting to $\alpha = (j_1 - s_1, 1^{s_1})$ and $\beta = (j_2 - s_2, 1^{s_2})$,

$$\begin{aligned}(6\text{-}26)\quad & \mathcal{O}_{j_1,j_2}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))(Q,\hbar) \\ &= \sum_{s_1=0}^{j_1-1}\frac{(-1)^{j_1+s_1+j_2+1}}{j_1 j_2} s_{(s_1+1,1^{j_1-s_1-1})}(q^\rho)\prod_{k=0}^{j_1-1}(1-q^kQq^{-s_1}) \\ &\qquad\qquad \times[p_{(j_2)}(-Qq^{\rho+(j_1-s_1,1^{s_1})},-q^{-\rho}) - p_{(j_2)}(-Qq^\rho,-q^{-\rho})] \\ &= \sum_{s_1=0}^{j_1-1}\frac{(-1)^{j_1+s_1+j_2+1}}{j_1 j_2} s_{(s_1+1,1^{j_1-s_1-1})}(q^\rho)\prod_{k=0}^{j_1-1}(1-q^kQq^{-s_1}) \\ &\qquad\qquad \times[p_{(j_2)}(-Qq^{\rho+(j_1-s_1,1^{s_1})}) - p_{(j_2)}(-Qq^\rho)] \\ &= \sum_{s_1=0}^{j_1-1}\frac{(-1)^{j_1+s_1+j_2+1}}{j_1 j_2} s_{(s_1+1,1^{j_1-s_1-1})}(q^\rho)\prod_{k=0}^{j_1-1}(1-q^kQq^{-s_1})(-Qq^{-s_1-\frac12})^{j_2}[q^{j_2 j_1}-1] \\ &= \frac{(-1)^{j_1+1}Q^{j_2}[j_1 j_2]_q}{j_1 j_2[j_2+m]_q}\sum_{m=0}^{j_1}\begin{bmatrix} j_1 \\ m\end{bmatrix}_q\begin{bmatrix} j_2+m \\ j_1\end{bmatrix}_q(-Q)^m,\end{aligned}$$

where in the first equality we have used (D-3) and (6-17), in the second the fact that for a 1–row partition $\alpha = (d)$, $p_{(d)}(x_1,\dots,x_n,\dots;y_1,\dots,y_n,\dots) = p_{(d)}(x_1,\dots,x_n,\dots) + p_{(d)}(y_1,\dots,y_n,\dots)$, and in the third equality the fact that the difference of infinite power sums in the term in square brackets telescopes to just two terms; the final calculations are repeated applications of the q–binomial theorem. Extracting the $\mathcal{O}(Q^{j_C})$ coefficient, we get

$$\mathsf{O}_{j_C;j_1,j_2}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0))(\hbar) = \frac{(-1)^{j_1+1+j_C+j_2}[j_1 j_2]_q}{j_1 j_2 [j_C]_q} \begin{bmatrix} j_1 \\ j_C - j_2 \end{bmatrix}_q \begin{bmatrix} j_C \\ j_1 \end{bmatrix}_q. \tag{6-27}$$

From Figure 15, the homomorphism of homology groups $\iota\colon \mathrm{H}_2^{\mathrm{rel}}(\mathrm{dP}_2^{\mathrm{op}}(1,0,0),\mathbb{Z}) \to \mathrm{H}_2(\mathrm{dP}_2,\mathbb{Z})$ is given by

$$\iota[D_1] = [E_1], \quad \iota[D_2] = [E_2], \quad \iota[C] = [H - E_1 - E_2], \tag{6-28}$$

and the resulting map of curve degrees is

$$d_0 \to j_C, \quad d_1 \to j_1, \quad d_2 \to j_2. \tag{6-29}$$

Together with the log results given by (4-39) in Theorem 4.9 and the blowup formulas of Propositions 4.3 and 6.9 for the log and open invariants, this proves Theorem 6.7 for $l = 3$.

6.3.3 $l = 4$: holomorphic pairs of pants

According to Example 6.5, for the only 4–component case $Y(D) = \mathbb{F}_0(0,0,0,0)$, we have that $Y^{\mathrm{op}}(D)$ is given by the 3–dimensional affine space with Aganagic–Vafa A–branes $L^{(i)}$ for $i = 1,2,3$ at framing shifted by -1, 0, and -1 ending on the three legs of the vertex, as in Figure 16. We will be concerned with counts of 3–holed open Gromov–Witten invariants of $\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)$, with winding numbers (j_1, j_1, j_2); see Example 6.6.

The connected generating function, by (6-5) and (6-6), is

$$\begin{aligned}
&\mathsf{O}_{j_1,j_1,j_2}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))(\hbar) \\
&\quad = \mathcal{Z}_{(j_1)(j_1)(j_2)}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)) - \mathcal{Z}_{(j_1)(j_1)\varnothing}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))\mathcal{Z}_{\varnothing\varnothing(j_2)}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)) \\
&\qquad - \mathcal{Z}_{(j_1)\varnothing(j_2)}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))\mathcal{Z}_{\varnothing(j_1)\varnothing}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)) \\
&\qquad - \mathcal{Z}_{\varnothing(j_1)(j_2)}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))\mathcal{Z}_{(j_1)\varnothing\varnothing}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)) \\
&\qquad + 2\mathcal{Z}_{(j_1)\varnothing\varnothing}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))\mathcal{Z}_{\varnothing(j_1)\varnothing}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))\mathcal{Z}_{\varnothing\varnothing(j_2)}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)),
\end{aligned} \tag{6-30}$$

where, by (6-17),

$$\begin{aligned}
&\mathcal{Z}_{(j_1),(j_1),(j_2)}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)) \\
&\qquad = \sum_{s_0,s_1,s_2} \frac{(-1)^{s_0+s_1+s_2}}{j_1^2 j_2} \mathcal{W}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))_{(j_1-s_0,1^{s_0}),(j_1-s_1,1^{s_1}),(j_2-s_2,1^{s_2})}
\end{aligned} \tag{6-31}$$

and, from (6-7) and (6-9),

$$\mathcal{W}_{\alpha\beta\gamma}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)) = (-1)^{|\alpha|+|\gamma|} \sum_{\delta} s_{\alpha^t/\delta}(q^{\rho+\gamma}) s_{\beta/\delta}(q^{\rho+\gamma^t}) s_{\gamma^t}(q^{\rho}). \tag{6-32}$$

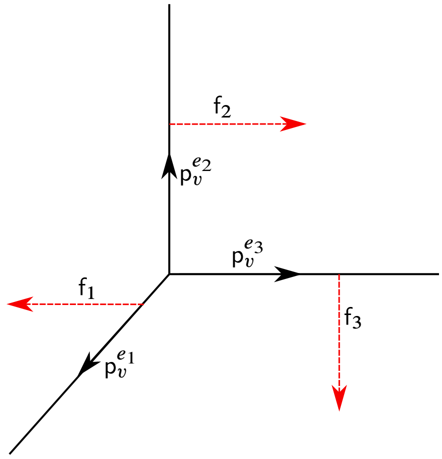


Figure 16: The toric CY3 graph of $\mathbb{F}_0^{\mathrm{op}}(0,0,0,0)$.

Elementary manipulations from (6-30)–(6-32) lead to

$$\text{(6-33)}\quad \mathsf{O}_{j_1,j_1,j_2}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))(\hbar) = \sum_{s_0,s_1,s_2} \frac{(-1)^{s_0+s_1+s_2+j_1+j_2}}{j_1^2 j_2}\Bigl[\bigl(s_{\alpha^t}(q^{\rho+\gamma})-s_{\alpha^t}(q^\rho)\bigr)\bigl(s_\beta(q^{\rho+\gamma^t})-s_\beta(q^\rho)\bigr)s_{\gamma^t}(q^\rho) + \sum_{\delta\neq\varnothing}\bigl(s_{\alpha^t/\delta}(q^{\rho+\gamma})s_{\beta/\delta}(q^{\rho+\gamma^t})-s_{\alpha^t/\delta}(q^\rho)s_{\beta/\delta}(q^\rho)\bigr)s_{\gamma^t}(q^\rho)\Bigr]$$

with

$$\alpha=(j_1-s_0,1^{s_0}),\quad \beta=(j_1-s_1,1^{s_1}),\quad \gamma=(j_2-s_2,1^{s_2}).$$

The part of the summation in the middle line, after carrying out the sums over s_0, s_1 and s_2 using (D-3) and (6-17), is equal to

$$\text{(6-34)}\quad \sum_{s_2=0}^{j_2-1}\frac{(-1)^{s_2+1+j_2}}{j_1^2 j_2}\bigl(p_{(j_1)}(q^{\rho+(j_2-s_2,1^{s_2})})-p_{(j_1)}(q^\rho)\bigr)\bigl(p_{(j_1)}(q^{\rho+(s_2+1,1^{j_2-s_2-1})})-p_{(j_1)}(q^\rho)\bigr)\times s_{(s_2+1,1^{j_2-s_2-1})}(q^\rho)$$
$$= \sum_{s_2=0}^{j_2-1}\frac{(-1)^{s_2}}{j_1^2 j_2}(q^{j_1(j_2-s_2-1/2)}-q^{j_1(-s_2-1/2)})(q^{j_1(s_2+1/2)}-q^{j_1(s_2+1/2-j_2)})s_{(j_2-s_2,1^{s_2})}(q^\rho)$$
$$= \frac{1}{j_1^2 j_2}\frac{[j_1 j_2]_q^2}{[j_2]_q},$$

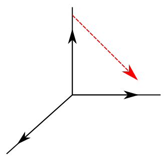
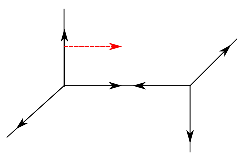
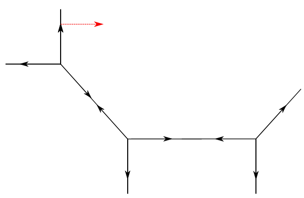
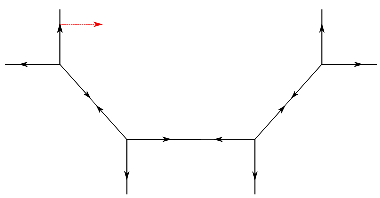
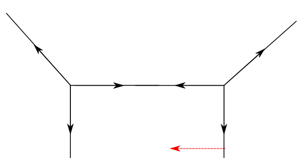
while the part in the last line is equal to zero. Indeed, when $\delta=\alpha^t$, we have $s_{\beta/\delta}(x)=\delta_{\beta\alpha^t}$ (since $|\alpha|=|\beta|=j_1$ in our case, $\alpha^t\preceq\beta$ implies $\alpha^t=\beta$), so the terms appearing in the difference in the second row of (6-33) are either individually zero or cancel out each other. When $\delta \neq \alpha^t$, we can use Lemma D.1 to expand $s_{\alpha^t/\delta}(x)$ in terms of ordinary Schur functions $s_\lambda(x)$ with $|\lambda| = |\alpha| - |\delta|$: it is easy to see that in the sum over s_0 the contribution labelled by each such Young diagram appears exactly twice and weighted with opposite signs. Therefore,

$$\mathsf{O}_{j_1,j_1,j_2}(\mathbb{F}_0^{\mathrm{op}}(0,0,0,0))(\hbar) = \frac{1}{j_1^2 j_2}\frac{[j_1 j_2]_q^2}{[j_2]_q}. \tag{6-35}$$

By construction from Examples 6.5 and 6.6,

$$d_1 \to j_2 \quad \text{and} \quad d_2 \to j_1, \tag{6-36}$$

and comparing with (4-41) gives (6-14), which concludes the proof of Theorem 6.7. □

$Y(D)$	$\Gamma_{Y^{\mathrm{op}}(D)}$
$\mathbb{P}^2(1,4)$	
$\mathrm{dP}_1(1,3)$ $\mathrm{dP}_1(0,4)$	
$\mathrm{dP}_2(1,2)$ $\mathrm{dP}_2(0,3)$	
$\mathrm{dP}_3(1,1)$ $\mathrm{dP}_3(0,2)$	
$\mathbb{F}_0(2,2)$ $\mathbb{F}_0(0,4)$	

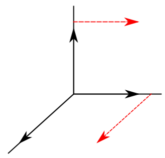
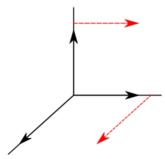
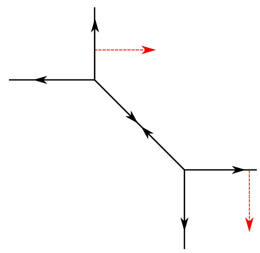
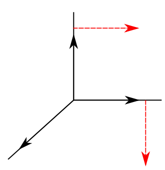
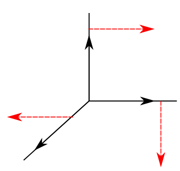
$Y(D)$	$\Gamma_{Y^{\mathrm{op}}(D)}$
$\mathbb{P}^2(1,1,1)$	
$\mathrm{dP}_1(1,1,0)$	
$\mathrm{dP}_2(1,0,0)$	
$\mathbb{F}_0(2,0,0)$	
$\mathbb{F}_0(0,0,0,0)$	

Table 3: $Y^{\mathrm{op}}(D)$ for l–component Looijenga pairs satisfying Property O.

7 KP and quiver DT invariants

7.1 Klemm–Pandharipande invariants of CY4–folds

Let Z be a smooth projective complex Calabi–Yau variety of dimension four and $d \in \mathrm{H}_2(Z,\mathbb{Z})$. Since $\mathrm{vdim}\overline{\mathcal{M}}_{g,n}(Z,d) = 1-g+n$, the only nonvanishing genus-zero primary Gromov–Witten invariants of Z without divisor insertions are[13]

$$\mathrm{GW}_{0,d;\gamma}(Z) := \int_{[\overline{\mathcal{M}}_{0,1}(Z,d)]^{\mathrm{vir}}} \mathrm{ev}_1^\star \gamma \quad \text{for } \gamma \in \mathrm{H}^4(Z,\mathbb{Z}). \tag{7-1}$$

The same considerations apply to the case of Z the Calabi–Yau total space of a rank $(4-r)$ concave vector bundle on an r–dimensional smooth projective variety. It was proposed by Greene, Morrison and Plesser in [50, Appendix B] and further elaborated upon by Klemm and Pandharipande in [68, Section 1.1] that a higher-dimensional version of the Aspinwall–Morrison should conjecturally produce integral invariants $\mathrm{KP}_{0,d}(Z)$, virtually enumerating rational degree-d curves incident to the Poincaré dual cycle of γ,

$$\mathrm{GW}_{0,d;\gamma}(Z) = \sum_{k|d} \frac{\mathrm{KP}_{0,d/k;\gamma}(Z)}{k^2}. \tag{7-2}$$

Conjecture 7.1 (Klemm–Pandharipande) $\mathrm{KP}_{0,d;\gamma}(Z) \in \mathbb{Z}$.

A symplectic proof of Conjecture 7.1 for projective Z, although likely adaptable to the noncompact setting, was given by Ionel and Parker in [62].

Our main focus will be on Z a noncompact CY4 local surface, ie $r=2$. In this case there is a single generator $\gamma = [\mathrm{pt}]$ for the fourth cohomology of Z, given by the Poincaré dual of the point class on the zero section, and we will henceforth use the simplified notation $\mathrm{KP}_{0,d}(Z) := \mathrm{KP}_{0,d;[\mathrm{pt}]}(Z)$.

7.2 Quiver Donaldson–Thomas theory

Let Q be a quiver with an ordered set Q_0 of n vertices $v_1,\dots v_n \in \mathsf{Q}_0$ and a set of oriented edges $\mathsf{Q}_1 = \{\alpha\colon v_i \to v_j\}$. We let $\mathbb{N}\mathsf{Q}_0$ be the free abelian semigroup generated by Q_0, and for $\mathsf{d} = \sum d_i v_i$ and $\mathsf{e} = \sum e_i v_i \in \mathbb{N}\mathsf{Q}_0$, we write $E_{\mathsf{Q}}(\mathsf{d},\mathsf{e})$ for the Euler form

$$E_{\mathsf{Q}}(\mathsf{d},\mathsf{e}) := \sum_{i=1}^n d_i e_i - \sum_{\alpha\colon v_i \to v_j} d_i e_j. \tag{7-3}$$

We assume in what follows that Q is symmetric; that is, for every i and j, the number of oriented edges from v_i to v_j is equal to the number of oriented edges from v_j to v_i. The Euler form is then a symmetric

[13] By the same formula, there are nonvanishing elliptic unpointed Gromov–Witten invariants for Z, which will not concern us in this paper. There are no Gromov–Witten invariants for a CY4 in genus $g>1$.

bilinear form. To C a symmetric bilinear pairing on $\mathbb{Z}^n$, we associate the generalised q–hypergeometric series

$$\Phi_C(q;x_1,\dots,x_n) := \sum_{\mathsf{d}\in\mathbb{N}^n}^{\infty} \frac{(-q^{1/2})^{C(\mathsf{d},\mathsf{d})}\mathsf{x}^{\mathsf{d}}}{\prod_{i=1}^n (q;q)_{d_i}}, \tag{7-4}$$

where $\mathsf{x}^{\mathsf{d}} = \prod_{i=1}^n x_i^{d_i}$. The motivic Donaldson–Thomas partition function associated to the cohomological Hall algebra of Q (without potential) is the generating function [34]

$$P_{\mathsf{Q}}(q;x_1,\dots,x_n) := \Phi_{E_{\mathsf{Q}}}(q;x_1,\dots,x_n), \tag{7-5}$$

and the motivic DT invariants $\mathrm{DT}_{\mathsf{d};i}(\mathsf{Q})$ of Q are the formal Taylor coefficients in the expansion of its plethystic logarithm [34; 70; 107],

$$\begin{aligned} P_{\mathsf{Q}}(q;x_1,\dots,x_n) &= \mathrm{Exp}\biggl(\frac{1}{[1]_q}\sum_{\mathsf{d}\neq 0}\sum_{i\in\mathbb{Z}} \mathrm{DT}_{\mathsf{d};i}(\mathsf{Q})\mathsf{x}^{\mathsf{d}}(-q^{1/2})^{-i}\biggr) \\ &= \exp\biggl[\sum_{n=1}^{\infty}\frac{1}{n[n]_q}\sum_{\mathsf{d}\neq 0}\sum_{i\in\mathbb{Z}} \mathrm{DT}^{\mathsf{Q}}_{\mathsf{d};i}\mathsf{x}^{n\mathsf{d}}(-q^{1/2})^{-ni}\biggr] \\ &= \prod_{\mathsf{d}\neq 0}\prod_{i\in\mathbb{Z}}\prod_{k\geq 0}\bigl(1-(-1)^i\mathsf{x}^{\mathsf{d}}q^{-k-(i+1)/2}\bigr)^{-\mathrm{DT}_{\mathsf{d};i}(\mathsf{Q})}. \end{aligned} \tag{7-6}$$

It will be of particular interest for us to consider a suitable semiclassical limit of (7-6)

$$\begin{aligned} y^{(i)}_{\mathsf{Q}}(x_1,\dots,x_n) &:= \lim_{q\to 1}\frac{P_{\mathsf{Q}}(q;x_1,\dots,q^{1/2}x_i,\dots,x_n)}{P_{\mathsf{Q}}(q;x_1,\dots,q^{-1/2}x_i,\dots x_n)} \\ &= \lim_{q\to 1}\mathrm{Exp}\biggl(\sum_{\mathsf{d}\neq 0}\frac{1}{[1]_q}\sum_{i\in\mathbb{Z}}[d_i]_q\mathrm{DT}^{\mathsf{Q}}_{\mathsf{d};i}\mathsf{x}^{\mathsf{d}}(-q^{1/2})^{-i}\biggr) \\ &= \prod_{\mathsf{d}\neq 0}\prod_{i\in\mathbb{Z}}(1-\mathsf{x}^{\mathsf{d}})^{-|\mathsf{d}|\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}}(\mathsf{Q})}, \end{aligned} \tag{7-7}$$

where

$$\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}}(\mathsf{Q}) := \sum_{i\in\mathbb{Z}}(-1)^i\mathrm{DT}_{\mathsf{d},i}(\mathsf{Q}) \tag{7-8}$$

are the numerical DT invariants. From (7-7), the numerical invariants can be extracted from the logarithmic primitive of $y^{(i)}_Q(\mathsf{x})$ with respect to x_i,

$$\int\frac{\mathrm{d}x_i}{x_i}\log y^{(i)}_Q(\mathsf{x}) =: \sum_{\mathsf{d}\neq 0}A_{\mathsf{d}}(\mathsf{Q})\mathsf{x}^{\mathsf{d}}, \tag{7-9}$$

as

$$A_{\mathsf{d}}(\mathsf{Q}) = \sum_{k|d}\frac{\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}/\mathsf{k}}(\mathsf{Q})}{k^2}. \tag{7-10}$$

The generating series

$$y_{\mathsf{Q}}(x_1,\dots,x_n) := \prod_{i=1}^{n} y_{\mathsf{Q}}^{(i)}(x_1,\dots,x_n)$$

has an interpretation as a generating function of Euler characteristics of certain noncommutative Hilbert schemes $\mathrm{Hilb}_{\mathsf{d}}(\mathsf{Q})$ attached to the moduli space of semistable representations of the quiver Q [37; 107],

$$y_{\mathsf{Q}}(x_1,\dots,x_n) = \sum_{\mathsf{d}\in\mathbb{Z}\mathsf{Q}_0} \chi(\mathrm{Hilb}_{\mathsf{d}}(\mathsf{Q}))\mathsf{x}^{\mathsf{d}} \in \mathbb{Z}[\![\mathsf{x}]\!]. \tag{7-11}$$

In particular, this implies that $\bigl(\sum_{i=1}^{n} d_i\bigr)\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}}(\mathsf{Q}) \in \mathbb{Z}$. More is true [34; 107] by the following theorem.

Theorem 7.2 (Efimov [34]) *The numerical Donaldson–Thomas invariants of a symmetric quiver* Q *without potential are positive integers*, $\mathrm{DT}^{\mathrm{num}}_{\mathsf{d}}(\mathsf{Q}) \in \mathbb{N}$.

7.3 KP integrality from DT theory

The genus-zero log-local and log-open correspondences of Theorem 6.7 imply that KP invariants of toric local surfaces are, up to a sign and possibly an integral shift, numerical DT invariants of a symmetric quiver. Combined with Theorem 7.2 this gives an algebrogeometric proof of Conjecture 7.1 for

$$Z = \mathrm{Tot}\bigl(\mathcal{O}(-D_1)\oplus\mathcal{O}(-D_2)\to Y\bigr).$$

Theorem 7.3 *Let* $Y(D)$ *be a* 2*–component quasi-tame Looijenga pair. Then there exists a symmetric quiver* $\mathsf{Q}(Y(D))$ *with* $\chi(Y)-1$ *vertices and a lattice isomorphism* $\kappa\colon \mathbb{Z}(\mathsf{Q}(Y(D)))_0 \xrightarrow{\sim} \mathrm{H}_2(Y,\mathbb{Z})$ *such that*

$$\mathrm{DT}^{\mathrm{num}}_{d}(\mathsf{Q}(Y(D))) = \Bigl|\mathrm{KP}_{\kappa(d)}(E_{Y(D)}) + \sum_i \alpha_i\delta_{d,v_i}\Bigr|, \tag{7-12}$$

with $\alpha_i \in \{-1,0,1\}$. *In particular*, $\mathrm{KP}_d(E_{Y(D)}) \in \mathbb{Z}$.

Proof The statement is a direct consequence of Theorem 6.7 combined with the strips–quivers correspondence of [103], which we briefly review here in our context. Since $Y(D)$ is a 2–component quasi-tame pair, it satisfies Property O by the discussion of Section 6.3. From Lemma D.1 and the proof of Theorem 6.7 (see in particular (6-19)), we have

$$\frac{\mathcal{W}_{(j_0)}(Y^{\mathrm{op}}(D))(Q,\hbar)}{\mathcal{W}_{\varnothing}(Y^{\mathrm{op}}(D))(Q,\hbar)} = \frac{(-1)^{fj_0}q^{(f+1/2)\binom{j_0}{2}}}{[j_0]_q!}\,\frac{\prod_{i=1}^{r}(\widetilde{Q}_i^{(1)};q)_{j_0}}{\prod_{k=1}^{s}(\widetilde{Q}_k^{(2)};q)_{j_0}}, \tag{7-13}$$

where f is the integral shift of f from canonical framing, (r,s) are nonnegative integers with $r+s+1=\chi(Y)-1$, and $\widetilde{Q}_i = \prod_{m=1}^{r+s} Q_m^{a_{m,i}}$ with $a_{m,i}\in\{-1,0,1\}$ for $i=1,\dots,r+s$. Elementary manipulations and use of the q–binomial theorem (see [103, Section 4.1]) show that

$$(7\text{-}14)\quad \psi_{Y(D)}(Q,\hbar,z) := \sum_{j_0\geq 0} \frac{\mathcal{W}_{(j_0)}(Y^{\mathrm{op}}(D))(Q,\hbar)}{\mathcal{W}_{\varnothing}(Y^{\mathrm{op}}(D))(Q,\hbar)} z^{j_0} = \frac{\prod_{i=1}^r (\widetilde{Q}_i;q)_\infty}{\prod_{k=1}^s (\widetilde{Q}_{r+k};q)_\infty} \cdot \Phi_{C(Y(D))}\big(q^{(r-s-1)/2}z, \widetilde{Q}_1^{(1)},\dots,\widetilde{Q}_r^{(1)}, q^{1/2}\widetilde{Q}_1^{(2)},\dots,q^{1/2}\widetilde{Q}_s^{(2)}\big),$$

where

$$(7\text{-}15)\quad C(Y(D)) = \left(\begin{array}{c|ccc|ccc} f+1 & \overbrace{1 & \cdots & 1}^{r} & \overbrace{1 & \cdots & 1}^{s} \\ \hline 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline 1 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{array}\right) \begin{array}{l} \\ \left.\begin{array}{c}\\ \\ \\ \end{array}\right\} r \\ \left.\begin{array}{c}\\ \\ \\ \end{array}\right\} s \end{array}$$

and moreover, the genus-zero limit of the logarithm of (7-14) is the generating function of disk invariants of $Y^{\mathrm{op}}(D)$ [5],

$$(7\text{-}16)\quad \lim_{\hbar\to 0} \hbar \log \psi_{Y(D)}(Q,\hbar,z) = \lim_{\hbar\to 0} \hbar\mathcal{O}(Y^{\mathrm{op}}(D))(Q,\hbar,\mathsf{x})\big|_{\mathsf{x}_{\vec{\mu}}=z^{j_0}\delta_{\vec{\mu},(j_0)}} = \sum_\beta O_{0;j_1,\dots,j_{r+s};(j_0)}(Y^{\mathrm{op}}(D)) z^{j_0} \prod_{i=1}^{r+s} Q_i^{j_i}.$$

The matrix C has nonnegative off-diagonal entries, and $\Phi_C(q;x_1,\dots,x_{r+s+1})$ cannot therefore be immediately interpreted as a motivic quiver DT partition function. However, writing $\mathsf{Q}(Y(D))$ for the symmetric quiver with adjacency matrix $C(Y(D))$, we have [103, Appendix A],

$$(7\text{-}17)\quad \Phi_{C(Y(D))}(q;x_1,\dots,x_{r+s+1}) = \prod_{\mathsf{d}\neq 0}\prod_{j\in\mathbb{Z}}\prod_{k\geq 0}\big(1-(-1)^j \mathsf{x}^{\mathsf{d}} q^{-k-(j+1)/2}\big)^{-\mathcal{E}_{\mathsf{d};j}^{C(Y(D))}} = \Phi_{E_{\mathsf{Q}(Y(D))}}(q^{-1};q^{-1/2}x_1,\dots,q^{-1/2}x_{r+s+1}).$$

The exponents $\mathcal{E}_{\mathsf{d};j}^{C(Y(D))}$ are then equal to the motivic DT invariants of $\mathsf{Q}(Y(D))$ up to sign. Furthermore, the numerical DT invariants also agree with the absolute value of $\mathcal{E}_{\mathsf{d}}^{C(Y(D)),\mathrm{num}} := \sum_j (-1)^j \mathcal{E}_{\mathsf{d};j}^{C(Y(D))}$ [103, Appendix A],

$$(7\text{-}18)\quad \mathrm{DT}_{\mathsf{d}}^{\mathrm{num}}(\mathsf{Q}(Y(D))) = |\mathcal{E}_{\mathsf{d}}^{C(Y(D)),\mathrm{num}}|.$$

For $j=(j_0,j_1,\dots,j_{r+s})$, define now the disk BPS invariants of $Y^{\mathrm{op}}(D)$ by

$$(7\text{-}19)\quad O_{0,j_1,\dots,j_{r+s};(j_0)}(Y^{\mathrm{op}}(D)) := \sum_{k\,|\,\gcd(j_0,\dots,j_{r+s})} \frac{1}{k^2}\mathfrak{D}_{j/k}(Y^{\mathrm{op}}(D)).$$

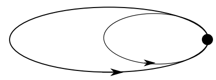
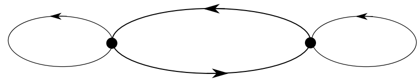
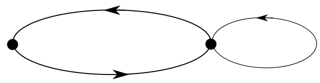
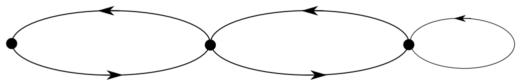
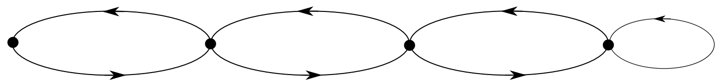
$Y(D)$	$\mathsf{Q}(Y(D))$
$\mathbb{P}^2(1,4)$	
$\mathbb{F}_0(2,2)$ $\mathbb{F}_0(0,4)$	
$\mathrm{dP}_1(1,3)$ $\mathrm{dP}_1(0,4)$	
$\mathrm{dP}_2(1,2)$ $\mathrm{dP}_2(0,3)$	
$\mathrm{dP}_3(1,1)$ $\mathrm{dP}_3(0,2)$	

Table 4: Quivers for 2–component quasi-tame Looijenga pairs.

From (7-13) and (7-16), we have that

$$\mathfrak{D}_{\tau(d)}(Y^{\mathrm{op}}(D)) + \sum_i \alpha_i \delta_{d,v_i} = \mathcal{E}_{\mathsf{d}}^{C(Y(D)),\mathrm{num}},$$

where

(7-20) $$\alpha_i = \begin{cases} 0 & \text{for } i = 1, \\ -1 & \text{for } i = 2, \dots, r+1, \\ +1 & \text{for } i = r+2, \dots, r+s+1, \end{cases}$$

(7-21) $$\tau(d_1, \dots, d_{r+s+1}) = \left(d_1, \sum_{m=1}^{r+s} a_{m,2} d_{m+1}, \dots, \sum_{m=1}^{r+s} a_{m,r+s} d_{m+1} \right).$$

But by (6-13), $O_j(Y^{\mathrm{op(D)}}) = N^{\mathrm{loc}}_{\iota(j)}(Y(D))$, and therefore $\mathfrak{D}_j(Y^{\mathrm{op}}(D)) = \mathrm{KP}_{\iota(j)}(E_{Y(D)})$, from which the claim follows by setting $\kappa := \iota \circ \tau$. □

Remark 7.4 Theorem 7.3, combined with Theorem 5.1, resembles previous correspondences identifying log GW invariants to DT invariants of quivers, and in particular [15], but it differs from them in a number of key respects: the quiver DT invariants here are identified with the (absolute value of the) BPS invariants of the local geometry, and therefore imply a finer integrality property of the log invariants via (5-1) and (7-2). Furthermore, unlike in [15], the motivic refinement is not expected to reconstruct the open Gromov–Witten count at higher genus, as the higher orders in $\hbar$ of (7-16) include contributions of open stable maps with more than one boundary component. A separate discussion of the open BPS structure of the higher-genus theory is the subject of the next section.

Example 7.1 Let $Y(D)=\mathbb{P}^2(1,4)$. In this case we have $r=s=0$, $f=1$, and $\mathsf{Q}(\mathbb{P}^2(1,4))$ is the 2–loop quiver. Moreover, the identification of dimension vectors with curve degrees is simply the identity, $\kappa=\mathrm{id}$, and the integral shift in (7-12) and (7-20) vanishes, $\alpha_1=0$. Then, by Theorem 7.3, the absolute value of the KP invariants of $E_{\mathbb{P}^2(1,4)}$ gives the unrefined DT invariants of $\mathsf{Q}(\mathbb{P}^2(1,4))$. We can in fact check directly that $\mathrm{KP}_d(E_{\mathbb{P}^2(1,4)})=(-1)^d\,\mathrm{DT}_d^{\mathrm{num}}(\mathsf{Q}(\mathbb{P}^2(1,4)))$: according to [107, Theorem 3.2],

$$\mathrm{DT}_d^{\mathrm{num}}(\mathsf{Q}(\mathbb{P}^2(1,4)))=\frac{(-1)^d}{d^2}\sum_{k|d}\mu\Big(\frac{d}{k}\Big)(-1)^k\binom{2k-1}{k-1}, \tag{7-22}$$

and the result follows from (3-16) and the equality

$$\frac{1}{2}\binom{2k}{k}=\frac{1}{2}\frac{(2k)!}{(k!)^2}=\frac{1}{2}\frac{2k}{k}\frac{(2k-1)!}{k!(k-1)!}=\binom{2k-1}{k-1}. \tag{7-23}$$

Remark 7.5 (non-quasi-tame pairs) The condition in Theorem 7.3 that $Y(D)$ is a 2–component quasi-tame pair is likely to be necessary. For example, for $Y(D)$ a non-quasi-tame pair, we do not expect that the result of the finite summation (3-21) can be further simplified down to a form akin to (3-8) as a ratio of products of factorials, unlike the case of the hypergeometric summations in the proof of Theorem 4.5. A little experimentation shows that, writing $N_{0,d}^{\mathrm{loc}}(\mathrm{dP}_5(0,0))=m(d)/n(d)$ with $\gcd(m(d),n(d))=1$, the numerator $m(d)$ is divisible by very large primes $\approx 10^7$ for low degrees $d_i\approx 10^1$ with $d_i\neq d_0$ for $i>0$. This creates a tension with $m(d)$ being a product of factorials with arguments linear in d_i with coefficients $\approx 10^1$, as those would be divisible by at most the largest prime in the range $\approx 10^1-10^2$. As generating functions of numerical DT invariants are always generalised hypergeometric functions [102], and their coefficients are therefore always products of ratios of factorials in the degrees, the KP/DT correspondence of Theorem 7.3 is unlikely to extend to the non-quasi-tame setting.

Remark 7.6 ($l>2$) For l–components pairs with $l>2$, a correspondence between quivers and $(l-1)$–holed open GW partition functions has received some preliminary investigation in the context of the links-quivers correspondence [72; 36], where open stable maps are considered with the same colouring by symmetric Young diagrams for all the connected components of the boundary. The general case of stable maps with arbitrary windings which is relevant for our purposes may, however, fall outside the remit of the open BPS/quiver DT correspondence. In particular, suppose that Q is a symmetric quiver such that

$$P_{\mathsf{Q}}(\alpha_1x,\dots,\alpha_rx,\beta_1y,\dots,\beta_2y)=\sum_{m,n}x^my^n\mathcal{W}_{(m),(n)}(X,L_1\cup L_2,\mathsf{f}_1,\mathsf{f}_2)$$

with $\alpha_i,\beta_i\in\mathbb{C}[q]$ and framed toric special Lagrangians L_1, L_2 in a Calabi–Yau threefold X. The simplest instance is $X=\mathbb{C}^3$ and L_1, L_2 framed toric Lagrangians on different legs: this arises for instance by considering $\mathrm{dP}_1^{\mathrm{op}}(1,1,0)$ and $\mathbb{F}_0^{\mathrm{op}}(2,0,0)$. It is easy to check that the analogue of the semiclassical limit (7-16) for the annulus generating function would be

$$\lim_{q\to 1}D_x^qD_y^q\log P_{\mathsf{Q}}(\alpha_1x,\dots,\alpha_rx,\beta_1y,\dots,\beta_sy)=\sum_{j_1,j_2,\beta}x^{j_1}y^{j_2}O_{0;j_1,j_2}(X,L_1\cup L_2,\mathsf{f}_1,\mathsf{f}_2), \tag{7-24}$$

where D_x^q denotes the q–derivative with respect to x. When $X = \mathbb{C}^3$, a natural guess in line with the disk case would be to take Q a quiver with two vertices, with dimension vectors in bijection with winding numbers along the homology circles in L_1 and L_2. However it is straightforward to verify from (7-6) that for $r+s=2$, the left-hand side of (7-24) does not have a limit as $q \to 1$ unless Q is disconnected, in which case the limit is identically zero, and hence disagrees with the right-hand side. Although this may not necessarily extend to quivers with higher number of vertices and finely tuned identifications of dimension vectors with winding degrees, it does suggest that a suitable generalisation of the correspondence might be required to encompass the counts of annuli as well.

8 Higher-genus BPS invariants

For $Y(D)$ a (not necessarily tame) l–component Looijenga pair satisfying Property O, we define

$$\Omega_d(Y(D))(q) := [1]_q^2 \left(\prod_{i=1}^{l-1} \frac{d \cdot D_i}{[d \cdot D_i]_q} \right) \sum_{k|d} \frac{\mu(k)}{k} \mathsf{O}_{\iota^{-1}(d/k)}(Y^{\mathrm{op}}(D))(-\mathrm{i}k \log q), \tag{8-1}$$

and for $Y(D)$ an l–component pair, not necessarily satisfying Property O, we write

$$\Omega_d(Y(D))(q) := [1]_q^2 \left(\prod_{i=1}^{l} \frac{1}{[d \cdot D_i]_q} \right) \sum_{k|d} (-1)^{d/k \cdot D + l} [k]_q^{l-2} k^{l-2} \mu(k)\, \mathsf{N}^{\log}_{d/k}(Y(D))(-\mathrm{i}k \log q). \tag{8-2}$$

The compatibility of (8-1) and (8-2) when $Y(D)$ satisfies both tameness and Property O follows from Theorem 6.7. From Table 1 and the discussion following Definition 6.3, any quasi-tame l–component Looijenga pair either satisfies Property O, or it is tame, or both: in this setting we will take $\Omega_d(Y(D))(q)$ to be either of the applicable definitions (8-1) or (8-2). We further write simply $\Omega_d(Y(D))$ for the genus-zero limit $\Omega_d(Y(D))(1)$,

$$\begin{aligned} \Omega_d(Y(D)) &:= \frac{1}{\prod_{i=1}^{l} (d \cdot D_i)} \sum_{k|d} (-1)^{\sum_{i=1}^{l} d/k \cdot D_i + 1} \frac{\mu(k)}{k^{4-2l}} N^{\log}_{0,d/k}(Y(D)) \\ &= \sum_{k|d} \frac{\mu(k)}{k^{4-l}} O_{0,\iota^{-1}(d/k)}(Y^{\mathrm{op}}(D)) \\ &= \sum_{k|d} \frac{\mu(k)}{k^{4-l}} N^{\mathrm{loc}}_{d/k}(Y(D)). \end{aligned} \tag{8-3}$$

A priori we can only expect $\Omega_d(Y(D)) \in \mathbb{Q}$ and $\Omega_d(Y(D))(q) \in \mathbb{Q}(q^{1/2})$. By (8-2) and (8-3), however, $\Omega_d(Y(D))$ and $\Omega_d(Y(D))(q)$ are amenable to a physical interpretation as Labastida–Mariño–Ooguri–Vafa (LMOV) partition functions [74; 73; 100; 88]. These heuristically count BPS domain walls in an M–theory compactification on $Y^{\mathrm{op}}(D)$ (see in particular [88, equation 2.10]): writing $\Omega_d(Y(D))(q) = \sum_j n_{d,j}(Y(D)) q^j$, the LMOV invariants $n_{d,j}(Y(D))$ would compute the net degeneracy of M2–branes with spin j and magnetic and bulk charge specified by d, ending on an M5–brane wrapped around the framed toric Lagrangian L in $Y^{\mathrm{op}}(D) = (X, L, \mathsf{f})$. From the vantage point of string theory, (8-2) (resp. (8-3)) are then expected to be integral Laurent polynomials (resp. integers: for $l=2$, since $\Omega_d(Y(D)) = \mathrm{KP}_d(E_{Y(D)}) = \mathfrak{D}(Y^{\mathrm{op}}(D))$ by (7-2), (7-19), (8-3) and (6-13), this is implied by Theorem 7.3). The next theorem shows that this is indeed the case.

Theorem 8.1 (the higher-genus open BPS property) *Let $Y(D)$ be a quasi-tame Looijenga pair. Then $\Omega_d(Y(D))(q) \in q^{-\frac{1}{2}g_{Y(D)}(d)}\mathbb{Z}[q]$ for an integral quadratic polynomial $g_{Y(D)}(d)$.*

Clearly, from (4-3) and (8-1)–(8-2), we have $\Omega_d(q) = \Omega_d(q^{-1})$, so Theorem 8.1 implies in particular that $\Omega_d(q)$ is a Laurent polynomial truncating at $\mathcal{O}(q^{\pm g_{Y(D)}(d)/2})$.

To prove Theorem 8.1 we shall need the following two lemmas. Let ω_d be a primitive d^{th} root of unity.

Lemma 8.2 (the q–Lucas theorem [99]) *Let $n \geq m$ be nonnegative integers. Then*

$$\begin{bmatrix} n \\ m \end{bmatrix}_{\omega_d} = \omega_d^{\frac{1}{2}m(m-n)} \binom{\lfloor n/d \rfloor}{\lfloor m/d \rfloor} \begin{bmatrix} n - d\lfloor n/d \rfloor \\ m - d\lfloor m/d \rfloor \end{bmatrix}_{\omega_d}. \tag{8-4}$$

In particular, if $d \,|\, m$ and $d \,|\, n$,

$$\begin{bmatrix} n \\ m \end{bmatrix}_{\omega_d} = \omega_d^{\frac{1}{2}m(m-n)} \binom{n/d}{m/d}.$$

Proof See eg [108, Theorem 2.2] for a proof. □

Lemma 8.3 *Let $d \,|\, m \,|\, n \in \mathbb{Z}^+$. Then*

$$\partial_q \begin{bmatrix} n \\ m \end{bmatrix}_q \bigg|_{q=\omega_d} = 0.$$

Proof For every $i < n$ with $d \nmid i$ we have $\left[\begin{smallmatrix} n \\ i \end{smallmatrix}\right]_{\omega_d} = 0$, since then

$$\begin{bmatrix} n - d\lfloor n/d \rfloor \\ i - d\lfloor i/d \rfloor \end{bmatrix}_{\omega_d} = \begin{bmatrix} 0 \\ i \bmod d \end{bmatrix}_{\omega_d} = 0. \tag{8-5}$$

The Cauchy binomial theorem,

$$\sum_{m=0}^{n} t^m q^{\frac{1}{2}m(n+1)} \begin{bmatrix} n \\ m \end{bmatrix}_q = \prod_{i=1}^{n} (1 + tq^i), \tag{8-6}$$

implies that

$$q^{\frac{1}{2}m(n+1)} \begin{bmatrix} n \\ m \end{bmatrix}_q = e_m(q, \dots, q^n), \tag{8-7}$$

where $e_j(x_1, \dots, x_n)$ is the j^{th} elementary symmetric polynomials in n variables. We differentiate (8-7) and evaluate at $q = \omega_d$, where now $d \,|\, m \,|\, n$. Write $n = abd$, $m = bd$ for $a, b \in \mathbb{Z}^+$. From (8-6) we find

$$\partial_q \prod_{i=1}^{n} (1+tq^i) = \prod_{i=1}^{n} (1+tq^i) \left(\sum_{j=1}^{n} \frac{jtq^{j-1}}{1+tq^j} \right) = \sum_{i=0}^{n} t^i e_i(q, \dots, q^n) \cdot t \sum_{j=1}^{n} \sum_{k=0}^{\infty} j(-t)^k q^{kj+j-1}. \tag{8-8}$$

Let us now evaluate at $q=\omega_d$ and take the $\mathcal{O}(t^m)$ coefficient on both sides. We have

$$(8\text{-}9)\quad \partial_q e_m(q,\dots,q^n)|_{q=\omega_d} = [t^m]\sum_{i=0}^{n} t^i e_i(\omega_d,\dots,\omega_d^n)\cdot t\sum_{j=1}^{n}\sum_{k=0}^{\infty} j(-t)^k\omega_d^{kj+j-1}$$
$$= [t^{bd}]\sum_{i=0}^{ab} t^{di}\omega_d^{id(n+1)/2}\omega_d^{id(id-n)/2}\binom{ab}{i}\cdot t\sum_{j=1}^{n}\sum_{k=0}^{\infty} j(-t)^k\omega_d^{kj+j-1}$$
$$= \sum_{i=0}^{b-1}\omega_d^{\frac{1}{2}id(d+1)}\binom{ab}{i}\sum_{j=1}^{abd} j(-1)^{bd-1-id}\omega_d^{bdj-1-idj}$$
$$= (-1)^{m+1}\frac{n(n+1)}{2\omega_d}\sum_{i=0}^{b-1}(-1)^i\binom{ab}{i} = \frac{(-1)^{b+m}n(n+1)}{2a\omega_d}\binom{ab}{b},$$

where we have used (8-5) and Lemma 8.2. On the other hand,

$$(8\text{-}10)\qquad \frac{\partial}{\partial q}q^{\frac{1}{2}m(n+1)}\begin{bmatrix} n\\ m\end{bmatrix}_q\Bigg|_{q=\omega_d} = \frac{m(n+1)}{2\omega_d}\omega_d^{m(m+1)/2}\binom{ab}{b} + \omega_d^{m(n+1)/2}\partial_q\begin{bmatrix} n\\ m\end{bmatrix}_q\Bigg|_{q=\omega_d}$$
$$= \frac{m(n+1)}{2\omega_d}(-1)^{b+m}\binom{ab}{b} + \omega_d^{m(n+1)/2}\partial_q\begin{bmatrix} n\\ m\end{bmatrix}_q\Bigg|_{q=\omega_d},$$

where in tracking down the last sign factor we have been mindful that $(-1)^{bm}=(-1)^m$ since $b|m$. The claim then follows by equating (8-9) to (8-10). □

Proof of Theorem 8.1 We break up the proof of the theorem by considering each value of l separately.

- $(l=2)$ It suffices to prove the theorem in the case $Y(D)=\mathrm{dP}_3(1,1)$, since $\Omega_d(\mathrm{dP}_3(1,1)) = \Omega_d(\mathrm{dP}_3(0,2))$ from (8-1) and the discussion of Section 6.3.1, and all other cases are then recovered from the blowup formulas of Propositions 4.3 and 6.9. Let $\tilde{d} := \gcd(d_0,d_1,d_2,d_3)$. We first plug (4-22) into (8-1),

$$(8\text{-}11)\qquad \Omega_d(\mathrm{dP}_3(1,1))(q) = [1]_q^2\sum_{k|\tilde{d}}\mu(k)\frac{(-1)^{(d_1+d_2+d_3)/k}}{[d_0/k]_{q^k}[(d_1+d_2+d_3-d_0)/k]_{q^k}}\Theta_{d/k}(q^k),$$
$$= \frac{[1]_q^2}{[d_0]_q[d_1+d_2+d_3-d_0]_q}\sum_{k|\tilde{d}}\mu(k)(-1)^{(d_1+d_2+d_3)/k}\Theta_{d/k}(q^k),$$

where

$$(8\text{-}12)\qquad \Theta_d(q) := \begin{bmatrix} d_0\\ d_1\end{bmatrix}_q\begin{bmatrix} d_1\\ d_0-d_2\end{bmatrix}_q\begin{bmatrix} d_1+d_2+d_3-d_0\\ d_1\end{bmatrix}_q\begin{bmatrix} d_1\\ d_0-d_3\end{bmatrix}_q.$$

It is immediate to verify that $\Omega_d(\mathrm{dP}_3(1,1))(q)\in q^{-\frac{1}{2}g_{\mathrm{dP}_3(1,1)}(d)}\mathbb{Z}[\![q]\!]$, with

$$(8\text{-}13)\qquad g_{\mathrm{dP}_3(1,1)}(d) = 2\,(d_1+d_2+d_3-d_0)\,d_0 - d_1^2 - d_2^2 - d_3^2 - d_1 - d_2 - d_3 + 2$$

since

$$\begin{bmatrix} n \\ m \end{bmatrix}_q \in q^{-m(n-m)/2}\mathbb{Z}[q], \quad \frac{1}{[n]_q} \in q^{n/2}\mathbb{Z}[\![q]\!] \quad \text{and} \quad [m]_q \in q^{-m/2}\mathbb{Z}[\![q]\!]$$

as formal Laurent series at $q=0$ with truncating principal part for any positive integers n, m. Furthermore, away from $q=0,\infty$, $\Omega_d(\mathrm{dP}_3(1,1))(q) \in \mathbb{Q}(q^{1/2})$ is a rational function of $q^{1/2}$ with at worst double poles possibly at the zeroes of $[d_0]_q[d_1+d_2+d_3-d_0]_q$, namely $q=\omega_{d_0}^j$ for $j=1,\dots,d_0-1$, and $q=\omega^j_{d_1+d_2+d_3-d_0}$ for $j=1,\dots,d_1+d_2+d_3-d_0-1$. We shall now prove that $\Omega_d(\mathrm{dP}_3(1,1))(q)$ is in fact regular on the unit circle.

First off, upon expanding all q–analogues in (8-11) in cyclotomic polynomials,

$$[n]_q = \prod_{d\mid n} \Phi_d(q), \tag{8-14}$$

it is straightforward to check that [59, Lemma 5.2]

$$\frac{[\gcd(n,m)]_q}{[n+m]_q}\begin{bmatrix} n+m \\ m \end{bmatrix}_q \in q^{\frac{1}{2}(n+m-nm-\gcd(n,m))}\mathbb{Z}[q], \tag{8-15}$$

which implies that $\Omega_d(\mathrm{dP}_3(1,1))(q)$ is regular on the unit circle outside of $\{\omega_{\tilde d}^j\}_{j=0}^{\tilde d}$, where we recall that $\tilde d := \gcd(d_0,d_1,d_2,d_3)$. Let now

$$\tilde\Omega_d(\mathrm{dP}_3(1,1))(q) := \frac{[d_0]_q[d_1+d_2+d_3-d_0]_q}{[1]_q^2}\Omega_d$$

and $\tilde d_i = d_i/\tilde d$. From Lemma 8.2, we have

$$\begin{aligned}\Theta_{d/k}(\omega_{\tilde d}^{kj}) \\ &= (-1)^{(d_1+d_2+d_3)j/k}\binom{\tilde d_1\epsilon_{k,j}}{\tilde d_2\epsilon_{k,j}}\binom{\tilde d_1\epsilon_{k,j}}{(\tilde d_0-\tilde d_2)\epsilon_{k,j}}\binom{(\tilde d_1+\tilde d_2+\tilde d_3-\tilde d_0)\epsilon_{k,j}}{\tilde d_1\epsilon_{k,j}}\binom{\tilde d_1\epsilon_{k,j}}{(\tilde d_0-\tilde d_3)\epsilon_{k,j}},\end{aligned} \tag{8-16}$$

where $\epsilon_{k,j} = \gcd(\tilde d/k, j)$. Then

$$\begin{aligned}\tilde\Omega_{\tilde d}(\mathrm{dP}_3(1,1))(\omega_{\tilde d}^j) \\ &= \sum_{k\mid\tilde d}\mu\Big(\frac{\tilde d}{k}\Big)(-1)^{(\tilde d_1+\tilde d_2+\tilde d_3)k(j+1)}\binom{\tilde d_1\epsilon_{\tilde d/k,j}}{\tilde d_2\epsilon_{\tilde d/k,j}}\binom{\tilde d_1\epsilon_{\tilde d/k,j}}{(\tilde d_0-\tilde d_2)\epsilon_{\tilde d/k,j}} \\ &\qquad\cdot\binom{(\tilde d_1+\tilde d_2+\tilde d_3-\tilde d_0)\epsilon_{\tilde d/k,j}}{\tilde d_1\epsilon_{\tilde d/k,j}}\binom{\tilde d_1\epsilon_{\tilde d/k,j}}{(\tilde d_0-\tilde d_3)\epsilon_{\tilde d/k,j}}.\end{aligned} \tag{8-17}$$

Consider first $\tilde d \neq 1$ and write $\nu_p(n)$ and $\mathrm{rad}(n)$ for the p–adic valuation and the radical of $n\in\mathbb{Z}^+$, respectively. Let $k\mid\tilde d$ and suppose without loss of generality that $\tilde d/k$ has no repeated prime factors, $\tilde d/k = \mathrm{rad}(\tilde d/k)$. Then, for $\omega_{\tilde d}\neq 1$, the following trichotomy holds:

(I) $\widetilde{d}/k \nmid j$ and there exists p' prime with $p' \,|\, \widetilde{d}/k$ and $p' \nmid j$.

Let $k' := kp'$. Then $k' \,|\, \widetilde{d}$, $\gcd(k', j) = \gcd(k, j)$, $\mu(\widetilde{d}/k') = -\mu(\widetilde{d}/k)$. Moreover $(-1)^{k'(j+1)} = (-1)^{k(j+1)}$, which is obvious when p' is odd, and it also holds when $p' = 2$ since in that case j must be odd. Then the contributions from k' and k to the sum (8-17) cancel each other.

(II) $\widetilde{d}/k \,|\, j$ and there exists $p' < \widetilde{d}$ such that $p' \,|\, d$ and $p' \nmid j$.

In this case we have $p' \nmid \widetilde{d}/k$, $p' \,|\, k$. Let $k' := k/p'$. Then as before $\mu(\widetilde{d}/k) = -\mu(\widetilde{d}/k')$, $(-1)^{k'(j+1)} = (-1)^{k'(j+1)}$ and $\gcd(k', j) = \gcd(k, j)$, and the summand corresponding to k' has opposite sign to the one corresponding to k in (8-17).

(III) $\widetilde{d}/k \,|\, j$ and $\widetilde{d}$ has no prime factor $p' \nmid j$.

Suppose for simplicity that $\mathrm{rad}(j)/\mathrm{rad}(\widetilde{d})$ is odd, the even case being essentially identical. Then (8-17) is unchanged upon replacing $j =: \prod_{p|j} p^{\nu_p(j)}$ with $\prod_{p|j,\, p|\widetilde{d}} p^{\nu_p(j)}$, so we may assume $\mathrm{rad}(j) = \mathrm{rad}(\widetilde{d})$. Let p' be such that $\nu_{p'}(\widetilde{d}) > \nu_{p'}(j)$ and let $k' := k/p'$. Then once again the contributions of k and k' to (8-17) cancel each other.

All in all, the above shows that $\widetilde{\Omega}_d(\mathrm{dP}_3(1,1))$ vanishes at $\omega_{\widetilde{d}}^j$ for all $\widetilde{d} > 1$ and $j = 1, \dots, \widetilde{d} - 1$. But by Lemma 8.3 these are all double zeroes, and therefore $\Omega_d(\mathrm{dP}_3(1,1))$ is regular therein. Moreover, $\Omega_d(\mathrm{dP}_3(1,1))$ is regular by construction at $q = 1$, where its value is given by replacing all q–expressions in (8-11) by their classical counterparts. Hence $\Omega_d(\mathrm{dP}_3(1,1)) \in \mathbb{Q}[q^{\pm 1/2}]$ is a rational Laurent polynomial; but we also know that $\Omega(\mathrm{dP}_3(1,1))_d \in q^{-\frac{1}{2}g_{\mathrm{dP}_3(1,1)}(d)}\mathbb{Z}[\![q]\!]$ is an integral Laurent series, which thus truncates at $\mathcal{O}(q^{\frac{1}{2}g_{\mathrm{dP}_3(1,1)}(d)})$. The statement of the theorem follows.

- ($l = 3$) As before, we prove the statement for $Y(D) = \mathrm{dP}_3(0,0,0)$ and recover all 3–component pairs by restriction in the degrees. Let

$$\widetilde{d} := \gcd(d_0, d_1, d_2, d_3) \quad \text{and} \quad \hat{d} := d_0^2 - d_0(d_1 + d_2 + d_3) + d_1 d_2 + d_1 d_3 + d_2 d_3.$$

From (4-39),

$$\Omega_d(\mathrm{dP}_3(0,0,0))(q) = [1]_q^2 \sum_{k|\widetilde{d}} \mu(k) k \frac{(-1)^{(d_0+d_1+d_2)/k+1}[\hat{d}/k^2]_{q^k}}{[d_1/k]_q [d_2/k]_q [d_3/k]_{q^k}} \Xi_{d/k}(q^k), \tag{8-18}$$

where

$$\Xi_d(q) := \begin{bmatrix} d_1 \\ d_0 - d_2 \end{bmatrix}_q \begin{bmatrix} d_2 \\ d_0 - d_3 \end{bmatrix}_q \begin{bmatrix} d_3 \\ d_0 - d_1 \end{bmatrix}_q. \tag{8-19}$$

Outside $q = 0, \infty$, the polynomial $\Omega_d(\mathrm{dP}_3(0,0,0))(q)$ has at worst double poles at $q = \omega_d^j$ only; also it is verified directly that $q^{\frac{1}{2}g_{\mathrm{dP}_3(0,0,0)}(d)}\Omega_d$ has a Taylor expansion at $q = 0$ with integer coefficients, where

$$g_{\mathrm{dP}_3(0,0,0)}(d) = g_{\mathrm{dP}_3(1,1)}(d). \tag{8-20}$$

For $q = 1$, the ratios of q–numbers in (8-18) limits to the corresponding classical counterparts, so $\Omega_d(\mathrm{dP}_3(0,0,0))(1)$ is well-defined. Suppose then $q = \omega_{\tilde{d}}^j \neq 1$. We have that

$$\frac{[\hat{d}/k^2]_{q^k}}{[d_1/k]_{q^k}[d_2/k]_{q^k}[d_3/k]_{q^k}} = \frac{\hat{d}}{k d_1 d_2 d_3}\left[\frac{\omega_{\tilde{d}}^{2j}}{(q-\omega_{\tilde{d}}^j)^2} + \frac{1}{q-\omega_{\tilde{d}}^j} + \mathcal{O}(1)\right]. \tag{8-21}$$

This is nearly k–independent, save for the factor of k that cancels the one present in the summand of (8-18). By the same arguments of the previous point, the resulting divisor sum

$$\sum_{k|\tilde{d}} \mu(k)(-1)^{(d_0+d_1+d_2)/k+1}\,\Xi_{d/k}(q^k)$$

vanishes quadratically at $\omega_{\tilde{d}}^j$, and therefore $\Omega_d(\mathrm{dP}_3(0,0,0))(q)$ is regular on the unit circle, concluding the proof.

- $(l = 4)$ This consists of the single case $Y(D) = \mathbb{F}_0(0,0,0,0)$. Let $\tilde{d} := \gcd(d_1, d_2)$. We have, from (6-35), that

$$\Omega_d(\mathbb{F}_0(0,0,0,0))(q) = \frac{[1]_q^2}{[d_1]_q^2[d_2]_q^2}\sum_{k|\tilde{d}} \mu(k)k^2[d_1d_2/k^2]_{q^k}^2. \tag{8-22}$$

In this case we have

$$g_{\mathbb{F}_0(0,0,0,0)}(d) = 2(d_1d_2 - d_1 - d_2 + 1). \tag{8-23}$$

As before, $\Omega_d(\mathbb{F}_0(0,0,0,0))(q)$ is a rational function with an integral Taylor–Laurent expansion at $q = 0$, order $g_{\mathbb{F}_0(0,0,0,0)}(d)/2$ singularities at $q = 0, \infty$ and possibly double poles at $q = \omega_{\tilde{d}}^j$. Expanding (8-22) at $\omega_{\tilde{d}}^j$ yields

$$\Omega_d(\mathbb{F}_0(0,0,0,0))(q) = \sum_{k|d}\mu(k)\left[\frac{\omega_{\tilde{d}}^j(\omega_{\tilde{d}}^j-1)^2}{(q-\omega_{\tilde{d}}^j)^2} + \frac{2\omega_{\tilde{d}}^j(\omega_{\tilde{d}}^j-1)}{q-\omega_{\tilde{d}}^j}\right] + \mathcal{O}(1), \tag{8-24}$$

which vanishes up to $\mathcal{O}(1)$ since $\sum_{k|d}\mu(k) = 0$, hence $\Omega_d(\mathbb{F}_0(0,0,0,0))(q) \in q^{-\frac{1}{2}g_{\mathbb{F}_0(0,0,0,0)}(d)}\mathbb{Z}[q]$. □

9 Orbifolds

In [18], we proposed in the context of toric pairs that the log-local principle should extend to Y a possibly singular $\mathbb{Q}$–factorial projective variety. We expect that this should also hold for nef Looijenga pairs, at least as long as the orbifold singularities are at the intersection of the divisors: the log GW theory is then well-defined since $Y(D)$ is log smooth, and the local GW theory makes sense by viewing Y and $E_{Y(D)}$ as smooth Deligne–Mumford stacks. In particular, introducing singularities gives new infinite lists of examples of nef/quasi-tame/tame Looijenga pairs.

We propose that also Theorems 6.7, 7.3 and 8.1 may extend to the orbifold setting. We present the simplest instance here, and defer a more in-depth discussion, including criteria for the validity of the orbifold versions of Theorems 6.7, 7.3 and 8.1 to [17].

Example 9.1 Let $Y = \mathbb{P}_{(1,1,n)}$ be the weighted projective plane with weights $(1,1,n)$, and $D = D_1 + D_2$ with D_1 a toric line passing through the orbifold point and D_2 a smooth member of the linear system given by the sum of the two other toric divisors. Since $D_1 \sim H/n$, $D_2 \sim (n+1)/nH$ and $H^2 = n$, we have $D_1^2 = 1/n$ and $D_2^2 = (n+1)^2/n$. Therefore $\mathbb{P}_{(1,1,n)}(1/n, (n+1)^2/n)$ is a tame orbifold Looijenga pair.

Local Gromov–Witten invariants of $Y(D)$ can be computed by the orbifold quantum Riemann–Roch theorem of [113]: when restricted to point insertions, it gives (3-8) specialised to the case at hand, and we get

$$N_{0,d}^{\mathrm{loc}}\left(\mathbb{P}_{(1,1,n)}\left(\frac{1}{n}, \frac{(n+1)^2}{n}\right)\right) = \frac{(-1)^{nd}}{(n+1)d^2}\binom{(n+1)d}{d}. \tag{9-1}$$

A toric model and a quantised scattering diagram for $Y(D)$ can be constructed as follows. The fan of $\mathbb{P}_{(1,1,n)}$ has 1–skeleton given by rays generated by $(-1,0)$, $(0,-1)$ and $(1,n)$. We may choose $D_1 = D_{(1,n)}$. Denote by $\widetilde{Y}$ the toric blowup obtained by adding a ray in the direction $(-1,1)$, and denote by E the corresponding divisor. Choose $\widetilde{D} = D_1 + D_2 + E$, where we identify D_1 and D_2 with their proper transforms. Then $\widetilde{Y}(\widetilde{D}) \to Y(D)$ is a corner blowup. The proper transform of $D_{(-1,0)}$ is a (-1)–curve, which we contract $\widetilde{Y}(\widetilde{D}) \to \overline{Y}(\overline{D})$. Then $\overline{Y} \setminus \overline{D}$ has Euler characteristic 0, hence is $(\mathbb{C}^*)^2$, and therefore $\overline{Y}(\overline{D})$ is toric, ie $\widetilde{Y}(\widetilde{D}) \to \overline{Y}(\overline{D})$ is a toric model and we are in the setting of Proposition 2.4. Identifying proper transforms, we have that $\overline{D} = D_1 + D_2 + E$, with D_1 corresponding to the ray $(1,n)$, D_2 to the ray $(0,-1)$ and E to the ray $(-1,1)$. Applying the $\mathrm{SL}(2,\mathbb{Z})$ transformation

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \tag{9-2}$$

we obtain the toric model depicted at left of Figure 17, for which the broken line calculation is straightforward. The result is

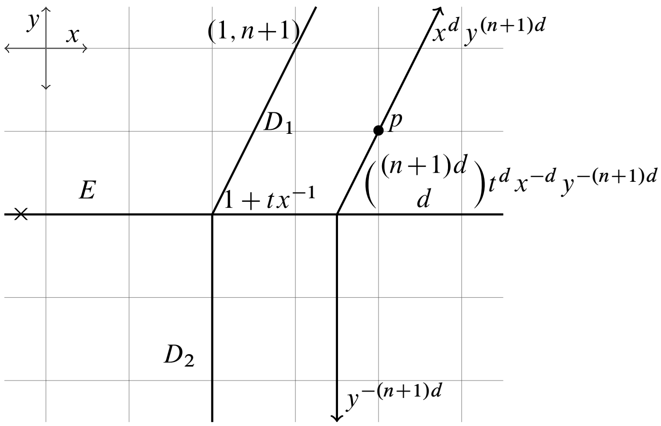
$$N_{0,d}^{\log}\left(\mathbb{P}_{(1,1,n)}\left(\frac{1}{n}, \frac{(n+1)^2}{n}\right)\right) = \binom{(n+1)d}{d}, \tag{9-3}$$

$$\mathsf{N}_{d}^{\log}\left(\mathbb{P}_{(1,1,n)}\left(\frac{1}{n}, \frac{(n+1)^2}{n}\right)\right) = \begin{bmatrix} (n+1)d \\ d \end{bmatrix}_q. \tag{9-4}$$

To construct

$$\mathbb{P}_{(1,1,n)}^{\mathrm{op}}\left(\frac{1}{n}, \frac{(n+1)^2}{n}\right),$$

we delete the line D_1. Then $\mathcal{O}(-D_2)$ is trivial on $\mathbb{P}_{(1,1,n)} \setminus D_1 = \mathbb{C}^2$, and $\mathrm{Tot}(K_{\mathbb{P}_{(1,1,n)} \setminus D_1}) = \mathbb{C}^3$, with an outer toric Lagrangian at framing shifted by n. A topological vertex calculation of higher-genus



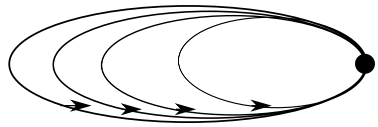


Figure 17: Left: Scatt $\mathbb{P}_{(1,1,n)}$. Right: The quiver for $Y(D) = \mathbb{P}(1,1,3)\bigl(\frac{1}{3}, \frac{16}{3}\bigr)$.

1–holed open Gromov–Witten invariants as in Section 6.3.1 shows that

$$\mathsf{O}_d\left(\mathbb{P}^{\mathrm{op}}_{(1,1,n)}\left(\frac{1}{n}, \frac{(n+1)^2}{n}\right)\right) = \frac{(-1)^{nd}}{d\,[(n+1)d]_q}\begin{bmatrix}(n+1)d\\ d\end{bmatrix}_q. \tag{9-5}$$

Equations (9-1), (9-3), (9-4) and (9-5) together imply that Theorems 5.1 and 6.7 extend to this case as well. The arguments in the proof of Theorem 7.3 also apply verbatim, with $\mathsf{Q}(\mathbb{P}_{(1,1,n)}(1/n, (n+1)^2/n))$ the $(n+1)$–loop quiver. An interesting consequence is that the integrality statement of Conjecture 7.1 appears to persist in the orbifold world too. The proof of the higher-genus open BPS property in Theorem 8.1 also carries through to this setting with no substantial modification.

Appendix A Proof of Theorem 3.3

Let Y be the toric surface given by the fan of Figure 18. It is described by the exact sequence

$$0 \to \mathbb{Z}^6 \xrightarrow{\begin{pmatrix}1&1&0&0&0&-2\\0&0&1&0&0&1\\1&0&0&1&0&0\\0&0&0&0&1&0\\0&1&0&1&0&0\\1&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&1\end{pmatrix}} \mathbb{Z}^8 \xrightarrow{\begin{pmatrix}1&1&0&-1&-1&-1&0&1\\0&1&1&1&0&-1&-1&-1\end{pmatrix}} \mathbb{Z}^2 \to 0 \tag{A-1}$$

showing that Y is a GIT quotient

$$\mathbb{C}^8/\!\!/(\mathbb{C}^\star)^6 = (\mathbb{C}^8 \setminus \{x_i x_j = 0\}_{(i,j)\neq(1,8),\, j\neq i+1})/(\mathbb{C}^\star)^6,$$

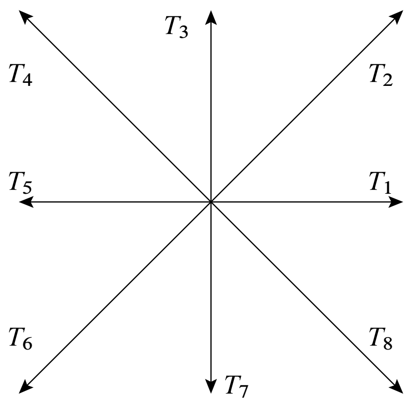


Figure 18: The fan of $\mathrm{Bl}_{4\,\mathrm{pts}}\mathbb{P}^1 \times \mathbb{P}^1$.

with $(\tau_1, \dots, \tau_6) \in (\mathbb{C}^\star)^6$ acting as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} \longrightarrow \begin{pmatrix} \tau_1\tau_2\tau_6^{-2}x_1 \\ \tau_3\tau_6 x_2 \\ \tau_1\tau_4 x_3 \\ \tau_5 x_4 \\ \tau_2\tau_4 x_5 \\ \tau_1\tau_3 x_6 \\ \tau_4 x_7 \\ \tau_5\tau_6 x_8 \end{pmatrix}. \tag{A-2}$$

There are dominant birational morphisms $Y \xrightarrow{\pi_1} \mathbb{P}^2$ and $Y \xrightarrow{\pi_2} \mathbb{P}^1 \times \mathbb{P}^1$, obtained by deleting the loci $\{x_i = 0\}_{i \in \{2,4,6,7,8\}}$ and $\{x_{2i} = 0\}$, respectively. Therefore $Y \simeq \mathrm{Bl}_{4\mathrm{pts}}\mathbb{P}^1 \times \mathbb{P}^1$, or equivalently, Y is a five-point toric blowup of $\mathbb{P}^2$, and deforms to dP_5 upon taking the points in general position. From (A-1) and Figure 18, in terms of the hyperplane H and exceptional classes $E_i \in \mathrm{Pic}(\mathrm{dP}_5)$, the toric divisors $T_i := \{x_i = 0\}$ read

$$\begin{aligned} &T_1 = H - E_1 - E_2 - E_4, \quad && T_3 = H - E_1 - E_3 - E_5, \quad && T_5 = E_2 - E_4, && T_7 = E_3 - E_5, \\ &T_2 = E_1, && T_4 = E_4, && T_6 = H - E_2 - E_3, \quad && T_8 = E_5. \end{aligned} \tag{A-3}$$

Under this identification the -2–curve classes T_{2k+1} do not belong to $\mathrm{NE}(\mathrm{dP}_5)$ (see the discussion of Section 2.3); however they do have, by construction, effective representatives in $A_1(Y)$, since they are prime toric divisors.

To write the I–function, we fix the following set of $\frac{1}{2}\mathbb{Z}$–generators of $A_1(Y)$:

$$C_i = \begin{cases} T_{2i} & \text{for } i = 1, \dots, 4, \\ D_{i+4} & \text{for } i = 1, 2, \end{cases} \tag{A-4}$$

where $D_1 = H - E_1 = T_1 + T_3 + 2T_6$ and $D_2 = 2H - E_2 - E_3 - E_4 - E_5 = T_2 + T_4 + T_5 + T_7 - T_8$. We will write φ_i with $(\varphi_i, C_j) = \delta_{ij}$ for their dual basis in cohomology, and denote curve classes in this basis as $d=\sum_i \frac{1}{2}\sigma_i\delta_i C_i$ with $\delta_i\in\mathbb{Z}$, and $\sigma_i=-1$ for $1\le i\le 4$ and $\sigma_i=1$ otherwise. To write the twisted I–function $I^{E_{Y(D)}}$, we need to expand $\theta_a=c_1(\mathcal{O}(T_a))$ and $\kappa_i=c_1(\mathcal{O}(D_i))$ in (3-11), yielding

$$I^{E_{Y(D)}}(y,z)= \sum_{\delta_i\in\mathbb{Z}}\Bigg[\frac{y_1^{-\frac{1}{2}\delta_1}y_2^{-\frac{1}{2}\delta_2}y_3^{-\frac{1}{2}\delta_3}y_4^{-\frac{1}{2}\delta_4}y_5^{\frac{1}{2}\delta_5}y_6^{\frac{1}{2}\delta_6}(-1)^{\delta_5+\delta_6}}{\left(1-\frac{2\varphi_1}{z}\right)_{\delta_1}\left(1-\frac{2\varphi_2}{z}\right)_{\delta_2}\left(1-\frac{2\varphi_3}{z}\right)_{\delta_3}\left(1-\frac{2\varphi_4}{z}\right)_{\delta_4}\left(\frac{z+\varphi_1+\varphi_2+\varphi_5}{z}\right)_{\frac{1}{2}(-\delta_1-\delta_2+\delta_5)}} \\ \frac{(2\varphi_6-\lambda_1)(2\varphi_5-\lambda_2)\left(\frac{z+2\varphi_6-\lambda_1}{z}\right)_{\delta_6-1}\left(\frac{z+2\varphi_5-\lambda_2}{z}\right)_{\delta_5-1}}{z\left(\frac{z+\varphi_3+\varphi_4+\varphi_5}{z}\right)_{\frac{1}{2}(-\delta_3-\delta_4+\delta_5)}\left(\frac{z+\varphi_1+\varphi_3+\varphi_6}{z}\right)_{\frac{1}{2}(-\delta_1-\delta_3+\delta_6)}\left(\frac{z+\varphi_2+\varphi_4+\varphi_6}{z}\right)_{\frac{1}{2}(-\delta_2-\delta_4+\delta_6)}}\Bigg], \tag{A-5}$$

where

$$(a)_n=a(a+1)\cdots(a+n-1) \tag{A-6}$$

is the Pochhammer symbol. By (3-12), the mirror map is extracted as the formal $\mathcal{O}(z^0)$ Taylor coefficient around $z=\infty$. We find that the sole contributions to the mirror map arise from multiple covers of our chosen generators C_i, that is when $\delta_i=2\sigma_i n$ for some $n\in\mathbb{N}^+$,

$$\tilde{t}^i(y)=\sum_{\delta_i=1}^{\infty}\frac{(2\delta_i-1)!}{(\delta_i!)^2}y^{\delta_i}, \tag{A-7}$$

which is closed-form inverted as

$$y_i(t)=\frac{\exp t^i}{(1-\exp t^i)^2}. \tag{A-8}$$

Then[14]

$$J^{E_{Y(D)}}_{\text{small}}=I^{E_{Y(D)}}(y(t),z),$$

and from (3-10) and (A-5) we find that whenever $d\neq 2\sigma_i n$ for $n\in\mathbb{N}^+$,

$$N^{\text{loc},\psi}_{\delta_1,\dots,\delta_6}(Y(D))=\frac{1}{\lambda_1\lambda_2}[z^{-1}\mathrm{e}^{\sum t_i\varphi_i/z}\mathbf{1}_{H_T(E_{Y(D)})}]I^{E_{Y(D)}}(y(t),z) \\ =[\mathrm{e}^{\sum_i\delta_i t_i}]\sum_{\delta_i'}^{\infty}\mathrm{S}^{[0]}_{\delta_1',\dots,\delta_6'}\prod_{i=1}^{6}y_i(t)^{\frac{1}{2}\sigma_i\delta_i'}, \tag{A-9}$$

[14]To obtain the small J–function, we should include a string-equation induced shift by multiplying the I–function by an overall factor of $\mathrm{e}^{\lambda_1\tilde{t}^5(y)+\lambda_2\tilde{t}^6(y)/z}$, in order to guarantee that the small J–function satisfies its defining property to be the unique family of Lagrangian cone elements with a Laurent expansion of the form $z+t+O(1/z)$ at $z=\infty$. These would result in a correction of the foregoing discussion for degrees $\delta_i=0$ when $i=1,\dots,4$. It is justified to ignore this for our purposes: since $d\cdot D_i=0$ and $\mathcal{O}(-D_i)$ is not a concave line bundle, the corresponding invariants are nonequivariantly ill-defined; and any sensible nonequivariant definition would satisfy automatically the log-local correspondence of Section 5, as the corresponding log invariants are trivially zero.

where

(A-10)
$$\mathsf{S}^{[0]}_{\delta'_1,\dots,\delta'_6} := \frac{(-1)^{\delta'_5+\delta'_6}(\delta'_5-1)!\,(\delta'_6-1)!}{\delta'_1!\delta'_2!\delta'_3!\delta'_4!\big(\frac{1}{2}(\delta'_5-\delta'_1-\delta'_2)\big)!\big(\frac{1}{2}(\delta'_5-\delta'_3-\delta'_4)\big)!\big(\frac{1}{2}(\delta'_6-\delta'_1-\delta'_3)\big)!\big(\frac{1}{2}(\delta'_6-\delta'_2-\delta'_4)\big)!}.$$

The arguments of the factorials in the denominator constrain the range of summation to extend over $\delta_i \neq 0$ alone; in particular the right-hand side is a Taylor series in $(y_1^{-1/2}, y_2^{-1/2}, y_3^{-1/2}, y_4^{-1/2}, y_5^{1/2}, y_6^{1/2})$, convergent in a ball centred at $y_i^{\sigma_i} = 0$.

We first perform the summation over δ'_6 to obtain

(A-11)
$$\sum_{\delta'_6=0}^{\infty} \mathsf{S}^{[0]}_{\delta'_1,\dots,\delta'_6} y_6^{\frac{1}{2}\delta'_6} = \frac{(-1)^{\delta'_2+\delta'_4+\delta'_5}(\delta'_2+\delta'_4-1)!\,(\delta'_5-1)!\left(\dfrac{e^{t_6}}{(e^{t_6}+1)^2}\right)^{\frac{1}{2}(\delta'_2+\delta'_4)}}{\delta'_1!\delta'_2!\delta'_3!\delta'_4!\big(\frac{1}{2}(-\delta'_1+\delta'_2-\delta'_3+\delta'_4)\big)!\big(\frac{1}{2}(-\delta'_1-\delta'_2+\delta'_5)\big)!\big(\frac{1}{2}(-\delta'_3-\delta'_4+\delta'_5)\big)!} \times {}_2F_1\left(\tfrac{1}{2}(\delta'_2+\delta'_4), \tfrac{1}{2}(\delta'_2+\delta'_4+1); \tfrac{1}{2}(-\delta'_1+\delta'_2-\delta'_3+\delta'_4+2); \frac{4e^{t_6}}{(e^{t_6}+1)^2}\right),$$

where

(A-12)
$${}_pF_r(a_1,\dots,a_p; b_1,\dots,b_r; z) := \sum_{k\geq 0} \frac{z^k}{k!} \frac{\prod_{j=1}^{p}(a_j)_k}{\prod_{j=1}^{r}(b_j)_k}$$

is the generalised hypergeometric function. Applying Kummer's quadratic transformation,

(A-13)
$${}_2F_1(a,b;a-b+1;z) = (z+1)^{-a}\,{}_2F_1\left(\frac{a}{2}, \frac{a+1}{2}; a-b+1; \frac{4z}{(z+1)^2}\right),$$

we obtain

(A-14)
$$N^{\mathrm{loc},\psi}_{\delta_1,\dots,\delta_6}(Y(D)) = [e^{\sum_{i=1}^{5}\delta_i t_i}] \sum_{\delta'_i}^{\infty} \mathsf{S}^{[1]}_{\delta'_1,\dots,\delta'_5,\delta_6} \prod_{i=1}^{5} y_i(t)^{\frac{1}{2}\sigma_i\delta'_i},$$

where

(A-15)
$$\mathsf{S}^{[1]}_{\delta'_1,\dots,\delta'_5,\delta_6} := \frac{(-1)^{\delta'_2+\delta'_4+\delta'_5}(\delta'_5-1)!\big(\frac{1}{2}(\delta'_1+\delta'_3)+\delta_6-1\big)!\big(\frac{1}{2}(\delta'_2+\delta'_4)+\delta_6-1\big)!}{\delta'_1!\delta'_2!\delta'_3!\delta'_4!\big(\frac{1}{2}(\delta'_1+\delta'_2+\delta'_3+\delta'_4-2)\big)!\big(\frac{1}{2}(-\delta'_1-\delta'_2+\delta'_5)\big)!} \times \frac{1}{\big(\frac{1}{2}(-\delta'_3-\delta'_4+\delta'_5)\big)!\big(-\frac{1}{2}\delta'_1-\frac{1}{2}\delta'_3+\delta_6\big)!\big(-\frac{1}{2}\delta'_2-\frac{1}{2}\delta'_4+\delta_6\big)!}.$$

Performing the same sequence of operations on the sum over δ'_5 yields

(A-16)
$$N^{\mathrm{loc},\psi}_{\delta_1,\dots,\delta_6}(Y(D)) = [e^{\sum_{i=1}^{4}\delta_i t_i}] \sum_{\delta'_i}^{\infty} \mathsf{S}^{[2]}_{\delta'_1,\dots,\delta'_4,\delta_5,\delta_6} \prod_{i=1}^{4} y_i(t)^{\frac{1}{2}\sigma_i\delta'_i},$$

where

$$\mathsf{S}^{[2]}_{\delta'_1,\ldots,\delta'_4,\delta_5,\delta_6} := \frac{(-1)^{\delta'_1+\delta'_4}\big(\frac{1}{2}(\delta'_1+\delta'_2)+\delta_5-1\big)!\big(\frac{1}{2}(\delta'_3+\delta'_4)+\delta_5-1\big)!\big(\frac{1}{2}(\delta'_1+\delta'_3)+\delta_6-1\big)!}{\delta'_1!\delta'_2!\delta'_3!\delta'_4!\big(\big(\frac{1}{2}(\delta'_1+\delta'_2+\delta'_3+\delta'_4-2)\big)!\big)^2\big(-\frac{1}{2}\delta'_1-\frac{1}{2}\delta'_2+\delta_5\big)!} \times \frac{\big(\frac{1}{2}(\delta'_2+\delta'_4)+\delta_6-1\big)!}{\big(-\frac{1}{2}\delta'_3-\frac{1}{2}\delta'_4+\delta_5\big)!\big(-\frac{1}{2}\delta'_1-\frac{1}{2}\delta'_3+\delta_6\big)!\big(-\frac{1}{2}\delta'_2-\frac{1}{2}\delta'_4+\delta_6\big)!}. \tag{A-17}$$

The final step is to now plug in the mirror maps (A-8) for $i = 1, \ldots, 4$. This gives

$$N^{\mathrm{loc},\psi}_{\delta_1,\ldots,\delta_6}(Y(D)) = \sum_{j_1,\ldots,j_4=0}^{\infty} \mathsf{S}^{[3]}_{\delta'_1+2j_1,\ldots,\delta'_4+2j_4,j_1,\ldots,j_4,\delta_5,\delta_6}, \tag{A-18}$$

where

$$\mathsf{S}^{[3]}_{\delta'_1,\ldots,\delta'_4,j_1,\ldots,j_4,\delta_5,\delta_6} := \mathsf{S}^{[2]}_{\delta'_1,\ldots,\delta'_4,\delta_5,\delta_6} \prod_{i=1}^{4} \binom{\delta'_i}{j_i}. \tag{A-19}$$

The change of basis $\{C_1, \ldots, C_6\} \to \{H - E_1 - \cdots - E_5, E_1, \ldots E_5\}$ in (A-3) and the corresponding change of variables in the curve degree parameters $\{\delta_1, \ldots, \delta_6\} \to \{d_0, \ldots, d_5\}$ finally leads to (3-21).

Appendix B Infinite scattering

We compute the invariants of Conjecture 4.7 for the geometries $\mathrm{dP}_1(0,4)$ and $\mathbb{F}_0(0,4)$. This application of our correspondences predicts new relations for q–hypergeometric sums in Conjecture B.3. We provide calculations by picture and leave the details to the reader.

Denote by E the exceptional divisor obtained by blowing up a point on D_1 in $\mathbb{P}^2(1,4)$. We write a curve class $d \in \mathrm{H}_2(\mathrm{dP}_1(0,4),\mathbb{Z})$ as $d = d_0(H-E) + d_1 E$. If $d_0 = 0$ or $d_1 = 0$, then the moduli space of stable log maps is empty and $\mathsf{N}^{\log}_d(\mathrm{dP}_1(0,4))(\hbar) = 0$. If $d_1 > d_0$, then there are no irreducible curve classes and $\mathsf{N}^{\log}_d(\mathrm{dP}_1(0,4))(\hbar) = 0$. The toric model of $\mathrm{dP}_1(0,4)$ is obtained from the toric model of $\mathbb{P}^2(1,4)$ by adding a focus–focus singularity in the direction of D_1. The opposite primitive vectors in the F_2 and D_1 directions are $\gamma_1 = (1,0)$ and $\gamma_2 = (-1,-2)$. Since the absolute value of their determinant is 2 and not 1, there is infinite scattering, which is described in Section 4.2. By choosing our broken lines to be sufficiently into the x–direction, we can restrict to walls that lie on the halfspace $x > 0$. Then these walls have slope $(n+1)\gamma_1 + n\gamma_2 = (1,-2n)$ for $n \geq 0$. The wallcrossing functions attached to them are $1 + t^{n+1}t_1^n x^{-1} y^{2n}$. The broken line computation is summarised in Figure 20.

Theorem B.1 *Let $d_0 > d_1 \geq 1$ and $d = d_0(H-E) + d_1 E$. Then $\mathsf{N}^{\log}_d(\mathrm{dP}_1(0,4))(\hbar)$ equals*

$$\sum_{m=1}^{d_1} \sum \begin{bmatrix} 2d_0 \\ k_1 \end{bmatrix}_q \begin{bmatrix} 2d_0 - 2(n_1-n_2)k_1 \\ k_2 \end{bmatrix}_q \cdots \begin{bmatrix} 2d_0 - 2\sum_{j=1}^{m-1}(n_j-n_m)k_j \\ k_m \end{bmatrix}_q \begin{bmatrix} 2d_1 \\ k_0 \end{bmatrix}_q, \tag{B-1}$$

the second summation being over $k_0 \geq 0$, $k_1, \ldots, k_m > 0$ and $n_1 > n_2 > \cdots > n_m > 0$ satisfying $k_0 + \sum_{j=1}^{m} k_j = d_1$ and $\sum_{j=1}^{m} n_j k_j = d_0 - d_1$.

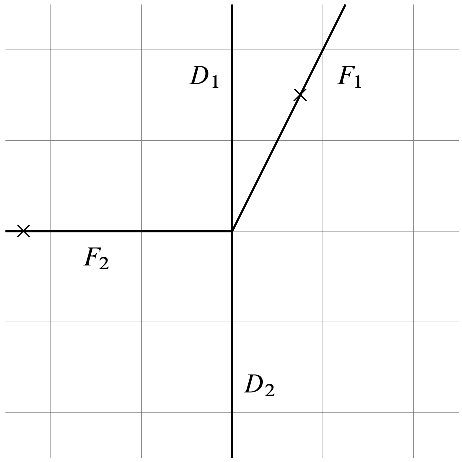


Figure 19: $\mathbb{F}_0(0,4)$.

For the case of $\mathbb{F}_0(0,4)$, let D_1 be a line of bidegree $(1,0)$ and let D_2 be a smooth divisor of bidegree $(1,2)$. Let d be a curve class of bidegree (d_1,d_2). We have $d\cdot D_1=d_2$ and $d\cdot D_2=2d_1+d_2$. Denote by pt_1 (resp. pt_2) their intersection points and by L_1 (resp. L_2) the lines of bidegree $(0,1)$ passing through pt_1 (resp. pt_2). We blow up pt_1 and pt_2, leading to exceptional divisors F_1 and F_2, and blow down the strict transforms of L_1 and L_2. The result is the Hirzebruch surface $\mathbb{F}_2$ with a focus–focus singularity on each of the fibrewise toric divisors, as in Figure 19.

Let d be a curve class of bidegree (d_1,d_2). The opposite primitive vectors in the F_2 and F_1 directions are $\gamma_1=(1,0)$ and $\gamma_2=(-1,-2)$. The absolute value of their determinant is 2, so there is infinite scattering as described in Section 4.2. We choose p to be in the lower left quadrant with coordinate (a,b) for $-1\ll a<0$ and $b\ll 0$. This depends on the degree and ensures that the broken lines are vertical at p. In particular, we can restrict to walls that lie on the halfspace $x<0$. Then these walls have slope $(n-1)\gamma_1+n\gamma_2=(-1,-2n)$ for $n\geq 1$. The wall-crossing functions attached to them are $1+t^{n-1}t_1^n xy^{2n}$. The broken line calculation is summarised in Figure 20.

Theorem B.2 *For $d_1\geq 1$, the generating function $\mathsf{N}^{\log}_{(d_1,d_2)}(\mathbb{F}_0(0,4))(\hbar)$ equals*

$$\sum_{m=1}^{\lfloor\frac{1}{2}(\sqrt{1+8d_1}-1)\rfloor}\ \sum_{\substack{d_1=\sum_{j=1}^m n_jk_j,\\ k_m,\dots,k_1>0,\\ n_m>\dots>n_1>0}} \begin{bmatrix} d_2+2d_1\\ k_m\end{bmatrix}_q \cdots \begin{bmatrix} d_2+2n_i\sum_{j=i}^m k_j+2\sum_{j=1}^{i-1}n_jk_j\\ k_i\end{bmatrix}_q \cdots \begin{bmatrix} d_2+2n_2\sum_{j=2}^m k_j+2n_1k_1\\ k_2\end{bmatrix}_q \begin{bmatrix} d_2+2n_1\sum_{j=1}^m k_j\\ k_1\end{bmatrix}_q \begin{bmatrix} d_2\\ \sum_{j=1}^m k_j\end{bmatrix}_q.$$

Conjecture 4.4 predicts that the multivariate q–hypergeometric sums of Theorems B.1 and B.2 dramatically simplify to remarkably compact q–binomial expressions. This is expressed by the following new conjectural q–binomial identities.

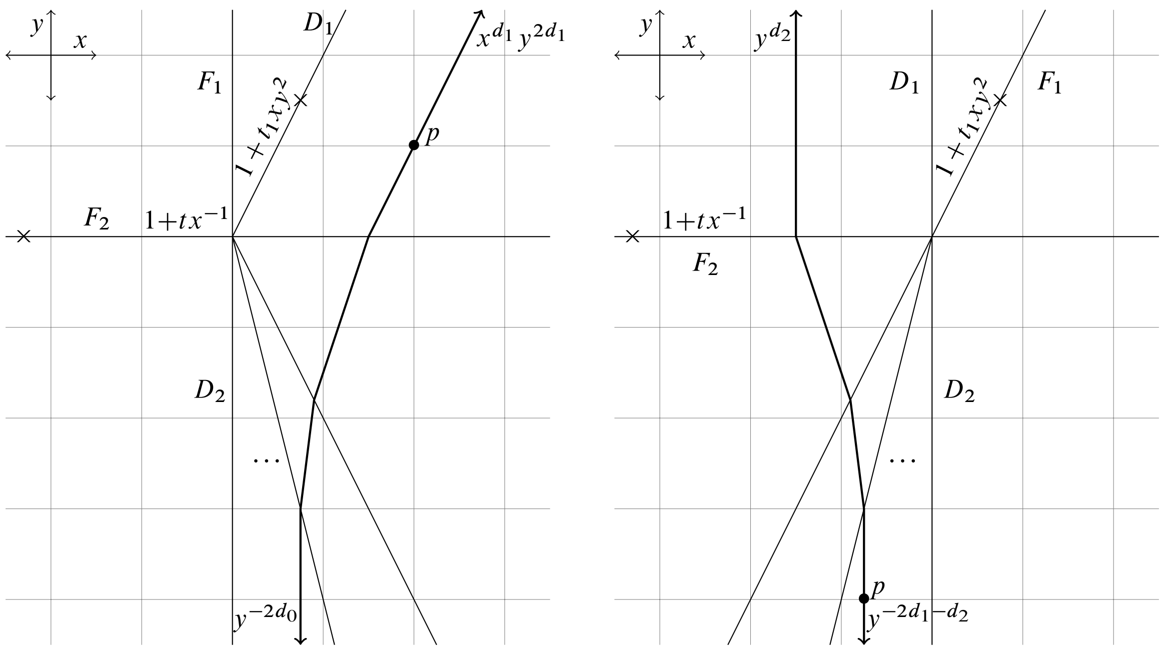


Figure 20: Scattering diagrams of $\mathrm{dP}_1(0,4)$, left, and $\mathbb{F}_0(0,4)$, right.

Conjecture B.3 *The q–hypergeometric sums of Theorems B.1 and B.2 are equal to*

$$\mathsf{N}_d^{\log}(\mathrm{dP}_1(0,4))(\hbar) = \frac{[2d_0]_q}{[d_0]_q}\begin{bmatrix} d_0 \\ d_1 \end{bmatrix}_q \begin{bmatrix} d_0+d_1-1 \\ d_0 \end{bmatrix}_q, \tag{B-2}$$

$$\mathsf{N}_d^{\log}(\mathbb{F}_0(0,4))(\hbar) = \frac{[2d_1+d_2]_q}{[d_2]_q}\begin{bmatrix} d_1+d_2-1 \\ d_1 \end{bmatrix}_q^2, \tag{B-3}$$

where $q = \mathrm{e}^{\mathrm{i}\hbar}$.

A proof of the identities of Conjecture B.3 was communicated to us by C Krattenthaler [71]. Note that the genus-zero log-local correspondence of Theorem 5.1 and the deformation invariance of local Gromov–Witten invariants give an entirely geometric proof of their classical limit at $q=1$.

Appendix C Proof of Theorem 5.4

Recall the notation of Section 5 and let $h\colon \Gamma \to \Delta$ be a rigid decorated parametrised tropical curve with $N_{0,d}^{\mathrm{loc},h}(Y(D)) \neq 0$. Our goal is to prove that $h = \overline{h}$. This will be done by a series of Lemmas constraining further and further the possible shape of h.

Recall that we are considering a degeneration with special fibre $\boldsymbol{P}_0^h$, which is a $(\mathbb{P}^1)^l$–bundle over the special fibre $\mathcal{Y}_0^h$ of a degeneration of the original surface Y. For every vertex V (resp. edge E) of Γ, the corresponding component (resp. node) of a stable log map with tropicalisation h maps to the irreducible component $\boldsymbol{P}_V$ (resp. divisor $\boldsymbol{P}_E$) of the special fibre $\boldsymbol{P}_0^h$. $\boldsymbol{P}_V$ (resp. $\mathbb{P}_E$) is a $(\mathbb{P}^1)^l$–bundle over a component (resp. divisor) Y_V (resp. Y_E) of $\mathcal{Y}_0^h$.

We are considering stable log maps to $\boldsymbol{P}_0^h$ with $l-1>0$ marked points mapping to the interior of $\boldsymbol{P}_{0,Y}$. So irreducible components containing these marked points map to $\boldsymbol{P}_{0,Y}$, and the corresponding vertices of Γ are mapped to v_Y by h. Hence, we can choose a flow on Γ such that unbounded edges are incoming, such that every vertex has at most one outgoing edge, and such that the sink V_0 satisfies $h(V_0)=v_Y$. For every vertex $V\neq V_0$, we denote by E_V the edge outgoing from V. Following the flow, the maps η_V define a cohomology class $\alpha_E\in \mathrm{H}^*(\boldsymbol{P}_E)$ for every edge E of Γ. The degeneration formula can be rewritten as

$$N_{0,d}^{\mathrm{loc},h}(Y(D))=\eta_{V_0}\biggl(\prod_{E\in\mathcal{E}_{\mathrm{in}}(V_0)}\alpha_E\biggr). \tag{C-1}$$

When used below, "descendant" and "ancestor" always refer to the ordering on the vertices of Γ induced by the flow: a vertex V is "older" than a vertex V' if the flow goes from V to V'.

The proof below consists of three steps. First, in Section C.1, we constrain the form of Γ near the boundary $\partial\Delta$ of the tropicalisation. Then we study the local structure of Γ near the vertex v_Y in Section C.2. Finally, in Section C.3, we combine together the local information obtained near the boundary and near v_Y to obtain global control on Γ.

C.1 Study near the boundary $\partial\Delta$

Recall from (5-8) that the boundary $\partial\Delta$ of Δ is the union of segments $(\partial\Delta)_j$ indexed by $1\le j\le l$.

Most of the analysis below involves the cohomology classes $\alpha_E\in H^\star(\mathbb{P}_E)$ recursively attached by the flow to the edges E of Γ. Geometrically, the class α_E captures the constraints on the position of the node dual to the edge E imposed by the ability to glue together the curve components corresponding to vertices coming before E in the flow. For every edge E of Γ, we denote by $H_{j,E}\in \mathrm{H}^2(\boldsymbol{P}_E)$ the first Chern class of the tautological line bundle $\mathcal{O}_{\mathbb{P}(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_E}\oplus\mathcal{O}_{Y_E})}(1)$. Geometrically, to have α_E positively proportional to $H_{j,E}$ means that the node corresponding to E is constrained to be in the preimage by the natural projection $\boldsymbol{P}_E\to\mathbb{P}(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_E}\oplus\mathcal{O}_{Y_E})$ of a section of $\mathbb{P}(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_E}\oplus\mathcal{O}_{Y_E})$.

We will use below the following facts on the classes $H_{j,E}$. We have $H_{j,E}^2=-c_1(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_E})H_{j,E}$. If $h(E)\not\subset(\partial\Delta)_j$, then $\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_E}=\mathcal{O}_{Y_E}$ and so $H_{j,E}^2=0$. If $h(E)\subset(\partial\Delta)_j$, then $\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_E}=\mathcal{O}(-1)$, and so $H_{j,E}^2=(\pi_E^*\mathrm{pt}_E)H_{j,E}$. If V is a vertex of Γ with $h(V)\in(\partial\Delta)_j$ and $d_V\cdot D_{j,V}^\partial>0$ for some $1\le j\le l$, then, as the line bundle $\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_V}=\mathcal{O}_{Y_V}(-D_{j,V}^\partial)$ has negative degree in restriction to the curve corresponding to V, this curve is constrained to lie in the zero section of $\mathbb{P}(\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_V}\oplus\mathcal{O}_{Y_V})$, and so α_{E_V} is a nonzero multiple of H_{j,E_V}.

Lemma C.1 *Let V be a vertex of Γ with $h(V)\in(\partial\Delta)_j$ for some $1\le j\le l$. Then we have $d_V\cdot D_{j,V}^\partial>0$ if and only if there is an edge E of Γ incident to V such that $h(E)\not\subset(\partial\Delta)_j$.*

Figure 21: Left: toric fan of Y_V for $V \in (\partial\Delta)_j - \{v_j\}$ obtained by adding rays in the lower part of the toric fan of $\mathbb{P}^1 \times \mathbb{P}^1$ (in thick). Right: toric fan of Y_V for $V = v_j$ obtained by adding rays in the lower part of the toric fan of $\mathbb{F}_{D_j^2}$ (in thick).

Proof First assume that $h(V) \neq v_j$. Then Y_V can be described as a toric blowup of $\mathbb{P}^1 \times \mathbb{P}^1$, where all the added rays are contained in the lower half-plane of the fan, and where the vertical ray corresponds to $D^{\partial}_{j,V}$; see Figure 21, left. The lower part of the fan gives a local picture of Δ^h near $h(V)$. By definition of the Δ^h, every edge E of Γ incident to V is mapped by h to one of the rays in the lower part of the fan. We have $h(E) \not\subset (\partial\Delta)_j$ if and only if E is contained in one of the rays in the strict lower part of the fan. The result then follows from toric homological balancing.

If $h(V) = v_j$, the argument is similar. Recall that we have $D_j \simeq \mathbb{P}^1$. The key point is that D_j is nef and so $D_j^2 \geq 0$. Therefore, Y_V can be described as a toric blowup of the Hirzebruch surface $\mathbb{F}_{D_j^2}$, where all the added rays are contained in the lower half-part of the fan, and where the vertical ray, with self-intersection D_j^2, corresponds to $D^{\partial}_{j,V}$; see Figure 21, right. The lower part of the fan gives a local picture of Δ^h near $h(V)$. By definition of the Δ^h, every edge E of Γ incident to V is mapped by h to one of the rays in the lower part of the fan. We have $h(E) \not\subset (\partial\Delta)_j$ if and only if $h(E)$ is contained in on of the rays in the strict lower part of the fan. As $D_j^2 \geq 0$, the lower part of the fan is convex and so the result follows from toric homological balancing. □

Lemma C.2 *Let V be a vertex of Γ with $h(V) \in (\partial\Delta)_j$. Assume that there exists an incoming edge E incident to V such that α_E is a nonzero multiple of $H_{j,E}$. Then $h(E_V) \subset (\partial\Delta)_j$ and α_{E_V} is a nonzero multiple of H_{j,E_V}.*

Proof If $h(E_V) \not\subset (\partial\Delta)_j$, then, by Lemma C.1, we have $d_V \cdot D^{\partial}_{j,V} > 0$ and so α_{E_V} is proportional to $H^2_{j,E_V} = 0$. Therefore, $\alpha_{E_V} = 0$, in contradiction with the assumption $N_d^{\mathrm{loc},h}(Y(D)) \neq 0$, and so this does not happen. Hence, we can assume that $h(E_V) \subset (\partial\Delta)_j$. If $d \cdot D^{\partial}_{j,V} > 0$, then α_{E_V} is a multiple of $H^2_{j,E_V} = (\pi^*_{E_V} \mathrm{pt}_{E_V}) H_{j,E_V}$. If $d \cdot D^{\partial}_{j,V} = 0$, then α_{E_V} is a multiple of H_{j,E_V}. □

Lemma C.3 *Let V be a vertex of Γ with $V \neq V_0$ and an incident incoming edge E with α_E a nonzero multiple of $H_{j,E}$. Then α_{E_V} is a nonzero multiple of H_{j,E_V}.*

Proof If $h(V) \in (\partial\Delta)_j$, then the result follows from Lemma C.2. If $h(V) \notin (\partial\Delta)_j$, then the result is clear as the line bundle $\mathcal{O}_{\mathcal{Y}}(-\mathcal{D}_j)|_{Y_V}$ is trivial. □

Lemma C.4 *Let V be a vertex of Γ such that $V \in (\partial\Delta)_j$ and such that there exists an incoming edge E incident to V with $h(E) \not\subset (\partial\Delta)_j$. Then α_{E_V} is a nonzero multiple of H_{j,E_V}, and we have $h(E_V) \not\subset (\partial\Delta)_j$.*

Proof By Lemma C.1, we have $d_V \cdot D^{\partial}_{j,V} > 0$, and so α_{E_V} is a nonzero multiple of H_{j,E_V}. If we had $h(E_V) \subset (\partial\Delta)_j$, then by iterative application of Lemma C.2, all the descendants of V would be mapped by h to $(\partial\Delta)_j$, in contradiction with the fact that the sink V_0 of Γ is mapped by h to v_Y. □

We say that a vertex V of Γ is a source if every bounded edge incident to V is outgoing. As we are assuming that every vertex of Γ has at most one outgoing edge, a source has a unique bounded incident edge. For a vertex V of Γ such that $h(V) \in \Delta - \partial\Delta - \{v_Y\}$, the toric balancing condition holds at $h(V)$. As the toric balancing condition cannot hold at a vertex with a unique incident bounded edge, we deduce that if V is a source of Γ, then either $h(V) = v_Y$ or $h(V) \in \partial\Delta$.

Lemma C.5 *Let V be a vertex of Γ such that V is a source and $h(V) \in \partial\Delta$. Then there exists $1 \le j \le l$ such that α_{E_V} is a nonzero multiple of H_{j,E_V}.*

Proof We know that $h(V) \in (\partial\Delta)_j$ for at least one j. Assume first that $h(V) \neq v_p$ for every $p \in D_j \cap D_{j'}$, that is, $h(V) \in (\partial\Delta)_j$ for a unique j. As V is a source, there is a single edge incident to V. By homological toric balancing (see Figure 21, left), this is possible only if $h(E_V)$ is contained in the ray opposite to the ray corresponding to $D^{\partial}_{j,V}$, and in particular we then have $d_V \cdot D^{\partial}_{j,V} > 0$.

It remains to treat the case where $h(V) = v_p$ for some $p \in D_j \cap D_{j'}$. In this case, we have $h(V) = v_p \in (\partial\Delta)_j \cap (\partial\Delta)_{j'}$. By homological toric balancing (see Figure 21, right), we necessarily have $d_V \cdot D^{\partial}_{k,V} > 0$ for some $k \in \{j, j'\}$. □

Lemma C.6 *Let V be a vertex of Γ such that V is a source, $h(V) \in (\partial\Delta)_j$ for some $1 \le j \le l$, and $h(V) \neq v_p$ for every $p \in D_j \cap D_{j'}$. Then d_V is a multiple of the class of a $\mathbb{P}^1$–fibre of Y_V and $h(E_V) \not\subset (\partial\Delta)_j$.*

Proof Similar to the proof of Lemma C.5. □

C.2 Study near the centre v_Y

Lemma C.7 *Let E be a bounded edge of Γ such that α_E is not a nonzero multiple of any $H_{j,E}$. Then we have $E = E_V$, where V is a source of Γ with $h(V) = v_Y$.*

Proof For a source V of Γ, we have either $h(V) = v_Y$ or $h(V) \in \partial\Delta$. If one of the source ancestors V of E had $h(V) \in \partial\Delta$, we would have, by combination of Lemma C.5 and Lemma C.3, that α_E is a nonzero multiple of $H_{j,E}$ for some $1 \le j \le l$. Therefore, for every source V that is an ancestor of E, we have $h(V) = v_Y$.

Assume by contradiction that there are at least two distinct source ancestors of E. Then there exists a vertex V which is an ancestor of E where at least two distinct source edges meet. As the source edges are emitted by sources mapped to v_Y by h, they can only meet if their images by h are contained in a common half-line in Δ with origin v_Y. If $h(V) \in (\partial\Delta)_j$ for some j, then α_{E_V}, and so α_E by Lemma C.3, would have been a nonzero multiple of $H_{j,E}$ by Lemma C.3. Therefore, $h(V) \in \Delta - \partial\Delta$. On the other hand, we have $h(V) \neq v_Y$. Therefore, the toric balancing condition applies at $h(V)$ and $h(E_V)$ is parallel to the direction of the incoming edges. Moving $h(V)$ along the common direction of all the edges incident to V produces a contradiction with the assumed rigidity of h.

Therefore, E admits a unique ancestor source V. So any other vertex of Γ along the flow from V to E would have to be a 2–valent vertex, in contradiction with the rigidity of h. We conclude that $E = E_V$. □

From now on, we assume that $l = 2$. In this case, Γ has a unique unbounded edge, and we choose the flow such that V_0 is the vertex V of Γ incident to this unbounded edge.

Lemma C.8 *The set of bounded edges of Γ incident to V_0 consists of two elements E_1 and E_2 with $\alpha_{E_1} = \lambda_1 H_{1,E_1}$ and $\alpha_{E_2} = \lambda_2 H_{2,E_2}$, where $\lambda_1, \lambda_2 \in \mathbb{Q} - \{0\}$.*

Proof It follows from Lemma C.7 that, for every bounded edge E incident to V_0, there exists a $1 \leq j \leq 2$ such that α_E is a nonzero multiple of $H_{j,E}$. As the moduli space M_{V_0} contains two $\mathbb{P}^1$–factors corresponding to the two extra directions $\mathcal{O}_Y^{\oplus 2}$, the condition $N_d^{\mathrm{loc},h}(Y(D)) \neq 0$ implies that for every $1 \leq j \leq 2$, there exists at least one bounded edge E incident to V_0 with α_E a nonzero multiple of $H_{j,E}$. As $H_1^2 = H_2^2 = 0$ on $\boldsymbol{P}_{V_0}$, for every $1 \leq j \leq 2$ there is at most one bounded edge incident to V_0 with α_E a nonzero multiple of $H_{j,E}$.

Therefore, we have two cases. Either the set of bounded edges incident to V_0 consists of one edge E with α_E a nonzero multiple of $H_{1,E} H_{2,E}$, or the set of bounded edges incident to V_0 consists of two edges E_1 and E_2 with α_{E_1} a nonzero multiple of $H_{1,E}$ but not of $H_{2,E}$, and α_{E_2} a nonzero multiple of $H_{2,E}$ but not $H_{1,E}$.

Let us show that the first case does not arise. If the set of bounded edges incident to V_0 consists of a single element, then the moduli space M_{V_0} has virtual dimension 2. Indeed, the virtual dimension of M_{V_0} is $0+2$, where 0 is the virtual dimension for rational curves in the log Calabi–Yau surface Y intersecting the boundary divisor D in a single point, and 2 comes from the two extra trivial directions $\mathcal{O}_Y^{\oplus 2}$. But we need to integrate over $[M_{V_0}]^{\mathrm{vir}}$ the pullbacks of the class $H_{1,E} H_{2,E}$ (coming from the bounded edge E incident to V_0) and the pullback of $\pi_{V_0}^* \mathrm{pt}_Y$ (coming from the unbounded edge incident to V_0). Therefore, the integrand is a class of degree at least $3 > 2$, and so this case does not arise if $N_d^{\mathrm{loc},h}(Y(D)) \neq 0$.

Thus, we are in the second case, where the set of bounded edges incident to V_0 consists of two edges E_1 and E_2 with α_{E_1} a nonzero multiple of $H_{1,E}$ but not of $H_{2,E}$, and α_{E_2} a nonzero multiple of $H_{2,E}$ but not $H_{1,E}$. In particular, the moduli space M_{V_0} has virtual dimension 3. Indeed, the virtual dimension of M_{V_0} is $1+2$, where 1 is the virtual dimension for rational curves in the log Calabi–Yau surface Y

intersecting the boundary divisor D in two points, and 2 comes from the two extra trivial directions $\mathcal{O}_Y^{\oplus 2}$. As we need to integrate over $[M_{V_0}]^{\mathrm{vir}}$ the pullbacks of the classes α_{E_1}, α_{E_2} and $\pi_{V_0}^* \mathrm{pt}_Y$, with $\deg \alpha_{E_1} \geq 1$ and $\deg \alpha_{E_2} \geq 1$, the condition $N_d^{\mathrm{loc},h}(Y(D)) \neq 0$ implies that $\deg \alpha_{E_1} = \deg \alpha_{E_2} = 1$ and so the classes α_{E_1} and α_{E_2} are scalar multiples of H_{1,E_1} and H_{2,E_2}, respectively. □

C.3 End of the proof

It follows from Lemma C.8 that there exists a unique decomposition

$$\Gamma = \Gamma_1 \cup \Gamma_2,$$

where Γ_1 and Γ_2 are connected subgraphs of Γ such that

(i) V_0 is a vertex of both Γ_1 and Γ_2,

(ii) if V is a vertex of Γ distinct from V_0, then V is a vertex of Γ_1 (resp. Γ_2) if and only if the flow starting at V ends at V_0 along the edge E_1 (resp. E_2).

As Γ is a graph of genus zero, the intersection $\Gamma_1 \cap \Gamma_2$ consists only of the common vertex V_0.

Lemma C.9 *For every $1 \leq j \leq 2$, there exists a unique vertex V_j of Γ such that $V_j \in \Gamma_j$ and $V_j \in \partial\Delta$. Moreover, $V_j \in (\partial\Delta)_j$ and $V_{j'} \notin (\partial\Delta)_{j'}$, where $\{j, j'\} = \{1, 2\}$.*

Proof By symmetry, we can assume $j = 1$ and $j' = 2$. We first remark that if V is a vertex of Γ_1 such that $V \in \partial\Delta$, then $V \in (\partial\Delta)_1$ and $V \notin (\partial\Delta)_2$. Otherwise, there would be a descendant V' of V_1 such that $h(V') \notin (\partial\Delta)_2$ and $h(E_{V'}) \not\subset (\partial\Delta)_2$, and so by Lemma C.1, $\alpha_{E_{V'}}$ would be a nonzero multiple of $H_{2,E_{V'}}$, and so by Lemma C.3, α_{E_1} would be a nonzero multiple of H_{2,E_1}, a contradiction.

As $\alpha_{E_1} = \lambda_1 H_{1,E_1}$ with $\lambda_1 \neq 0$, it follows from Lemma C.1 that there exists a vertex V_1 of Γ_1 such that $V_1 \in (\partial\Delta)_1$ and $h(E_{V_1}) \not\subset (\partial\Delta)_1$. Moreover, there exists a unique vertex with these properties: else, by Lemma C.3, α_{E_1} would be proportional to $H_{1,E_1}^2 = 0$, a contradiction. Our first remark applied to V_1 shows that $V_1 \notin (\partial\Delta)_2$.

It remains to show that if V is a vertex of Γ_1 such that $h(V) \in (\partial\Delta)_1 \setminus (\partial\Delta)_2$, then $V = V_1$. Assume by contradiction that there exists a vertex V of Γ_1 such that $h(V) \in \partial\Delta$ and $V \neq V_1$. Up to replacing V by one of its ancestors, we can assume that no ancestor of V is contained in $\partial\Delta$. There are now two cases. First, if V is a source, then, by Lemma C.6, $h(E_V) \not\subset (\partial\Delta)_1$, and so $V = V_1$ by the uniqueness of V_1, a contradiction. Second, if V is not a source, then there exists an edge E incident to V such that $h(E) \not\subset (\partial\Delta)_1$, and so by Lemma C.4, $h(E_V) \not\subset (\partial\Delta)_1$, and so $V = V_1$ by the uniqueness of V_1, a contradiction again. □

We now explain how to conclude the proof of Theorem 5.4, that is, show that $h = \bar{h}$. We say that an edge E of Γ is *radial* if $h(E) \not\subset \partial\Delta$ and the direction of $h(E)$ passes through v_Y. We claim that all edges of Γ_1 are radial. Indeed, let V be a vertex of Γ_1 such that $V \neq V_1$ and $h(V) \neq v_Y$. Then $h(V) \notin \partial\Delta$, so $H_{1,V}^2 = 0$, and so there exists at most one edge E incident to V such that α_E is a nonzero multiple of $H_{1,E}$. On the other hand, edges E such that α_E is not a nonzero multiple of $H_{1,E}$ are radial by Lemma C.7. As $h(V) \notin \partial\Delta$ and $h(V) \neq v_Y$, the toric balancing condition holds for V, and so if all incident edges to E except possibly one are radial, then they are in fact all radial.

As all edges of Γ_1 are radial, every vertex V of Γ_1 satisfies either $V = V_1$ or $h(V) = v_Y$: else, moving V in the radial direction would contradict the assumed rigidity of h. In other words, the graph Γ_1 has a very simple form: a vertex V_1 connected by some vertices V such as $h(V) = v_Y$. On the other hand, as all the edges through V_1 are radial, it follows from toric homological balancing that the curve class d_{V_1} is a multiple of the class of a $\mathbb{P}^1$–fibre of Y_{V_1} and that $h(V_1) = v_1$. In this context, the dimension argument of [43, Lemma 5.4] shows that a nonzero Gromov–Witten invariant is only possible if the curve component corresponding to the vertex V_1 has maximal tangency, that is, if there is a single edge incident to V_1. It follows in particular that V_0 is the single vertex of Γ_1 whose image by h is v_Y. Replacing Γ_1 by Γ_2 in the previous arguments, we finally obtain that $h = \overline{h}$. □

Appendix D Symmetric functions

D.0.1 Partitions and representations of S_n A partition $\lambda \vdash d$ of a nonnegative integer $d \in \mathbb{N}$ is a monotone nonincreasing sequence $\lambda := \{\lambda_i\}_{i=1}^r$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ such that $\sum_{i=1}^r \lambda_i = d$; when $d = 0$ we write $\lambda = \varnothing$ for the empty partition. We will often use the shorthand notation

$$\{\lambda_1^{n_1}, \dots, \lambda_k^{n_k}\} := \{\overbrace{\lambda_1, \dots, \lambda_1}^{n_1 \text{ times}}, \dots, \overbrace{\lambda_k, \dots, \lambda_k}^{n_k \text{ times}}\} \tag{D-1}$$

for partitions with repeated entries.

With notation as in the beginning of Section 6, a partition λ is bijectively associated to:

- A Young diagram Y_λ with $m_j(\lambda)$ rows of boxes of length j; there is a natural involution in the space of partitions, $\lambda \to \lambda^t$, given by transposition of the corresponding Young diagram.
- A conjugacy class $C_\lambda \in \mathrm{Conj}(S_{|\lambda|})$ of the symmetric group $S_{|\lambda|}$ with automorphism group of order $|\mathrm{Aut}_{C_\lambda}| = |\lambda|!\, z_\lambda$, with
$$z_\lambda := \prod_j m_j(\lambda)!\, j^{m_j(\lambda)}.$$
- An irreducible representation $\rho_\lambda \in \mathrm{Rep}(S_d)$. For $\eta \in \mathrm{Conj}(S_d)$, we write $\chi_\lambda(\eta)$ for the irreducible character $\mathrm{Tr}_{\rho_\lambda}(\eta)$.
- By Schur–Weyl duality, an irreducible representation $R_\lambda \in \mathrm{Rep}(\mathrm{GL_n}(\mathbb{C}))$ for $n \geq \ell_\lambda$.

We will be concerned with two linear bases of the ring of integral symmetric polynomials in n variables, $\Lambda_n := \mathbb{Z}[x_1, \dots, x_n]^{S_n}$, labelled by partitions with $\ell(\lambda) \leq n$. Write $x := (x_1, \dots, x_n)^{S_n} \in \mathbb{C}^n / S_n$ for an orbit x of the adjoint action of $\mathrm{GL}_n(\mathbb{C})$ (equivalently, the Weyl group action on $\mathbb{C}^n$), and g_x for any element of the orbit. We write

$$p_\lambda(x) := \prod_i \mathrm{Tr}_{\mathbb{C}^n}\, g_x^{m_i(\lambda)} \quad \text{and} \quad s_\lambda(x) := \mathrm{Tr}_{R_\lambda}(g_x) \tag{D-2}$$

for, respectively, the symmetric power function and the Schur function determined by λ; we have $\Lambda_n = \mathrm{span}_{\mathbb{Z}}\{p_\lambda\}_{\{\lambda\in\mathcal{P},\ell_\lambda\le n\}} = \mathrm{span}_{\mathbb{Z}}\{s_\lambda\}_{\{\lambda\in\mathcal{P},\ell\lambda\le n\}}$. These two bases are related as

$$s_\mu(x) = \sum_{|\lambda|=|\mu|} \frac{\chi_\mu(\lambda)}{z_\lambda} p_\lambda(x) \quad \text{and} \quad p_\mu(x) = \sum_{|\lambda|=|\mu|} \chi_\mu(\lambda) s_\lambda(x). \tag{D-3}$$

For λ, μ a pair of partitions, the skew Schur polynomials $s_{\lambda/\mu}(x)$ are defined by

$$s_{\lambda/\mu}(x) = \sum_{\nu\in\mathcal{P}} \mathsf{LR}^{\lambda}_{\mu\nu} s_\nu(x), \tag{D-4}$$

where $\mathsf{LR}^{\lambda}_{\mu\nu}$ are the Littlewood–Richardson coefficients $R_\mu \otimes R_\nu =: \bigoplus_{\lambda\vdash(|\mu|+|\nu|)} \mathsf{LR}^{\lambda}_{\mu\nu} R_\lambda$.

Let $\rho\colon \Lambda_n \to \Lambda_{n+1}$ be the monomorphism of rings defined by $\rho(p_{(i)}(x_1,\dots,x_n)) = p_{(i)}(x_1,\dots,x_{n+1})$. We define the ring of symmetric functions $\Lambda := \varinjlim \Lambda_n$ as the direct limit under these inclusions, and denote by the same symbols p_λ, s_λ and $s_{\lambda/\mu}$ the symmetric functions obtained as the images of the power sums, Schur polynomials and skew Schur polynomials under the direct limit. In the next sections it will be of importance to formally expand the infinite product $\prod_{i,j}(1-x_i y_j) \in \Lambda \otimes_{\mathbb{Z}} \Lambda$ around $(x,y)=(0,0)$, and it is a classical result in the theory of symmetric functions out that this expansion can be cast in multiple ways in terms of an average over partitions of bilinear expressions of linear generators of Λ. In particular, we have the Cauchy identities

$$\sum_{\lambda\in\mathcal{P}} s_\lambda(x)s_\lambda(y) = \prod_{i,j}(1-x_i y_j)^{-1} \quad \text{and} \quad \sum_{\lambda\in\mathcal{P}} s_\lambda(x)s_{\lambda^t}(y) = \prod_{i,j}(1+x_i y_j). \tag{D-5}$$

A skew generalisation of these [80, Section I.5] is

$$\begin{aligned} \sum_{\lambda\in\mathcal{P}} s_{\lambda/\mu}(x)s_{\lambda/\nu}(y) &= \prod_{i,j}(1-x_i y_j)^{-1} \sum_{\eta\in\mathcal{P}} s_{\nu/\eta}(x)s_{\mu/\eta}(y), \\ \sum_{\lambda\in\mathcal{P}} s_{\lambda^t/\mu}(x)s_{\lambda/\nu}(y) &= \prod_{i,j}(1+x_i y_j) \sum_{\eta\in\mathcal{P}} s_{\nu^t/\eta}(x)s_{\mu^t/\eta^t}(y). \end{aligned} \tag{D-6}$$

Another noteworthy sum we will need is [80, Section I.5]

$$\sum_{\delta\in\mathcal{P}} s_{\lambda/\delta}(x)s_{\delta/\nu}(y) = s_{\lambda/\mu}(x,y), \tag{D-7}$$

where $s_{\lambda/\mu}(x,y)$ denotes the skew Schur function in the variables $(x_1,x_2,\dots,x_i,\dots,y_1,y_2,\dots,y_i,\dots)$.

D.0.2 Shifted symmetric functions and the principal stable specialisation From these ingredients and $\mu \in \mathcal{P}$, we define a class of Laurent series of a single variable $q^{1/2}$ obtained by the *principal stable specialisation*

$$\mathsf{q}\colon \Lambda \to \mathbb{Q}[\![q^{-1/2}]\!], \quad f(x_1,\dots,x_i,\dots) \mapsto f(x_1 = q^{-1+1/2},\dots,x_n = q^{-i+1/2},\dots). \tag{D-8}$$

As is customary in the topological vertex literature, and since $-i+\frac{1}{2}$ is the component of the Weyl vector ρ of A_n with respect to the fundamental weight ω_{n-i}, we use the shorthand notation $f(q^\rho) := f(x_i = q^{-i+1/2})$. For f a power sum or Schur function, $f(q^\rho)$ converges to a rational function of $q^{1/2}$.

In particular,

$$p_{(d_1,\dots,d_n)}(q^\rho) = \prod_{i=1}^{n} \frac{1}{[d_i]_q}, \tag{D-9}$$

and, for Schur functions, Stanley [110; 111] proved the product formula

$$s_\lambda(q^\rho) = \frac{q^{\frac{1}{4}\kappa(\lambda)}}{\prod_{(i,j)\in\lambda}[h(i,j)]}, \tag{D-10}$$

where $h(i,j)$ is the number of squares directly below or to the right of a cell (i,j) (counting (i,j) once) in the Young diagram of λ. For example, when $\lambda = (i-j, 1^j)$ is a hook Young diagram with i boxes and $j+1$ rows, this gives

$$s_{(i-j,1^j)}(q^\rho) = \frac{q^{\frac{1}{2}(\binom{i}{2}-ij)}}{[i]_q[i-j-1]_q![j]_q!}. \tag{D-11}$$

More generally, for $\mu \in \mathcal{P}$ we will consider the shifted power, Schur and skew Schur functions,

$$\begin{aligned} p_\lambda(q^{\rho+\mu}) &:= p_\lambda(x_i = q^{-i+\mu_i+1/2}), \\ s_\lambda(q^{\rho+\mu}) &:= s_\lambda(x_i = q^{-i+\mu_i+1/2}), \\ s_{\lambda/\delta}(q^{\rho+\mu}) &:= \sum_{\nu\in\mathcal{P}} \mathsf{LR}^{\lambda}_{\delta\nu} s_\nu(q^{\rho+\mu}). \end{aligned} \tag{D-12}$$

The identities

$$s_\lambda(q^\rho) = q^{\kappa(\rho)/2} s_{\lambda^t}(q^\rho), \tag{D-13}$$

$$s_{\lambda/\mu}(q^{\rho+\alpha}) = s_{\lambda^t/\mu^t}(-q^{-\rho-\alpha^t}) \tag{D-14}$$

follow easily from (D-11), (D-12) and the fact that Littlewood–Richardson coefficients are invariant under simultaneous transposition of their arguments. Following [63], we introduce the following notation for the Cauchy infinite products (D-5) in the principal stable specialisation:

$$\begin{aligned} \{\alpha,\beta\}_Q := \prod_{i,j\geq 1}(1 - Qq^{-i-j+1+\alpha_i+\beta_i}) &= \sum_{\lambda\in\mathcal{P}} s_\lambda(q^{\rho+\alpha}) s_{\lambda^t}(-Qq^{\rho+\beta}) \\ &= \left[\sum_{\lambda\in\mathcal{P}} s_\lambda(q^{\rho+\alpha}) s_\lambda(Qq^{\rho+\beta})\right]^{-1}. \end{aligned} \tag{D-15}$$

Finally, we will need to specialise expressions involving skew Schur functions and Cauchy products to the case of hook Young diagrams. These can be given closed-form q–factorial expressions, as follows.

Lemma D.1 *We have*

$$\mathsf{LR}^{(i-r,1^r)}_{\beta,\gamma} = \begin{cases} \delta_{i,j+k}(\delta_{r,s+t} + \delta_{r,s+t+1}) & \text{if } \beta = (j-s,1^s) \text{ and } \gamma = (k-t,1^t), \\ 0 & \text{else.} \end{cases} \tag{D-16}$$

Moreover,

(D-17) $$s_{(i-j,1^j)/\gamma}(q^\rho) = \begin{cases} \dfrac{q^{\frac{1}{4}(i-k-1)(i-2j-k+2l)}}{[i-k+l-j]_q![j-l]_q!} & \text{if } \gamma = (k-l, 1^l), \\ 0 & \text{else}, \end{cases}$$

(D-18) $$\frac{\{(i-j,1^j),\varnothing\}_Q}{\{\varnothing,\varnothing\}_Q} = \prod_{k=0}^{i-1}(1-q^k Q q^{-j}) = (Qq^{-j};q)_i.$$

The content of the lemma follows from a straightforward application of the Littlewood–Richardson rule in the case of hook partitions $(i-r, 1^r)$. The product formula[15] for the hook skew-Schur functions (D-17) follows then immediately from (D-11). Finally, (D-18) follows from a straightforward calculation from (D-15); see [63, Section 3.4] for details.

References

[1] **D Abramovich**, **Q Chen**, *Stable logarithmic maps to Deligne–Faltings pairs, II*, Asian J. Math. 18 (2014) 465–488 MR Zbl

[2] **D Abramovich**, **Q Chen**, **M Gross**, **B Siebert**, *Decomposition of degenerate Gromov–Witten invariants*, Compos. Math. 156 (2020) 2020–2075 MR Zbl

[3] **D Abramovich**, **J Wise**, *Birational invariance in logarithmic Gromov–Witten theory*, Compos. Math. 154 (2018) 595–620 MR Zbl

[4] **M Aganagic**, **C Beem**, *The geometry of D–brane superpotentials*, J. High Energy Phys. (2011) 060, 25 MR Zbl

[5] **M Aganagic**, **R Dijkgraaf**, **A Klemm**, **M Mariño**, **C Vafa**, *Topological strings and integrable hierarchies*, Comm. Math. Phys. 261 (2006) 451–516 MR Zbl

[6] **M Aganagic**, **A Klemm**, **M Mariño**, **C Vafa**, *The topological vertex*, Comm. Math. Phys. 254 (2005) 425–478 MR Zbl

[7] **M Aganagic**, **C Vafa**, *Mirror symmetry*, *D–branes and counting holomorphic discs*, preprint (2000) arXiv hep-th/0012041

[8] **D Auroux**, *Mirror symmetry and T–duality in the complement of an anticanonical divisor*, J. Gökova Geom. Topol. 1 (2007) 51–91 MR Zbl

[9] **L J Barrott**, *Explicit equations for mirror families to log Calabi–Yau surfaces*, Bull. Korean Math. Soc. 57 (2020) 139–165 MR Zbl

[10] **F Block**, **L Göttsche**, *Refined curve counting with tropical geometry*, Compos. Math. 152 (2016) 115–151 MR Zbl

[11] **D Borisov**, **D Joyce**, *Virtual fundamental classes for moduli spaces of sheaves on Calabi–Yau four-folds*, Geom. Topol. 21 (2017) 3231–3311 MR Zbl

[15] Unlike the Schur case of (D-10), closed q–formulas for principally specialised skew-Schur functions are generally difficult to find, and (D-17) is not listed in the most recent literature about them [94; 66], although it can be seen to follow easily from existing results; see eg [93, Theorem 1.4].

[12] **P Bousseau**, *Tropical refined curve counting from higher genera and lambda classes*, Invent. Math. 215 (2019) 1–79 MR Zbl

[13] **P Bousseau**, *Quantum mirrors of log Calabi–Yau surfaces and higher-genus curve counting*, Compos. Math. 156 (2020) 360–411 MR Zbl

[14] **P Bousseau**, *The quantum tropical vertex*, Geom. Topol. 24 (2020) 1297–1379 MR Zbl

[15] **P Bousseau**, *On an example of quiver Donaldson–Thomas/relative Gromov–Witten correspondence*, Int. Math. Res. Not. 2021 (2021) 11845–11888 MR Zbl

[16] **P Bousseau**, *Strong positivity for the skein algebras of the* 4*–punctured sphere and of the* 1*–punctured torus*, Comm. Math. Phys. 398 (2023) 1–58 MR Zbl

[17] **P Bousseau**, **A Brini**, **M van Garrel**, *Stable maps to Looijenga pairs: orbifold examples*, Lett. Math. Phys. 111 (2021) art. id. 109 MR Zbl

[18] **P Bousseau**, **A Brini**, **M van Garrel**, *On the log-local principle for the toric boundary*, Bull. Lond. Math. Soc. 54 (2022) 161–181 MR Zbl

[19] **P Bousseau**, **H Fan**, **S Guo**, **L Wu**, *Holomorphic anomaly equation for* $(\mathbb{P}^2, E)$ *and the Nekrasov–Shatashvili limit of local* $\mathbb{P}^2$, Forum Math. Pi 9 (2021) art. id. e3 MR Zbl

[20] **J Bryan**, **R Pandharipande**, *Curves in Calabi–Yau threefolds and topological quantum field theory*, Duke Math. J. 126 (2005) 369–396 MR Zbl

[21] **Y Cao**, **M Kool**, **S Monavari**, *Stable pair invariants of local Calabi–Yau* 4*–folds*, Int. Math. Res. Not. 2022 (2022) 4753–4798 MR Zbl

[22] **Y Cao**, **N C Leung**, *Donaldson–Thomas theory for Calabi–Yau* 4*–folds*, preprint (2014) arXiv 1407.7659

[23] **Y Cao**, **D Maulik**, **Y Toda**, *Genus zero Gopakumar–Vafa type invariants for Calabi–Yau* 4*–folds*, Adv. Math. 338 (2018) 41–92 MR Zbl

[24] **Y Cao**, **D Maulik**, **Y Toda**, *Stable pairs and Gopakumar–Vafa type invariants for Calabi–Yau* 4*–folds*, J. Eur. Math. Soc. (JEMS) 24 (2022) 527–581 MR Zbl

[25] **Q Chen**, *Stable logarithmic maps to Deligne–Faltings pairs I*, Ann. of Math. 180 (2014) 455–521 MR Zbl

[26] **T-M Chiang**, **A Klemm**, **S-T Yau**, **E Zaslow**, *Local mirror symmetry: calculations and interpretations*, Adv. Theor. Math. Phys. 3 (1999) 495–565 MR Zbl

[27] **J Choi**, **M van Garrel**, **S Katz**, **N Takahashi**, *Local BPS invariants: enumerative aspects and wall-crossing*, Int. Math. Res. Not. 2020 (2020) 5450–5475 MR Zbl

[28] **J Choi**, **M van Garrel**, **S Katz**, **N Takahashi**, *Log BPS numbers of log Calabi–Yau surfaces*, Trans. Amer. Math. Soc. 374 (2021) 687–732 MR Zbl

[29] **J Choi**, **M van Garrel**, **S Katz**, **N Takahashi**, *Sheaves of maximal intersection and multiplicities of stable log maps*, Selecta Math. 27 (2021) art. id. 61 MR Zbl

[30] **T Coates**, **A Corti**, **H Iritani**, **H-H Tseng**, *Computing genus-zero twisted Gromov–Witten invariants*, Duke Math. J. 147 (2009) 377–438 MR Zbl

[31] **T Coates**, **A Givental**, *Quantum Riemann–Roch, Lefschetz and Serre*, Ann. of Math. 165 (2007) 15–53 MR Zbl

[32] **B Davison**, **T Mandel**, *Strong positivity for quantum theta bases of quantum cluster algebras*, Invent. Math. 226 (2021) 725–843 MR Zbl

[33] **S Di Rocco**, *k–very ample line bundles on del Pezzo surfaces*, Math. Nachr. 179 (1996) 47–56 MR Zbl

[34] **A I Efimov**, *Cohomological Hall algebra of a symmetric quiver*, Compos. Math. 148 (2012) 1133–1146 MR Zbl

[35] **T Ekholm**, **P Kucharski**, **P Longhi**, *Multi-cover skeins, quivers, and* $3d$ $\mathcal{N}=2$ *dualities*, J. High Energy Phys. (2020) art. id. 018 MR Zbl

[36] **T Ekholm**, **P Kucharski**, **P Longhi**, *Physics and geometry of knots–quivers correspondence*, Comm. Math. Phys. 379 (2020) 361–415 MR Zbl

[37] **J Engel**, **M Reineke**, *Smooth models of quiver moduli*, Math. Z. 262 (2009) 817–848 MR Zbl

[38] **P Engel**, *Looijenga's conjecture via integral-affine geometry*, J. Differential Geom. 109 (2018) 467–495 MR Zbl

[39] **B Fang**, **C-C M Liu**, *Open Gromov–Witten invariants of toric Calabi–Yau* 3*–folds*, Comm. Math. Phys. 323 (2013) 285–328 MR Zbl

[40] **R Friedman**, *On the geometry of anticanonical pairs*, preprint (2015) arXiv 1502.02560

[41] **R Friedman**, **F Scattone**, *Type III degenerations of* $K3$ *surfaces*, Invent. Math. 83 (1986) 1–39 MR Zbl

[42] **W Fulton**, *Introduction to toric varieties*, Annals of Mathematics Studies 131, Princeton Univ. Press (1993) MR Zbl

[43] **M van Garrel**, **T Graber**, **H Ruddat**, *Local Gromov–Witten invariants are log invariants*, Adv. Math. 350 (2019) 860–876 MR Zbl

[44] **G Gasper**, **M Rahman**, *Basic hypergeometric series*, 2nd edition, Encyclopedia of Mathematics and its Applications 96, Cambridge Univ. Press (2004) MR Zbl

[45] **P Georgieva**, *Open Gromov–Witten disk invariants in the presence of an anti-symplectic involution*, Adv. Math. 301 (2016) 116–160 MR Zbl

[46] **A B Givental**, *Equivariant Gromov–Witten invariants*, Int. Math. Res. Not. (1996) 613–663 MR Zbl

[47] **R Gopakumar**, **C Vafa**, *M–theory and topological strings*, *I*, preprint (1998) arXiv hep-th/9809187

[48] **R Gopakumar**, **C Vafa**, *M–theory and topological strings*, *II*, preprint (1998) arXiv hep-th/9812127

[49] **T Graber**, **R Pandharipande**, *Localization of virtual classes*, Invent. Math. 135 (1999) 487–518 MR Zbl

[50] **B R Greene**, **D R Morrison**, **M R Plesser**, *Mirror manifolds in higher dimension*, Comm. Math. Phys. 173 (1995) 559–597 MR Zbl

[51] **M Gross**, *Tropical geometry and mirror symmetry*, CBMS Regional Conference Series in Mathematics 114, Amer. Math. Soc., Providence, RI (2011) MR Zbl

[52] **M Gross**, **P Hacking**, **S Keel**, *Birational geometry of cluster algebras*, Algebr. Geom. 2 (2015) 137–175 MR Zbl

[53] **M Gross**, **P Hacking**, **S Keel**, *Mirror symmetry for log Calabi–Yau surfaces, I*, Publ. Math. Inst. Hautes Études Sci. 122 (2015) 65–168 MR Zbl

[54] **M Gross**, **P Hacking**, **S Keel**, *Moduli of surfaces with an anti-canonical cycle*, Compos. Math. 151 (2015) 265–291 MR Zbl

[55] **M Gross**, **P Hacking**, **B Siebert**, *Theta functions on varieties with effective anti-canonical class*, Mem. Amer. Math. Soc. 1367, Amer. Math. Soc., Providence, RI (2022) MR Zbl

[56] **M Gross**, **R Pandharipande**, **B Siebert**, *The tropical vertex*, Duke Math. J. 153 (2010) 297–362 MR Zbl

[57] **M Gross**, **B Siebert**, *From real affine geometry to complex geometry*, Ann. of Math. 174 (2011) 1301–1428 MR Zbl

[58] **M Gross**, **B Siebert**, *Logarithmic Gromov–Witten invariants*, J. Amer. Math. Soc. 26 (2013) 451–510 MR Zbl

[59] **V J W Guo**, **C Krattenthaler**, *Some divisibility properties of binomial and q–binomial coefficients*, J. Number Theory 135 (2014) 167–184 MR Zbl

[60] **P Hacking**, **A Keating**, *Homological mirror symmetry for log Calabi–Yau surfaces*, Geom. Topol. 26 (2022) 3747–3833 MR Zbl

[61] **T Hausel**, **B Sturmfels**, *Toric hyperKähler varieties*, Doc. Math. 7 (2002) 495–534 MR Zbl

[62] **E-N Ionel**, **T H Parker**, *The Gopakumar–Vafa formula for symplectic manifolds*, Ann. of Math. 187 (2018) 1–64 MR Zbl

[63] **A Iqbal**, **A-K Kashani-Poor**, *The vertex on a strip*, Adv. Theor. Math. Phys. 10 (2006) 317–343 MR Zbl

[64] **S Katz**, *Genus zero Gopakumar–Vafa invariants of contractible curves*, J. Differential Geom. 79 (2008) 185–195 MR Zbl

[65] **S Katz**, **C-C M Liu**, *Enumerative geometry of stable maps with Lagrangian boundary conditions and multiple covers of the disc*, from "The interaction of finite-type and Gromov–Witten invariants" (D Auckly, J Bryan, editors), Geom. Topol. Monogr. 8, Geom. Topol. Publ., Coventry (2006) 1–47 MR Zbl

[66] **J S Kim**, **M Yoo**, *Product formulas for certain skew tableaux*, European J. Combin. 84 (2020) art. id. 103038 MR Zbl

[67] **T Kimura**, **M Panfil**, **Y Sugimoto**, **P Sułkowski**, *Branes, quivers and wave-functions*, SciPost Phys. 10 (2021) art. id. 051 MR

[68] **A Klemm**, **R Pandharipande**, *Enumerative geometry of Calabi–Yau 4–folds*, Comm. Math. Phys. 281 (2008) 621–653 MR Zbl

[69] **Y Konishi**, **S Minabe**, *Flop invariance of the topological vertex*, Internat. J. Math. 19 (2008) 27–45 MR Zbl

[70] **M Kontsevich**, **Y Soibelman**, *Cohomological Hall algebra, exponential Hodge structures and motivic Donaldson–Thomas invariants*, Commun. Number Theory Phys. 5 (2011) 231–352 MR Zbl

[71] **C Krattenthaler**, *Proof of two multivariate q–binomial sums arising in Gromov–Witten theory*, preprint (2021) arXiv 2102.02360

[72] **P Kucharski**, **M Reineke**, **M Stošić**, **P Sułkowski**, *Knots–quivers correspondence*, Adv. Theor. Math. Phys. 23 (2019) 1849–1902 MR Zbl

[73] **J M F Labastida**, **M Mariño**, *Polynomial invariants for torus knots and topological strings*, Comm. Math. Phys. 217 (2001) 423–449 MR Zbl

[74] **J M F Labastida**, **M Mariño**, **C Vafa**, *Knots, links and branes at large N*, J. High Energy Phys. (2000) art. id. 7 MR Zbl

[75] **J Li**, *A degeneration formula of GW–invariants*, J. Differential Geom. 60 (2002) 199–293 MR Zbl

[76] **J Li**, **C-C M Liu**, **K Liu**, **J Zhou**, *A mathematical theory of the topological vertex*, Geom. Topol. 13 (2009) 527–621 MR Zbl

[77] **J Li**, **Y S Song**, *Open string instantons and relative stable morphisms*, Adv. Theor. Math. Phys. 5 (2001) 67–91 MR Zbl

[78] **C-C M Liu**, **K Liu**, **J Zhou**, *A formula of two-partition Hodge integrals*, J. Amer. Math. Soc. 20 (2007) 149–184 MR Zbl

[79] **E Looijenga**, *Rational surfaces with an anticanonical cycle*, Ann. of Math. 114 (1981) 267–322 MR Zbl

[80] **I G Macdonald**, *Symmetric functions and Hall polynomials*, 2nd edition, Oxford Univ. Press (1995) MR Zbl

[81] **T Mandel**, *Tropical theta functions and log Calabi–Yau surfaces*, Selecta Math. 22 (2016) 1289–1335 MR Zbl

[82] **T Mandel**, *Classification of rank* 2 *cluster varieties*, Symmetry Integrability Geom. Methods Appl. 15 (2019) art. id. 042 MR Zbl

[83] **T Mandel**, *Scattering diagrams, theta functions, and refined tropical curve counts*, J. Lond. Math. Soc. 104 (2021) 2299–2334 MR Zbl

[84] **T Mandel**, *Theta bases and log Gromov–Witten invariants of cluster varieties*, Trans. Amer. Math. Soc. 374 (2021) 5433–5471 MR Zbl

[85] **T Mandel**, **H Ruddat**, *Descendant log Gromov–Witten invariants for toric varieties and tropical curves*, Trans. Amer. Math. Soc. 373 (2020) 1109–1152 MR Zbl

[86] **T Mandel**, **H Ruddat**, *Tropical quantum field theory, mirror polyvector fields, and multiplicities of tropical curves*, Int. Math. Res. Not. 2023 (2023) 3249–3304 MR Zbl

[87] **C Manolache**, *Virtual pull-backs*, J. Algebraic Geom. 21 (2012) 201–245 MR Zbl

[88] **M Mariño**, **C Vafa**, *Framed knots at large N*, from "Orbifolds in mathematics and physics" (A Adem, J Morava, Y Ruan, editors), Contemp. Math. 310, Amer. Math. Soc., Providence, RI (2002) 185–204 MR Zbl

[89] **D Maulik**, **A Oblomkov**, **A Okounkov**, **R Pandharipande**, *Gromov–Witten/Donaldson–Thomas correspondence for toric* 3*–folds*, Invent. Math. 186 (2011) 435–479 MR Zbl

[90] **D Maulik**, **Y Toda**, *Gopakumar–Vafa invariants via vanishing cycles*, Invent. Math. 213 (2018) 1017–1097 MR Zbl

[91] **P Mayr**, $\mathcal{N}=1$ *mirror symmetry and open/closed string duality*, Adv. Theor. Math. Phys. 5 (2001) 213–242 MR Zbl

[92] **G Mikhalkin**, *Enumerative tropical algebraic geometry in* $\mathbb{R}^2$, J. Amer. Math. Soc. 18 (2005) 313–377 MR Zbl

[93] **A H Morales**, **I Pak**, **G Panova**, *Hook formulas for skew shapes, I: q–analogues and bijections*, J. Combin. Theory Ser. A 154 (2018) 350–405 MR Zbl

[94] **A H Morales**, **I Pak**, **G Panova**, *Hook formulas for skew shapes, III: Multivariate and product formulas*, Algebr. Comb. 2 (2019) 815–861 MR Zbl

[95] **D Mumford**, *Towards an enumerative geometry of the moduli space of curves*, from "Arithmetic and geometry, II" (M Artin, J Tate, editors), Progr. Math. 36, Birkhäuser, Boston, MA (1983) 271–328 MR Zbl

[96] **N Nabijou**, **D Ranganathan**, *Gromov–Witten theory with maximal contacts*, Forum Math. Sigma 10 (2022) art. id. e5 MR Zbl

[97] **T Nishinou**, **B Siebert**, *Toric degenerations of toric varieties and tropical curves*, Duke Math. J. 135 (2006) 1–51 MR Zbl

[98] **J Oh**, **R P Thomas**, *Counting sheaves on Calabi–Yau* 4*–folds*, *I*, preprint (2020) arXiv 2009.05542

[99] **G Olive**, *Generalized powers*, Amer. Math. Monthly 72 (1965) 619–627 MR Zbl

[100] **H Ooguri**, **C Vafa**, *Knot invariants and topological strings*, Nuclear Phys. B 577 (2000) 419–438 MR Zbl

[101] **R Pandharipande**, **R P Thomas**, *Stable pairs and BPS invariants*, J. Amer. Math. Soc. 23 (2010) 267–297 MR Zbl

[102] **M Panfil**, **M Stošić**, **P Sułkowski**, *Donaldson–Thomas invariants, torus knots, and lattice paths*, Phys. Rev. D 98 (2018) art. id. 026022 MR

[103] **M Panfil**, **P Sułkowski**, *Topological strings, strips and quivers*, J. High Energy Phys. (2019) art. id. 124 MR Zbl

[104] **B Parker**, *Gluing formula for Gromov–Witten invariants in a triple product*, preprint (2015) arXiv 1511.00779

[105] **D Ranganathan**, *Logarithmic Gromov–Witten theory with expansions*, Algebr. Geom. 9 (2022) 714–761 MR Zbl

[106] **S Rayan**, *Aspects of the topology and combinatorics of Higgs bundle moduli spaces*, Symmetry Integrability Geom. Methods Appl. 14 (2018) art. id. 129 MR Zbl

[107] **M Reineke**, *Degenerate cohomological Hall algebra and quantized Donaldson–Thomas invariants for m–loop quivers*, Doc. Math. 17 (2012) 1–22 MR Zbl

[108] **B E Sagan**, *Congruence properties of q–analogs*, Adv. Math. 95 (1992) 127–143 MR Zbl

[109] **J P Solomon**, **S B Tukachinsky**, *Point-like bounding chains in open Gromov–Witten theory*, preprint (2016) arXiv 1608.02495

[110] **R P Stanley**, *Theory and application of plane partitions, I*, Studies in Appl. Math. 50 (1971) 167–188 MR Zbl

[111] **R P Stanley**, *Theory and application of plane partitions, II*, Studies in Appl. Math. 50 (1971) 259–279 MR Zbl

[112] **R P Stanley**, *Enumerative combinatorics, II*, Cambridge Studies in Advanced Mathematics 62, Cambridge Univ. Press (1999) MR Zbl

[113] **H-H Tseng**, *Orbifold quantum Riemann–Roch, Lefschetz and Serre*, Geom. Topol. 14 (2010) 1–81 MR Zbl

[114] **T Y Yu**, *Enumeration of holomorphic cylinders in log Calabi–Yau surfaces, I*, Math. Ann. 366 (2016) 1649–1675 MR Zbl

[115] **T Y Yu**, *Enumeration of holomorphic cylinders in log Calabi–Yau surfaces, II: Positivity, integrality and the gluing formula*, Geom. Topol. 25 (2021) 1–46 MR Zbl

[116] **D Zeilberger**, *A q–Foata proof of the q–Saalschütz identity*, European J. Combin. 8 (1987) 461–463 MR Zbl

[117] **Y Zhou**, *Weyl groups and cluster structures of families of log Calabi–Yau surfaces*, preprint (2019) arXiv 1910.05762

Department of Mathematics, University of Georgia
Athens, GA, United States

School of Mathematics and Statistics, University of Sheffield
Sheffield, United Kingdom

School of Mathematics, University of Birmingham
Birmingham, United Kingdom

pierrick.bousseau@uga.edu, a.brini@sheffield.ac.uk, m.vangarrel@bham.ac.uk

Proposed: Richard P Thomas Received: 8 April 2022
Seconded: Mark Gross, Dan Abramovich Revised: 8 April 2022

Geometry & Topology Publications, an imprint of **mathematical sciences publishers** msp

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 28 Issue 1 (pages 1–496) 2024