Geometry & Topology

Volume 28 (2024)

Issue 3 (pages 1005–1499)

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**:3 (2024) 1005–1063
DOI: 10.2140/gt.2024.28.1005
Published: 10 May 2024

Homological mirror symmetry for hypertoric varieties I: Conic equivariant sheaves

MICHAEL MCBREEN
BEN WEBSTER

We consider homological mirror symmetry in the context of hypertoric varieties, showing that an appropriate category of B–branes (that is, coherent sheaves) on an additive hypertoric variety matches a category of A–branes on a Dolbeault hypertoric manifold for the same underlying combinatorial data. For technical reasons, the A–branes we consider are modules over a deformation quantization (that is, DQ–modules). We consider objects in this category equipped with an analogue of a Hodge structure, which corresponds to a $\mathbb{G}_m$–action on the dual side of the mirror symmetry.

This result is based on hands-on calculations in both categories. We analyze coherent sheaves by constructing a tilting generator, using the characteristic p approach of Kaledin; the result is a sum of line bundles, which can be described using a simple combinatorial rule. The endomorphism algebra H of this tilting generator has a simple quadratic presentation in the grading induced by $\mathbb{G}_m$–equivariance. In fact, we can confirm it is Koszul, and compute its Koszul dual $H^!$.

We then show that this same algebra appears as an Ext–algebra of simple A–branes in a Dolbeault hypertoric manifold. The $\mathbb{G}_m$–equivariant grading on coherent sheaves matches a Hodge grading in this category.

14J33, 47A67

1 Introduction

Toric varieties have proven many times in algebraic geometry to be a valuable testing ground. Their combinatorial flavor and concrete nature has been extremely conducive to calculation. Certainly this is the case in the domain of homological mirror symmetry; see Abouzaid [1] and Fukaya, Oh, Ohta and Ono [19; 20].

Toric varieties have a natural hyperkähler analogue, which we call *hypertoric varieties*; in some other places in the literature, they are called "toric hyperkähler varieties". Just as toric varieties can be written as Kähler quotients of complex vector spaces, hypertoric varieties are hyperkähler quotients by tori; see Definition 2.1.

Despite their name, hypertoric varieties are almost never toric. Rather, they are *conical symplectic resolutions*: the natural map $\pi : \mathfrak{M} \to \operatorname{Spec} H^0(\mathfrak{M}, \mathscr{O}_{\mathfrak{M}})$ is a proper resolution of singularities, and there is an action of $\mathbb{G}_m$ on $\mathfrak{M}$ which dilates the algebraic symplectic form and contracts $\operatorname{Spec} H^0(\mathfrak{M}, \mathscr{O}_{\mathfrak{M}})$ to a point o. Among symplectic resolutions, hypertoric varieties are distinguished by the presence of an effective complex hamiltonian action of a half-dimensional complex torus.

In this paper, we study homological mirror symmetry for hypertoric varieties. This is typically understood to mean an equivalence between the derived category of coherent sheaves (or B–branes) on an algebraic variety and the Fukaya category (the A–branes) of a related symplectic manifold. We will instead prove a different, but closely related, equivalence.

On the B side, we consider the derived category of coherent sheaves on the hypertoric variety $\mathfrak{M}$. For the statement of our equivalence, it is most natural to impose some finiteness conditions. The simplest version of our equivalence concerns the category $\mathsf{Coh}(\mathfrak{M})_o$ of sheaves set-theoretically supported on the fiber $\pi^{-1}(o)$ over the cone point of $\operatorname{Spec} H^0(\mathfrak{M}, \mathscr{O}_{\mathfrak{M}})$.

On the A–side, we take our mirror space to be a *Dolbeault hypertoric manifold* $\mathfrak{D}$, as defined by Hausel and Proudfoot. This is a multiplicative analogue of $\mathfrak{M}$, equipped with a fibration $q : \mathfrak{D} \to \mathbb{C}^d$ by Lagrangian abelian subvarieties degenerating to a union of toric varieties over $0 \in \mathbb{C}^d$. We prove in the sequel paper [21], joint with Ben Gammage, that $q^{-1}(0)$ is the skeleton of a suitable Liouville structure on $\mathfrak{D}$. When we need to distinguish, we will call usual hypertoric varieties *additive*.

We define a certain category dq of deformation quantization modules on $\mathfrak{D}$, quantizing the irreducible components of $q^{-1}(0) \subset \mathfrak{D}$. Let DQ be the (dg enhanced) derived category of dq. We prove:

Theorem A (Theorem 4.36) *There is an equivalence of dg categories* $D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{C}})_o) \to \mathsf{DQ}$.

The simples of dq may be thought of as certain distinguished objects in the Fukaya category of $\mathfrak{D}$. We do not attempt to make this precise here; the exact relation of DQ to $\mathrm{Fuk}(\mathfrak{D})$ is described in [21].

The left-hand category has an important extra structure: the conical $\mathbb{G}_m$ action. To understand its mirror, we consider an abelian category $\mu\mathsf{m}$ consisting of DQ–modules endowed with a "microlocal mixed Hodge structure", along with its derived category $\mu\mathsf{M}$. We have the following graded version of Theorem A:

Theorem B (Corollary 5.10) *There is an equivalence of dg categories* $D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})_o) \to \mu\mathsf{M}$, *such that tensoring with the weight* 1 *representation of* $\mathbb{G}_m$ *corresponds to a* $\frac{1}{2}$ *Tate twist.*

This equivalence may be thought of as homological mirror symmetry for two subcategories of the A– and B–branes, both of which are enriched with suitable notions of $\mathbb{G}_m$–equivariance. The reader may compare with Braverman, Maulik and Okounkov [12] and Maulik and Okounkov [28] and their sequels, where the same $\mathbb{G}_m$–action plays a key role.

In fact, we construct a family of equivalences, which are best understood in terms of special t–structures on both sides. On the one hand, the Dolbeault space $\mathfrak{D}$ depends on a choice of parameter $\zeta \in \mathfrak{t}_{\mathbb{R}}^{\vee} \cong H^2(\mathfrak{M}, \mathbb{R})$, in the complement of a periodic hyperplane arrangement. As ζ crosses these hyperplanes, components of the central fiber $q^{-1}(0)$ may appear or disappear. Thus, different chambers yield different abelian categories $\mu\mathsf{m}$, which are nevertheless derived equivalent.

On the other hand, a choice of ζ in the complement of the arrangement determines a *tilting generator* of $D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})_o)$. This is a vector bundle $\mathcal{T}^{\zeta}$ such that $\mathrm{Ext}(\mathcal{T}^{\zeta}, -)$ defines an equivalence of dg–categories

$$D^b(\mathsf{Coh}(\mathfrak{M})) \cong D^b(H^{\zeta}\text{–mod}^{\mathrm{op}}),$$

where $H^{\zeta} = \mathrm{End}(\mathcal{T}^{\zeta})$. In particular, the natural t–structure on the right-hand side defines an "exotic" t–structure on the left-hand side.

Our construction of $\mathcal{T}^{\zeta}$ follows a recipe of Kaledin [25]. The algebra H^{ζ} is thus an analogue in our context of Bezrukavnikov's noncommutative Springer resolution [5]. Its significance can be understood as follows. Both $\mathfrak{M}$ and H^{ζ} are naturally defined over $\mathbb{Z}$. Given a field $\mathbb{K}$ of characteristic p, let $\mathfrak{M}_{\mathbb{K}}$ and $H_{\mathbb{K}}^{\zeta}$ be the corresponding $\mathbb{K}$–forms. Suppose $p\zeta \in H^2(\mathfrak{M}; \mathbb{Z})$, in which case it defines a class $\lambda \in \mathrm{Pic}(\mathfrak{M}_{\mathbb{K}})$. There is an associated Frobenius-constant quantization of the variety $\mathfrak{M}_{\mathbb{K}}$ in the sense of Bezrukavnikov and Kaledin [6]. We write $A_{\mathbb{K}}^{\lambda}$ for the resulting noncommutative algebra, which deforms $H^0(\mathfrak{M}_{\mathbb{K}}, \mathscr{O}_{\mathfrak{M}})$. By Theorem 3.18, there is equivalence of abelian categories between the category of $A_{\mathbb{K}}^{\lambda}$–modules with special central character, and the category of finite-dimensional representations of $H_{\mathbb{K}}^{\lambda}$ satisfying a nilpotence condition.

While this construction springs from geometry in characteristic p, and the tilting property is checked using this approach, the tilting generators we consider are sums of line bundles and have a simple combinatorial construction, as does the endomorphism ring H^{ζ}. This endomorphism ring inherits a grading from a $\mathbb{G}_m$–equivariant structure on $\mathcal{T}^{\zeta}$ and is Koszul with respect to it. Thus, the category of $\mathbb{G}_m$–equivariant coherent sheaves on $\mathfrak{M}$ is controlled by the derived category of graded H^{ζ}–modules, or equivalently by graded modules over $(H^{\zeta})^!$, its Koszul dual. It is this Koszul dual that has a natural counterpart on the mirror side.

Theorem 5.9 of this paper explains the relevance of these structures to our mirror equivalence. It can be paraphrased as follows:

Theorem C *Under the equivalence of Theorem B, the natural t–structure on deformation quantization modules on $\mathfrak{D}$ corresponds to the exotic t–structure on coherent sheaves on $\mathfrak{M}$ arising from the tilting bundle $\mathscr{T}^{\zeta}$.*

There are many directions one can go from here. For instance, it is natural to expect different t–structures should fit together into a real variation of stability in the sense of Anno, Bezrukavnikov and Mirković [2], in particular, as predicted by [2, Conjecture 1]. In [37], the second author will show this in the more general context of Coulomb branches.

As a result of our use of DQ–modules as a substitute for the Fukaya category, this paper contains little about Lagrangian branes, pseudoholomorphic disks and other staples of symplectic geometry. The reader may wish to compare with the interesting recent preprint by Lau and Zheng [27], which appeared a few days before this paper and treats the problem of nonequivariant mirror symmetry for hypertoric varieties from the perspective of SYZ fibrations.

The variety $\mathfrak{M}$ is the Coulomb branch (in the sense of Braverman, Finkelberg and Nakajima [11]) with gauge group given by a torus, and that $\mathfrak{D}$ is expected to be a hyperkähler rotation of the K–theoretic version of this construction. Thus, it is natural to consider how these constructions can be generalized to that case. The analogous calculation of a tilting bundle with explicit endomorphism ring can be generalized in this case, as the second author will show in [36], but it is very difficult to even conjecture the correct category to consider on the A side.

One key source of interest in hypertoric varieties is that they provide excellent examples of conic symplectic singularities (see [10; 9]), which can be understood in combinatorial terms. Considerations in 3-d mirror symmetry [9] and calculations in the representation theory of its quantization led Braden, Licata, Proudfoot and the second author to suggest that hypertoric varieties should be viewed as coming in dual pairs, corresponding to Gale dual combinatorial data. In particular, the categories $\mathcal{O}$ attached to these two varieties are Koszul dual [7; 8]. An obvious question in this case is how the categories we have considered, such as coherent sheaves, can be interpreted in terms of the dual variety (they are certainly not equivalent or Koszul dual to the coherent sheaves on the dual variety, as some very simple examples show). Some calculations in quantum field theory suggest that they are the representations of a vertex algebra constructed by a BRST analogue of the hyperkähler reduction, but this is definitely a topic which will need to wait for future research.

Detailed outline of the argument

Part 1 Coherent sheaves and characteristic p quantizations of the additive hypertoric variety

Section 2.1 defines the additive hypertoric variety $\mathfrak{M}$. In Section 2.2 we fix a field $\mathbb{K}$ of characteristic p, and review the relation between the quantization of $\mathfrak{M}_{\mathbb{K}}$, called $A^{\lambda}_{\mathbb{K}}$, and coherent sheaves on $\mathfrak{M}_{\mathbb{K}}$. In Section 3.1 we introduce a category of modules $A^{\lambda}_{\mathbb{K}}$–mod_o, along with its graded counterpart $A^{\lambda}_{\mathbb{K}}$–mod^D_o. All these objects depend on a quantization parameter λ. In Sections 3.3, 3.2 and 3.5 we classify the projective pro-objects $P_{\boldsymbol{x}}$ of $A^{\lambda}_{\mathbb{K}}$–mod^D_o, which also yields a classification of simple objects $L_{\boldsymbol{x}}$.

Both projectives and simples are indexed by the chambers of a periodic hyperplane arrangement $\mathfrak{A}^{\mathrm{per}}_{\lambda}$, defined in Definition 3.8. We compute the endomorphism algebra $\bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\widetilde{\Lambda}} \mathrm{Hom}(P_{\boldsymbol{x}}, P_{\boldsymbol{y}})$ in Theorem 3.13. The latter contains a ring of power series $\widehat{S}$ as a central subalgebra, and we define a variant $\widetilde{H}^{\lambda}_{\mathbb{K}}$ (Definition 3.14) in which $\widehat{S}$ is replaced by the corresponding polynomial ring S. We find that $A^{\lambda}_{\mathbb{K}}$–mod^D_o is equivalent to the subcategory of $\widetilde{H}^{\lambda}_{\mathbb{K}}$–modules on which S acts nilpotently.

The algebra $\widetilde{H}^{\lambda}_{\mathbb{K}}$ has a natural lift to $\mathbb{Z}$, written $\widetilde{H}^{\lambda}_{\mathbb{Z}}$, which we will use to compare with characteristic-zero objects on the mirror side. Corollary 3.22 shows that $\widetilde{H}^{\lambda}_{\mathbb{Z}}$ is Koszul. We compute the Koszul dual algebra $\widetilde{H}^{!}_{\lambda,\mathbb{K}} = \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\widetilde{\Lambda}} \operatorname{Ext}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}})$ (Definition 3.23 and Theorem 3.24).

In Section 3.9 we describe the ungraded category $A^{\lambda}_{\mathbb{K}}$–mod$_o$ in terms of the graded one. Its simples and projectives are indexed by the toroidal hyperplane arrangement $\mathfrak{A}^{\mathrm{tor}}_{\lambda}$ obtained as the quotient of $\mathfrak{A}^{\mathrm{per}}_{\lambda}$ by certain translations. We describe the corresponding algebras $H^{\lambda}_{\mathbb{K}} = \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\Lambda} \operatorname{Hom}(P_{\boldsymbol{x}}, P_{\boldsymbol{y}})$ and $H^{!}_{\lambda,\mathbb{K}} = \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\Lambda} \operatorname{Ext}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}})$, where the sums now range over simples (resp. projectives) for $A^{\lambda}_{\mathbb{K}}$–mod$_o$.

In Section 3.10 we use the above results to produce a tilting bundle $\mathscr{T}^{\lambda}$ on $\mathfrak{M}$ with endomorphism ring $\operatorname{End}(\mathscr{T}^{\lambda}) = H^{\lambda}$. Passing to characteristic zero, and replacing λ by a parameter $\zeta \in \mathfrak{t}^{*}_{\mathbb{R}}$, we obtain equivalences (from Corollary 3.41 and Proposition 3.43, respectively)

$$D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})) \cong D^b(H^{\zeta,\mathrm{op}}_{\mathbb{Q}}\text{–mod}) \quad \text{and} \quad H^{!}_{\zeta,\mathbb{Q}}\text{–perf} \cong D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})_o), \tag{1-1}$$

where $H^{!}_{\zeta,\mathbb{Q}}$–perf is the category of perfect dg–modules over this ring.

Remark 1.1 Throughout, we will always endow the bounded derived category D^b of an abelian category with its usual dg–enhancement using injective resolutions; thus if we write $D^b(\mathsf{a}) \cong \mathsf{C}$ for an abelian category a and a dg–category C, we really mean that this dg–enhancement is quasiequivalent to C.

Part 2 Deformation quantization and microlocal mixed Hodge modules on the Dolbeault hypertoric manifold The second half of our paper begins with a definition of the Dolbeault hypertoric manifold $\mathfrak{D}$ (Definition 4.3), depending on a moment map parameter ζ. The complex manifold $\mathfrak{D}$ is a complex integrable system, with a "central fiber" consisting of a collection of complex Lagrangian submanifolds $\mathfrak{X}_{\boldsymbol{x}}$ indexed by the chambers of a toroidal hyperplane arrangement $\mathfrak{B}^{\mathrm{tor}}_{\zeta}$ (Definition 4.10 and Proposition 4.11).

The universal cover $\widetilde{\mathfrak{D}}$ of the Dolbeault space is an infinite-type complex symplectic manifold, whose geometry is described by a periodic hyperplane arrangement $\mathfrak{B}^{\mathrm{per}}_{\zeta}$. In turn, $\widetilde{\mathfrak{D}}$ is an open submanifold of an infinite-type algebraic symplectic variety $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$. The latter has a key additional structure: an action of a torus $\mathbb{S} \cong \mathbb{C}^*$ dilating both the complex symplectic form and the base of the integrable system, and preserving the central fiber.

In Section 4.5, we define a sheaf $\mathcal{O}^{\hbar}_{\phi}$ of $\mathbb{C}((\hbar))$–algebras on $\mathfrak{D}$ quantizing the structure sheaf, and for each $\mathfrak{X}_{\boldsymbol{x}}$ we define a module $\mathcal{L}_{\boldsymbol{x}}$ over $\mathcal{O}^{\hbar}_{\phi}$ supported on $\mathfrak{X}_{\boldsymbol{x}}$.

Although $\mathbb{S}$ does not preserve $\widetilde{\mathfrak{D}} \subset \widetilde{\mathfrak{D}}^{\mathrm{alg}}$, we can nevertheless make sense of $\mathbb{S}$–equivariant DQ–modules on $\widetilde{\mathfrak{D}}$ and $\mathfrak{D}$, and we show that $\mathcal{L}_{\boldsymbol{x}}$ has a natural $\mathbb{S}$–equivariant structure.

We define a subcategory dq of $\mathbb{S}$–equivariant $\mathcal{O}^{\hbar}_{\phi}$–modules on $\mathfrak{D}$ generated by the simple DQ–modules $\mathcal{L}_{\boldsymbol{x}}$, together with the category dg–category DQ of complexes in dq. The $\mathbb{S}$–equivariance yields a category with $\mathbb{C}$ (rather than $\mathbb{C}((\hbar))$) coefficients. We write $\widetilde{\mathsf{dq}}$ and $\widetilde{\mathsf{DQ}}$ for the corresponding categories on $\widetilde{\mathfrak{D}}$.

When λ is the reduction of $p\zeta$, the arrangements $\mathfrak{A}_\lambda^{\mathrm{tor}}$ and $\mathfrak{B}_\zeta^{\mathrm{tor}}$ are identified. We hence have a bijection of chambers, and a corresponding bijection of isomorphism classes of simple objects for the categories dq and $A_{\mathbb{K}}^\lambda$–mod$_o$. Moreover, Theorem 4.27 shows that the Ext–algebras of the simples in both categories share a common integral form: $H^!_{\lambda,\mathbb{C}} \cong E^!_{\mathbb{C}} := \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\Lambda} \operatorname{Ext}(\mathcal{L}_{\boldsymbol{x}}, \mathcal{L}_{\boldsymbol{x}})$.

Unfortunately, some care is needed about concluding that this isomorphism induces an equivalence of categories $\mathsf{DQ} \to D^b(\mathsf{Coh}(\mathfrak{M}))_o$, since a priori it is not clear that $E^!_{\mathbb{C}}$ is formal as a dg–algebra, which we would need to define a fully faithful functor. We prove this equivalence by constructing projective objects in dq, and showing that $H_{\lambda,\mathbb{C}}$ appears as their automorphism algebra. This shows that we have the desired derived equivalence (Theorem 4.36).

We can further account for the grading on $H_{\lambda,\mathbb{Q}}$ and reduce the structure ring to $\mathbb{Q}$ from $\mathbb{C}$ by considering a new graded abelian category $\mu\mathsf{m}$ (Definition 5.8), and a corresponding triangulated category $D^b(\mu\mathsf{M})$. Each object of $\mu\mathsf{m}$ is a $\mathcal{O}_\phi^\hbar$–module, such that for each lagrangian $\mathfrak{X}_{\boldsymbol{x}}$, the restriction to a Weinstein neighborhood of $\mathfrak{X}_{\boldsymbol{x}}$ is equipped with the structure of a mixed Hodge module. These structures are required to be compatible in a natural sense whenever two components intersect. We define $\mu\mathsf{m}$ as the category generated by a special collection of such objects.

Each object $\mathcal{L}_{\boldsymbol{x}}$ has a natural lift to $\mu\mathsf{m}$, and moreover any simple object of $\mu\mathsf{m}$ is isomorphic to such a lift. This allows us to conclude that the equivalence $D^b_{\mathrm{perf}}(\mathsf{Coh}(\mathfrak{M}_{\mathbb{C}})_o) \to \mathsf{DQ}$ can be upgraded to an equivalence of graded categories $D^b_{\mathrm{perf}}(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})_o) \to \mu\mathsf{M}$ in the spirit of equivariant mirror symmetry.

Remark 1.2 In an earlier version of this paper, the proof of the main result depended on the use of this Hodge structure. In revisions responding to a referee's comments, we found a proof that avoids the use of it, so we have moved all discussion of Hodge topics to Section 5, after the proof of Theorem 4.36. We have left the discussion of Hodge structures in the paper, since we believe it is of some interest in understanding how $\mathbb{C}^*$–actions translate through mirror symmetry.

Acknowledgements We would like to thank Andrei Okounkov for suggesting a multiplicative analogue of the hypertoric variety as a mirror, Tamás Hausel and Nick Proudfoot for sharing their unpublished results on multiplicative hypertoric varieties, and Roman Bezrukavnikov, Ben Gammage, Sam Gunningham, Paul Seidel, Vivek Shende and Michael Thaddeus for helpful conversations. We would also like to thank the referee for many helpful comments.

McBreen performed part of this work at the Massachusetts Institute of Technology, the École Polytechnique Fédérale de Lausanne, and during the Junior Trimester on *Symplectic geometry and representation theory* at the Hausdorff Research Institute for Mathematics. He gratefully acknowledges the hospitality of all these institutions. While at the EPFL, he was supported by the Advanced Grant 320593 *Arithmetic and physics of Higgs moduli spaces* of the European Research Council. This work was also partly supported by the Simons Foundation, as part of a Simons Investigator award.

Webster was supported during the course of this work by the NSF under grant DMS-1151473, the Alfred Sloan Foundation, by Discovery Grant RGPIN-2018-03974 from the Natural Sciences and Engineering Research Council of Canada, and by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

2 Hypertoric enveloping algebras

2.1 Additive hypertoric varieties

For a general introduction to hypertoric varieties, see [33].

Consider a split algebraic torus T over $\mathbb{Z}$ of dimension k (that is, an algebraic group isomorphic to $\mathbb{G}_m^k$) and a faithful linear action of T on the affine space $\mathbb{A}_{\mathbb{Z}}^n$, which we may assume is diagonal in the usual basis. We let $D \cong \mathbb{G}_m^n$ be the group of diagonal matrices in this basis, and write $G := D/T$.

We have an induced action of T on the cotangent bundle $T^*\mathbb{A}_{\mathbb{Z}}^n \cong \mathbb{A}_{\mathbb{Z}}^{2n}$. We use z_i for the usual coordinates on $\mathbb{A}_{\mathbb{Z}}^n$, and w_i for the dual coordinates. This action has an algebraic moment map $\mu\colon T^*\mathbb{A}_{\mathbb{Z}}^n \to \mathfrak{t}_{\mathbb{Z}}^*$, defined by a map of polynomial rings $\mathbb{Z}[\mathfrak{t}_{\mathbb{Z}}] \to \mathbb{Z}[\mathsf{z}_1, \dots, \mathsf{z}_n, \mathsf{w}_1, \dots, \mathsf{w}_n]$ sending a cocharacter χ to the sum $\sum_{i=1}^n \langle \epsilon_i, \chi \rangle \mathsf{z}_i \mathsf{w}_i$, where ϵ_i is the character on D defined by the action on the i^{th} coordinate line, and $\langle -, - \rangle$ is the usual pairing between characters and cocharacters of D.

For us, the main avatar of this action is the (*additive*) *hypertoric variety*. This is an algebraic hamiltonian reduction of $T^*\mathbb{A}_{\mathbb{Z}}^n$ by T. It comes in affine and smooth flavors, these being the categorical and GIT quotients (respectively) of the scheme-theoretic fiber $\mu^{-1}(0)$ by the group T. More precisely, fix a character $\alpha\colon T \to \mathbb{G}_m$ whose kernel does not fix a coordinate line.

Definition 2.1 For a commutative ring $\mathbb{K}$, we let

$$\mathfrak{N}_{\mathbb{K}} := \operatorname{Spec}(\mathbb{K}[\mathsf{z}_1, \dots, \mathsf{z}_n, \mathsf{w}_1, \dots, \mathsf{w}_n]^T / \langle \mu^*(\chi) \mid \chi \in \mathfrak{t}_{\mathbb{Z}} \rangle),$$
$$\mathfrak{M}_{\mathbb{K}} := \operatorname{Proj}(\mathbb{K}[\mathsf{z}_1, \dots, \mathsf{z}_n, \mathsf{w}_1, \dots, \mathsf{w}_n, t]^T / \langle \mu^*(\chi) \mid \chi \in \mathfrak{t}_{\mathbb{Z}} \rangle),$$

where t is an additional variable of degree 1 with T–weight $-\alpha$.

Both varieties carry a residual action of the torus $G = D/T$, and an additional commuting action of a rank-one torus $\mathbb{S} := \mathbb{G}_m$ which scales the coordinates w_i linearly while fixing z_i.

We say that the sequence $T \to D \to G$ is *unimodular* if the image of any tuple of coordinate cocharacters in $\mathfrak{d}_{\mathbb{Z}} := \operatorname{Lie}(D)_{\mathbb{Z}}$ forming a $\mathbb{Q}$–basis of $\mathfrak{g}_{\mathbb{Q}} := \operatorname{Lie}(G)_{\mathbb{Q}}$ also forms a $\mathbb{Z}$–basis of $\mathfrak{g}_{\mathbb{Z}}$.

Let $\pi\colon \mathfrak{M}_{\mathbb{C}} \to \mathfrak{N}_{\mathbb{C}}$ be the natural map. If we assume unimodularity, then $\mathfrak{M}_{\mathbb{C}}$ is a smooth scheme and π defines a proper $T \times \mathbb{S}$–equivariant resolution of singularities of $\mathfrak{N}_{\mathbb{C}}$. Together with the algebraic symplectic form on $\mathfrak{M}_{\mathbb{C}}$ arising from Hamiltonian reduction, this makes $\mathfrak{M}_{\mathbb{C}}$ a *symplectic resolution*.

Many elements of this paper make sense in the broader context of symplectic resolutions, although we will not press this point here. In the nonunimodular case, $\mathfrak{M}_{\mathbb{C}}$ may have orbifold singularities.

In the description given above, $\mathfrak{N}_{\mathbb{C}}$ appears as the Higgs branch of the $\mathcal{N}=4$ three-dimensional gauge theory attached to the representation of $T_{\mathbb{C}}$ on $\mathbb{C}^n$. However, it is more natural from the perspective of what is to follow to see $\mathfrak{N}_{\mathbb{C}}$ as the Coulomb branch of the theory attached to the dual action of $(D/T)^\vee$ on $\mathbb{C}^n$, in the sense of Braverman, Finkelberg and Nakajima [11; 32]. This leads to a different presentation of the hypertoric enveloping algebra, which will be useful for understanding its representation theory. In particular, the multiplicative hypertoric varieties we'll discuss later appear naturally from this perspective as the Coulomb branches of related 4–dimensional theories.

2.2 Quantizations

The ring of functions on the hypertoric variety $\mathfrak{N}_{\mathbb{Z}}$ has a quantization which we call the *hypertoric enveloping algebra*. We construct it by a quantum analogue of the Hamiltonian reduction that defines $\mathfrak{M}_{\mathbb{Z}}$. Consider the Weyl algebra W_n generated over $\mathbb{Z}$ by the elements $z_1, \dots, z_n, \partial_1, \dots, \partial_n$ modulo the relations

$$[z_i, z_j] = 0, \quad [\partial_i, \partial_j] = 0, \quad [\partial_i, z_j] = \delta_{ij}.$$

It is a quantization of the ring of functions on $T^*\mathbb{A}^n_{\mathbb{Z}}$. The torus D acts on W_n, scaling z_i by the character ϵ_i and ∂_i by ϵ_i^{-1}. It thus determines a decomposition into weight spaces

$$W_n = \bigoplus_{\boldsymbol{a} \in \mathbb{Z}^n} W_n[\boldsymbol{a}].$$

Let

$$h_i^+ := z_i \partial_i, \quad h_i^- := \partial_i z_i = h_i^+ + 1, \quad h_i^{\mathrm{mid}} := \tfrac{1}{2}(h_i^+ + h_i^-) = h_i^+ + \tfrac{1}{2} = h_i^- - \tfrac{1}{2}.$$

Each of the tuples $h_i^+, h_i^-, h_i^{\mathrm{mid}}$ generate the same subalgebra, ie the D–fixed subalgebra $\mathbb{Z}[h_i^{\pm}] = W_n[0]$.

Via the embedding $T \to D$, W_n carries an action of the torus T. To this action one can associate a *noncommutative moment map*, ie a map $\mu_q \colon \mathbb{Z}[\mathfrak{t}_{\mathbb{Z}}] \to W_n$ such that $[\mu_q(\chi), -]$ coincides with the action of the Lie algebra $\mathfrak{t}_{\mathbb{Z}}$. This property uniquely determines μ_q up to the addition of a character in $\mathfrak{t}^*_{\mathbb{Z}}$. We make the following choice:

$$\mu_q(\chi) := \sum_{i=1}^{n} \langle \epsilon_i, \chi \rangle h_i^+.$$

It's worth nothing that in the formula above, we have broken the symmetry between z_i and ∂_i; it would arguably be more natural to use h_i^{mid}, but this requires inserting a lot of annoying factors of $\frac{1}{2}$ into formulas, not to mention being a bit confusing in positive characteristic.

Definition 2.2 The hypertoric enveloping algebra $A_{\mathbb{Z}}$ is the subring $W_n^T \subset W_n$ invariant under T. We'll also consider the central quotients of this algebra associated to a character $\lambda \in \mathfrak{t}^*_{\mathbb{Z}}$, given by

$$A^\lambda_{\mathbb{Z}} := A_{\mathbb{Z}} / \langle \mu_q(\chi) - \lambda(\chi) \mid \chi \in \mathfrak{t}_{\mathbb{Z}} \rangle.$$

We will often abbreviate “hypertoric enveloping algebra” to HEA.

Let $A_{\mathbb{K}} := A_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{K}$ be the base change of this algebra to a commutative ring $\mathbb{K}$. The algebra $A_{\mathbb{C}}$ was studied extensively in [8; 31]. The algebra $A_{\mathbb{K}}$ when $\mathbb{K}$ has characteristic p was studied in work of Stadnik [35]. Fix a field $\mathbb{K}$ of characteristic p for the rest of the paper.

Unlike W_n itself, or its base change to a characteristic 0 field, the ring $W_n \otimes_{\mathbb{Z}} \mathbb{F}_p$ has a “big center” generated by the elements z_i^p, ∂_i^p. This central subring can be identified with the function ring $H^0(X^{(1)}, \mathscr{O}_{X^{(1)}})$, where $X = T^*\mathbb{A}^n_{\mathbb{F}_p}$.

2.3 Coulomb presentation

The algebra $A_{\mathbb{K}}$ has a different presentation which is more compatible with the subalgebra $\mathbb{K}[h_i^{\pm}]$. The action of D on $A_{\mathbb{K}}$ determines a decomposition into weight subspaces. Since $A_{\mathbb{K}} = W_n^T$, its weights lie in $\mathfrak{t}_{\mathbb{Z}}^{\perp} = \mathfrak{g}_{\mathbb{Z}}^*$:

$$A_{\mathbb{K}} = \bigoplus_{\boldsymbol{a} \in \mathfrak{t}_{\mathbb{Z}}^{\perp}} A_{\mathbb{K}}[\boldsymbol{a}].$$

For each $\boldsymbol{a} \in \mathfrak{t}_{\mathbb{Z}}^{\perp}$, we let

$$m(\boldsymbol{a}) := \prod_{a_i>0} z_i^{a_i} \prod_{a_i<0} \partial_i^{-a_i}. \tag{2-1}$$

Up to scalar multiplication, this is the unique element in $A_{\mathbb{K}}[\boldsymbol{a}]$ in of minimal degree.

Each weight space $A_{\mathbb{K}}[\boldsymbol{a}]$ is a module over the D–invariant subalgebra generated by the h_i^+. Let

$$[h_i]^{(a)} := \begin{cases} 1 & \text{if } a = 0, \\ z_i^a \partial_i^a = (h_i^- - 1)(h_i^- - 2)\cdots(h_i^- - a) & \text{if } a > 0, \\ \partial_i^{-a} z_i^{-a} = (h_i^+ + 1)(h_i^+ + 2)\cdots(h_i^+ - a) & \text{if } a < 0. \end{cases}$$

Theorem 2.3 [14, (6.21b)] *The algebra $A_{\mathbb{K}}$ is generated by $\mathbb{K}[h_1^{\pm}, \dots, h_n^{\pm}]$ and $m(\boldsymbol{a})$ for $\boldsymbol{a} \in \mathfrak{g}_{\mathbb{Z}}^*$, subject to the relations*

$$(h_i^{\pm} - a_i)m(\boldsymbol{a}) = m(\boldsymbol{a})h_i^{\pm}, \tag{2-2}$$

$$m(\boldsymbol{a})m(\boldsymbol{b}) = \prod_{\substack{a_i b_i < 0 \\ |a_i| \le |b_i|}} [h_i]^{(a_i)} \cdot m(\boldsymbol{a} + \boldsymbol{b}) \cdot \prod_{\substack{a_i b_i < 0 \\ |a_i| > |b_i|}} [h_i]^{(-b_i)}. \tag{2-3}$$

We call this is the *Coulomb* presentation, since it matches the presentation of the abelian Coulomb branch in [11, (4.7)], and shows that the algebra $A_{\mathbb{K}}$ can also be realized using this dual approach. As mentioned in the introduction, the techniques of this paper generalize to Coulomb branches with nonabelian gauge group as well, whereas it seems very challenging to generalize them to Higgs branches with nonabelian gauge group (that is, hyperkähler reductions by noncommutative groups).

2.4 Characteristic p localization

Following [35], in this section we exploit the large center of quantizations in characteristic p so as to relate modules over $A^\lambda_{\mathbb{K}}$ with coherent sheaves on $\mathfrak{M}^{(1)}_{\mathbb{K}}$. Roughly speaking, upon restriction to fibers of $\pi\colon \mathfrak{M}^{(1)}_{\mathbb{K}} \to \mathfrak{N}^{(1)}_{\mathbb{K}}$, the quantization becomes the algebra of endomorphisms of a vector bundle, and thus Morita-equivalent to the structure sheaf of the fiber.

Theorem 2.4 [35, Theorems 4.3.1 and 4.3.4] *For any $\lambda \in \mathfrak{t}^*_{\mathbb{F}_p}$, there exists a coherent sheaf $\mathscr{A}^\lambda$ of algebras Azumaya over the structure sheaf on $\mathfrak{M}^{(1)}_{\mathbb{K}}$ such that $\Gamma(\mathfrak{M}^{(1)}_{\mathbb{K}}, \mathscr{A}^\lambda) \cong A^\lambda_{\mathbb{K}}$.*

This theorem includes the existence of an injection $H^0(\mathfrak{N}^{(1)}_{\mathbb{K}}, \mathscr{O}_{\mathfrak{N}^{(1)}_{\mathbb{K}}}) \to A^\lambda_{\mathbb{K}}$; this is induced by the map $H^0((T^*\mathbb{A}^n_{\mathbb{K}})^{(1)}, \mathscr{O}_{(T^*\mathbb{A}^n_{\mathbb{K}})^{(1)}}) \to W_n \otimes \mathbb{K}$ sending

$$\mathsf{z}_i \mapsto z_i^p \quad \text{and} \quad \mathsf{w}_i \mapsto \partial_i^p.$$

Consider the moment map $\mu\colon \mathfrak{M}^{(1)}_{\mathbb{K}} \to \mathfrak{d}^{(1)}_{\mathbb{K}}$. Work of Stadnik shows that the Azumaya algebra $\mathscr{A}^\lambda$ splits on fibers of this map after field extension. Fix $\xi \in \mathfrak{d}^{(1)}_{\mathbb{K}}$. Possibly after extending $\mathbb{K}$, we can choose ν such that $\nu^p - \nu = \xi$, and define the splitting bundle as the quotient $\mathscr{A}^\lambda / \sum_{i=1}^n \mathscr{A}^\lambda (h_i^+ - \nu_i)$; this left module is already supported on the fiber $\mu^{-1}(\xi)$, since

$$(h_i^+ - \nu)^p - (h_i^+ - \nu) = z_i^p \partial_i^p - \nu^p + \nu = z_i^p \partial_i^p - \xi.$$

We can thicken this to the formal neighborhood of the fiber $\mu^{-1}(\hat{\xi})$ by taking the inverse limit $\mathscr{Q}_\nu := \varinjlim \mathscr{A}^\lambda / \sum_{i=1}^n \mathscr{A}^\lambda (h_i^+ - \nu_i)^N$.

Theorem 2.5 [35, Theorem 4.3.8] *The natural map*

$$\mathscr{A}^\lambda|_{\mu^{-1}(\hat{\xi})} \to \mathrm{End}_{\mathscr{O}_{\mathfrak{M}^{(1)}}}(\mathscr{Q}_\nu, \mathscr{Q}_\nu)$$

is an isomorphism.

The sheaf $\mathscr{A}^\lambda$ is not globally split; it has no global zero-divisor sections. It still has a close relationship with a tilting vector bundle on $\mathfrak{M}^{(1)}_{\mathbb{K}}$. We'll fix our attention on the case where $\xi = 0$, so $\nu_i \in \mathbb{F}_p$.

Let $\mathscr{T}_{\mathbb{K}}$ be an $\mathbb{S}$–equivariant locally free coherent sheaf on $\mathfrak{M}^{(1)}_{\mathbb{K}}$ such that $\mathscr{T}_{\mathbb{K}}|_{\mu^{-1}(\hat{0})} \cong \mathscr{Q}_\nu$. Such a sheaf exists by [25, Theorem 1.8(ii)]. As coherent sheaves, we have isomorphisms

$$\mathscr{A}^\lambda_{\mathbb{K}} \cong \mathrm{Fr}_* \mathscr{O}_{\mathfrak{M}_{\mathbb{K}}} \cong \mathscr{T}_{\mathbb{K}} \otimes \mathscr{T}^*_{\mathbb{K}} \cong \mathrm{End}(\mathscr{T}_{\mathbb{K}}).$$

By [35, Corollary 4.4.2], these sheaves have vanishing higher cohomology. Furthermore, combining with results of Kaledin [25, Theorem 1.4], this shows that:

Proposition 2.6 *For p sufficiently large and ν_i generic, the sheaf $\mathscr{T}_{\mathbb{K}}$ is a tilting generator on $\mathfrak{M}_{\mathbb{K}}$ and has a lift $\mathscr{T}_{\mathbb{Q}}$ which is a tilting generator on $\mathfrak{M}_{\mathbb{Q}}$; that is, $\mathrm{Ext}^i(\mathscr{T}_{\mathbb{Q}}, \mathscr{T}_{\mathbb{Q}}) = 0$ for $i > 0$, and $\mathrm{Ext}^i(\mathscr{T}_{\mathbb{Q}}, \mathscr{F}) = 0$ implies $\mathscr{F} = 0$ for any coherent sheaf on $\mathfrak{M}_{\mathbb{Q}}$.*

We will later calculate the sheaf $\mathscr{T}_{\mathbb{K}}$, once we understand $\mathscr{A}^\lambda_{\mathbb{K}}$ a bit better.

3 The representation theory of hypertoric enveloping algebras

3.1 Module categories and weight functors

Recall that we have a short exact sequence of tori $T \to D \to G$. $A^{\lambda}_{\mathbb{K}}$ is a quotient of W_n^T, and thus carries a residual action of G, which we will now use to study its modules.

Let $o \in \mathfrak{N}^{(1)}$ be the point defined by $z_i = w_i = 0$, ie the unique $\mathbb{K}$–valued $\mathbb{S}$–fixed point of $\mathfrak{N}^{(1)}$. The following category will play a central role in this paper.

Definition 3.1 Let $A^{\lambda}_{\mathbb{K}}\text{–mod}_o$ be the category of finitely generated $A^{\lambda}_{\mathbb{K}}$–modules that are set-theoretically supported at o when viewed as modules over $H^0(\mathfrak{N}^{(1)}_{\mathbb{K}}, \mathscr{O}_{\mathfrak{N}^{(1)}_{\mathbb{K}}})$.

In fact, we will first study the following closely related category.

Definition 3.2 Let $A^{\lambda}_{\mathbb{K}}\text{–mod}_o^D$ be the category of modules in $A^{\lambda}_{\mathbb{K}}\text{–mod}_o$ which are additionally endowed with a compatible D–action such that T acts via the character λ, and the action of $d_i \in \mathfrak{d}_{\mathbb{Z}}$ satisfies

$$(h_i^+ - d_i)^N v = 0 \quad \text{for } N \gg 0. \tag{3-1}$$

The difference $\mathsf{s}_i = h_i^+ - d_i$ acts centrally on such a module, since the adjoint action of h_i^+ on $A^{\lambda}_{\mathbb{K}}$ agrees with the action of d_i. The operator s_i is thus the nilpotent part of the Jordan decomposition of $h_i^{\pm}$. The operators s_i define an action of the polynomial ring $U_{\mathbb{K}}(\mathfrak{d})$, which factors through $U_{\mathbb{K}}(\mathfrak{g})$ since elements of $\mathfrak{t}$ act by zero. This extends to an action of the completion of $U_{\mathbb{K}}(\mathfrak{g})$, since s_i acts nilpotently by (3-1).

Definition 3.3 Let $S := U_{\mathbb{K}}(\mathfrak{g})$, and let $\widehat{S}$ be its completion at zero.

Let $\mathfrak{g}^{*,\lambda}_{\mathbb{Z}} \subset \mathfrak{d}^*_{\mathbb{Z}}$ be the $\mathfrak{g}^*_{\mathbb{Z}}$–coset of characters of D whose restriction to T coincides with λ. It indexes the D–weights which can occur in an object of $A^{\lambda}_{\mathbb{K}}\text{–mod}_o^D$.

We can construct projectives objects in a slight enlargement of $A^{\lambda}_{\mathbb{K}}\text{–mod}_o^D$ by working with the exact functors picking out weight spaces. That is, for each $\boldsymbol{a} \in \mathfrak{g}^{*,\lambda}_{\mathbb{Z}}$, we consider the functor which associates to an object $M \in A^{\lambda}_{\mathbb{K}}\text{–mod}_o^D$ the vector space

$$W_{\boldsymbol{a}}(M) := \{m \in M \mid m \text{ has } D\text{–weight } \boldsymbol{a}\}.$$

Note that even though we are working in characteristic p, the D–weights are valued in $\mathfrak{g}^{*,\lambda}_{\mathbb{Z}} \subset \mathbb{Z}^n$. This functor is exact, and we will show that it is pro-representable.

3.2 Projectives representing the weight functors

To construct the projective object that represents this functor, we consider the filtration of it by

$$W^N_{\boldsymbol{a}}(M) := \{m \in W_{\boldsymbol{a}}(M) \mid (h_i^+ - a_i)^N m = 0 \text{ for all } i\}.$$

Proposition 3.4 *We have a canonical isomorphism*

$$W_{\boldsymbol{a}}^N(M) \cong \operatorname{Hom}_{A_{\mathbb{K}}^{\lambda}\text{–mod}_o^D}\Big(A_{\mathbb{K}}^{\lambda} \Big/ \sum_{i=1}^{n} A_{\mathbb{K}}^{\lambda}(h_i^+ - a_i)^N, M\Big),$$

where D acts on $A_{\mathbb{K}}^{\lambda} / \sum_{i=1}^{n} A_{\mathbb{K}}^{\lambda}(h_i^+ - a_i)^N$ so that the image $1_{\boldsymbol{a}}$ of 1 has weight $\boldsymbol{a}$.

Since $W_{\boldsymbol{a}}(M) = \varinjlim W_{\boldsymbol{a}}^N(M)$, we have that $W_{\boldsymbol{a}}(M)$ is represented by the module

(3-2) $$Q_{\boldsymbol{a}} := \varprojlim A_{\mathbb{K}}^{\lambda} / A_{\mathbb{K}}^{\lambda}(h_i^+ - a_i)^N$$

with its induced D–action. Note that $Q_{\boldsymbol{a}} = \Gamma(\mu^{-1}(\widehat{0}); \mathscr{Q}_{\boldsymbol{a}})$.

This is endowed with the usual induced topology, and it is a pro-weight module in the sense that its weight spaces are pro-finite dimensional. This is a projective object in the category $\widehat{A_{\mathbb{K}}^{\lambda}\text{–mod}}^D$ of complete topologically finitely generated $A_{\mathbb{K}}^{\lambda}$–modules M with compatible D–action in the sense that

$$\lim_{N \to \infty} (h_i^+ - d_i)^N v = 0.$$

That is, s_i acts topologically nilpotently on each D–weight space. This is equivalent to (3-1) if the topology on M is discrete.

In the arguments below, Hom and End will be interpreted to mean continuous homomorphisms compatible with D; all objects in $A_{\mathbb{K}}^{\lambda}\text{–mod}_o^D$ will be given the discrete topology, so continuity is a trivial condition for homomorphisms between them.

Lemma 3.5 *If $\boldsymbol{b}$ is a character of D/T, then $W_{\boldsymbol{a}}(Q_{\boldsymbol{a}+\boldsymbol{b}}) \cong \widehat{S}$. Otherwise, this weight space is* 0.

Proof For any character $\boldsymbol{b}$ of D which vanishes on T, the $\boldsymbol{b}$ weight space $A_{\mathbb{K}}^{\lambda}[\boldsymbol{b}]$ is a free rank-one module over S (acting via multiplication by s_i), generated by $m(\boldsymbol{b})$. Thus, the $\boldsymbol{a} + \boldsymbol{b}$ weight space of $A_{\mathbb{K}}^{\lambda} / A_{\mathbb{K}}^{\lambda}(h_i^+ - a_i)^N$ is generated by $m(\boldsymbol{b})$, subject to the relations

$$\mathsf{s}_i^N m(\boldsymbol{b}) \cdot 1_{\boldsymbol{a}} = m(\boldsymbol{b}) \mathsf{s}_i^N \cdot 1_{\boldsymbol{a}} = m(\boldsymbol{b})(h_i^+ - a_i)^N \cdot 1_{\boldsymbol{a}} = 0,$$

and is thus free over the quotient ring $S / \sum S \cdot \mathsf{s}_i^N$. Taking the inverse limit, we see that every weight space of $Q_{\boldsymbol{a}}$ is a free module of rank one over $\widehat{S}$. □

Corollary 3.6 *We have an isomorphism of rings*

$$\operatorname{End}(Q_{\boldsymbol{a}}) \cong W_{\boldsymbol{a}}(Q_{\boldsymbol{a}}) \cong \widehat{S}.$$

Since $\widehat{S}$ is local, the module $Q_{\boldsymbol{a}}$ is indecomposable (in the category $\widehat{A_{\mathbb{K}}^{\lambda}\text{–mod}}^D$).

3.3 Isomorphisms between projectives

In this section, we determine the distinct isomorphism classes of weight functors, ie we determine all isomorphisms between the pro-projectives $Q_{\boldsymbol{a}}$. As we will see, there are typically many distinct weights $\boldsymbol{a} \in \mathfrak{g}_{\mathbb{Z}}^{*,\lambda}$ that give isomorphic functors.

By the results of the previous section, the space $W_{\boldsymbol{a}}(Q_{\boldsymbol{a}+\boldsymbol{b}}) = \mathrm{Hom}(Q_{\boldsymbol{a}}, Q_{\boldsymbol{a}+\boldsymbol{b}})$ is free of rank one over $\widehat{S}$, with generator $m(\boldsymbol{b})$. Likewise, $\mathrm{Hom}(Q_{\boldsymbol{a}+\boldsymbol{b}}, Q_{\boldsymbol{a}})$ is generated by $m(-\boldsymbol{b})$. Thus in order to verify whether $Q_{\boldsymbol{a}}$ and $Q_{\boldsymbol{a}+\boldsymbol{b}}$ are isomorphic, it is enough to check whether the composition $m(-\boldsymbol{b})m(\boldsymbol{b})$, viewed as an endomorphism of $Q_{\boldsymbol{a}}$, is an invertible element of the local ring $\mathrm{End}(Q_{\boldsymbol{a}}) \cong \widehat{S}$.

By (2-3), we have that

$$m(-\boldsymbol{b})m(\boldsymbol{b}) = \prod_{i=1}^{n} [h_i]^{(-b_i)},$$

where the right-hand side is a product of factors of the form $h_i^+ + k$ with k an integer between $\frac{1}{2}$ and $b_i + \frac{1}{2}$. To check whether $h_i^+ + k$ defines an invertible element of $\widehat{S}$, it is enough to compute its action on the weight space of weight $\boldsymbol{a}$, on which h_i acts by $a_i + \mathsf{s}_i$. The resulting endomorphism $h_i^+ + k = \mathsf{s}_i + (a_i + k)$ is invertible if and only if $k + a_i \not\equiv 0 \pmod p$.

The number of noninvertible factors (each equal to s_i) in $[h_i]^{(-b_i)}$ is therefore the number of integers k divisible by p lying between $a_i + \frac{1}{2}$ and $a_i + b_i + \frac{1}{2}$. We denote it by $\delta_i(\boldsymbol{a}, \boldsymbol{a} + \boldsymbol{b})$.

We can sum up the above computations as follows. Put

$$q(y,k) = \begin{cases} 1 & \text{if } k = 0, \\ \dfrac{1}{y+k} & \text{if } k \neq 0, \end{cases}$$

where y is a formal variable and $k \in \mathbb{K}$. Note that $q(\mathsf{s}_i, a_i + j)(h_i^+ + j)$ acts on a D–weight space of weight $\boldsymbol{a}$ by 1 if $a_i + j$ is not divisible by p, and by s_i if it is. Let

$$c_{\boldsymbol{a}}^{\boldsymbol{b}} = m(\boldsymbol{b}) \prod_{i=1}^{n} \prod_{j=1}^{b_i} q(\mathsf{s}_i, a_i + j) \in W_{\boldsymbol{a}}(Q_{\boldsymbol{a}+\boldsymbol{b}}) = \mathrm{Hom}(Q_{\boldsymbol{a}}, Q_{\boldsymbol{a}+\boldsymbol{b}}). \tag{3-3}$$

It is a generator of the $\widehat{S}$–module $W_{\boldsymbol{a}}(Q_{\boldsymbol{a}+\boldsymbol{b}})$. Note that this expression breaks the symmetry between positive and negative; if $b_i \le 0$ for all i, then $c_{\boldsymbol{a}}^{\boldsymbol{b}} = m(\boldsymbol{b})$, since all the products in the definition are over empty sets.

Lemma 3.7

$$c_{\boldsymbol{a}+\boldsymbol{b}}^{-\boldsymbol{b}} c_{\boldsymbol{a}}^{\boldsymbol{b}} = \prod_{i=1}^{n} \mathsf{s}_i^{\delta_i(\boldsymbol{a}, \boldsymbol{a}+\boldsymbol{b})}.$$

Proof We have

$$\begin{aligned} c_{\boldsymbol{a}+\boldsymbol{b}}^{-\boldsymbol{b}} c_{\boldsymbol{a}}^{\boldsymbol{b}} &= m(-\boldsymbol{b}) \cdot \prod_{i=1}^{n} \prod_{j=1}^{-b_i} q(\mathsf{s}_i, a_i + b_i + j) \cdot m(\boldsymbol{b}) \cdot \prod_{i=1}^{n} \prod_{j=1}^{b_i} q(\mathsf{s}_i, a_i + j) \\ &= m(-\boldsymbol{b}) m(\boldsymbol{b}) \prod_{i=1}^{n} \Bigl(\prod_{j=1}^{-b_i} q(\mathsf{s}_i, a_i + b_i + j) \cdot \prod_{j=1}^{b_i} q(\mathsf{s}_i, a_i + j) \Bigr) \\ &= \prod_{i=1}^{n} [h_i]^{(-b_i)} \prod_{j=1}^{-b_i} q(\mathsf{s}_i, a_i + b_i + j) \prod_{j=1}^{b_i} q(\mathsf{s}_i, a_i + j). \end{aligned}$$

Note that for each index i only one of the products is nonunital, depending on the sign, and in either case, we obtain the product of $q(\mathsf{s}_i, a_i + j)$ ranging over integers lying between $a_i + \frac{1}{2}$ and $a_i + b_i + \frac{1}{2}$. As we noted earlier, $[h_i]^{(-b_i)}$ is the product of $h_i + j$ with j ranging over this set. Thus, we obtain the product over this same set of $(h_i + j)q(\mathsf{s}_i, a_i + j)$, which is precisely $\mathsf{s}_i^{\delta_i(\boldsymbol{a},\boldsymbol{a}+\boldsymbol{b})}$. □

It remains for us to describe which pairs $\boldsymbol{a}, \boldsymbol{a}'$ satisfy $\delta_i(\boldsymbol{a}, \boldsymbol{a}') = 0$ for all i, and thus index isomorphic projective modules.

Definition 3.8 Let $\mathfrak{A}_\lambda^{\mathrm{per}}$ be the periodic hyperplane arrangement in $\mathfrak{g}_{\mathbb{Z}}^{*,\lambda}$ defined by the hyperplanes $d_i = kp - \frac{1}{2}$ for $k \in \mathbb{Z}$ and $i = 1, \dots, n$.

By definition, $\delta_i(\boldsymbol{a}, \boldsymbol{a}')$ is the minimal number of hyperplanes $d_i = kp - \frac{1}{2}$ crossed when traveling from $\boldsymbol{a}$ to $\boldsymbol{a}'$. Given $\boldsymbol{x} \in \mathbb{Z}^n$, let

$$\Delta_{\boldsymbol{x}} = \{\boldsymbol{a} \in \mathfrak{g}_{\mathbb{Z}}^{*,\lambda} \mid px_i \le a_i < px_i + p\} \quad \text{and} \quad \Delta_{\boldsymbol{x}}^{\mathbb{R}} = \{\boldsymbol{a} \in \mathfrak{g}_{\mathbb{Z}}^{*,\lambda} \otimes \mathbb{R} \mid px_i \le a_i < px_i + p\}.$$

We have shown:

Theorem 3.9 *We have an isomorphism $Q_{\boldsymbol{a}} \cong Q_{\boldsymbol{a}'}$ if and only if we have $\boldsymbol{a}, \boldsymbol{a}' \in \Delta_{\boldsymbol{x}}$ for some $\boldsymbol{x}$.*

Let

$$\widetilde{\Lambda}(\lambda) = \{\boldsymbol{x} \in \mathbb{Z}^n \mid \Delta_{\boldsymbol{x}} \neq \varnothing\} \quad \text{and} \quad \widetilde{\Lambda}^{\mathbb{R}}(\lambda) = \{\boldsymbol{x} \in \mathbb{Z}^n \mid \Delta_{\boldsymbol{x}}^{\mathbb{R}} \neq \varnothing\}.$$

Thus, $\widetilde{\Lambda}(\lambda)$ canonically parametrizes the set of indecomposable projective modules in the pro-completion of $A_{\mathbb{K}}^{\lambda}\text{–mod}_o^{\boldsymbol{D}}$. It follows that $\widetilde{\Lambda}(\lambda)$ also canonically parametrizes the simple modules in this category.

Let us call the parameter λ *smooth* if there is a neighborhood U of λ in $\mathbb{R} \otimes \mathfrak{g}_{\mathbb{Z}}^{*,\lambda}$ such that for all $\lambda' \in U$, we have $\widetilde{\Lambda}(\lambda) = \widetilde{\Lambda}^{\mathbb{R}}(\lambda')$. In particular, if λ is smooth, then the hyperplanes in $\mathfrak{A}_\lambda^{\mathrm{per}}$ must intersect generically.

3.4 A taxicab metric

We can endow $\widetilde{\Lambda}(\lambda)$ with a metric given by the taxicab distance $|\boldsymbol{x} - \boldsymbol{y}|_1 = \sum_i |x_i - y_i|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \widetilde{\Lambda}(\lambda)$. We can add a graph structure to $\widetilde{\Lambda}(\lambda)$ by adding in a pair of edges between any two chambers satisfying $|\boldsymbol{x} - \boldsymbol{y}|_1 = 1$; generically, this is the same as requiring that $\Delta_{\boldsymbol{x}}^{\mathbb{R}}$ and $\Delta_{\boldsymbol{y}}^{\mathbb{R}}$ are adjacent across a hyperplane.

We say that this adjacency is *across* i if $\boldsymbol{x}$ and $\boldsymbol{y}$ differ in the i^{th} coordinate. For every $\boldsymbol{x}$, let $\alpha(\boldsymbol{x})$ be the set of neighbors of $\boldsymbol{x}$ in $\widetilde{\Lambda}(\lambda)$. Generically, this is the same as the number of facets of $\Delta_{\boldsymbol{x}}^{\mathbb{R}}$; we let $\alpha_i(\boldsymbol{x}) \subset \alpha(\boldsymbol{x})$ be those facets adjacent across i. Note that in some degenerate cases, we may have that $\boldsymbol{x}, \boldsymbol{x} + \epsilon_i, \boldsymbol{x} - \epsilon_i \in \widetilde{\Lambda}(\lambda)$, so the size of $\alpha_i(\boldsymbol{x})$ is typically 0 or 1, but could be 2.

3.5 Weights of simple modules

Definition 3.10 For any $\boldsymbol{x} \in \mathbb{Z}^n$ such that $\Delta_{\boldsymbol{x}} \neq \varnothing$, we let $P_{\boldsymbol{x}} := Q_{\boldsymbol{b}}$ for some $\boldsymbol{b} \in \Delta_{\boldsymbol{x}}$.

Lemma 3.11 *The module $P_{\boldsymbol{x}}$ has a unique simple quotient $L_{\boldsymbol{x}}$, and the $L_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \mathbb{Z}^n$ such that $\Delta_{\boldsymbol{x}} \neq \varnothing$ are a complete irredundant list of simple modules in $A^{\lambda}_{\mathbb{K}}\text{–mod}^{D}_{o}$.*

Furthermore, the $\boldsymbol{a}$–weight space of $L_{\boldsymbol{x}}$ is one-dimensional if $\boldsymbol{a} \in \Delta_{\boldsymbol{x}}$, and 0 otherwise.

Proof We show that $Q_{\boldsymbol{a}}$ has a unique simple quotient by showing that the sum of two proper submodules is proper; this then shows that there is a unique maximal proper submodule, and $L_{\boldsymbol{x}}$ is the quotient by it. A submodule $M \subset Q_{\boldsymbol{a}}$ is proper if and only if $W_{\boldsymbol{a}}(M) \subset W_{\boldsymbol{a}}(Q_{\boldsymbol{a}}) \cong \widehat{S}$ is a proper submodule; that is, if it lies in the unique maximal ideal $\mathfrak{m} \subset \widehat{S}$. This shows that the sum of two proper submodules is proper, and so $L_{\boldsymbol{x}}$ is well-defined.

Using the isomorphism $Q_{\boldsymbol{a}} \cong Q_{\boldsymbol{b}}$ if $\boldsymbol{a}, \boldsymbol{b} \in \Delta_{\boldsymbol{x}}$, we can extend this to the observation that a submodule $M \subset P_{\boldsymbol{x}}$ is proper if and only if $W_{\boldsymbol{a}}(M) \subset \mathfrak{m} W_{\boldsymbol{a}}(P_{\boldsymbol{x}})$ for all $\boldsymbol{a} \in \Delta_{\boldsymbol{x}}$.

By Lemma 3.7, we can check that there is a unique submodule M in $P_{\boldsymbol{x}}$ such that

$$W_{\boldsymbol{a}}(M) = \begin{cases} \mathfrak{m} W_{\boldsymbol{a}}(P_{\boldsymbol{x}}) & \text{if } \boldsymbol{a} \in \Delta_{\boldsymbol{x}}, \\ W_{\boldsymbol{a}}(P_{\boldsymbol{x}}) & \text{if } \boldsymbol{a} \notin \Delta_{\boldsymbol{x}}. \end{cases}$$

By the observation above, this must be the maximal proper submodule, so $L_{\boldsymbol{x}} = P_{\boldsymbol{x}}/M$. This shows that $L_{\boldsymbol{x}}$ has the claimed dimensions of weight spaces. Furthermore, this shows that we can recover the set $\Delta_{\boldsymbol{x}}$ for $L_{\boldsymbol{x}}$, so we must have $L_{\boldsymbol{x}} \ncong L_{\boldsymbol{y}}$ if $\boldsymbol{x} \neq \boldsymbol{y}$.

For any simple L, we must have $W_{\boldsymbol{a}}(L) \neq 0$ for some $\boldsymbol{a}$. This induces a map $P_{\boldsymbol{x}} \to L$ where $\boldsymbol{a} \in \Delta_{\boldsymbol{x}}$. Since $L_{\boldsymbol{x}}$ is the unique simple quotient of $P_{\boldsymbol{x}}$, this shows that $L_{\boldsymbol{x}} \cong L$. This shows that they give a complete list and completes the proof. □

Example 3.12 An interesting example to keep in mind is the following. Let T be the scalar matrices acting on $\mathbb{A}^3$. In this case, $n = 3$ and $k = 1$. The space $\mathfrak{g}^{*,\lambda}_{\mathbb{Z}}$ is an affine space on which d_1 and d_2 give a set of coordinates, with d_3 related by the relation $d_3 = -d_1 - d_2 + \lambda$ for some $\lambda \in \mathbb{Z}$. Thus, the hyperplane arrangement that interests us is given by

$$d_1 = kp - \tfrac{1}{2}, \quad d_2 = kp - \tfrac{1}{2}, \quad -d_1 - d_2 + \lambda = kp - \tfrac{1}{2}.$$

In particular, we have that $\Delta_{\boldsymbol{x}} \neq \varnothing$ if and only if there exist integers a_1 and a_2 such that

$$x_1 p \le a_1 < x_1 p + p, \quad x_2 p \le a_2 < x_2 p + p, \quad x_3 p \le -a_1 - a_2 + \lambda < x_3 p + p.$$

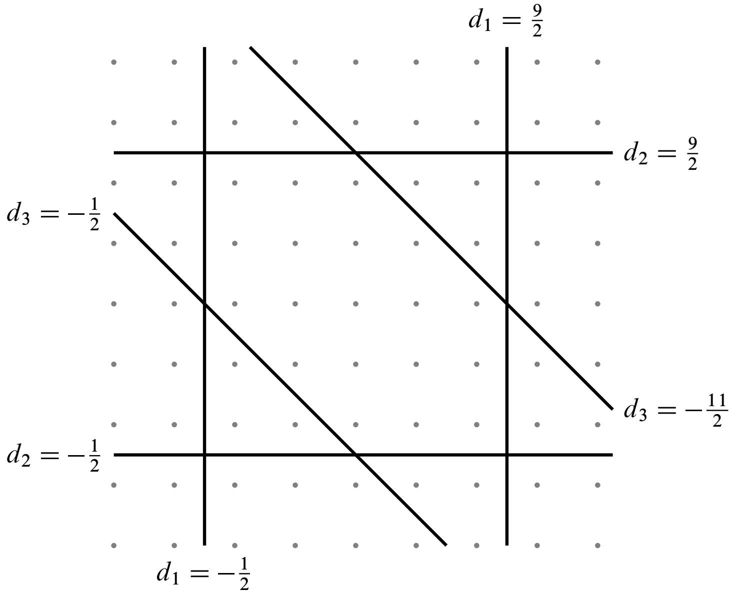
The values of $-a_1 - a_2 + \lambda$ for a_1, a_2 satisfying the first two inequalities range from $-(x_1+x_2+2)p+2+\lambda$ to $-(x_1+x_2)p+\lambda$. Thus, x_3 is a possibility if $-x_1-x_2-2+\lfloor(\lambda+3)/p\rfloor x_3 \le -x_1-x_2+\lfloor\lambda/p\rfloor$. Thus, there are three such x_3 if $\lfloor(\lambda+3)/p\rfloor = \lfloor\lambda/p\rfloor$, that is, if $\lambda \not\equiv -1, -2 \bmod p$. If $\lambda \equiv -1, -2 \bmod p$, then there are two, and the parameter λ is not smooth.

Of course, the numbers -1 and -2 have another significance in terms of $\mathbb{P}^2$: the line bundles $\mathscr{O}(-1)$ and $\mathscr{O}(-2)$ on $\mathbb{P}^2$ are the unique ones that have trivial pushforward. This is not coincidence. Let λ_+ be the unique integer in the range $0 \le \lambda_+ < p$ congruent to $\lambda \pmod p$, and λ_- the unique such integer in $-p \le \lambda_- < 0$. The simples $(x_1, x_2, -x_1 - x_2 + \lfloor \lambda/p \rfloor)$ and $(x_1, x_2, -x_1 - x_2 - 2 + \lfloor \lambda/p \rfloor)$ over $A^{\lambda}_{\mathbb{K}}$ can be identified with $H^0(\mathbb{P}^2; \mathscr{O}(\lambda_+))$ and $H^1(\mathbb{P}^2; \mathscr{O}(\lambda_-))$. If $\lambda \cong -1, -2 \bmod p$, then the latter group is trivial, so one of the simple representations is "missing".

The final simple can be identified with the first cohomology of the kernel of the map $\mathscr{O}(\lambda_+)^{\oplus 3} \to \mathscr{O}(\lambda_+ + p)$ defined by (z_1^p, z_2^p, z_3^p) (in characteristic p, this is a map of twisted D–modules); this map is surjective on sheaves, but *injective* on sections, with the desired simple module its cokernel.

Let's assume for simplicity that $0 \le \lambda \le p - 3$. In this case, the "picture" of these representations when $p = 5$ and $\lambda = 1$ is as follows:

(3-4)



The three chambers shown (read SW to NE) are $\Delta_{(0,0,0)}$, $\Delta_{(0,0,-1)}$, $\Delta_{(0,0,-2)}$.

3.6 The endomorphism algebra of a projective generator

Having developed this structure theory, we can easily give a presentation of our category. For each pair $\boldsymbol{x}, \boldsymbol{y}$ with $\Delta_{\boldsymbol{x}} \neq \varnothing$ and $\Delta_{\boldsymbol{y}} \neq \varnothing$, we can define $c_{\boldsymbol{x},\boldsymbol{y}}$ to be $c_{\boldsymbol{a}}^{\boldsymbol{a}'-\boldsymbol{a}}$ for $\boldsymbol{a} \in \Delta_{\boldsymbol{y}}$ and $\boldsymbol{a}' \in \Delta_{\boldsymbol{x}}$. For each i, let $\eta_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}) = \frac{1}{2}(|x_i - y_i| + |y_i - u_i| - |x_i - u_i|)$.

Theorem 3.13 *The algebra $\bigoplus_{\boldsymbol{x},\boldsymbol{y} \in \widetilde{\Lambda}} \operatorname{Hom}(P_{\boldsymbol{x}}, P_{\boldsymbol{y}})$ is generated by the idempotents $1_{\boldsymbol{x}}$ and the elements $c_{\boldsymbol{x},\boldsymbol{y}}$ over $\widehat{S}$, modulo the relation*

(3-5)
$$c_{\boldsymbol{x},\boldsymbol{y}} c_{\boldsymbol{y},\boldsymbol{u}} = \prod_i \mathsf{s}_i^{\eta_i(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u})} c_{\boldsymbol{x},\boldsymbol{u}}.$$

Note that this relation is homogeneous if $\deg c_{\boldsymbol{x},\boldsymbol{y}} = |\boldsymbol{x} - \boldsymbol{y}|_1$ and $\deg \mathsf{s}_i = 2$.

Proof The relation holds by an easy extension of Lemma 3.7. To see that these elements and relations are sufficient, note that in the algebra $\widehat{H}$ with this presentation, the Hom-space $1_{\boldsymbol{x}}\widehat{H}1_{\boldsymbol{y}}$ is cyclically generated over $\widehat{S}$ by $c_{\boldsymbol{x},\boldsymbol{y}}$. The image of $c_{\boldsymbol{x},\boldsymbol{y}}$ under the induced map $1_{\boldsymbol{x}}\widehat{H}1_{\boldsymbol{y}} \to \operatorname{Hom}(P_{\boldsymbol{y}}, P_{\boldsymbol{x}})$ generates the target space over $\widehat{S}$. Since the target is free of rank 1 as a $\widehat{S}$–module, the map must be an isomorphism. □

Definition 3.14 Let $S_{\mathbb{Z}} := U_{\mathbb{Z}}(\mathfrak{g})$. Let $\widetilde{H}^{\lambda}_{\mathbb{Z}}$ be the graded algebra over $S_{\mathbb{Z}}$ generated by $1_{\boldsymbol{x}}$ and $c_{\boldsymbol{x},\boldsymbol{y}}$, with presentation given in Theorem 3.13. Let $\widetilde{H}^{\lambda}_{\mathbb{K}} := \widetilde{H}^{\lambda}_{\mathbb{Z}} \otimes \mathbb{K}$.

This algebra is isomorphic to its opposite via the anti-isomorphism which acts by the identity on $S_{\mathbb{Z}}$ and $c_{\boldsymbol{x},\boldsymbol{y}} \mapsto c_{\boldsymbol{y},\boldsymbol{x}}$. Since this algebra has a left action on the sum $\bigoplus_{\boldsymbol{x}} P_{\boldsymbol{x}}$, it naturally has a *right* action on $\bigoplus_{\boldsymbol{x}} \operatorname{Hom}(P_{\boldsymbol{x}}, M)$ for any $A^{\lambda}_{\mathbb{K}}$–module, which we will turn into a left module structure using the anti-automorphism above.

It may concern the reader that $\widetilde{H}^{\lambda}_{\mathbb{K}}$ is not a unital algebra, but it has a structure which can serve as a replacement. We follow the terminology and notation of [13] in this section. We call a $\mathbb{K}$–algebra A *locally unital* if there are idempotents 1_α indexed by some set $\aleph$ such that $A = \bigoplus_{\alpha,\beta\in\aleph} 1_\alpha A 1_\beta$.

Definition 3.15 Given a locally unital algebra A, let $\mathscr{P}(A)$ be the category where the objects are the set $\aleph$, and the morphism spaces are given by $\operatorname{Hom}(\alpha,\beta) = 1_\alpha A 1_\beta$.

Note that $\mathscr{P}$ is equivalent to the subcategory of left projective modules with objects $A1_\alpha$.

We call a module M over A *locally unital* if $M = \bigoplus_{\alpha\in\aleph} 1_\alpha M$; this is automatic if $\mathbb{K}$ is a field and M is finite-dimensional over $\mathbb{K}$. We can think of $\alpha \mapsto 1_\alpha M$ as a functor $\mathscr{P}^{\mathrm{op}} \to \mathbb{K}$–mod, and conversely, every locally unital left A–module arises from a unique such functor. In particular, the results of [29], which are formulated in terms of representations of categories, also apply to locally unital algebras.

The algebra $\widetilde{H}^{\lambda}_{\mathbb{K}}$ may not be left or right Noetherian as a ring, since it is not finitely generated as a module over itself. However, it can be *locally left Noetherian*,[1] meaning that left submodules of $A1_{\boldsymbol{x}}$ are finitely generated.

Proposition 3.16 *If $\mathbb{K}$ is Noetherian, then the algebra $\widetilde{H}^{\lambda}_{\mathbb{K}}$ is locally left Noetherian.*

Proof Consider a submodule $U \subset \widetilde{H}^{\lambda}_{\mathbb{K}}1_{\boldsymbol{x}}$. The intersection $U \cap 1_{\boldsymbol{y}}\widetilde{H}^{\lambda}_{\mathbb{K}}1_{\boldsymbol{x}}$ must be of the form $I_{\boldsymbol{y}}c_{\boldsymbol{y},\boldsymbol{x}}$ for some ideal $I_{\boldsymbol{y}} \subset S_{\mathbb{K}}$. Since $c_{\boldsymbol{z},\boldsymbol{y}}I_{\boldsymbol{y}}c_{\boldsymbol{y},\boldsymbol{x}} \subset I_{\boldsymbol{z}}c_{\boldsymbol{z},\boldsymbol{x}}$, we have $I_{\boldsymbol{y}} \subset I_{\boldsymbol{z}}$ if for each i, either $x_i \le y_i \le z_i$ or $z_i \le y_i \le x_i$. For any subset B of $\mathbb{Z}^n_{\ge 0}$, there is a finite list of points $b^{(1)}, \dots, b^{(r)}$ such that for any $b \in B$, there is some r such that $b_i \ge b_i^{(r)}$ for all i. This means that for any finitely generated ideal $I \subset S_{\mathbb{K}}$ and any subset $B \subset \widetilde{\Lambda}$, the submodule generated by $Ic_{\boldsymbol{y},\boldsymbol{x}}$ for $\boldsymbol{y} \in B$ is finitely generated, since it is generated by $Ic_{\boldsymbol{y},\boldsymbol{x}}$ for finitely many choices of $\boldsymbol{y}$. Since $\mathbb{K}$ is Noetherian, so is $S_{\mathbb{K}}$, and thus every ideal in I is finitely generated.

[1] This is not identical to the notion of "locally Noetherian" found in scheme theory, but is related: the spectrum of a commutative locally Noetherian ring will be a possibly infinite disjoint union of Noetherian schemes, which is thus locally Noetherian as a scheme.

Thus, if U is not finitely generated, then infinitely many different ideals must appear as $I_{\boldsymbol{y}}$. By standard methods, we can choose an infinite sequence $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \dots$ such that the ideals $I_{\boldsymbol{y}^{(i)}}$ are all different, and for each i, the difference of coordinates $y_i^{(k)} - x_i$ are either all positive and weakly increasing with respect to k, or negative and weakly decreasing. In either case, we have $I_{\boldsymbol{y}^{(1)}} \subset I_{\boldsymbol{y}^{(2)}} \subset \cdots$. Since $S_{\mathbb{K}}$ is Noetherian, the existence of such a chain of ideals which are all distinct contradicts the ascending chain condition, proving that U is finitely generated. □

If an algebra A is locally left Noetherian, then its category of finitely generated, locally unital modules is an abelian category A–lu-mod. The objects $A1_{\alpha_i}$ form a *nearly resolving set of projectives* in the sense of Freyd [18, Section 1], ie every object in this category is a quotient of a finite sum of these projectives.

Lemma 3.17 *Assume A is locally Noetherian. If $\mathscr{C}$ is an abelian category and $\alpha \mapsto P_\alpha \colon \mathscr{P} \to \mathscr{C}$ is a fully faithful functor such that the set of projectives $\{P_\alpha\}_{\alpha\in\aleph}$ is nearly resolving, then the functor*

$$\mathsf{M}\colon \mathscr{C} \to A\text{–lu-mod}, \quad \mathsf{M}(M) = \bigoplus_{\alpha\in\aleph} \mathrm{Hom}(P_\alpha, M),$$

is an equivalence.

Proof In the terms of [18], this functor M sends an object in $\mathscr{C}$ to the corresponding representation of the category $\mathscr{P}^{\mathrm{op}}$. By a small modification of [18, Theorem 1.2] (stated above [18, Theorem 1.3], with the proof left to the reader), this functor is an equivalence to the subcategory of representations which are the cokernel of a map of the form $\mathsf{M}\big(\bigoplus_{i=1}^r P_{\alpha_i}\big) \to \mathsf{M}\big(\bigoplus_{j=1}^s P_{\beta_j}\big)$, that is of the form $\bigoplus_{i=1}^r A1_{\alpha_i} \to \bigoplus_{j=1}^s A1_{\beta_j}$. Since A is locally Noetherian, the modules of this form are exactly the finitely generated, locally unital modules. □

Let $\widetilde{H}^\lambda_{\mathbb{K}}$–mod$_o$ denote the category of finite-dimensional representations of $\widetilde{H}^\lambda_{\mathbb{K}}$, on which each s_i acts nilpotently. As discussed above, such a module is necessarily locally unital.

Theorem 3.18 *The functor*

$$\bigoplus_{\boldsymbol{x}\in\widetilde{\Lambda}(\lambda)} \mathrm{Hom}(P_{\boldsymbol{x}}, -)\colon A^\lambda_{\mathbb{K}}\text{–mod}^D_o \to \widetilde{H}^\lambda_{\mathbb{K}}\text{–mod}_o$$

defines an equivalence of categories between $A^\lambda_{\mathbb{K}}$–mod^D_o and the category of finite-dimensional representations of $\widetilde{H}^\lambda_{\mathbb{K}}$, on which each s_i acts nilpotently.

Proof First, consider the category $\widehat{A^\lambda_{\mathbb{K}}\text{–mod}}^D$. Since any module M in this category is topologically finitely generated over $A^\lambda_{\mathbb{K}}$, we can assume that the generators are generalized weight vectors for D. These weight vectors induce a surjection $\bigoplus_{i=1}^k Q_{\boldsymbol{a}_i} \to M$. This shows that the $\{P_{\boldsymbol{x}}\}$ are a nearly resolving set of projectives in this category, and we have an equivalence of the category $\widehat{A^\lambda_{\mathbb{K}}\text{–mod}}^D$ to the category of modules over the completion $\widehat{H}^\lambda_{\mathbb{K}} \cong \widetilde{H}^\lambda_{\mathbb{K}} \otimes_S \widehat{S}$ by Lemma 3.17; note that we have used the anti-automorphism of $\widetilde{H}^\lambda_{\mathbb{K}}$ to switch between left and right modules.

Now, we will show that restricting this functor gives the desired equivalence. A finite-dimensional representation of $\widetilde{H}^{\lambda}_{\mathbb{K}}$ on which each s_i acts nilpotently can be inflated to a $\widehat{H}^{\lambda}_{\mathbb{K}}$–module, and thus sent to a $A^{\lambda}_{\mathbb{K}}$–module by this equivalence. The nilpotent condition and finite dimensionality imply that this module is a sum of finitely many generalized weight spaces, so it is supported on a finite union of points in $\mathfrak{N}^{(1)}_{\mathbb{K}}$. The functions z_i and w_i must act nilpotently for weight reasons, so the only point in the support must be o. On the other hand, if the corresponding $A^{\lambda}_{\mathbb{K}}$–module is supported on o, then by coherence, it must be finite-dimensional, and thus give a finite-dimensional $\widehat{H}^{\lambda}_{\mathbb{K}}$–module, and s_i acts nilpotently on any finite-dimensional $\widehat{H}^{\lambda}_{\mathbb{K}}$–module. □

In fact, we will see that when λ is smooth, $\widetilde{H}^{\lambda}_{\mathbb{K}}$ admits a presentation as a quadratic algebra. We begin by producing some generators.

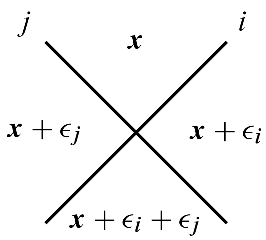
Let $\epsilon_i = (0, \dots, 0, 1, 0, \dots, 0)$ by the i^{th} unit vector. Let $c^{\pm i}_{\boldsymbol{x}} = c_{\boldsymbol{x}\pm\epsilon_i,\boldsymbol{x}}$; note that $\deg c^{\pm i}_{\boldsymbol{x}} = 1$. These elements correspond to the adjacencies in the graph structure of $\widetilde{\Lambda}(\lambda)$. Thus, we have a homomorphism from the path algebra of $\widetilde{\Lambda}(\lambda)$ sending each length 0 path to the corresponding $1_{\boldsymbol{x}}$ and each edge to the corresponding $c^{\pm i}_{\boldsymbol{x}}$.

We'll be interested in the particular cases of (3-5) which relate these length 1 paths. If $\boldsymbol{x}, \boldsymbol{x}+\epsilon_i \in \widetilde{\Lambda}(\lambda)$, then

$$c^{-i}_{\boldsymbol{x}+\epsilon_i} c^{+i}_{\boldsymbol{x}} = \mathsf{s}_i 1_{\boldsymbol{x}+\epsilon_i} \quad \text{and} \quad c^{+i}_{\boldsymbol{x}} c^{-i}_{\boldsymbol{x}+\epsilon_i} = \mathsf{s}_i 1_{\boldsymbol{x}+\epsilon_i}. \tag{3-6a}$$

We can view this as saying that the length 2 paths that cross a hyperplane and return satisfy the same linear relations as the normal vectors to the corresponding hyperplanes.

If $\boldsymbol{x}, \boldsymbol{x}+\epsilon_i, \boldsymbol{x}+\epsilon_j, \boldsymbol{x}+\epsilon_i+\epsilon_j \in \widetilde{\Lambda}(\lambda)$, then the corresponding chambers fit together as pictured:



In this situation, we find that either way of going around the codimension 2 subspace gives the same result, and that more generally any two paths between chambers that never cross the same hyperplane twice give equal elements of the algebra:

$$c^{+j}_{\boldsymbol{x}+\epsilon_i} c^{+i}_{\boldsymbol{x}} = c^{+i}_{\boldsymbol{x}+\epsilon_j} c^{+j}_{\boldsymbol{x}}, \quad c^{-j}_{\boldsymbol{x}+\epsilon_j} c^{-i}_{\boldsymbol{x}+\epsilon_i+\epsilon_j} = c^{-i}_{\boldsymbol{x}+\epsilon_i} c^{-j}_{\boldsymbol{x}+\epsilon_i+\epsilon_j}, \tag{3-6b}$$

$$c^{-j}_{\boldsymbol{x}+\epsilon_i+\epsilon_j} c^{+i}_{\boldsymbol{x}+\epsilon_j} = c^{+i}_{\boldsymbol{x}} c^{-j}_{\boldsymbol{x}+\epsilon_j}, \quad c^{-i}_{\boldsymbol{x}+\epsilon_i+\epsilon_j} c^{+j}_{\boldsymbol{x}+\epsilon_i} = c^{+j}_{\boldsymbol{x}} c^{-i}_{\boldsymbol{x}+\epsilon_i}. \tag{3-6c}$$

If λ is a smooth parameter, then as the following theorem shows, these are the only relations needed.

Theorem 3.19 *If λ is a smooth parameter, then the algebra $\bigoplus_{\boldsymbol{x},\boldsymbol{y}} \operatorname{Hom}(P_{\boldsymbol{x}}, P_{\boldsymbol{y}})$ is generated by the idempotents $1_{\boldsymbol{x}}$ and the elements $c^{\pm i}_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \widetilde{\Lambda}(\lambda)$ over $\widehat{S}$ modulo the relations* (3-6a)–(3-6c).

Proof Since these relations are a consequence of Theorem 3.13, it suffices to show that the elements $c_{\boldsymbol{x}}^{\pm i}$ generate, and that the relations (3-5) are a consequence of (3-6a)–(3-6c).

We show that $c_{\boldsymbol{x}}^{\pm i}$ generate $c_{\boldsymbol{x},\boldsymbol{y}}$ by induction on the L_1–norm $|\boldsymbol{x}-\boldsymbol{y}|$. If $|\boldsymbol{x}-\boldsymbol{y}|_1=1$, then $c_{\boldsymbol{x},\boldsymbol{y}}=c_{\boldsymbol{y}}^{\pm i}$. On the other hand, if $|\boldsymbol{x}-\boldsymbol{y}|_1>1$, then there is some $\boldsymbol{x}'\neq \boldsymbol{x},\boldsymbol{y}$ such that $|\boldsymbol{x}-\boldsymbol{x}'|_1+|\boldsymbol{x}'-\boldsymbol{y}|_1=|\boldsymbol{x}-\boldsymbol{y}|_1$. Choosing a generic parameter λ' such that $\widetilde{\Lambda}^{\mathbb{R}}(\lambda')=\widetilde{\Lambda}(\lambda)$, we can consider the line segment joining generic points in $\overline{\Delta}_{\boldsymbol{x}}$ and $\overline{\Delta}_{\boldsymbol{y}}$, and let $\boldsymbol{x}'$ be any chamber this line segment passes through. The smoothness hypothesis is needed to conclude that there is such a chamber that lies in $\widetilde{\Lambda}(\lambda)$. Since $c_{\boldsymbol{x},\boldsymbol{y}}=c_{\boldsymbol{x},\boldsymbol{x}'}c_{\boldsymbol{x}',\boldsymbol{y}}$, this proves generation by induction.

We must now check that the relations (3-5) are satisfied. First, consider the situation where $\boldsymbol{x}^{(0)}=\boldsymbol{x},\dots,$ $\boldsymbol{x}^{(m)}=\boldsymbol{y}$ is a path with $|\boldsymbol{x}^{(i)}-\boldsymbol{x}^{(i+1)}|_1=1$ with $\boldsymbol{x}^{(i)}\in\widetilde{\Lambda}(\lambda)$, and $\boldsymbol{y}^{(0)}=\boldsymbol{x},\dots,\boldsymbol{y}^{(m)}=\boldsymbol{y}$ is a path with the same conditions. These two paths differ by a finite number of applications of the relations (3-6b)–(3-6c).

It remains to show that if $\boldsymbol{x}^{(0)}=\boldsymbol{x},\dots,\boldsymbol{x}^{(m)}=\boldsymbol{y}$ is a path of minimal length between these points with $|\boldsymbol{x}^{(i)}-\boldsymbol{x}^{(i+1)}|_1=1$, and we have similar paths $\boldsymbol{y}^{(0)}=\boldsymbol{y},\dots,\boldsymbol{y}^{(n)}=\boldsymbol{u}$ and $\boldsymbol{u}^{(0)}=\boldsymbol{x},\dots,\boldsymbol{u}^{(p)}=\boldsymbol{u}$, then

$$c_{\boldsymbol{x}^{(0)},\boldsymbol{x}^{(1)}}\cdots c_{\boldsymbol{x}^{(m-1)},\boldsymbol{x}^{(m)}}c_{\boldsymbol{y}^{(0)},\boldsymbol{y}^{(1)}}\cdots c_{\boldsymbol{y}^{(n-1)},\boldsymbol{y}^{(n)}}=c_{\boldsymbol{u}^{(0)},\boldsymbol{u}^{(1)}}\cdots c_{\boldsymbol{u}^{(p-1)},\boldsymbol{u}^{(p)}}\prod_{i=1}^{n}\mathsf{s}_i^{\eta_i(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u})}. \tag{3-7}$$

We'll prove this by induction on $\min(m,n)$. If $m=0$ or $n=0$, then this is tautological. Assume $m=1$, and $\boldsymbol{x}=\boldsymbol{y}+\sigma\epsilon_j$ for $\sigma\in\{1,-1\}$. If $\sigma(y_j-u_j)\geq 0$, then $\eta_j(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u})=0$, so this follows from the statement about minimal length paths. If $\sigma(y_j-u_j)<0$, then $\eta_j(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u})=1$, and we can assume that $\boldsymbol{y}^{(1)}=\boldsymbol{x},\dots,\boldsymbol{y}^{(n)}$ is a minimal length path from $\boldsymbol{x}$ to $\boldsymbol{u}$. Thus

$$c_{\boldsymbol{x},\boldsymbol{y}}c_{\boldsymbol{y},\boldsymbol{x}}\cdots c_{\boldsymbol{y}^{(n-1)},\boldsymbol{y}^{(n)}}=c_{\boldsymbol{x},\boldsymbol{y}^{(2)}}\cdots c_{\boldsymbol{y}^{(n-1)},\boldsymbol{y}^{(n)}}\mathsf{s}_j,$$

as desired. The argument if $n=1$ is analogous.

Now consider the general case. Assume for simplicity that $n\geq m$. Consider the path $\boldsymbol{x}^{(m-1)},\boldsymbol{y}^{(0)},\dots,$ $\boldsymbol{y}^{(n)}=\boldsymbol{u}$. Either this is a minimal path, or by induction, we have that

$$c_{\boldsymbol{x}^{(m-1)},\boldsymbol{y}}c_{\boldsymbol{y},\boldsymbol{y}^{(1)}}\cdots c_{\boldsymbol{y}^{(n-1)},\boldsymbol{y}^{(n)}}=c_{\boldsymbol{w}^{(1)},\boldsymbol{w}^{(2)}}\cdots c_{\boldsymbol{w}^{(n-2)},\boldsymbol{w}^{(n-1)}}\mathsf{s}_j$$

for a minimal path $\boldsymbol{w}^{(0)}=\boldsymbol{x}^{(m-1)},\boldsymbol{w}^{(2)},\dots,\boldsymbol{w}^{(n-1)}=\boldsymbol{y}^{(n)}$ with j being the index that changes from $\boldsymbol{x}^{(m-1)}$ to $\boldsymbol{y}$.

In the former case, by induction, relation (3-7) for the paths $\boldsymbol{x}^{(0)}=\boldsymbol{x},\dots,\boldsymbol{x}^{(m-1)}$ and $\boldsymbol{x}^{(m-1)},\boldsymbol{y}^{(0)},\dots,$ $\boldsymbol{y}^{(n)}=\boldsymbol{u}$ holds. This is just a rebracketing of the desired case of (3-7). In the latter, after rebracketing, we have

$$\begin{aligned}&(c_{\boldsymbol{x}^{(0)},\boldsymbol{x}^{(1)}}\cdots c_{\boldsymbol{x}^{(m-2)},\boldsymbol{x}^{(m-1)}})(c_{\boldsymbol{x}^{(m-1)},\boldsymbol{x}^{(m)}}\cdots c_{\boldsymbol{y}^{(n-1)},\boldsymbol{y}^{(n)}})\\ &=(c_{\boldsymbol{x}^{(0)},\boldsymbol{x}^{(1)}}\cdots c_{\boldsymbol{x}^{(m-2)},\boldsymbol{x}^{(m-1)}})(c_{\boldsymbol{w}^{(0)},\boldsymbol{w}^{(1)}}\cdots c_{\boldsymbol{w}^{(n-2)},\boldsymbol{w}^{(n-1)}})\mathsf{s}_j=c_{\boldsymbol{u}^{(0)},\boldsymbol{u}^{(1)}}\cdots c_{\boldsymbol{u}^{(p-1)},\boldsymbol{u}^{(p)}}\prod_{i=1}^{n}\mathsf{s}_i^{\eta_i(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u})},\end{aligned}$$

applying (3-7) to the shorter paths. $\square$

3.7 Quadratic duality and the Ext–algebra of the sum of all simple modules.

The algebra $\widetilde{H}^\lambda_{\mathbb{Z}}$ for smooth parameters has already appeared in the literature in [7]; it is the "A–algebra" of the hyperplane arrangement defined by $d_i = pk - \frac{1}{2}$ for all $k \in \mathbb{Z}$. This is slightly outside the scope of that paper, since only finite hyperplane arrangements were considered there, but the results of that paper are easily extended to the locally finite case. In particular, we have that the algebra $\widetilde{H}^\lambda_{\mathbb{Z}}$ is quadratic, and its quadratic dual also has a geometric description, given by the "B–algebra". We will use this to produce a description of the Ext–algebra of the sum of all simple representations of $\widetilde{H}^\lambda_{\mathbb{Z}}$.

If we fix an integer m, we may consider the hyperplane arrangement given by $d_i = pk - \frac{1}{2}$ for $k \in [-m, m]$. Let $H^{[m]}$ be the A–algebra associated to this arrangement as in [8, Section 8.3] — in that paper, it is denoted by $A(\eta, -)$. We leave the dependence on λ and the ground ring implicit.

By definition, $H^{[m]}$ is obtained by considering the chambers of the arrangement we have fixed above, putting a quiver structure on this set by connecting chambers adjacent across a hyperplane, and then imposing the same local relations (3-6a)–(3-6c). One result which will be extremely important for us is:

Theorem 3.20 [8, Lemma 8.25] *The algebra $H^{[m]}$ is finite-dimensional in each graded degree, with finite global dimension $\leq 2n$.*

There's a natural map of $H^{[m]}$ to $\widetilde{H}^\lambda$, sending the idempotents for chambers to $1_{\boldsymbol{x}}$ for $x_i \in [-m-1, m]$.

Proposition 3.21 *Fix $\boldsymbol{x}, \boldsymbol{y}$ and an integer q. For m sufficiently large, the map $H^{[m]} \to \widetilde{H}^\lambda$ induces an isomorphism $(1_{\boldsymbol{x}} H^{[m]} 1_{\boldsymbol{y}})_q \cong (1_{\boldsymbol{x}} \widetilde{H}^\lambda 1_{\boldsymbol{y}})_q$ between homogeneous elements of degree q and an isomorphism $\operatorname{Ext}_{H^{[m]}}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}) \cong \operatorname{Ext}_{\widetilde{H}^\lambda}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}})$.*

Proof An element of $(1_{\boldsymbol{x}} H^{[m]} 1_{\boldsymbol{y}})_q$ can be written as a sum of length n paths from x to y. Thus, it can only pass through $\boldsymbol{u}$ if $|\boldsymbol{x} - \boldsymbol{u}| + |\boldsymbol{u} - \boldsymbol{y}| \leq q$. Thus, if $m > q + |\boldsymbol{x}|_1 + |\boldsymbol{y}_1|$, then no hyperplane crossed by this path is excluded in $H^{[m]}$. The map $(1_{\boldsymbol{x}} H^{[m]} 1_{\boldsymbol{y}})_q \to (1_{\boldsymbol{x}} \widetilde{H}^\lambda 1_{\boldsymbol{y}})_q$ is clearly surjective in this case, and injective as well, since any relation used in H is also a relation in $H^{[m]}$.

Thus, if we take a projective resolution of $L_{\boldsymbol{x}}$ over $H^{[m]}$ and tensor it with $\widetilde{H}^\lambda$, we can choose m sufficiently large that the result is still exact in degrees below $2q$. Since $H^{[m]}$ is Koszul, with global dimension $\leq 2n$, every simple over $H^{[m]}$ has a linear resolution of length less than $\leq 2n$. This establishes that the tensor product complex is a projective resolution for $m \gg 0$.

This establishes that we have an isomorphism $\operatorname{Ext}_{H^{[m]}}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}) \to \operatorname{Ext}_{\widetilde{H}^\lambda}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}})$ for $m \gg 0$. □

Corollary 3.22 *The algebra $\widetilde{H}^\lambda$ is Koszul with global dimension $\leq 2n$.*

Note that in the language of [29, Section 5.4], we should say that the category $\mathscr{P}(\widetilde{H}^\lambda)$ is Koszul. By [29, Theorem 30], the Koszul dual of $\widetilde{H}^\lambda$ is its quadratic dual. Thus, let us calculate its quadratic dual.

Continue to assume that $\widetilde{\Lambda}(\lambda)$ is smooth. If we dualize the short exact sequence

$$0 \to \mathfrak{t}_{\mathbb{Z}} \to \mathfrak{d}_{\mathbb{Z}} \to \mathfrak{g}_{\mathbb{Z}} \to 0$$

we obtain a dual sequence

$$0 \leftarrow \mathfrak{t}^*_{\mathbb{Z}} \leftarrow \mathfrak{d}^*_{\mathbb{Z}} \leftarrow \mathfrak{g}^*_{\mathbb{Z}} \leftarrow 0.$$

Let t_i be the image in $\mathfrak{t}^*_{\mathbb{Z}}$ of the i^{th} coordinate weight of $\mathfrak{t}^*_{\mathbb{Z}}$.

Definition 3.23 Let $\widetilde{H}^!_{\lambda,\mathbb{Z}}$ (resp. $\widetilde{H}^!_{\lambda,\mathbb{K}}$) be the dg–algebra generated over $U_{\mathbb{Z}}(\mathfrak{t}^*)$ (resp. $U_{\mathbb{K}}(\mathfrak{t}^*)$) by elements $e_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \widetilde{\Lambda}(\lambda)$, $d^{\pm i}_{\boldsymbol{x}}$ for $\boldsymbol{x}, \boldsymbol{x} \pm \epsilon_i \in \widetilde{\Lambda}(\lambda)$ with trivial differential and subject to the following quadratic relations:

- Write $d_{\boldsymbol{x},\boldsymbol{u}} := d^{\pm i}_{\boldsymbol{x}}$ where $\boldsymbol{u} = \boldsymbol{x} \pm \epsilon_i$. For each $\boldsymbol{x}$ and each i, we have

(3-8a)
$$\sum_{\boldsymbol{u} \in \alpha_i(\boldsymbol{x})} d_{\boldsymbol{x},\boldsymbol{u}} d_{\boldsymbol{u},\boldsymbol{x}} = \mathsf{t}_i e_{\boldsymbol{x}}.$$

  Note that this implies that if $\alpha_i(\boldsymbol{x}) = \varnothing$, then $\mathsf{t}_i e_{\boldsymbol{x}} = 0$.

- If $\boldsymbol{x}, \boldsymbol{x}+\epsilon_i, \boldsymbol{x}+\epsilon_j, \boldsymbol{x}+\epsilon_i+\epsilon_j \in \widetilde{\Lambda}(\lambda)$, then

(3-8b)
$$d^{+j}_{\boldsymbol{x}+\epsilon_i} d^{+i}_{\boldsymbol{x}} = -d^{+i}_{\boldsymbol{x}+\epsilon_j} d^{+j}_{\boldsymbol{x}}, \quad d^{-j}_{\boldsymbol{x}+\epsilon_j} d^{-i}_{\boldsymbol{x}+\epsilon_i+\epsilon_j} = -d^{-i}_{\boldsymbol{x}+\epsilon_i} d^{-j}_{\boldsymbol{x}+\epsilon_i+\epsilon_j},$$

(3-8c)
$$d^{-j}_{\boldsymbol{x}+\epsilon_i+\epsilon_j} d^{+i}_{\boldsymbol{x}+\epsilon_j} = -d^{+i}_{\boldsymbol{x}} d^{-j}_{\boldsymbol{x}+\epsilon_j}, \quad d^{-i}_{\boldsymbol{x}+\epsilon_i+\epsilon_j} d^{+j}_{\boldsymbol{x}+\epsilon_i} = -d^{+j}_{\boldsymbol{x}} d^{-i}_{\boldsymbol{x}+\epsilon_i}.$$

- If $\boldsymbol{x}$ and $\boldsymbol{u}$ are chambers such that $|\boldsymbol{x}-\boldsymbol{u}| = 2$ and there is only one length 2 path $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u})$ in $\widetilde{\Lambda}(\lambda)$ from $\boldsymbol{x}$ to $\boldsymbol{u}$, then

(3-8d)
$$d_{\boldsymbol{x},\boldsymbol{y}} d_{\boldsymbol{y},\boldsymbol{u}} = 0.$$

For example, if $\boldsymbol{x} \notin \widetilde{\Lambda}$ but $\boldsymbol{x}+\epsilon_i, \boldsymbol{x}+\epsilon_j, \boldsymbol{x}+\epsilon_i+\epsilon_j \in \widetilde{\Lambda}(\lambda)$, then $d^{-j}_{\boldsymbol{x}+\epsilon_i+\epsilon_j} d^{+i}_{\boldsymbol{x}+\epsilon_j} = 0$. We suppress the dependence of $\widetilde{H}$ and $\widetilde{H}^!$ on λ and the ground ring, to avoid clutter. The following holds over both $\mathbb{K}$ and $\mathbb{Z}$:

Theorem 3.24 *The algebras $\widetilde{H}$ and $\widetilde{H}^!$ are quadratically dual, with the pairing $\widetilde{H}_1 \times \widetilde{H}^!_1$ given by*

$$\langle c^{\sigma i}_{\boldsymbol{x}}, d^{\sigma' j}_{\boldsymbol{y}} \rangle = \delta_{\boldsymbol{x},\boldsymbol{y}} \delta_{i,j} \delta_{\sigma,\sigma'}.$$

Again, in the notation of [29], we would say that the categories $\mathscr{P}(\widetilde{H})$ and $\mathscr{P}(\widetilde{H}^!)$ are quadratically dual.

Proof What we must show is that the quadratic relations of $\widetilde{H}$ in $\widetilde{H}_1 \otimes_{\widetilde{H}_0} \widetilde{H}_1$ are the annihilator of the relations of $\widetilde{H}^!$ in $\widetilde{H}^!_1 \otimes_{\widetilde{H}_0} \widetilde{H}^!_1$. It is enough to consider $e_{\boldsymbol{x}} \widetilde{H}_1 \otimes_{\widetilde{H}_0} \widetilde{H}_1 e_{\boldsymbol{y}}$ for any pair of idempotents $e_{\boldsymbol{x}}$ and $e_{\boldsymbol{y}}$. This space can only be nonzero if $|\boldsymbol{x}-\boldsymbol{y}| = 2$ or 0. Let us first assume that $|\boldsymbol{x}-\boldsymbol{y}| = 2$. If there is one path between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\widetilde{\Lambda}$, then $e_{\boldsymbol{x}} \widetilde{H}_1 \otimes_{\widetilde{H}_0} \widetilde{H}_1 e_{\boldsymbol{y}} \cong e_{\boldsymbol{x}} \widetilde{H}_2 e_{\boldsymbol{y}}$ and there are no relations. On the other hand, in $\widetilde{H}^!$, by (3-8d) we have that all elements of $e_{\boldsymbol{x}} \widetilde{H}^!_1 \otimes_{\widetilde{H}^!_0} \widetilde{H}^!_1 e_{\boldsymbol{y}}$ are relations.

If there are two paths, through $\boldsymbol{u}$ and $\boldsymbol{u}'$, then the element $c_{\boldsymbol{x},\boldsymbol{u}} \otimes c_{\boldsymbol{u},\boldsymbol{y}} - c_{\boldsymbol{x},\boldsymbol{u}'} \otimes c_{\boldsymbol{u}',\boldsymbol{y}}$ spans the set of relations. Its annihilator is $d_{\boldsymbol{x},\boldsymbol{u}} \otimes d_{\boldsymbol{u},\boldsymbol{y}} + d_{\boldsymbol{x},\boldsymbol{u}'} \otimes d_{\boldsymbol{u}',\boldsymbol{y}}$, which spans the relations in $\widetilde{H}^!$ by (3-8b)–(3-8c). This deals with the case where $|\boldsymbol{x}-\boldsymbol{y}| = 2$.

Now, assume that $\boldsymbol{x} = \boldsymbol{y}$. The space $e_{\boldsymbol{x}} \widetilde{H}_1 \otimes_{\widetilde{H}_0} \widetilde{H}_1 e_{\boldsymbol{x}}$ is spanned by $c_{\boldsymbol{x},\boldsymbol{u}} c_{\boldsymbol{u},\boldsymbol{x}}$ for $\boldsymbol{u} \in \alpha(\boldsymbol{x})$. Thus, $e_{\boldsymbol{x}} \widetilde{H}_1 \otimes_{\widetilde{H}_0} \widetilde{H}_1 e_{\boldsymbol{x}} \cong \mathbb{K}^{\alpha(\boldsymbol{x})}$. We can map this to $\mathfrak{d}_{\mathbb{K}}$ by sending the unit vector corresponding to $\boldsymbol{u}$ to s_i, where $\boldsymbol{x} = \boldsymbol{u} \pm \epsilon_i$. The relations are the preimage of $\mathfrak{t}_{\mathbb{K}}$.

By standard linear algebra, the annihilator of a preimage is the image of the annihilator under the dual map. Thus, we must consider the dual map $\mathfrak{t}_{\mathbb{K}}^{\perp} \subset \mathfrak{d}_{\mathbb{K}}^* \to \mathbb{K}^{\alpha(\boldsymbol{x})}$, and identify its image with the relations in $\widetilde{H}^!$. These are exactly the relations imposed by taking linear combinations of the relations in (3-8a) such that the right-hand side is 0. □

Corollary 3.25 *We have a quasi-isomorphism of dg–algebras*

$$\bigoplus_{\boldsymbol{x},\boldsymbol{y}} \operatorname{Ext}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}) \cong \widetilde{H}^!_{\lambda,\mathbb{K}},$$

with individual summands given by $\operatorname{Ext}(L_{\boldsymbol{y}}, L_{\boldsymbol{x}}) \cong e_{\boldsymbol{x}} \widetilde{H}^!_{\lambda} e_{\boldsymbol{y}}$.

Proof Here, we apply Theorem 3.18; this equivalence of abelian categories implies that we can replace the computation of $\operatorname{Ext}_{A^{\lambda}}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}})$ with that of the corresponding one-dimensional simple modules over $\widetilde{H}^{\lambda}$ in the subcategory of modules on which s_i acts nilpotently.

If we instead did the same computation in the bounded derived category of all finitely generated modules, then we would know the result is $e_{\boldsymbol{x}} \widetilde{H}^!_{\lambda} e_{\boldsymbol{y}}$ by Koszul duality. The formality of the Ext–algebra follows from the consistency of A_∞–operations with the internal grading, so this is a quasi-isomorphism of dg–algebras. Thus, we need to know that the inclusion of the category on which s_i acts nilpotently induces a fully faithful functor on derived categories.

For this, it's enough to show that every pair of objects A, B has an object C (all in the subcategory) and a surjective morphism $\psi\colon C \to A$ such that the induced map $\operatorname{Ext}^n(A, B) \to \operatorname{Ext}^n(C, B)$ is trivial for all n. We can accomplish this with C a sum of quotients of $\widetilde{H}^{\lambda} 1_{\boldsymbol{z}}$'s by the ideal generated by s_i^N for $N \gg 0$; this is clear for degree reasons if A and B are gradable, and since gradable objects dg–generate, this is enough. □

This gives us a combinatorial realization of the Ext–algebra of the simple modules in this category. We can restate it in terms of Stanley–Reisner rings as follows.

For every pair $\boldsymbol{x}, \boldsymbol{y}$, we have a polytope $\overline{\Delta}^{\mathbb{R}}_{\boldsymbol{x}} \cap \overline{\Delta}^{\mathbb{R}}_{\boldsymbol{y}}$, which has an associated Stanley–Reisner ring $\mathrm{SR}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{K}}$. The latter is the quotient of $\mathbb{K}[\mathsf{t}_1, \dots, \mathsf{t}_n]$ by the relation that $\mathsf{t}_{i_1} \cdots \mathsf{t}_{i_k} = 0$ if the intersection of $\overline{\Delta}^{\mathbb{R}}_{\boldsymbol{x}} \cap \overline{\Delta}^{\mathbb{R}}_{\boldsymbol{y}}$ with the hyperplanes defined by $a_{i_j} = pn$ for $n \in \mathbb{Z}$ is empty. Let $\overline{\mathrm{SR}}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{K}}$ be its quotient modulo the system of parameters defined by the image of $\mathfrak{t}_{\mathbb{K}}^{\perp}$.

We can define $\mathrm{SR}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{Z}}$ and $\overline{\mathrm{SR}}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{Z}}$ by the same prescription, replacing $\mathbb{K}$ by $\mathbb{Z}$ everywhere. In [7, Definition 4.1], the authors define a product on the sum $\overline{\mathrm{SR}}_{\mathbb{Z}} \cong \bigoplus_{\boldsymbol{x},\boldsymbol{y} \in \widetilde{\Lambda}} \overline{\mathrm{SR}}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{Z}}$, which they call the "$B$–algebra". The same definition works over $\mathbb{K}$.

The result [7, 4.14] shows that this algebra is isomorphic to the "A–algebra" — that defined by the relations (3-6a)–(3-6c) — for a Gale dual hyperplane arrangement. Unfortunately, for a periodic arrangement, the Gale dual is an arrangement on an infinite-dimensional space, which we will not consider. We can easily restate this theorem in a way which will generalize for us. Assume that λ is a smooth parameter.

Proposition 3.26 [7, Theorem A] *The algebra $\overline{\mathrm{SR}}_{\mathbb{K}}$ is quadratic dual to $\widetilde{H}^{\lambda}_{\mathbb{K}}$. That is, it is isomorphic to $\widetilde{H}^{!}_{\lambda,\mathbb{K}}$. In particular, we have the canonical isomorphisms*

$$\mathrm{Ext}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}) \cong e_{\boldsymbol{x}} \widetilde{H}^{!}_{\lambda,\mathbb{K}} e_{\boldsymbol{y}} \cong \overline{\mathrm{SR}}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{K}}[-|\boldsymbol{x}-\boldsymbol{y}|_1].$$

3.8 Interpretation as the cohomology of a toric variety

For our purposes, the key feature of the quadratic dual of $\widetilde{H}^{\lambda}_{\mathbb{Z}}$ is its topological interpretation, which is exactly as in [7, Section 4.3]. This interpretation will allow us to match the Ext–algebras which appear on the mirror side, in the second half of this paper.

Indeed, the periodic hyperplane arrangement $\mathfrak{A}^{\mathrm{per}}_{\lambda}$ defines a tiling of $\mathfrak{g}^{*,\lambda}_{\mathbb{R}}$ by the polytopes $\overline{\Delta}^{\mathbb{R}}_{\boldsymbol{x}}$.

To each such polytope we can associate a G–toric variety $\mathfrak{X}_{\boldsymbol{x}}$; see [15, Chapter XI]. Each facet of the polytope defines a toric subvariety of $\mathfrak{X}_{\boldsymbol{x}}$. In particular, the facet $\Delta^{\mathbb{R}}_{\boldsymbol{x}} \cap \Delta^{\mathbb{R}}_{\boldsymbol{y}}$ defines a toric subvariety $\mathfrak{X}_{\boldsymbol{x},\boldsymbol{y}}$ of both $\mathfrak{X}_{\boldsymbol{x}}$ and $\mathfrak{X}_{\boldsymbol{y}}$.

Furthermore, the Stanley–Reisner ring $\mathrm{SR}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{K}}$ is identified with $H^*_G(\mathfrak{X}_{\boldsymbol{x},\boldsymbol{y}}; \mathbb{K})$, and the quotient $\overline{\mathrm{SR}}(\boldsymbol{x}, \boldsymbol{y})_{\mathbb{K}}$ is identified with $H^*(\mathfrak{X}_{\boldsymbol{x},\boldsymbol{y}}; \mathbb{K})$. Composing this identification with Proposition 3.26, we have an identification

$$e_{\boldsymbol{x}} \widetilde{H}^{!}_{\lambda,\mathbb{K}} e_{\boldsymbol{y}} \cong H^*(\mathfrak{X}_{\boldsymbol{x},\boldsymbol{y}}; \mathbb{K})[-|\boldsymbol{x}-\boldsymbol{y}|_1].$$

In this presentation, multiplication in the Ext–algebra is given by a natural convolution on cohomology groups [7, Section 4.3].

3.9 Degrading

So far, we have only considered $A^{\lambda}_{\mathbb{K}}$–modules which are endowed with a D–action. Now, we use the results of the preceding sections to describe the category $A^{\lambda}_{\mathbb{K}}$–mod$_o$ of modules without this extra structure.

Proposition 3.27 *Assume that L is a simple module in the category $A^{\lambda}_{\mathbb{K}}$–mod$_o$. Then we have an isomorphism of $A^{\lambda}_{\mathbb{K}}$–modules $L \cong L_{\boldsymbol{x}}$ for some $\boldsymbol{x}$.*

Proof On the subcategory $A^{\lambda}_{\mathbb{K}}$–mod$_o$, the central element

$$z_i^p \partial_i^p = h_i^+(h_i^+ - 1)(h_i^+ - 2) \cdots (h_i^+ - p + 1)$$

acts nilpotently, so h_i^+ has spectrum in $\mathbb{F}_p$. In L, there thus must exist a simultaneous eigenvector v for all h_i^+'s, and $\boldsymbol{a}$ such that $h_i^+ v = a_i v$. Thus, $W^1_{\boldsymbol{a}}(L) \neq 0$, which shows that there is a nonzero map $Q_{\boldsymbol{a}} \cong P_{\boldsymbol{x}} \to L$, so we must have $L \cong L_{\boldsymbol{x}}$. □

This shows that $L_{\boldsymbol{x}}$ gives a complete list of simples. The module $P_{\boldsymbol{x}}$ represents the $\boldsymbol{a}$ generalized eigenspace of h_i^+, and thus still projective. In fact, there are redundancies in this list, but they are easy to understand.

Definition 3.28 Let $\Lambda(\lambda)$ be the quotient of $\widetilde{\Lambda}(\lambda)$ by the equivalence relation that $\boldsymbol{x} \sim \boldsymbol{y}$ if and only if $\boldsymbol{x}|_{\mathfrak{t}_\mathbb{Z}} = \boldsymbol{y}|_{\mathfrak{t}_\mathbb{Z}}$. Equivalently, $\boldsymbol{x} \sim \boldsymbol{y}$ if $\boldsymbol{y} = \boldsymbol{x} + \gamma$, where γ lies in $\mathfrak{t}_\mathbb{Z}^\perp = \mathfrak{g}_\mathbb{Z}^*$.

We write $\bar{\boldsymbol{x}}$ for the image of $\boldsymbol{x}$ in $\Lambda(\lambda)$. Recall that $\mathfrak{g}_\mathbb{Z}^{*,\lambda}$ is a torsor for the lattice $\mathfrak{g}_\mathbb{Z}^*$. The action of the sublattice $p\cdot\mathfrak{g}_\mathbb{Z}^*$ preserves the periodic arrangement $\mathfrak{A}_\lambda^{\mathrm{per}}$. The quotient $\mathfrak{A}_\lambda^{\mathrm{tor}} = \mathfrak{A}_\lambda^{\mathrm{per}}/p\cdot\mathfrak{g}_\mathbb{Z}^*$ is an arrangement on the quotient $\mathfrak{g}_\mathbb{Z}^{*,\lambda}/p\cdot\mathfrak{g}_\mathbb{Z}^*$, and $\Lambda(\lambda)$ is the set of chambers of $\mathfrak{A}_\lambda^{\mathrm{tor}}$.

Example 3.29 In the setting of Example 3.12, $\mathfrak{A}_\lambda^{\mathrm{tor}}$ has three chambers. A set of representatives is given by those chambers of the periodic arrangement lying within the pictured square.

Theorem 3.30 *As $A_\mathbb{K}^\lambda$–modules, $L_{\boldsymbol{x}} \cong L_{\boldsymbol{y}}$ if and only if $\boldsymbol{x} \sim \boldsymbol{y}$. That is, the simple modules in $A_\mathbb{K}^\lambda$–mod_o are in bijection with $\Lambda(\lambda)$.*

Proof If $\boldsymbol{x} \sim \boldsymbol{y}$, then $P_{\boldsymbol{x}}$ and $P_{\boldsymbol{y}}$ are canonically isomorphic as $A_\mathbb{K}^\lambda$–modules, since (3-2) is only sensitive to the coset of $\boldsymbol{a}$ under the action of $p\cdot\mathfrak{t}_\mathbb{Z}^\perp$. It follows that $L_{\boldsymbol{x}} \cong L_{\boldsymbol{y}}$ as $A_\mathbb{K}^\lambda$–modules. On the other hand, if $L_{\boldsymbol{x}} \cong L_{\boldsymbol{y}}$ as $A_\mathbb{K}^\lambda$–modules, their weights modulo p must agree. This is only possible if $\boldsymbol{x}|_{\mathfrak{t}_\mathbb{Z}} = \boldsymbol{y}|_{\mathfrak{t}_\mathbb{Z}}$. □

When convenient, we will write $L_{\bar{\boldsymbol{x}}}$ for the simple attached to $\bar{\boldsymbol{x}} \in \Lambda(\lambda)$. We can understand the Ext–algebra of simples using the degrading functor $\mathbb{D}\colon A_\mathbb{K}^\lambda\text{–mod}_o^D \to A_\mathbb{K}^\lambda\text{–mod}_o$ which forgets the action of D.

Theorem 3.31 *We have a canonical isomorphism of algebras*

$$\operatorname{Ext}_{A_\mathbb{K}^\lambda\text{–mod}_o}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}) \cong \bigoplus_{\boldsymbol{x}|_{\mathfrak{t}_\mathbb{Z}} = \boldsymbol{u}|_{\mathfrak{t}_\mathbb{Z}}} \operatorname{Ext}_{A_\mathbb{K}^\lambda\text{–mod}_o^D}(L_{\boldsymbol{u}}, L_{\boldsymbol{y}}) \cong \bigoplus_{\boldsymbol{y}|_{\mathfrak{t}_\mathbb{Z}} = \boldsymbol{u}|_{\mathfrak{t}_\mathbb{Z}}} \operatorname{Ext}_{A_\mathbb{K}^\lambda\text{–mod}_o^D}(L_{\boldsymbol{x}}, L_{\boldsymbol{u}}).$$

Proof This is immediate from the fact that $P_{\boldsymbol{x}}$ remains projective in $A_\mathbb{K}^\lambda$–mod_o, so the degrading of a projective resolution of $L_{\boldsymbol{x}}$ remains projective. □

One can easily see that this implies that, just like $A_\mathbb{K}^\lambda$–mod_o^D, the category $A_\mathbb{K}^\lambda$–mod_o has a Koszul graded lift, since the coincidence of the homological and internal gradings is unchanged.

We can deduce a presentation of

$$H_{\lambda,\mathbb{K}}^! = \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\Lambda(\lambda)} \operatorname{Ext}_{A_\mathbb{K}^\lambda\text{–mod}_o}(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}).$$

Indeed, we think of $\widetilde{H}_{\lambda,\mathbb{K}}^!$ as the path algebra of the quiver $\widetilde{\Lambda}(\lambda)$ (over the base ring $U_\mathbb{K}(\mathfrak{t}^*)$) satisfying the relations in Definition 3.23, and then apply the quotient map to $\Lambda(\lambda)$, keeping the arrows and relations in place. This is well-defined since the relations (3-8a)–(3-8c) are unchanged by adding a character of G to $\boldsymbol{x}$.

Likewise, we have the following description of the endomorphism algebra of the projectives. Let $H^{\lambda}_{\mathbb{K}}$ be the algebra generated by the idempotents $1_{\boldsymbol{x}}$ and the elements $c^{\pm i}_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \Lambda(\lambda)$ over S modulo the relations (3-6a)–(3-6c). Let $H^{\lambda}_{\mathbb{Z}}$ be the natural lift to $\mathbb{Z}$.

Proposition 3.32
$$H_{\lambda,\mathbb{K}} = \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\Lambda(\lambda)} \operatorname{Hom}_{A^{\lambda}_{\mathbb{K}}\text{-mod}_o}(P_{\boldsymbol{x}}, P_{\boldsymbol{y}}).$$

Example 3.33 We continue Example 3.12. The set $\Lambda(\lambda)$ has 3 elements corresponding to the chambers A, where $x_1+x_2+x_3 = \lfloor \lambda/p \rfloor$, B, where $x_1+x_2+x_3 = \lfloor \lambda/p \rfloor - 1$, and C, where $x_1+x_2+x_3 = \lfloor \lambda/p \rfloor - 2$. We have adjacencies between A and B across 3 hyperplanes, and between B and C across 3 hyperplanes, with none between A and C.

Thus, our quiver is
$$\begin{tikzcd} A \arrow[r, dash, bend left] \arrow[r, dash] \arrow[r, dash, bend right] & B \arrow[r, dash, bend left] \arrow[r, dash] \arrow[r, dash, bend right] & C. \end{tikzcd}$$

We use x_i to the path from A to B across the d_i hyperplane, and y_i the path from C to B across the d_i hyperplane. Our relations thus become
$$x_1^* x_1 = x_2^* x_2 = x_3^* x_3, \quad y_1^* y_1 = y_2^* y_2 = y_3^* y_3,$$
$$x_1 x_1^* + y_1 y_1^* = x_2 x_2^* + y_2 y_2^* = x_3 x_3^* + y_3 y_3^*,$$
$$x_i^* y_j = -x_j^* y_i, \quad y_i^* x_j = -y_j^* x_i, \quad x_i x_j^* = -y_j y_i^* \quad \text{when } i \neq j,$$
$$y_1^* x_1 = y_2^* x_2 = y_3^* x_3 = x_1^* y_1 = x_2^* y_2 = x_3^* y_3 = 0,$$
$$x_i^* x_j = y_i^* y_j = 0 \quad \text{when } i \neq j.$$

Note that there are only finitely many elements of $\Lambda(\lambda)$. In fact, the number of such elements has an explicit upper bound. A *basis* of the inclusion $T \subset D$ is a set of coordinates such that the corresponding coweights form a basis of $\mathfrak{d}_{\mathbb{Q}}/\mathfrak{t}_{\mathbb{Q}}$. For generic parameters, taking the intersection of the corresponding coordinate subtori defines a bijection of the bases with the vertices of $\mathfrak{A}^{\mathrm{tor}}_{\lambda}$.

Lemma 3.34 *The number of elements of $\Lambda^{\mathbb{R}}(\lambda)$ is less than or equal to the number of bases for the inclusion $T \subset D$.*

Proof Choose a generic cocharacter $\xi \in \mathfrak{t}^{\perp}_{\mathbb{Q}} \subset \mathfrak{d}^*_{\mathbb{Q}}$. Note that a real number c satisfies the equations $x_i p \le c < x_i p + p$ if and only if it satisfies $x_i p - \epsilon < c < x_i p + p - \epsilon$ for ϵ sufficiently small. Thus, we will have no fewer nonempty regions if we consider the chambers
$$E^{\mathbb{R}}_{\boldsymbol{x}} = \{\boldsymbol{a} \in \mathfrak{g}^{*,\lambda}_{\mathbb{Z}} \otimes \mathbb{R} \mid p x_i - \epsilon_i < a_i < p x_i + p - \epsilon_i\}$$
for some sufficiently small $\epsilon_i > 0$ chosen generically. Note that $E^{\mathbb{R}}_{\boldsymbol{x}}$ is open. For any $\overline{E}^{\mathbb{R}}_{\boldsymbol{x}}$, there is a maximal point for this cocharacter, that is, a point $\boldsymbol{a}$ such that for all $\boldsymbol{b} \neq \boldsymbol{a} \in \overline{\Delta}^{\mathbb{R}}_{\boldsymbol{x}}$, we have $\xi(\boldsymbol{b} - \boldsymbol{a}) < 0$. By standard convex geometry, this is only possible if there are hyperplanes in our arrangement passing through $\boldsymbol{a}$ defined by coordinates that are a basis. In fact, by the genericity of the elements ϵ_i, we can

assume that the point $\boldsymbol{a}$ is hit by exactly a basis of hyperplanes. This gives a map from $\Lambda(\lambda)^{\mathbb{R}}$ to the set of bases and this map is injective, since all but one of the chambers that contain $\boldsymbol{a}$ in its closure will contain points higher than $\boldsymbol{a}$. $\square$

Since the number of elements of $\Lambda^{\mathbb{R}}(\lambda)$ is lower semicontinuous in λ, we see immediately that λ is smooth if the size of $\Lambda(\lambda)$ is the number of bases.

3.10 Tilting generators for coherent sheaves

We can also interpret these results in terms of coherent sheaves. In particular, we can consider the coherent sheaf $\mathscr{Q}_{\boldsymbol{a}} = \varprojlim \mathscr{A}_{\mathbb{K}}^{\lambda}/\mathscr{A}_{\mathbb{K}}^{\lambda}(h_i^+ - a_i)^N$ on the formal completion of the fiber $\mu^{-1}(0)$. Here, as before, we assume that $a_i \in \mathbb{F}_p$, so $a_i^p - a_i = 0$. On this formal subscheme, this is an equivariant splitting bundle for the Azumaya algebra $\mathscr{A}_{\mathbb{K}}^{\lambda}$ by [35, Theorem 4.3.4].

If we think of $\mathscr{A}_{\mathbb{K}}^{\lambda}|_{\mu^{-1}(\hat{0})}$ as a left module over itself, it decomposes according the eigenvalues of h_i^+ acting on the right. By construction, each generalized eigenspace defines a copy of $\mathscr{Q}_{\boldsymbol{a}}$ for some weight $\boldsymbol{a}$. If we let $\mathfrak{g}_{\mathbb{F}_p}^{*,\lambda}$ be the set of characters of $\mathfrak{d}_{\mathbb{F}_p}$ which agree with $\lambda \pmod p$ on $\mathfrak{t}_{\mathbb{F}_p}$, then these are precisely the simultaneous eigenvalues of the Euler operators h_i^+ that occur. Thus, we have

$$\mathscr{A}_{\mathbb{K}}^{\lambda}|_{\mu^{-1}(\hat{0})} \cong \bigoplus_{\boldsymbol{b} \in \mathfrak{g}_{\mathbb{F}_p}^{*,\lambda}} \mathscr{Q}_{\boldsymbol{b}}.$$

In particular, given an $\mathscr{A}_{\mathbb{K}}^{\lambda}$–module $\mathscr{M}$ over the formal neighborhood of $\mu^{-1}(\hat{0})$, we have an isomorphism of coherent sheaves

$$\mathscr{M} \cong \bigoplus_{\boldsymbol{b} \in \mathfrak{g}_{\mathbb{F}_p}^{*,\lambda}} \mathscr{H}om_{\mathscr{A}_{\mathbb{K}}^{\lambda}}(\mathscr{Q}_{\boldsymbol{b}}, \mathscr{M}). \tag{3-9}$$

The elements of $\mathscr{A}_{\mathbb{K}}^{\lambda}$ act on $\mathscr{Q}_{\boldsymbol{a}}$ on the left as endomorphisms of the underlying coherent sheaf; in particular, $\mathscr{Q}_{\boldsymbol{a}}$ naturally decomposes as the sum of the generalized eigenspaces for the Euler operators h_i^+.

In fact, each eigenspace for the action of h_i^+ defines a line bundle, so that the sheaf $\mathscr{Q}_{\boldsymbol{a}}$ is the sum of these line bundles. The next few results will provide a description of these line bundles. We begin with some preliminaries. Recall that $\mathfrak{M}_{\mathbb{K}}$ is defined as a free quotient of a D–stable subset of $T^*\mathbb{A}_{\mathbb{K}}^n$ by T. Given any character of $\boldsymbol{x} \in D$, the associated bundle construction defines a D–line bundle on $\mathfrak{M}_{\mathbb{K}}$. If we forget the D–equivariance, then the underlying line bundle depends only on the image $\bar{\boldsymbol{x}}$ of $\boldsymbol{x}$ in $\mathfrak{d}_{\mathbb{Z}}^*/\mathfrak{t}_{\mathbb{Z}}^{\perp}$.

Definition 3.35 Given $\boldsymbol{x} \in \mathfrak{d}_{\mathbb{Z}}^*$, let $\ell(\boldsymbol{x})$ be the associated D–equivariant line bundle line bundle on $\mathfrak{M}_{\mathbb{K}}$. We sometimes write $\ell(\bar{\boldsymbol{x}})$ for the underlying nonequivariant line bundle.

Recall that the Weyl algebra $W_{\mathbb{K}}$ defines a coherent sheaf over the spectrum of its center, namely $(T^*\mathbb{A}_{\mathbb{K}}^{(1)})^n$. As a coherent sheaf, it is simply a direct sum of copies of the structure sheaf. Consider a monomial $m(\boldsymbol{k}, \boldsymbol{l}) := \prod_{i=1}^n \partial_i^{\boldsymbol{k}_i} z_i^{\boldsymbol{l}_i}$, viewed as a section of the structure sheaf. We have the following description of its D–weight $\boldsymbol{x} \in \mathfrak{d}_{\mathbb{Z}}^*$.

Write ϵ_i for the generators of $\mathfrak{d}^*_{\mathbb{Z}}$, so that $\boldsymbol{x} = \sum_{i=1}^n \delta_i \epsilon_i$. Let δ_i^+ be the maximal power of z_i^p dividing $m(\boldsymbol{l}, \boldsymbol{k})$, and let δ_i^- be the maximal power of ∂_i^p dividing $m(\boldsymbol{l}, \boldsymbol{k})$. Then $\delta_i = \delta_i^+ - \delta_i^-$. In the notation of Section 3.3, we can write this as $\boldsymbol{x} = \sum_{i=1}^n \delta_i(0, \boldsymbol{l} - \boldsymbol{k})\epsilon_i$. We conclude the following.

Lemma 3.36 *The monomial $m(\boldsymbol{l}, \boldsymbol{k})$ descends to a section of the line bundle $\ell\big(\sum_{i=1}^n \delta_i(0, \boldsymbol{l} - \boldsymbol{k})\epsilon_i\big)$ on $\mathfrak{M}^{(1)}_{\mathbb{K}}$.*

The following proposition holds over the formal neighborhood $\pi^{-1}(\hat{0})$.

Proposition 3.37 *We have isomorphisms*

$$\mathscr{H}om_{\mathscr{A}^\lambda_{\mathbb{K}}}(\mathscr{Q}_{\boldsymbol{b}}, \mathscr{Q}_{\boldsymbol{a}}) \cong \ell\bigg(\sum_{i=1}^n \delta_i(\boldsymbol{b}, \boldsymbol{a})\epsilon_i\bigg), \tag{3-10}$$

$$\mathscr{Q}_{\boldsymbol{a}} \cong \bigoplus_{\boldsymbol{b} \in \mathfrak{g}^{*,\lambda}_{\mathbb{F}_p}} \ell\bigg(\sum_{i=1}^n \delta_i(\boldsymbol{b}, \boldsymbol{a})\epsilon_i\bigg). \tag{3-11}$$

Note that the image of $\sum_{i=1}^n \delta_i(\boldsymbol{b}, \boldsymbol{a})\epsilon_i$ in $\mathfrak{d}^*_{\mathbb{Z}}/\mathfrak{t}^{\perp}_{\mathbb{Z}}$ depends only on the class of $\boldsymbol{b}$ in $\mathfrak{d}^*_{\mathbb{Z}}/p\cdot\mathfrak{t}^{\perp}_{\mathbb{Z}}$, so that the sum is well defined. The different isomorphism classes of line bundles that appear are in bijection with the chambers of $\Lambda(\lambda)$, but not canonically so, since we must choose $\boldsymbol{a}$.

Proof The second isomorphism follows from the first by (3-9). To construct the first isomorphism, we recall that $\mathscr{A}^\lambda_{\mathbb{K}} \cong \operatorname{End}(\mathscr{Q}_{\boldsymbol{a}}) \cong \operatorname{End}(\mathscr{Q}_{\boldsymbol{b}})$. Thus $\mathscr{H}om_{\mathscr{A}^\lambda_{\mathbb{K}}}(\mathscr{Q}_{\boldsymbol{a}}, \mathscr{Q}_{\boldsymbol{b}})$ is a line bundle. It has a section given by the element $m(\boldsymbol{b} - \boldsymbol{a}) \in A^\lambda_{\mathbb{K}}$. By Lemma 3.36, it is the line bundle defined via the associated bundle construction by the character $\big(\sum_{i=1}^n \delta_i(\boldsymbol{b}, \boldsymbol{a})\epsilon_i\big)$ of T. The proposition follows. □

We now pass from characteristic p to characteristic zero. The first step is to replace the parameter $\lambda \in \mathfrak{t}^*_{\mathbb{F}_p}$ by a parameter $\zeta \in \mathfrak{t}^*_{\mathbb{R}}$.

Definition 3.38 Let $\mathfrak{A}^{\mathrm{per}}_\zeta$ be the periodic hyperplane arrangement in $\mathfrak{g}^{*,\zeta}_{\mathbb{R}}$ defined by the hyperplanes $d_i = k$ for $k \in \mathbb{Z}$ and $i = 1, \dots, n$.

This is the arrangement obtained from Definition 3.8 by sending $p \to \infty$ and rescaling by $1/p$. We can define $\tilde{\Lambda}(\zeta)$, $\Lambda(\zeta)$ as before. If the element $p\zeta$ lies in $\mathfrak{t}^*_{\mathbb{Z}}$, then its image λ in $\mathfrak{t}^*_{\mathbb{F}_p}$ satisfies $\tilde{\Lambda}(\lambda) = \tilde{\Lambda}(\zeta)$ and $\Lambda(\lambda) = \Lambda(\zeta)$. The parameter ζ is smooth if and only if λ is smooth.

Let

$$\mathscr{T}^\zeta_{\mathbb{Z}} \cong \bigoplus_{\bar{\boldsymbol{x}} \in \Lambda(\zeta)} \ell(\bar{\boldsymbol{x}}).$$

For another commutative ring R, let $\mathscr{T}^\zeta_R$ be the corresponding bundle on $\mathfrak{M}_R$, the base change to $\operatorname{Spec}(R)$. Every line bundle which appears has a canonical $\mathbb{S}$–equivariant structure (induced from the trivial $\mathbb{S}$–equivariant structure on $\mathscr{O}_{T^*\mathbb{A}^n_{\mathbb{Z}}}$), and we endow $\mathscr{T}^\zeta_{\mathbb{Z}}$ with the induced $\mathbb{S}$–equivariant structure. Note that any lift of $\Lambda(\zeta)$ to $\widetilde{\Lambda}(\zeta)$ determines a $D \times \mathbb{S}$–equivariant structure, although we do not need it here. The $\mathbb{S}$–weights make $\operatorname{End}(\mathscr{T}_{\mathbb{Z}}^{\lambda})$ into a $\mathbb{Z}_{\geq 0}$ graded algebra. Let $\boldsymbol{x} \in \mathfrak{d}_{\mathbb{Z}}^{*}$.

Consider the monomial

$$m(\boldsymbol{x}) := \prod_{x_i > 0} \mathsf{z}_i^{x_i} \prod_{x_i < 0} \mathsf{w}_i^{-x_i}.$$

Note the similarity with (2-1), with the key difference that we do not require $\boldsymbol{x} \in \mathfrak{t}_{\mathbb{Z}}^{\perp}$. After Hamiltonian reduction, this defines a section of $\ell(\boldsymbol{x})$ with $\mathbb{S}$–weight equal to $|\boldsymbol{x}|_1$. By the same token, it defines an element of $\operatorname{Hom}(\ell(\boldsymbol{y}), \ell(\boldsymbol{y}'))$ whenever $\boldsymbol{y}' = \boldsymbol{y} + \boldsymbol{x}$.

Proposition 3.39 *For all λ, we have an isomorphism of graded algebras $H_{\mathbb{Z}}^{\lambda} \cong \operatorname{End}_{\mathsf{Coh}(\mathfrak{M})}(\mathscr{T}_{\mathbb{Z}}^{\lambda})$ sending $c_{\boldsymbol{x},\boldsymbol{y}} \mapsto m(\boldsymbol{y} - \boldsymbol{x})$ and $\mathsf{s}_i \mapsto \mathsf{z}_i \mathsf{w}_i$.*

Proof We first check that the map is well-defined. The map $\mathsf{s}_i \mapsto \mathsf{z}_i \mathsf{w}_i$ is well-defined since the linear relations satisfied by s_i exactly match the relations on $\mathsf{z}_i \mathsf{w}_i$ coming from restriction to the zero fiber of the T–moment map. The map $c_{\boldsymbol{x},\boldsymbol{y}} \mapsto m(\boldsymbol{y} - \boldsymbol{x})$ is well-defined if the elements $m(\boldsymbol{y} - \boldsymbol{x})$ satisfy relations (3-6a) and (3-6b)–(3-6c). Relation (3-6a) is satisfied if $m(\epsilon_i) m(-\epsilon_i) = \mathsf{z}_i \mathsf{w}_i$. This is immediate from the definition.

The relations (3-6b)–(3-6c) are clear from the commutativity of multiplication. Thus, we have defined an algebra map $H_{\mathbb{Z}}^{\zeta} \to \operatorname{End}(\mathscr{T}_{\mathbb{Z}}^{\zeta})$. This is a map of graded algebras, since both $c_{\boldsymbol{x},\boldsymbol{y}}$ and $m(\boldsymbol{y} - \boldsymbol{x})$ have degree $|\boldsymbol{y} - \boldsymbol{x}|_1$.

This map is a surjection, since homomorphisms from one line bundle to another are spanned over $\mathbb{Z}[\mathsf{z}_1 \mathsf{w}_1, \dots, \mathsf{z}_n \mathsf{w}_n]$ by $m(\boldsymbol{x})$. Since $H_{\mathbb{Z}}^{\zeta}$ is torsion-free over $\mathbb{Z}$, it's enough to check that it is injective modulo sufficiently large primes, which follows from Theorems 3.13 and 3.19. $\square$

This allows us to understand more fully the structure of the bundle $\mathscr{T}_{\mathbb{Z}}^{\zeta}$. Note that the bundle $\mathscr{T}_{\mathbb{Z}}^{\zeta}$ depends on ζ, but only through the structure of the set $\Lambda(\zeta)$.

Proposition 3.40 *The bundle $\mathscr{T}_{\mathbb{Q}}^{\zeta}$ is a tilting generator on $\mathfrak{M}_{\mathbb{Q}}$ if and only if ζ is smooth.*

Proof The bundle $\mathscr{T}_{\mathbb{Q}}^{\zeta}$ is tilting by Theorem 2.5, so we need only check if it is a generator. In order to check this over $\mathbb{Q}$, it is enough to check it modulo a large prime p. Fix an affine line Z in $\mathfrak{g}_{\mathbb{Z}}^{*,\zeta}$. By [25, Proposition 4.2], there is an integer N, independent of p, such that the set of $\lambda \in Z_{\mathbb{F}_p}$ such that $\mathscr{T}_{\mathbb{F}_p}^{\lambda}$ is *not* a generator has size $\leq N$.

If ζ is smooth, then for all sufficiently large p we can find smooth ζ' satisfying $p\zeta' \in \mathfrak{t}_{\mathbb{Z}}^{*}$ and such that $\Lambda(\zeta') = \Lambda(\zeta)$. It follows that moreover $\mathscr{T}_{\mathbb{Q}}^{\zeta'} = \mathscr{T}_{\mathbb{Q}}^{\zeta}$.

Since ζ' is smooth, $\lambda' = p\zeta'$ is also smooth. The number of $\lambda \in Z_{\mathbb{F}_p}$ such that $\Lambda(\lambda) = \Lambda(\lambda')$ is asymptotic to Ap, where A is the volume in $Z_{\mathbb{R}/\mathbb{Z}}$ of the real points such that $\Lambda(\lambda/p)^{\mathbb{R}} = \Lambda(\zeta)^{\mathbb{R}}$. Thus, whenever $p \geq N/A$, there must be some choice of λ such that $\mathscr{T}_{\mathbb{Q}}^{\lambda/p} = \mathscr{T}_{\mathbb{Q}}^{\zeta}$ is a tilting generator.

If ζ is not smooth, then $H^{\zeta}_{\mathbb{Q}}$ has fewer simple modules than at a smooth parameter, so $\mathscr{T}^{\zeta}_{\mathbb{Q}}$ cannot be a generator. □

Combining the above results yields the following equivalence of categories. In the following, we view $\mathscr{T}^{\zeta}_{\mathbb{Q}}$ as a coherent sheaf of $H^{\zeta}_{\mathbb{Q}}$–modules.

Corollary 3.41 *For smooth ζ, the adjoint functors*

$$-\overset{L}{\otimes}_{H^{\zeta}_{\mathbb{Q}}} \mathscr{T}^{\zeta}_{\mathbb{Q}} \colon D^b(H^{\zeta,\mathrm{op}}_{\mathbb{Q}}\text{–mod}) \to D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})),$$
$$\mathbb{R}\mathrm{Hom}(\mathscr{T}^{\zeta}_{\mathbb{Q}}, -) \colon D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})) \to D^b(H^{\zeta,\mathrm{op}}_{\mathbb{Q}}\text{–mod}),$$

define equivalences between the derived categories of coherent sheaves over $\mathfrak{M}_{\mathbb{Q}}$ and finitely generated right $H^{\zeta}_{\mathbb{Q}}$–modules.

The same functors define an equivalence between the derived categories of graded modules and equivariant sheaves

$$-\overset{L}{\otimes}_{H^{\zeta}_{\mathbb{Q}}} \mathscr{T}^{\zeta}_{\mathbb{Q}} \colon D^b(H^{\zeta,\mathrm{op}}_{\mathbb{Q}}\text{–gmod}) \to D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})),$$
$$\mathbb{R}\mathrm{Hom}(\mathscr{T}^{\zeta}_{\mathbb{Q}}, -) \colon D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})) \to D^b(H^{\zeta,\mathrm{op}}_{\mathbb{Q}}\text{–gmod}).$$

Finally, identical statements hold if we replace H by $\widetilde{H}$, and replace $\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})$ by $\mathsf{Coh}_G(\mathfrak{M}_{\mathbb{Q}})$, and $\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})$ by $\mathsf{Coh}_{\mathbb{G}_m \times G}(\mathfrak{M}_{\mathbb{Q}})$.

Since $H^{\zeta,\mathrm{op}}_{\mathbb{Q}}$ is defined as a path algebra modulo relations, its graded simple modules are just the one-dimensional modules $L^{\mathrm{op}}_{\boldsymbol{x}} := \mathrm{Hom}\big(\bigoplus_{\boldsymbol{y} \in \Lambda(\zeta)} L_{\boldsymbol{y}}, L_{\boldsymbol{x}}\big)$; we denote the corresponding complexes of coherent sheaves by

$$\mathscr{L}_{\boldsymbol{x}} := L^{\mathrm{op}}_{\boldsymbol{x}} \overset{L}{\otimes}_{H^{\zeta}_{\mathbb{Q}}} \mathscr{T}^{\zeta}_{\mathbb{Q}}.$$

The induced t–structure on $D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}))$ is what's often called an "exotic t–structure".

We also have a Koszul dual description of coherent sheaves as dg–modules over the quadratic dual $H^{!}_{\zeta,\mathbb{Q}}$. Since $H^{\zeta}_{\mathbb{Q}}$ is an infinite-dimensional algebra, we have to be a bit careful about finiteness properties here. We let $\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})_o$ be the category of coherent sheaves set-theoretically supported on the fiber $\pi^{-1}(o)$, and $H^{\zeta,\mathrm{op}}_{\mathbb{Q}}\text{–mod}_o$ denote the corresponding category of $H^{\mathrm{op}}_{\mathbb{Q}}$–modules; one characterization of these modules is that for some integer N, they are killed by all algebra elements of degree $> N$.

Lemma 3.42 *A complex of coherent sheaves lies in $D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})_o)$ if and only if it is in the triangulated envelope of the complexes $\mathscr{L}_{\boldsymbol{x}}$.*

Proof The complex M is in the subcategory $D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})_o)$ if and only if it is sent to a complex of modules over $H^{\zeta,\mathrm{op}}_{\mathbb{Q}}$ killed up to homotopy by a sufficiently high power of the two-sided ideal generated by the elements of positive degree in $H^0(\mathfrak{M}_{\mathbb{Q}}, \mathscr{O}_{\mathfrak{M}_{\mathbb{Q}}})$. This ideal contains all elements of sufficiently large degree (since the quotient by it is finite-dimensional and graded), so each cohomology module of the image is a finite extension of the graded simples. Thus the complex itself is an iterated extension of shifts of these modules. □

Let $H^!_{\zeta,\mathbb{Q}}$ –perf be the category of perfect dg–modules over $H^!_{\zeta,\mathbb{Q}}$. As usual, we abuse notation and let $D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}}))$ to denote the usual dg–enhancement of this category, and similarly with $D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}}))$. Combining the equivalence of Corollary 3.41 with Koszul duality:

Proposition 3.43 (1) *We have an equivalence of dg–categories* $H^!_{\zeta,\mathbb{Q}}\text{–perf} \cong D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})_o)$ *induced by* $\bigoplus_{y\in\Lambda(\zeta)} \operatorname{Ext}(\mathscr{L}_y, -)$.

(2) *We have an equivalence of dg–categories* $D^b_{\mathrm{perf}}(H^!_{\zeta,\mathbb{Q}}\text{–gmod}) \cong D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})_o)$ *induced by* $\bigoplus_{y\in\Lambda(\zeta)} \operatorname{Ext}(\mathscr{L}_y, -)$.

Proof Since elements of $D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{Q}})_o)$ are finite extensions of $\mathscr{L}_y$ for different y, they are sent by $\bigoplus_{y\in\Lambda(\zeta)} \operatorname{Ext}(\mathscr{L}_y, -)$ to perfect complexes and vice versa. This proves item (1).

Item (2) is just the graded version of this statement, which corresponds to Corollary 3.41 via the usual Koszul duality; see [4, Theorem 2.12.1]. □

This shows that smooth parameters also have an interpretation in terms of $\mathscr{A}^{\lambda}_{\mathbb{K}}$; this is effectively a restatement of Proposition 3.40, so we will not include a proof.

Proposition 3.44 *The functor* $\mathbb{R}\Gamma\colon D^b(\mathscr{A}^{\lambda}_{\mathbb{K}}\text{–mod}_0) \mapsto D^b(A^{\lambda}_{\mathbb{K}}\text{–mod}_0)$ *is an equivalence of categories if and only if the parameter* λ *is smooth.*

4 Mirror symmetry via microlocal sheaves

In the previous sections, the conical $\mathbb{G}_m$–action on hypertoric varieties played a key role in our study of coherent sheaves. This is what allowed us to construct a tilting bundle based on a quantization in characteristic p. This conic action also plays a crucial role in the study of enumerative invariants of these varieties [12; 28; 30]. The quantum connection and quantum cohomology which appear in those papers lose almost all of their interesting features if one does not work equivariantly with respect to the conic action. We are thus interested in a version of mirror symmetry which remembers this conic action.

We expect the relevant A–model category to be a subcategory of a Fukaya category of the *Dolbeault hypertoric manifold* $\mathfrak{D}$, built from Lagrangian branes endowed with an extra structure corresponding to the conical $\mathbb{G}_m$–action on $\mathfrak{M}$. However, rather than working directly with the Fukaya category, we will replace it below by a category of DQ–modules on $\mathfrak{D}$. The calculations presented there should also be valid in the Fukaya category. The reader is referred to the sequel [21] to this paper for more discussion of this point.

After defining the relevant spaces and categories of DQ–modules, we state our main equivalence in Theorems 5.9 and 4.36.

There are a few obvious related questions. What corresponds to the category of all (not necessarily equivariant) coherent sheaves on $\mathfrak{M}$? What corresponds to the full category of DQ–modules of $\mathfrak{D}$? We plan to address these questions in a future publication.

4.1 Dolbeault hypertoric manifolds

In this section, we introduce Dolbeault hypertoric manifolds, whose definition we learned from unpublished work of Hausel and Proudfoot.

Dolbeault hypertoric manifolds are complex manifolds attached to the data of a toric hyperplane arrangement (ie a collection of codimension-one affine subtori), in much the same way that an additive hypertoric variety is attached to an affine hyperplane arrangement, and a toric variety is attached to a polytope. They carry a complex symplectic form, and a proper fibration whose generic fibers are complex lagrangian abelian varieties.

Our construction of Dolbeault manifolds parallels the construction of toric varieties as Hamiltonian reductions of powers of a basic building block.

For toric varieties, this building block is $\mathbb{C}$ with the usual Hamiltonian action of $\mathbb{U}_1$. Its polytope is a ray in $\mathbb{R}$. Other toric varieties are constructed by taking the Hamiltonian reduction of $\mathbb{C}^n$ by a subtorus of $\mathbb{U}_1^n$. Additive hypertoric varieties are similarly constructed from the basic building block $T^*\mathbb{C}$ with its hyperhamiltonian action of $\mathbb{U}_1$. The affine hyperplane arrangement associated to this building block is a single point in $\mathbb{R}$. For Dolbeault manifolds, our basic building block will be the Tate curve $\mathfrak{Z}$ with a (quasi)-hyperhamiltonian action of $\mathbb{U}_1$. Its toric hyperplane arrangement is a single point in $\mathbb{U}_1$.

We give a construction of $\mathfrak{Z}$ suited to our purposes below, culminating in Definition 4.2.

Let $\mathbb{C}^* = \operatorname{Spec} \mathbb{C}[q, q^{-1}]$, and let $\mathbb{D}^*$ be the punctured disk defined by $0 < q < 1$. Let $\mathfrak{Z}^*$ be the family of elliptic curves over $\mathbb{D}^*$ defined by $(\mathbb{C}^* \times \mathbb{D}^*)/\mathbb{Z}$, where $1 \in \mathbb{Z}$ acts by multiplication by $q \times 1$.

We will define an extension of $\mathfrak{Z}^*$ to a family $\mathfrak{Z}$ over $\mathbb{D}$ with central fiber equal to a nodal elliptic curve.

Let $\mathfrak{W}_n := \operatorname{Spec} \mathbb{C}[x, y]$ for $n \in \mathbb{Z}$. Consider the birational map $f : \mathfrak{W}_n \to \mathfrak{W}_{n+1}$ defined by $f^*(x) = 1/y$, $f^*(y) = xy^2$. This defines an automorphism of the subspace $\mathfrak{W}_0 \setminus \{xy = 0\}$, and identifies the y–axis in $\mathfrak{W}_n$ with the x–axis in $\mathfrak{W}_{n+1}$ birationally, so they glue to a $\mathbb{P}^1$. If we let $q := xy$, then we can rewrite this automorphism as $(x, y) \mapsto (q^{-1}x, qy)$. Note that this map preserves the product xy and commutes with the $\mathbb{C}^*$–action on $\mathfrak{W}_n$ defined by $\tau \cdot x = \tau x$, $\tau \cdot y = \tau^{-1} y$; we let $\mathbb{T}$ denote this copy of $\mathbb{C}^*$.

Definition 4.1 Let $\mathfrak{W}$ be the quotient of the union $\bigsqcup_{n \in \mathbb{Z}} \mathfrak{W}_n$ by the equivalence relations that identify the points $x \in \mathfrak{W}_n$ and $f(x) \in \mathfrak{W}_{n+1}$.

The variety $\mathfrak{W}$ is smooth of infinite type, with a map $q := xy : \mathfrak{W} \to \mathbb{C}$ and an action of $\mathbb{C}^*$ preserving the fibers of q. The map $\mathfrak{W}_0 \setminus \{xy = 0\} \to \mathfrak{W} \setminus q^{-1}(0)$ is easily checked to be an isomorphism.

$\mathfrak{W}$ carries a $\mathbb{Z}$–action defined by sending $\mathfrak{W}_n$ to $\mathfrak{W}_{n+1}$ via the identity map. The action of $n \in \mathbb{Z}$ is the unique extension of the automorphism of $\mathfrak{W}_0 \setminus \{xy = 0\}$ given by $(x, y) \mapsto (q^{-n}x, q^n y)$. Thus, n fixes a point (x, y) if and only if q is an n^{th} root of unity. In particular, the action of $\mathbb{Z}$ on

$$\tilde{\mathfrak{Z}} := q^{-1}(\mathbb{D}) \tag{4-1}$$

is free. Combining this with the paragraph above, we see that $q^{-1}(\mathbb{D}^*) = \{(x, y) \in \mathfrak{W}_0 \mid xy \in \mathbb{D}^*\}$; since we can choose $x \in \mathbb{C}^*$ and $q \in \mathbb{D}^*$, with $y = q/x$ uniquely determined, we have an isomorphism $q^{-1}(\mathbb{D}^*) \cong \mathbb{C}^* \times \mathbb{D}^*$. Transported by this isomorphism, the $\mathbb{C}^*$–action we have defined acts by scalar multiplication on the first factor, and trivially on the second.

Thus, we obtain the following commutative diagram of spaces:

$$\begin{tikzcd} \mathbb{C}^* \times \mathbb{D}^* \cong q^{-1}(\mathbb{D}^*) \arrow[r, "j"] \arrow[d, "q"'] & \tilde{\mathfrak{Z}} \arrow[d, "q"] \\ \mathbb{D}^* \arrow[r] & \mathbb{D} \end{tikzcd} \tag{4-2}$$

The fiber $\tilde{\mathfrak{Z}}_0 := q^{-1}(0)$ is an infinite chain of $\mathbb{CP}^1$'s with each link connected to the next by a single node. The action of $\mathbb{C}^*$ on $\tilde{\mathfrak{Z}}_0$ scales each component, matching the usual action of scalars on $\mathbb{CP}^1$, thought of as the Riemann sphere. The action of the generator of $\mathbb{Z}$ translates the chain by one link.

Definition 4.2 Let $\mathfrak{Z} := \tilde{\mathfrak{Z}}/\mathbb{Z}$.

The manifold $\mathfrak{Z}$ will be our basic building block. We now study various group actions and moment maps for $\mathfrak{Z}$, in order to eventually define a symplectic reduction of $\mathfrak{Z}^n$.

The action of $\mathbb{C}^*$ on $\tilde{\mathfrak{Z}}$ descends to an action on $\mathfrak{Z}$; note that on any nonzero fiber of the map to $\mathbb{D}$, it factors through a free action of the quotient group $\mathbb{C}^*/q^{\mathbb{Z}}$, which is transitive unless $q = 0$. Thus the generic fiber of q is an elliptic curve. The fiber $\mathfrak{Z}_0 := q^{-1}(0)$ is a nodal elliptic curve. We write $\boldsymbol{n}$ for the node.

The action of $\mathbb{U}_1 \subset \mathbb{C}^*$ on $\tilde{\mathfrak{Z}}$ is Hamiltonian with respect to a hyperkähler symplectic form and metric described in [24, Proposition 3.2], where one also finds a description of the $\mathbb{Z}$–equivariant moment map. This moment map descends to

$$\mu\colon \mathfrak{Z} \to \mathbb{R}/\mathbb{Z} = \mathbb{U}_1.$$

Hence μ is the quasihamiltonian moment map for the action of $\mathbb{U}_1$ on $\mathfrak{Z}$. We may arrange that $\mu(\boldsymbol{n}) = \mathbf{1} \in \mathbb{U}_1$. The nodal fiber $\mathfrak{Z}_0$ is the image of a $\mathbb{U}_1$–equivariant immersion $\iota\colon \mathbb{CP}^1 \to \mathfrak{Z}$, which is an embedding except that 0 and ∞ are both sent to $\boldsymbol{n}$. We have a commutative diagram

$$\begin{tikzcd} \mathbb{CP}^1 \arrow[r, "\iota"] \arrow[d, "\mu_{\mathbb{CP}^1}"'] & \mathfrak{Z} \arrow[d, "\mu"] \\ {[0,1]} \arrow[r] & \mathbb{R}/\mathbb{Z} \end{tikzcd} \quad \text{where } \mu_{\mathbb{CP}^1}(z) = \frac{|z|^2}{1+|z|^2}\colon \mathbb{CP}^1 \to [0,1]. \tag{4-3}$$

The action of $\mathbb{U}_1$ and map $\mu \times q$ form a kind of "multiplicative hyperkähler hamiltonian action" of $\mathbb{U}_1$. In particular, $(\mu \times q)^{-1}(a,b)$ is a single $\mathbb{U}_1$ orbit, which is free unless $a = 1$ and $b = 0$, in which case it is just the node $\boldsymbol{n}$. It's worth comparing this with the hyperkähler moment map on $T^*\mathbb{C}$ for the action of $\mathbb{U}_1$: this is given by the map

$$T^*\mathbb{C} \to \mathbb{R} \times \mathbb{C}, \quad (z,w) \mapsto (|z|^2 - |w|^2, zw).$$

The fibers over nonzero elements of $\mathbb{R} \times \mathbb{C}$ are circles, and the fiber over zero is the origin. In a neighborhood of $\boldsymbol{n}$, $\mu \times q$ is analytically isomorphic to this map.

Without seeking to formalize the notion, we will simply mimic the notion of hyperkähler reduction in this setting. Recall that a hypertoric variety $\mathfrak{M}$ is defined using an embedding of tori $(\mathbb{C}^*)^k = T \to D = (\mathbb{C}^*)^n$. Let $T_\mathbb{R}$ and $D_\mathbb{R}$ be the corresponding compact tori in these groups, and $T^\vee\mathbb{R} \cong \mathfrak{t}^*_\mathbb{R}/\mathfrak{t}^*_\mathbb{Z}$ the Langlands dual torus; the usual inner product induces an isomorphism $D_\mathbb{R} \cong D^\vee_\mathbb{R}$, which we will leave implicit. Thus, we have an action of $T_\mathbb{R}$ on $\mathfrak{Z}^n$ and a $T_\mathbb{R}$–invariant map

$$\Phi\colon \mathfrak{Z}^n \to T^\vee_\mathbb{R} \times \mathfrak{t}^*. \tag{4-4}$$

Given $\zeta \in T^\vee_\mathbb{R}$, let $\zeta' = \zeta \times 0 \in T^\vee_\mathbb{R} \times \mathfrak{t}^*$. For generic ζ the action of $T_\mathbb{R}$ on $\Phi^{-1}(\zeta')$ is locally free.

For the rest of this paper, we make the additional assumption that the torus embedding $T \to D$ is *unimodular*, meaning that if e_k are the coordinate basis of $\mathfrak{d}_\mathbb{Z}$, then any collection of e_k whose image spans $\mathfrak{d}_\mathbb{Q}/\mathfrak{t}_\mathbb{Q}$ also spans $\mathfrak{d}_\mathbb{Z}/\mathfrak{t}_\mathbb{Z}$. As with toric varieties, this guarantees that for generic ζ the action of T on $\Phi^{-1}(\zeta')$ is actually free. We expect that this assumption can be lifted without significant difficulties, but it will help alleviate notation in what follows.

The following definition is due to Hausel and Proudfoot.

Definition 4.3 Let $\mathfrak{D} := \Phi^{-1}(\zeta')/T_\mathbb{R}$.

Proposition 4.4 *$\mathfrak{D}$ is a $2d = (2n-2k)$–dimensional holomorphic symplectic manifold.*

We will also need to consider the universal cover $\widetilde{\mathfrak{D}}$; this can also be constructed as a reduction. We have a hyperkähler moment map $\widetilde{\Phi}\colon \widetilde{\mathfrak{Z}}^n \to \mathfrak{t}^*_\mathbb{R} \oplus \mathfrak{t}^*$. Let $\widetilde{\zeta}'$ be a preimage of ζ'.

Definition 4.5 Let $\widetilde{\mathfrak{D}} := \widetilde{\Phi}^{-1}(\widetilde{\zeta}')/T_\mathbb{R}$.

$\widetilde{\mathfrak{D}}$ carries a natural action of $\mathfrak{g}^*_\mathbb{Z}$, the subgroup of $\mathbb{Z}^n$ which preserves the level $\widetilde{\Phi}^{-1}(\widetilde{\zeta}')$. The quotient by this map is $\mathfrak{D}$, and the quotient map $\nu\colon \widetilde{\mathfrak{D}} \to \mathfrak{D}$ is a universal cover. Note that $\widetilde{\mathfrak{D}}$ is a (nonmultiplicative) hyperkähler reduction, and the action of $\mathfrak{g}^*_\mathbb{Z}$ preserves the resulting complex symplectic form. This gives one way of defining the complex symplectic form on $\mathfrak{D}$.

The T action on $\widetilde{\mathfrak{Z}}^n$ and the holomorphic part of the hyperkähler moment map $\widetilde{\Phi}_\mathbb{C}$ both extend to the infinite-type algebraic variety $\mathfrak{W}^n$.

Definition 4.6 Let $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$ be the holomorphic symplectic reduction $\widetilde{\Phi}_\mathbb{C}^{-1}(0)/\!\!/_{\widetilde{\zeta}'}T$, where we take the GIT quotient by T with linearization determined by $\widetilde{\zeta}'$.

As opposed to $\widetilde{\mathfrak{D}}$, the space $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$ is naturally an infinite-type but finite-dimensional algebraic variety. Its construction and properties are described in detail in [23]. It contains the complex manifold $\widetilde{\mathfrak{D}}$ as an (analytic) open subset.

Let $q_{\mathfrak{D}} \colon \mathfrak{D} \to \mathfrak{t}_{\mathbb{C}}^{\perp} \cong \mathfrak{g}_{\mathbb{C}}^*$ be the map induced by $q^n \colon \mathfrak{Z}^n \to \mathbb{C}^n \cong \mathfrak{d}_{\mathbb{C}}^*$. Its fibers are complex Lagrangians. The action of $\mathbb{C}^*$ on $\mathfrak{Z}$ defines an action of $G = D/T$ on $\mathfrak{D}$, which preserves the complex symplectic form and the fibers of the map $q_{\mathfrak{D}}$, and acts transitively on fibers over values $(q_1, \dots, q_n)$ with $q_i \neq 0$ for all i. Such fibers are d–dimensional abelian varieties.

Definition 4.7 We define the *core* of $\mathfrak{D}$ to be $\mathfrak{C} := q_{\mathfrak{D}}^{-1}(0)$, and denote by $\widetilde{\mathfrak{C}}$ its preimage in $\widetilde{\mathfrak{D}}$.

We thus have inclusions $\widetilde{\mathfrak{C}} \xrightarrow{\text{closed}} \widetilde{\mathfrak{D}} \xrightarrow{\text{open}} \widetilde{\mathfrak{D}}^{\mathrm{alg}}$. The lattice $\mathfrak{g}_{\mathbb{Z}}^*$ acts compatibly on all three spaces, but the quotient only makes sense for the first two, where it gives the inclusion $\mathfrak{C} \to \mathfrak{D}$.

Whereas $\mathfrak{D}$ is merely a complex manifold, we will see that $\mathfrak{C}$ is naturally an algebraic variety. It is a free quotient of $\widetilde{\mathfrak{D}}$, whose components, as we shall see, are smooth complex Lagrangians. We can give an explicit description of $\mathfrak{C}$ as follows, in the spirit of the combinatorial description of toric varieties in terms of their moment polytopes. In our setting, polytopes are replaced by toroidal arrangements.

We have the map $D_{\mathbb{R}}^* \to T_{\mathbb{R}}^*$; let $G_{\mathbb{R}}^{*,\zeta}$ be the preimage of ζ. It is a torsor over $G_{\mathbb{R}}^*$. The preimage of $G_{\mathbb{R}}^{*,\zeta}$ under the quotient $\mathfrak{d}_{\mathbb{R}}^* \to D_{\mathbb{R}}^*$ is given by $\mathfrak{g}_{\mathbb{R}}^{*,\zeta} := \widetilde{\zeta}' + \mathfrak{g}_{\mathbb{R}}^*$.

Definition 4.8 Let $\mathfrak{B}_{\zeta}^{\mathrm{per}} \subset \mathfrak{g}_{\mathbb{R}}^{*,\zeta}$ be the periodic hyperplane arrangement defined by the preimage of $\mathfrak{B}_{\zeta}^{\mathrm{tor}}$ in $\widetilde{\zeta}' + \mathfrak{g}_{\mathbb{R}}^*$. Let $\widetilde{\Lambda}^{\mathbb{R}}(\zeta)$ be the set of chambers of $\mathfrak{B}_{\zeta}^{\mathrm{per}}$. We write $\Delta_{\boldsymbol{x}}^{\mathbb{R}} \subset \mathfrak{g}_{\mathbb{R}}^{*,\zeta}$ for the (closed) chamber indexed by $\boldsymbol{x} \in \widetilde{\Lambda}^{\mathbb{R}}(\zeta)$.

As in Section 3.8, let $\mathfrak{X}_{\boldsymbol{x}}$ be the toric variety obtained from the polytope $\Delta_{\boldsymbol{x}}^{\mathbb{R}}$ by the Delzant construction.

Proposition 4.9 (1) *The irreducible components of $\widetilde{\mathfrak{C}}$ are smooth toric varieties $\mathfrak{X}_{\boldsymbol{x}}$ indexed by $\boldsymbol{x} \in \widetilde{\Lambda}^{\mathbb{R}}(\zeta)$.*

(2) *The intersection $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$ is the toric subvariety of either component indexed by $\Delta_{\boldsymbol{x}}^{\mathbb{R}} \cap \Delta_{\boldsymbol{y}}^{\mathbb{R}}$.*

(3) *The image under the $G_{\mathbb{R}}$–moment map of $\mathfrak{X}_{\boldsymbol{x}}$ is precisely the polytope $\Delta_{\boldsymbol{x}}^{\mathbb{R}}$.*

(4) *All components meet with normal crossings.*

Proof We begin by noting that $\widetilde{\mathfrak{C}}$ is the image in $\widetilde{\mathfrak{D}}$ of $\Phi^{-1}(\zeta') \cap \widetilde{\mathfrak{Z}}_0^n$. The irreducible components of $\widetilde{\mathfrak{Z}}_0^n$ are copies of $(\mathbb{CP}^1)^n$ indexed by $\boldsymbol{x} \in \mathbb{Z}^n$. The moment map $\mu^n \colon \widetilde{\mathfrak{Z}}_0^n \to \mathbb{R}^n$, restricted to the component $(\mathbb{CP}^1)_{\boldsymbol{x}}^n$, has image the translation $[0,1]_{\boldsymbol{x}}^n$ of the unit cube by $\boldsymbol{x}$. We write $\widetilde{\Phi}_{\boldsymbol{x}} \colon (\mathbb{CP}^1)^n \to \mathfrak{t}_{\mathbb{R}}^{\vee}$ for the restriction the $T_{\mathbb{R}}$ moment map. It is given by be the composition of $\mu_{\mathbb{CP}^1}^n \colon (\mathbb{CP}^1)^n \to [0,1]_{\boldsymbol{x}}^n$ with the projection $p \colon [0,1]_{\boldsymbol{x}}^n \subset \mathfrak{d}_{\mathbb{R}}^* \to \mathfrak{t}_{\mathbb{R}}^{\vee}$.

The preimage $p^{-1}(\zeta) \subset [0,1]_{\boldsymbol{x}}^n$ is a polytope, given by $\mathfrak{g}_{\mathbb{R}}^{*,\zeta} \cap [0,1]_{\boldsymbol{x}}^n$. It is nonempty precisely when $\boldsymbol{x} \in \widetilde{\Lambda}^{\mathbb{R}}(\zeta)$, in which case it is the chamber $\Delta_{\boldsymbol{x}}^{\mathbb{R}}$.

The irreducible components of $\widetilde{\mathfrak{C}}$ are thus the quotients $\widetilde{\Phi}_{\boldsymbol{x}}^{-1}(\zeta)/T_{\mathbb{R}}$ for $\boldsymbol{x} \in \widetilde{\Lambda}^{\mathbb{R}}(\zeta)$. The claims (1), (2) and (3) now follow from standard toric geometry.

Claim (4) follows from the corresponding property for $\widetilde{\mathfrak{Z}}_0^n$. In fact, the singular points of $\mathfrak{C}$ are analytically locally a product of m nodes, and a $(d-m)$–dimensional affine space. □

Definition 4.10 Let $\mathfrak{B}_\zeta^{\mathrm{tor}} \subset G_{\mathbb{R}}^{*,\zeta}$ be the toric hyperplane arrangement defined by the coordinate subtori of $D_{\mathbb{R}}^*$. Let $\Lambda^{\mathbb{R}}(\zeta)$ be the set of chambers of $\mathfrak{B}_\zeta^{\mathrm{tor}}$. Given $\boldsymbol{x} \in \Lambda^{\mathbb{R}}(\zeta)$, we write $\Delta_{\boldsymbol{x}}^{\mathbb{R}} \subset G_{\mathbb{R}}^{*,\zeta}$ for the corresponding chamber.

The toric arrangement $\mathfrak{B}_\zeta^{\mathrm{tor}}$ is simply the quotient of the periodic arrangement $\mathfrak{B}_\zeta^{\mathrm{per}}$ by the action of the lattice $\mathfrak{g}_{\mathbb{Z}}^*$. The restriction of the quotient map to a fixed chamber $\Delta_{\boldsymbol{x}} \subset \mathfrak{g}_{\mathbb{R}}^{*,\zeta}$ is one-to-one on the interior, but may identify certain smaller strata. Correspondingly, the composition $\mathfrak{X}_{\boldsymbol{x}} \to \widetilde{\mathfrak{D}} \to \mathfrak{D}$ is in general only an immersion. The following is easily deduced from Proposition 4.9.

Proposition 4.11 (1) *The irreducible components of $\mathfrak{C}$ are immersed toric varieties $\overline{\mathfrak{X}}_{\boldsymbol{x}}$ indexed by $\boldsymbol{x} \in \Lambda^{\mathbb{R}}(\zeta)$. Any lift of $\boldsymbol{x} \in \Lambda^{\mathbb{R}}(\zeta)$ to $\widetilde{\Lambda}^{\mathbb{R}}(\zeta)$ determines a birational map $\mathfrak{X}_{\boldsymbol{x}} \to \overline{\mathfrak{X}}_{\boldsymbol{x}}$ with finite fibers.*

(2) *The intersection $\overline{\mathfrak{X}}_{\boldsymbol{x}} \cap \overline{\mathfrak{X}}_{\boldsymbol{y}}$ is the (immersed) toric subvariety of either component indexed by $\Delta_{\boldsymbol{x}}^{\mathbb{R}} \cap \Delta_{\boldsymbol{y}}^{\mathbb{R}}$.*

(3) *The image under the $G_{\mathbb{R}}$–moment map of $\overline{\mathfrak{X}}_{\boldsymbol{x}}$ is precisely the toric chamber $\Delta_{\boldsymbol{x}}^{\mathbb{R}}$.*

(4) *All components meet with normal crossings.*

Example 4.12 We continue with Example 3.12. In this case, for generic $\widetilde{\zeta}'$, we arrive at a picture like in (3-4). The three chambers shown in total there correspond to the three core components of C: two of these are isomorphic to $\mathbb{CP}^2$, and one to $\mathbb{CP}^1 \times \mathbb{CP}^1$ blown up at $(0,0)$ and (∞,∞). We join these by joining the lines at ∞ in the first $\mathbb{CP}^2$ to the exceptional locus of the blowup at $(0,0)$, and its coordinate lines to the unique lifts of $\mathbb{CP}^1 \times \{\infty\}$ and $\{\infty\} \times \mathbb{CP}^1$ to lines in the blowup (note that in the blowup, these lines don't intersect). With the second $\mathbb{CP}^2$ we do the same gluing with 0 and ∞ reversed.

Note that in $\mathfrak{C}$, the two $\mathbb{CP}^2$'s are embedded, but the third component is only immersed: it intersects itself transversely at each torus fixed point.

4.2 Weinstein neighborhoods and scaling actions

Let $\boldsymbol{x}$ be a chamber of the periodic arrangement $\mathfrak{B}_\zeta^{\mathrm{per}}$, and let $\mathfrak{X}_{\boldsymbol{x}}$ be a component of the periodic core. We will construct an open neighborhood $\widetilde{\mathfrak{D}}_{\boldsymbol{x}}^{\mathrm{alg}} \cong T^*\mathfrak{X}_{\boldsymbol{x}}$ of $\mathfrak{X}_{\boldsymbol{x}}$ in $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$. Its intersection

$$\widetilde{\mathfrak{D}}_{\boldsymbol{x}} := \widetilde{\mathfrak{D}}_{\boldsymbol{x}}^{\mathrm{alg}} \cap \widetilde{\mathfrak{D}}$$

is an open neighborhood of $\mathfrak{X}_{\boldsymbol{x}}$ in $\widetilde{\mathfrak{D}}$, which maps by an immersion to an open neighborhood of $\overline{\mathfrak{X}}_{\boldsymbol{x}}$ in $\mathfrak{D}$.

Consider the union $\mathfrak{W}_0 \cup \mathfrak{W}_1 \subset \mathfrak{W}$. This is a Zariski open subset of $\mathfrak{W}$ isomorphic to $T^*\mathbb{CP}^1$. Let $\widetilde{\mathfrak{Z}}_0$ be its intersection with $\widetilde{\mathfrak{Z}}$. This is an open submanifold, isomorphic to a tubular neighborhood of $\mathbb{CP}^1$ in its cotangent bundle.

These identifications map the function q to the function induced by the vector field $z\,d/dz$ for z the usual coordinate on $\mathbb{CP}^1$. The induced map $\widetilde{U} \to \mathfrak{Z}$ is an immersion.

Applying the action of $\mathbb{Z}$ gives neighborhoods $\mathfrak{W}_k$ of each component of $q^{-1}(0) \subset \mathfrak{W}$. Repeating the same construction for the product $\mathfrak{W}^n$, we obtain for each $\boldsymbol{x} \in \mathbb{Z}^n$ an open neighborhood $\widetilde{\mathfrak{W}}_{\boldsymbol{x}}$ of $(\mathbb{CP}^1)^n_{\boldsymbol{x}}$ in $\mathfrak{W}^n$, isomorphic to $T^*(\mathbb{CP}^1)^n$. This neighborhood is preserved by the (complex hamiltonian) action of $T_{\mathbb{R}}$. Consider its complex symplectic reduction

$$\widetilde{\mathfrak{D}}^{\mathrm{alg}}_{\boldsymbol{x}} := \mathfrak{W}_{\boldsymbol{x}} /\!/_{\tilde{\zeta}'} T.$$

It is an open neighborhood of $\mathfrak{X}_{\boldsymbol{x}}$ in $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$, naturally symplectomorphic to $T^*\mathfrak{X}_{\boldsymbol{x}}$. Intersecting with $\widetilde{\mathfrak{D}} \subset \widetilde{\mathfrak{D}}^{\mathrm{alg}}$, we obtain an open neighborhood $\widetilde{\mathfrak{D}}_{\boldsymbol{x}}$ of the zero section in $T^*\mathfrak{X}_{\boldsymbol{x}}$ mapping by a symplectic immersion

$$\iota_{\boldsymbol{x}} : \widetilde{\mathfrak{D}}_{\boldsymbol{x}} \to \mathfrak{D} \tag{4-5}$$

to an open subset of $\mathfrak{D}$ extending the immersion $\mathfrak{X}_{\boldsymbol{x}} \to \mathfrak{D}$ and a corresponding lift $\tilde{\iota}_{\boldsymbol{x}} : \widetilde{\mathfrak{D}}_{\boldsymbol{x}} \to \widetilde{\mathfrak{D}}$, which is a symplectomorphism onto an open subset of $\widetilde{\mathfrak{D}}$. The set of such lifts is a torsor over $\mathfrak{g}^*_{\mathbb{Z}}$.

4.3 Scaling actions

The scaling $\mathbb{C}^*$–action on $T^*\mathfrak{X}_{\boldsymbol{x}}$ extends to an action of $\mathbb{C}^*$ on $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$, which does not preserve $\widetilde{\mathfrak{D}}$. We first describe this action in the basic case of $\mathfrak{W}$. Fix $p \in \mathbb{Z}$ and let $\mathbb{S}_p$ be the copy of $\mathbb{C}^*$ which acts on $\mathfrak{W}_k$ giving x degree $1-k+p$ and y degree $k-p$. One can easily check that this action descends to an action on $\mathfrak{W}$ and gives the Poisson bracket degree one. On $\mathfrak{W}_p \cup \mathfrak{W}_{p+1} \cong T^*\mathbb{CP}^1$, it acts by the scaling action on the fibers. Note that $\mathbb{S}_p$ does not preserve the open subset $\widetilde{\mathfrak{Z}} \subset \mathfrak{W}$.

The action of $\mathbb{S}_p \times \mathbb{T}$ does not commute with the translation action of $\mathbb{Z}$. Instead, the $\mathbb{Z}$–action intertwines the actions of $\mathbb{S}_p \times \mathbb{T}$ for different p. In particular, all such actions are given by precomposing an isomorphism $\mathbb{S}_p \times \mathbb{T} \to \mathbb{S}_0 \times \mathbb{T}$ with the action of the latter torus on $\mathfrak{W}$.

We can upgrade all these structures to the general case: for each $\boldsymbol{x}$, we have a copy $\mathbb{S}_{\boldsymbol{x}}$ of $\mathbb{C}^*$ which acts on $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$ such that on $\widetilde{\mathfrak{D}}^{\mathrm{alg}}_{\boldsymbol{x}} \subset T^*\mathfrak{X}_{\boldsymbol{x}}$ it matches the scaling action. As before, these actions do not commute with the $\mathfrak{g}^*_{\mathbb{Z}}$–action. Instead, they are intertwined by this action. In particular, all such actions factor through an isomorphism $\mathbb{S}_{\boldsymbol{x}} \times G \to \mathbb{S}_0 \times G$ with the action of the latter torus on $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$. We make the following (purely notational) definition, to emphasize this independence of choices.

Definition 4.13 Let $\mathbb{S}G := \mathbb{S}_0 \times G$, with its action on $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$.

4.4 Other flavors of multiplicative hypertoric manifold

In this paper, starting from the data of an embedding of tori $T \to \mathbb{G}_m^n$, we have constructed both an additive hypertoric variety $\mathfrak{M}$ and a Dolbeault hypertoric manifold $\mathfrak{D}$. We view the latter as a multiplicative analogue of $\mathfrak{M}$. One can attach to the same data another, better known multiplicative analogue $\mathfrak{B}$, which however plays only a motivational role in this paper. For a definition, see [22]. $\mathfrak{B}$ is often simply known as a multiplicative hypertoric variety. For generic parameters, it is a smooth affine variety, of the same dimension as $\mathfrak{M}$ and $\mathfrak{D}$. In fact, work of Zsuzsanna Dancso, Vivek Shende and the first author [16] constructs a smooth open embedding $\mathfrak{D} \to \mathfrak{B}$, such that $\mathfrak{B}$ retracts smoothly onto the image. The embedding does not, however, respect complex structures; for instance, the complex Lagrangians considered here map to real submanifolds of the multiplicative hypertoric variety. Instead, $\mathfrak{B}$ and $\mathfrak{D}$ play roles analogous to the Betti and Dolbeault moduli of a curve.

In the sequel [21] to this paper, joint with Ben Gammage, we show that the core $\mathfrak{C} \subset \mathfrak{D}$ becomes the Liouville skeleton of $\mathfrak{B}$, thought of as a Liouville manifold with respect to the affine Liouville structure. Microlocal sheaves on this skeleton compute the wrapped Fukaya category of $\mathfrak{B}$. In the next section, we will introduce a category of deformation quantization modules on $\mathfrak{D}$, which roughly corresponds to microlocal sheaves on $\mathfrak{B}$ with an extra $\mathbb{G}_m$ equivariant structure. This helps place our main results in the usual context of homological mirror symmetry. The relationship between the two papers is explained in more detail in [21].

4.5 Deformation quantization of $\mathfrak{D}$

In the next few sections, we define a deformation quantization of $\mathfrak{D}$ over $\mathbb{C}((\hbar^{1/2}))$, and compare modules over this quantization with the category $A_{\mathbb{K}}^{\lambda}$–mod_o from the first half of the paper. We'll also discuss how the structure of $\mathbb{G}_m$–equivariance of coherent sheaves can be recaptured by considering a category $\mu\mathsf{m}$ of deformation quantization modules equipped with the additional structure of a "microlocal mixed Hodge module".

Consider the sheaf of analytic functions $\mathcal{O}_{\mathfrak{W}_n}$ on $\mathfrak{W}_n$. We'll endow the sheaf $\mathcal{O}_{\mathfrak{W}_n}^{\hbar} := \mathcal{O}_{\mathfrak{W}_n}((\hbar^{1/2}))$ with the Moyal product multiplication

$$f \star g := fg + \sum_{n=1}^{\infty} \frac{\hbar^n}{2^n n!}\left(\frac{\partial^n f}{dx^n}\frac{\partial^n g}{dy^n} - \frac{\partial^n g}{dx^n}\frac{\partial^n f}{dy^n}\right).$$

If f or g is a polynomial this formula only has finitely many terms, but for a more general meromorphic function, we will have infinitely many. Following the conventions of [10], we let $\mathcal{O}_{\mathfrak{W}_n}^{\hbar}(0) = \mathcal{O}_{\mathfrak{W}_n}[\![\hbar^{1/2}]\!]$, which is clearly a subalgebra. We'll clarify later why we have adjoined a square root of $\hbar$.

Sending $x \mapsto 1/y$ and $y \mapsto xy^2$ induces an algebra automorphism of this sheaf on the subset $\mathfrak{W}_n \setminus \{xy=0\}$, since

$$\frac{1}{y} \star xy^2 = xy + \frac{\hbar}{2} \quad \text{and} \quad xy^2 \star \frac{1}{y} = xy - \frac{\hbar}{2}.$$

This shows that we have an induced star product on the sheaf $\mathcal{O}^{\hbar}_{\mathfrak{W}}$, and thus on $\mathcal{O}^{\hbar}_{\mathfrak{Z}}$. We now use noncommutative Hamiltonian reduction to define a star product on $\mathcal{O}^{\hbar}_{\mathfrak{D}}$. This depends on a choice of noncommutative moment map $\kappa_{\hbar}\colon \mathfrak{t} \to \mathcal{O}^{\hbar}_{\mathfrak{Z}^n}$. We fix $\phi \in \mathfrak{d}^*$. Given $(a_1, \dots, a_n) \in \mathfrak{d}$, define

$$\kappa_{\hbar}(a_1, \dots, a_n) := \sum a_i x_i y_i + \hbar\phi(\boldsymbol{a}).$$

Our quantum moment map is the restriction of $\kappa_{\hbar}$ to $\mathfrak{t} \subset \mathfrak{d}$. Note that this agrees mod $\hbar$ with the pullback of functions from $\mathfrak{t}^*$ under Φ.

Let $\mathcal{C}_{\phi} = \mathcal{O}^{\hbar}_{\mathfrak{Z}^n}/\mathcal{O}^{\hbar}_{\mathfrak{Z}^n} \cdot \kappa_{\hbar}(\mathfrak{t})$ be the quotient of $\mathcal{O}^{\hbar}_{\mathfrak{Z}^n}$ by the left ideal generated by these functions. This is supported on the subset $\Phi^{-1}(T^{\vee}_{\mathbb{R}} \times \{0\})$. We have an endomorphism sheaf $\mathscr{E}nd(\mathcal{C}_{\phi})$ of this sheaf of modules over $\mathcal{O}^{\hbar}_{\mathfrak{Z}^n}$.

Definition 4.14 Let $\mathcal{O}^{\hbar}_{\phi}$ be the sheaf of algebras on $\mathfrak{D}$ defined by restricting $\mathscr{E}nd(\mathcal{C}_{\phi})$ to $\Phi^{-1}(\zeta')$ and pushing the result forward to $\mathfrak{D}$.

One can easily check, as in [26], that $\mathcal{O}^{\hbar}_{\phi}$ defines a deformation quantization of $\mathfrak{D}$, that is, this sheaf is free and complete over $\mathbb{C}[[\hbar]]$, we have an isomorphism of algebra sheaves $\mathcal{O}^{\hbar}_{\phi}(0)/\hbar\mathcal{O}^{\hbar}_{\phi}(0) \cong \mathscr{O}_{\mathfrak{D}}$, and given two meromorphic sections f, g, we have

$$f \star g - g \star f \equiv \hbar\{f, g\} \pmod{\hbar}.$$

4.6 G-equivariant modules

By a $\mathcal{O}^{\hbar}_{\phi}$–module, we will always mean a sheaf $\mathcal{M}$ of $\mathcal{O}^{\hbar}_{\phi}$–modules which admits a good lattice $\mathcal{M}(0) \subset \mathcal{M}$.

By construction, the map $\kappa_{\hbar}\colon \mathfrak{d} \to \mathcal{O}^{\hbar}_{\mathfrak{Z}^n}$ descends to a map $\mathfrak{g} \to \mathcal{O}^{\hbar}_{\phi}$, which quantizes the moment map for the action of G on $\mathfrak{D}$.

Definition 4.15 We call a $\mathcal{O}^{\hbar}_{\phi}$–module *pre-weakly G–equivariant* if the action of $\mathfrak{g}$ via left multiplication by $\hbar^{-1}\kappa_{\hbar}$ on the sections on any G–invariant open set is locally finite, ie it is spanned by its generalized weight spaces for this torus.

A pre-weak equivariant structure can be upgraded to a weak equivariant structure as follows: we can assume that $\mathcal{M}$ is indecomposable, so all weights appearing are in a single coset of the character lattice of G. We can take the semisimple part of the action of each element of $\mathfrak{g}$, and globally shift by a character of the Lie algebra to make all weights appearing integral. The resulting action integrates to a weak G–equivariant structure (but we do not want to fix a specific one); we call such an action compatible with the $\mathcal{O}^{\hbar}_{\phi}$–module structure. Note that pre-weakly G–equivariant modules are a Serre subcategory.

Lemma 4.16 *Any pre-weakly G–equivariant module $\mathcal{M}$ is supported on $q_{\mathfrak{D}}^{-1}(0)$.*

Proof Given any nonzero $X \in \mathfrak{g}$, consider the action of $\hbar^{-1}k := \hbar^{-1}\kappa_{\hbar}(X)$ on $\mathcal{M}(U)$ for U a G–equivariant open subset. By the assumption of local finiteness, for each $m \in \mathcal{M}(U)$, there is a monic polynomial $p(u) = u^d + p_{d-1}u^{d-1} + \dots + p_0 \in \mathbb{C}[u]$ such that $p(\hbar^{-1}k)m = 0$.

If $U \cap q_{\mathfrak{D}}^{-1}(0) = \varnothing$, then k is invertible in $\mathcal{O}_\phi^\hbar(0)$, and so we have

$$m = \hbar(-p_{d-1}k^{-1} - \cdots - p_0 p_{d-1}\hbar^{d-1}k^{-d})m.$$

Thus, for any choice of good lattice $\mathcal{M}(0) \subset \mathcal{M}$, we have $\mathcal{M}(0)(U) \subset \hbar\mathcal{M}(0)(U)$. Nakayama's lemma then implies that $\mathcal{M}(0)(U) = 0$, so $\mathcal{M}(U) = 0$. □

Unfortunately, the action of $\mathbb{S}G$ on $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$ does not preserve $\widetilde{\mathfrak{D}}$. We can nevertheless speak of $\mathbb{S}G$–equivariance on $\widetilde{\mathfrak{D}}$ and $\mathfrak{D}$, as follows.

Let $\mathcal{M}$ be a pre-weakly G–equivariant $\mathcal{O}_\phi^\hbar$–module. Let $\nu^*\mathcal{M}$ be the pullback of this module to $\widetilde{\mathfrak{D}}$. We write $(\nu^*\mathcal{M})^{\mathrm{alg}}$ for the pushforward of $\nu^*\mathcal{M}$ along the inclusion $\widetilde{\mathfrak{D}} \to \widetilde{\mathfrak{D}}^{\mathrm{alg}}$. By Lemma 4.16, $\nu^*\mathcal{M}$ is supported on $\widetilde{\mathfrak{C}} \subset \widetilde{\mathfrak{D}}$, and this subset remains closed in $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$. Thus the support is not enlarged.

Fix $\boldsymbol{x}_0 \in \widetilde{\Lambda}$.

Definition 4.17 A *pre-weakly* $\mathbb{S}G$*–equivariant* structure on a pre-weakly G–equivariant $\mathcal{O}_\phi^\hbar$–module $\mathcal{M}$ is an action of the Lie algebra $\mathrm{Lie}(\mathbb{S}_{\boldsymbol{x}_0})$ commuting with $\mathfrak{g}$ which integrates to an equivariant structure for $\mathbb{S}_{\boldsymbol{x}_0}$ on $(\nu^*\mathcal{M})^{\mathrm{alg}}$.

We write $\mathcal{O}_\phi^\hbar\text{–mod}^{\mathbb{S}G}$ for the category of such modules. Since a homomorphism between pre-weakly $\mathbb{S}G$–equivariant modules is $\mathrm{Lie}(\mathbb{S}_{\boldsymbol{x}_0})$–equivariant, multiplication by $\hbar$ is not a morphism in this category, so this category is $\mathbb{C}$–linear, not $\mathbb{C}((\hbar))$–linear.

As with pre-weakly G–equivariant modules, after making some auxiliary choices, we can endow a pre-weakly $\mathbb{S}G$–equivariant-module with a "compatible" action of the torus $\mathbb{S}G$, which integrates the semisimple part of (a shift of) the infinitesimal action.

Lemma 4.18 *Let $\mathcal{M} \in \mathcal{O}_\phi^\hbar\text{–mod}^{\mathbb{S}G}$. Fix a compatible action of $\mathbb{S}G$. The action of $\mathbb{C}^*$ on $\nu^*\mathcal{M}^{\mathrm{alg}}$ induced by the composition $\mathbb{C}^* \cong \mathbb{S}_{\boldsymbol{y}} \to \mathbb{S}G$ does not depend on the G–equivariant structure, up to isomorphism.*

Proof Again, we can reduce to the case where $\mathcal{M}$ is indecomposable. By construction, any two compatible G–equivariant structures on $\mathcal{M}$ differ by tensor product with a character of the group G, so the induced $\mathbb{S}_{\boldsymbol{y}}$ structures differ by tensor product with a character of $\mathbb{S}_{\boldsymbol{y}}$, which we can think of as the integer weight w. Since $\hbar$ has weight 1 under $\mathbb{S}_{\boldsymbol{y}}$, multiplication by $\hbar^w$ intertwines these two actions, and gives an isomorphism between the two $\mathbb{S}_{\boldsymbol{y}}$–equivariant structures. □

4.7 The deformation quantization near a component of $\mathfrak{C}$

Given $\phi \in \mathfrak{t}^*_{\mathbb{Q}}$, we can define a fractional line bundle ℓ_ϕ on any quotient by a free T–action. The component $\mathfrak{X}_{\boldsymbol{x}}$ was defined by a free $T_{\mathbb{R}}$–action; by standard toric geometry, it also carries a canonical presentation as a free T–quotient. Applying this construction to $\mathfrak{X}_{\boldsymbol{x}}$ thus yields a bundle $\ell_{\phi,\boldsymbol{x}}$. If $\phi \in \mathfrak{t}^*_{\mathbb{Z}}$, the set of honest characters, then this is an honest line bundle; otherwise, it gives a line bundle over a gerbe, but we can still define an associated Picard groupoid, and thus a sheaf of twisted differential operators (TDO) on $\mathfrak{X}_{\boldsymbol{x}}$. Let $\Omega_{\boldsymbol{x}}$ be the canonical line bundle on $\mathfrak{X}_{\boldsymbol{x}}$, and $\Omega_{\boldsymbol{x}}^{1/2}$ the half-density fractional line bundle. It is a classical fact that $\Omega_{\boldsymbol{x}} = l_{-\phi_0,\boldsymbol{x}}$, where ϕ_0 is the sum of all T–characters of $\mathbb{C}^n$ induced by the map $T \to D$.

We let $D_{\phi,\boldsymbol{x}}$ denote the TDO associated to the fractional line bundle $\ell_{\phi,\boldsymbol{x}} \otimes \Omega_{\boldsymbol{x}}^{1/2}$, and let $\mathcal{R}_{\phi,\boldsymbol{x}}$ be its microlocalization on $T^*\mathfrak{X}_{\boldsymbol{x}}$. That is, $\mathcal{R}_{\phi,\boldsymbol{x}}$ is a sheaf in the classical topology on $T^*\mathfrak{X}_{\boldsymbol{x}}$ whose sections on T^*U for $U \subset \mathfrak{X}_{\boldsymbol{x}}$ are the Rees algebra for the order filtration on $D_{\phi,\boldsymbol{x}}(U)$; for an open subset $V \subset T^*U$ (where we can assume without loss of generality that U is affine), we further invert any element of the Rees algebra whose image under the map $\mathcal{R}_{\phi,\boldsymbol{x}}(U)/\hbar\mathcal{R}_{\phi,\boldsymbol{x}}(U) \cong \mathcal{O}_{T^*U}(T^*U)$ is invertible on V. The construction of this algebra is discussed in more detail in [10, Section 4.1]. We'll be more interested in its localization:

Definition 4.19 $$\mathcal{W}_{\phi,\boldsymbol{x}} := \mathcal{R}_{\phi,\boldsymbol{x}}[\hbar^{-1/2}].$$

If we equip a module $\mathcal{M}$ over the TDO $D_{\phi,\boldsymbol{x}}$ with a good filtration, which for technical reasons we'll index with $\frac{1}{2}\mathbb{Z}$, its Rees module $\mathcal{M}(0)$ generated by $\hbar^{-k}\mathcal{M}_{\leq k}$ for $k \in \frac{1}{2}\mathbb{Z}$ is a coherent module over the Rees algebra; we can use this as a definition of good filtration. That is, it is a coherent sheaf of $\mathcal{R}_{\phi,\boldsymbol{x}}$–modules, equipped with a $\mathbb{C}^*$–equivariant structure for the squared scaling $\mathbb{C}^*$–action (or equivalently, a grading of its sections on T^*U). Inverting $\hbar$, we obtain a $\mathcal{W}_{\phi,\boldsymbol{x}}$–module $\mathfrak{M} = \mathcal{M}(0)[\hbar^{-1/2}]$ which is independent of the choice of good filtration, which is *good* in the sense of [10, Section 4], that is, it admits a coherent, $\mathbb{C}^*$–equivariant $\mathcal{R}_{\phi,\boldsymbol{x}}$–lattice. By [10, Proposition 4.5], this is an equivalence between coherent $D_{\phi,\boldsymbol{x}}$–modules and good $\mathcal{W}_{\phi,\boldsymbol{x}}$–modules.

Theorem 4.20 *We have an isomorphism of algebra sheaves $\iota_{\boldsymbol{x}}^*\mathcal{O}_\phi^\hbar \cong \mathcal{W}_{\phi,\boldsymbol{x}}|_{\tilde{\mathfrak{D}}_{\boldsymbol{x}}}$.*

Proof First, we check that this holds in the base case, ie when $\mathfrak{D} = \mathfrak{Z}$. It is convenient to check this on the universal cover $\tilde{\mathfrak{Z}}$. By the $\mathbb{Z}$–symmetry of the latter, it is enough to check for a single component of the core. Hence, consider the copy of $\mathbb{CP}^1$ in the union of $\mathfrak{W}_0 \cup \mathfrak{W}_1$. Using superscripts to indicate which $\mathfrak{W}_*$ we work on, we have birational coordinates $y^{(0)} = 1/x^{(1)}$ and $x^{(0)} = y^{(1)}(x^{(1)})^2$. We thus have an isomorphism of $\mathfrak{W}_0 \cup \mathfrak{W}_1$ to $T^*\mathbb{CP}^1$ with coordinate z and dual coordinate ξ sending

$$x^{(1)} \mapsto z, \quad y^{(1)} \mapsto \xi, \quad x^{(0)} \mapsto z^2\xi, \quad y^{(0)} \mapsto \frac{1}{z}.$$

We can quantize this to a map from $\mathcal{O}_{\mathfrak{W}}^\hbar$ to $\mathcal{R}_\phi$ by the corresponding formulas

$$x^{(1)} \mapsto z, \quad y^{(1)} \mapsto \hbar\frac{d}{dz}, \quad x^{(0)} \mapsto \hbar z^2\frac{d}{dz}, \quad y^{(0)} \mapsto \frac{1}{z}.$$

This induces an isomorphism of sheaves, which in turn restricts to an isomorphism $\iota_{\boldsymbol{x}}^*\mathcal{O}_{\mathfrak{Z}}^\hbar \to \mathcal{W}_{\phi,\boldsymbol{x}}|_{\mathfrak{Z}_0}$. Under this isomorphism,

$$q = x^{(1)}y^{(1)} \mapsto \hbar z\frac{d}{dz} - \frac{\hbar}{2}.$$

To proceed to the general case, we consider $\widetilde{\mathfrak{Z}}^n$ and its quantized T–moment map $\kappa_\hbar$. Fix as above an open subset of isomorphic to $(T^*\mathbb{P}^1)^n$. Applying the above morphism to the image of $\kappa_\hbar$, we obtain

$$\kappa_\hbar(a_1,\dots,a_n) \mapsto \sum_i a_i z_i \frac{d}{dz_i} - \frac{\hbar}{2} + \hbar\phi(\boldsymbol{a}) = \sum_i \frac{a_i}{2}\Big(z_i \frac{d}{dz_i} + \frac{d}{dz_i} z_i\Big) + \hbar\phi(\boldsymbol{a}).$$

The result then follows from the compatibility of twisted microlocal differential operators with symplectic reduction as in [10, Proposition 3.16]. We can identify the twist of a TDO from its period by [10, Proposition 4.4]. ◻

Thus, given a $\widetilde{\mathcal{O}}^\hbar_\phi$–module $\mathcal{M}$, we can pull it back to an $\mathcal{W}_{\phi,\boldsymbol{x}}|_{\widetilde{\mathfrak{D}}_{\boldsymbol{x}}}$–module $\mathcal{M}|_{\widetilde{\mathfrak{D}}_{\boldsymbol{x}}}$ on $\widetilde{\mathfrak{D}}_{\boldsymbol{x}} \subset T^*\mathfrak{X}_{\boldsymbol{x}}$.

If we additionally choose a $\mathbb{S}_{\boldsymbol{x}}$–equivariant structure which makes $\mathcal{M}|_{\widetilde{\mathfrak{D}}_{\boldsymbol{x}}}$ into a good module, then the equivalence of [10, Proposition 4.5] will give a corresponding module over the TDO $D_{\phi,\boldsymbol{x}}$, with a choice of good filtration.

Definition 4.21 Given $\mathcal{M} \in \mathcal{O}^\hbar_\phi\text{–mod}^{\mathbb{S}G}$, let $\sigma_{\boldsymbol{x}}(\mathcal{M}) \in D_{\phi,\boldsymbol{x}}$–mod be the module defined as above for some choice of $\mathbb{S}_{\boldsymbol{x}}$–equivariant structure.

The resulting D–module does not depend on the choice of compatible $\mathbb{S}G$–equivariant structure, by Lemma 4.18. It does carry a good filtration which depends on this choice, but only up to a shift on each indecomposable summand of $\sigma_{\boldsymbol{x}}(\mathcal{M})$.

The modules $\sigma_{\boldsymbol{x}}(\mathcal{M})$ for different $\boldsymbol{x}$ are compatible in the following sense. As discussed previously, the intersection $\mathfrak{X}_{\boldsymbol{x}} \cap \widetilde{\mathfrak{D}}^{\mathrm{alg}}_{\boldsymbol{y}}$ is precisely the conormal bundle $N_{\boldsymbol{x},\boldsymbol{y}} = N^*_{\mathfrak{X}_{\boldsymbol{y}}}(\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}})$ to $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$ in $T^*\mathfrak{X}_{\boldsymbol{y}}$. Thus the intersection $\widetilde{\mathfrak{D}}^{\mathrm{alg}}_{\boldsymbol{x},\boldsymbol{y}} = \widetilde{\mathfrak{D}}^{\mathrm{alg}}_{\boldsymbol{x}} \cap \widetilde{\mathfrak{D}}^{\mathrm{alg}}_{\boldsymbol{y}}$ can be identified with $T^*(N_{\boldsymbol{x},\boldsymbol{y}})$ or swapping the roles of $\boldsymbol{x}$, $\boldsymbol{y}$ with $T^*(N_{\boldsymbol{y},\boldsymbol{x}})$.

Since the vector bundles $N^*_{\mathfrak{X}_{\boldsymbol{y}}}(\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}})$ and $N^*_{\mathfrak{X}_{\boldsymbol{x}}}(\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}})$ are dual, so Fourier transform $\mathcal{F}_{\boldsymbol{y},\boldsymbol{x}}$ gives an equivalence between the categories of pre-weakly G–equivariant D–modules on these spaces, and between constructible sheaves with $\mathbb{R}$–coefficients, which are compatible with respect to the solution functor. By construction, we thus have

$$\mathcal{F}_{\boldsymbol{y},\boldsymbol{x}}\sigma_{\boldsymbol{x}}(\mathcal{M})|_{N_{\boldsymbol{x},\boldsymbol{y}}} \cong \sigma_{\boldsymbol{y}}(\mathcal{M})|_{N_{\boldsymbol{y},\boldsymbol{x}}}. \tag{4-6}$$

4.8 Preliminaries on the Ext–algebra of the simples

Assume that ϕ is chosen so that $\ell_\phi \otimes \Omega^{1/2}$ is an honest line bundle for all $\boldsymbol{x}$. From now on, we use the abbreviations $\mathcal{W}_{\boldsymbol{x}} := \mathcal{W}_{\phi,\boldsymbol{x}}$, $D_{\boldsymbol{x}} := D_{\phi,\boldsymbol{x}}$ and $\ell_{\boldsymbol{x}} := \ell_{\phi,\boldsymbol{x}}$, since the dependence on ϕ will not play any further role in this paper.

Remark 4.22 Recall that $\Omega^{1/2}_{\boldsymbol{x}}$ equals $\ell_{-\phi_0/2,\boldsymbol{x}}$ where ϕ_0 is the sum of all T–characters induced by the embedding $T \to D$. Thus our assumption will be satisfied whenever $\phi(\boldsymbol{a}) \in \mathbb{Z} + \frac{1}{2}\sum a_i$ for all $\boldsymbol{a} \in \mathfrak{t}_{\mathbb{Z}}$.

For example, we can let ϕ be the restriction of the element $\left(\frac{1}{2},\dots,\frac{1}{2}\right)\in\mathfrak{d}^*$. Nothing we do will depend on this choice; in fact, the categories of $\mathcal{O}_\phi^\hbar$–modules for ϕ in a fixed coset of $\mathfrak{t}_\mathbb{Z}^*$ are all equivalent via tensor product with quantizations of line bundles on $\mathfrak{D}$ (as in [10, Section 5.1]), so our calculations will be independent of this choice.

In this case, the sheaf $\mathcal{W}_{\boldsymbol{x}}$ naturally acts on $\mathcal{L}'_{\boldsymbol{x}} := \ell_{\boldsymbol{x}}\otimes\Omega_{\boldsymbol{x}}^{1/2}((\hbar))$ as a sheaf on $\mathfrak{X}_{\boldsymbol{x}}$ pushed forward into $\widetilde{\mathfrak{D}}_{\boldsymbol{x}}$; under the equivalence of [10, Proposition 4.5] mentioned above, this corresponds to the twisted D–module $\ell_{\boldsymbol{x}}\otimes\Omega_{\boldsymbol{x}}^{1/2}$. Of course, this sheaf is equivariant for the action of $\mathbb{S}_{\boldsymbol{x}}$, and pre-weakly G–equivariant.

Via the maps

$$\iota_{\boldsymbol{x}}:\widetilde{\mathfrak{D}}_{\boldsymbol{x}}\to\mathfrak{D}\quad\text{and}\quad\tilde{\iota}_{\boldsymbol{x}}:\widetilde{\mathfrak{D}}_{\boldsymbol{x}}\to\widetilde{\mathfrak{D}},$$

we can define modules over $\mathcal{O}_\phi^\hbar$ and $\widetilde{\mathcal{O}}_\phi^\hbar$:

Definition 4.23 Let $\mathcal{L}_{\boldsymbol{x}}=\iota_*\mathcal{L}'_{\boldsymbol{x}}$ and $\widetilde{\mathcal{L}}_{\boldsymbol{x}}=\tilde{\iota}_*\mathcal{L}'_{\boldsymbol{x}}$.

Using $\mathbb{S}_{\boldsymbol{x}}$–equivariance, and the pre-weak G–equivariant of this module, we obtain a twisted D–module $\sigma_{\boldsymbol{y}}\widetilde{\mathcal{L}}_{\boldsymbol{x}}$. Recall that we have a universal cover map $\nu:\widetilde{\mathfrak{D}}\to\mathfrak{D}$.

Proposition 4.24 *We have isomorphisms* $\nu_*\widetilde{\mathcal{L}}_{\boldsymbol{x}}\cong\mathcal{L}_{\boldsymbol{x}}$ *and* $\nu^*\mathcal{L}_{\boldsymbol{x}}\cong\bigoplus_{\boldsymbol{z}\in\mathfrak{g}_\mathbb{Z}^*}\widetilde{\mathcal{L}}_{\boldsymbol{x}+\boldsymbol{z}}$.

The first category we will consider on the A side of our correspondence is DQ, the dg–subcategory of $\mathcal{O}_\phi^\hbar\text{–mod}^{\mathbb{S}G}$ generated by $\mathcal{L}_{\boldsymbol{x}}$ for all $\boldsymbol{x}$. As observed before, since weakly G–equivariant modules form a Serre subcategory, any finite-length object in this category is pre-weakly G–equivariant, and so we can define the D–modules $\sigma_{\boldsymbol{y}}(\mathcal{M})$ for any module M in this category.

This has a natural t–structure, whose heart is an abelian category dq. We similarly let $\widetilde{\mathsf{DQ}}$ be the dg–subcategory generated by $\widetilde{\mathcal{L}}_{\boldsymbol{x}}$, and $\widetilde{\mathsf{dq}}$ the heart of the natural t–structure. This definition might seem slightly ad hoc, but we will later see that it is motivated by our notion of microlocal mixed Hodge modules.

Since the Ext sheaf between $\widetilde{\mathcal{L}}_{\boldsymbol{x}}$ and $\widetilde{\mathcal{L}}_{\boldsymbol{y}}$ is supported on the intersection between $\mathfrak{X}_{\boldsymbol{x}}$ and $\mathfrak{X}_{\boldsymbol{y}}$, we have

$$\operatorname{Ext}_{\widetilde{\mathsf{DQ}}}(\widetilde{\mathcal{L}}_{\boldsymbol{y}},\widetilde{\mathcal{L}}_{\boldsymbol{x}})\cong\operatorname{Ext}_{\mathcal{W}_{\boldsymbol{x}}|_{\widetilde{\mathfrak{D}}_{\boldsymbol{x}}}\text{–mod}^{\mathbb{S}}}(\tilde{\iota}_{\boldsymbol{x}}^*\widetilde{\mathcal{L}}_{\boldsymbol{y}},\mathcal{L}'_{\boldsymbol{x}}).$$

In fact, if we replace $\widetilde{\mathcal{L}}_{\boldsymbol{y}}$ by an injective resolution, we see that this induces a homotopy equivalence between the corresponding Ext complexes. Since $\mathscr{L}'_{\boldsymbol{x}}$ is supported on the zero section, Ext to it is unchanged by passing to an open subset containing this support and

$$\operatorname{Ext}_{\widetilde{\mathsf{DQ}}}(\widetilde{\mathcal{L}}_{\boldsymbol{y}},\widetilde{\mathcal{L}}_{\boldsymbol{x}})\cong\operatorname{Ext}_{D(\mathfrak{X}_{\boldsymbol{x}})}(\sigma_{\boldsymbol{x}}\widetilde{\mathcal{L}}_{\boldsymbol{y}},\ell_{\boldsymbol{x}}\otimes\Omega_{\boldsymbol{x}}^{1/2}),\tag{4-7}$$

where the latter Ext is computed in the category of D–modules on the toric variety $\mathfrak{X}_{\boldsymbol{x}}$. In the toric variety $\mathfrak{X}_{\boldsymbol{x}}$, the preimage of the intersection with the image of $\mathfrak{X}_{\boldsymbol{y}}$ is a toric subvariety corresponding to the intersection of the corresponding chambers in $\mathfrak{B}_\zeta^{\text{per}}$.

Lemma 4.25 *The microlocalization $\sigma_{\boldsymbol{x}}\widetilde{\mathcal{L}}_{\boldsymbol{y}}$ is the line bundle $\ell_{\boldsymbol{y}} \otimes \Omega_{\boldsymbol{y}}^{1/2}$ pulled back to $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$ and pushed forward to $\mathfrak{X}_{\boldsymbol{x}}$ as a $D_{\boldsymbol{x}}$–module.*

Proof Consider the intersection of $\mathfrak{X}_{\boldsymbol{y}}$ with $\widetilde{\mathfrak{D}}_{\boldsymbol{x}}^{\mathrm{alg}}$. This is a closed $\mathbb{S}_{\boldsymbol{x}}$–invariant Lagrangian closed subset, so it is the conormal to its intersection with the zero section $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$. The D–module $\sigma_{\boldsymbol{x}}\widetilde{\mathcal{L}}_{\boldsymbol{y}}$ has singular support on this subvariety, and thus must be a local system on $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$, which is necessarily $\ell_{\boldsymbol{y}} \otimes \Omega_{\boldsymbol{y}}^{1/2}$. □

Since $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$ is a smooth toric subvariety, the sheaf Ext between $\sigma_{\boldsymbol{x}}\widetilde{\mathcal{L}}_{\boldsymbol{y}}$ and $\ell_{\boldsymbol{x}} \otimes \Omega_{\boldsymbol{x}}^{1/2}$ is $\mathbb{C}_{\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}}[-k]$, where k is the codimension of $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$ in $\mathfrak{X}_{\boldsymbol{x}}$. This shows that we have an isomorphism

$$\operatorname{Ext}^m_{\mathsf{DQ}}(\widetilde{\mathcal{L}}_{\boldsymbol{y}}, \widetilde{\mathcal{L}}_{\boldsymbol{x}}) \cong H^{m-k}(\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}; \mathbb{C}). \tag{4-8}$$

We are interested in the class $\boldsymbol{d}_{\boldsymbol{y},\boldsymbol{x}}$ in the left-hand space corresponding to the identity in $H^*(\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}; \mathbb{C})$. Unfortunately, this is only well-defined up to scalar. We will only need the case where $|\boldsymbol{x}-\boldsymbol{y}|_1 = 1$. In this case, we can define $\boldsymbol{d}_{\boldsymbol{y},\boldsymbol{x}}$ (without scalar ambiguity) as follows.

Consider the inclusions $\mathfrak{X}_{\boldsymbol{x}} \setminus (\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}) \xhookrightarrow{j} \mathfrak{X}_{\boldsymbol{x}} \xhookleftarrow{i} \mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}$ and the corresponding sequence of D–modules

$$0 \to \mathscr{O}_{\mathfrak{X}_{\boldsymbol{x}}} \to j_*\mathscr{O}_{\mathfrak{X}_{\boldsymbol{x}} \setminus (\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}})} \to j_*\mathscr{O}_{\mathfrak{X}_{\boldsymbol{x}} \setminus (\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}})}/\mathscr{O}_{\mathfrak{X}_{\boldsymbol{x}}} \to 0. \tag{4-9}$$

Any identification of the right-hand D–module with $i_!\mathscr{O}_{\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}}$ defines a class

$$\boldsymbol{d}_{\boldsymbol{y},\boldsymbol{x}} \in \operatorname{Ext}^1(\mathscr{O}_{\mathfrak{X}_{\boldsymbol{x}}}, i_!\mathscr{O}_{\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}}) = \operatorname{Ext}^1_{\mathsf{DQ}}(\widetilde{\mathcal{L}}_{\boldsymbol{y}}, \widetilde{\mathcal{L}}_{\boldsymbol{x}}).$$

Such an identification is obtained by picking the germ of a function g on $\mathfrak{X}_{\boldsymbol{x}}$ in the formal neighborhood of $\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}$ that vanishes on this divisor with order 1. Given such a function, the map $f \mapsto \tilde{f}/g$, where $\tilde{f}$ is an extension of a meromorphic function on $\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}$ to the formal neighborhood, defines an isomorphism of D–modules $i_!\mathscr{O}_{\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}} \to j_*\mathscr{O}_{\mathfrak{X}_{\boldsymbol{x}} \setminus (\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}})}/\mathscr{O}_{\mathfrak{X}_{\boldsymbol{x}}}$.

We can arrange our choice of chart in $\widetilde{\mathfrak{Z}}^n$ so that $\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}$ is defined by the vanishing of one of the coordinate functions; note that in this case, $\mathfrak{X}_{\boldsymbol{y}} \cap \mathfrak{X}_{\boldsymbol{x}}$ is defined inside $\mathfrak{X}_{\boldsymbol{y}}$ by the vanishing of the symplectically dual coordinate function (for instance, if the first is defined by the vanishing of x_i, then the latter will be defined by y_i). We choose this as the function to define $\boldsymbol{d}_{\boldsymbol{x},\boldsymbol{y}}$.

Definition 4.26 For any $\boldsymbol{x}, \boldsymbol{y}$ such that $|\boldsymbol{x}-\boldsymbol{y}|_1 = 1$, let $\boldsymbol{d}_{\boldsymbol{y},\boldsymbol{x}} \in \operatorname{Ext}^1_{\mathsf{DQ}}(\widetilde{\mathcal{L}}_{\boldsymbol{y}}, \widetilde{\mathcal{L}}_{\boldsymbol{x}})$ be the class defined by the above prescription.

4.9 Mirror symmetry

We are almost ready to compare the first and second halves of this paper. First, we need to match the parameters entering into our constructions. Recall that $\mathfrak{D}$ depends on a choice of generic stability parameter $\zeta \in T^*_{\mathbb{R}}$, Likewise, the hypertoric enveloping algebra in characteristic p depends on a central character $\lambda \in \mathfrak{t}^*_{\mathbb{F}_p}$. The algebra $\widetilde{H}^{\lambda,!}$ which describes the Ext groups of its simple modules (Definition 3.23) thereby also depends on λ. In order to match ζ and λ, we identify $\mathfrak{t}^*_{\mathbb{F}_p}$ with $\mathfrak{t}^*_{\mathbb{Z}}/p\mathfrak{t}^*_{\mathbb{Z}}$ and thereby embed it in $T^*_{\mathbb{R}} = \mathfrak{t}^*_{\mathbb{R}}/\mathfrak{t}^*_{\mathbb{Z}}$ via $\lambda \mapsto (1/p)\lambda$. From now on we suppose that λ is smooth, and that ζ is its image in $T^*_{\mathbb{R}}$. It follows that $\mathfrak{B}^{\mathrm{per}}_{\zeta}$ is the real form of $\mathfrak{A}^{\mathrm{per}}_{\lambda}$, and their sets of chambers are naturally in bijection. Since the chambers of $\mathfrak{A}^{\mathrm{per}}_{\lambda}$ index the simples of $A^{\lambda}_{\mathbb{K}}$–mod$_o$ and the chambers of $\mathfrak{B}^{\mathrm{per}}_{\zeta}$ index the simples of DQ, we obtain a bijection of simple objects.

We now show that the Ext–algebras of these simples share an integral form.

Theorem 4.27 *We have isomorphisms of algebras*

$$\widetilde{H}^!_{\zeta,\mathbb{C}} \cong \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\widetilde{\Lambda}(\zeta)} \mathrm{Ext}_{\widetilde{\mathsf{DQ}}}(\widetilde{\mathcal{L}}_{\boldsymbol{x}}, \widetilde{\mathcal{L}}_{\boldsymbol{y}}) \quad \text{and} \quad H^!_{\zeta,\mathbb{C}} \cong \bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\Lambda(\zeta)} \mathrm{Ext}_{\mathsf{DQ}}(\mathcal{L}_{\boldsymbol{x}}, \mathcal{L}_{\boldsymbol{y}}),$$

sending $d_{\boldsymbol{x},\boldsymbol{y}} \mapsto \boldsymbol{d}_{\boldsymbol{x},\boldsymbol{y}}$ when $\boldsymbol{y} \in \alpha_i(\boldsymbol{x})$.

Proof We need to check that the rule $d_{\boldsymbol{x},\boldsymbol{y}} \mapsto \boldsymbol{d}_{\boldsymbol{x},\boldsymbol{y}}$ defines a homomorphism, ie that the relations (3-8a)–(3-8d) hold in $\bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\widetilde{\Lambda}(\zeta)} \mathrm{Ext}_{\widetilde{\mathsf{DQ}}}(\widetilde{\mathcal{L}}_{\boldsymbol{x}}, \widetilde{\mathcal{L}}_{\boldsymbol{y}})$.

(1) The relation (3-8a) follows from the fact that when $|\boldsymbol{x}-\boldsymbol{y}|_1 = 1$, the element $\boldsymbol{d}_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{d}_{\boldsymbol{y},\boldsymbol{x}}$ is the class in $H^2(\mathfrak{X}_{\boldsymbol{x}};\mathbb{Q})$ dual to the divisor $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$, while the class t_i is defined by the Chern class of the corresponding line bundle, for which a natural section vanishes with order one on $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$ for $\boldsymbol{y} \in \alpha(i)$ and nowhere else.

(2) The relations (3-8b) and (3-8c) equate two elements of the one-dimensional space $\mathrm{Ext}^2(\mathcal{L}_{\boldsymbol{x}}, \mathcal{L}_{\boldsymbol{w}}) \cong H^0(\mathfrak{X}_x \cap \mathfrak{X}_w;\mathbb{C})$. Thus, we only need check that we have the scalars right, and this can be done after restricting to any small neighborhood where all the classes under consideration have nonzero image.

Thus ultimately we can reduce to assuming $\mathfrak{X}_{\boldsymbol{x}} = \mathbb{C}^2$, and $\mathfrak{X}_{\boldsymbol{y}}$ and $\mathfrak{X}_{\boldsymbol{w}}$ are the conormals to the coordinate lines, and $\mathfrak{X}_{\boldsymbol{z}}$ the cotangent fiber over 0. Let r_1, r_2 be the usual coordinates on $\mathbb{C}^2$, and ∂_1, ∂_2 be the directional derivatives for these coordinates. Thus, we are interested in comparing the Ext^2's given by the sequences in the first and third row of the diagram below. Both sequences are quotients of the free Koszul resolution in the second row:

$$\begin{tikzcd}
\dfrac{D}{D\cdot(r_1,r_2)} \arrow[r, "1\mapsto 1"] & \dfrac{D_{\partial_2}}{D_{\partial_2}\cdot(r_1,r_2)} \arrow[r, "1/\partial_2\mapsto 1"] & \dfrac{D_{\partial_1}}{D_{\partial_1}\cdot(r_1,\partial_2)} \arrow[r, "1/\partial_1\mapsto 1"] & \dfrac{D}{D\cdot(\partial_1,\partial_2)} \\
D \arrow[u, "1\mapsto 1"'] \arrow[d, "1\mapsto -1"] \arrow[r, "1\mapsto\left[\begin{smallmatrix}\partial_2\\ -\partial_1\end{smallmatrix}\right]"] & D\oplus D \arrow[u, "\left[\begin{smallmatrix}a\\ b\end{smallmatrix}\right]\mapsto a/\partial_2"'] \arrow[d, "\left[\begin{smallmatrix}a\\ b\end{smallmatrix}\right]\mapsto b/\partial_1"] \arrow[r, "\left[\begin{smallmatrix}a\\ b\end{smallmatrix}\right]\mapsto a\partial_1+b\partial_2"] & D \arrow[u, "1\mapsto 1/\partial_1"'] \arrow[d, "1\mapsto 1/\partial_2"] \arrow[r, "1\mapsto 1"] & \dfrac{D}{D\cdot(\partial_1,\partial_2)} \arrow[u, "\cong"'] \arrow[d, "\cong"] \\
\dfrac{D}{D\cdot(r_1,r_2)} \arrow[r, "1\mapsto 1"] & \dfrac{D_{\partial_1}}{D_{\partial_1}\cdot(r_1,r_2)} \arrow[r, "1/\partial_1\mapsto 1"] & \dfrac{D_{\partial_2}}{D_{\partial_2}\cdot(r_2,\partial_1)} \arrow[r, "1/\partial_2\mapsto 1"] & \dfrac{D}{D\cdot(\partial_1,\partial_2)}
\end{tikzcd}$$

The opposite signs in the leftmost column confirm that we have $\boldsymbol{d}_{z,w}\boldsymbol{d}_{w,x} = -\boldsymbol{d}_{z,y}\boldsymbol{d}_{y,x}$. Hence the elements $\boldsymbol{d}_{x,y}$ satisfy the relations (3-8b)–(3-8c).

(3) The relations (3-8d) follows from the fact that in this case $\mathfrak{X}_x \cap \mathfrak{X}_z = \varnothing$.

Recall that the complex dimension of $e_x \widetilde{H}^!_{\zeta,\mathbb{C}} e_y$ coincides with that of $H^*(\mathfrak{X}_x \cap \mathfrak{X}_y; \mathbb{C})$, as we discussed in Section 3.8. Thus the spaces $e_x \widetilde{H}^!_{\zeta,\mathbb{C}} e_y$ and $H^*(\mathfrak{X}_x \cap \mathfrak{X}_y; \mathbb{C})$ are vector spaces of the same rank. Thus, in order to show that our map is an isomorphism, it is enough to show that it is surjective.

By Kirwan surjectivity, the fundamental class generates $H^*(\mathfrak{X}_x \cap \mathfrak{X}_y; \mathbb{C})$ as a module over the Chern classes of line bundles associated to representations of T. Since the fundamental classes are images of $\pm\boldsymbol{d}_{x,y}$ and the Chern classes are images of $\mathbb{C}[\mathsf{t}_1, \dots, \mathsf{t}_n]$, we have a surjective map. As noted before, comparing dimensions shows that it is also injective, which concludes the proof. $\square$

Comparing this result with Proposition 3.43, we see that the categories DQ and $D^b(\mathsf{Coh}(\mathfrak{M}))$ are rather similar. We would immediately obtain a fully faithful functor $\mathsf{DQ} \to D^b(\mathsf{Coh}(\mathfrak{M}))$ if we knew that $\bigoplus_{x,y \in \Lambda(\lambda)} \mathrm{Ext}_{\mathsf{DQ}}(\mathcal{L}_x, \mathcal{L}_y)$ were formal as a dg–algebra, but it is not clear that this is the case.

To show this formality, we need to use a different approach to construct this functor, using projective objects in the category $\widetilde{\mathsf{dq}}$. This approach also naturally leads to a structure on DQ that corresponds to the $\mathbb{G}_m$–action on $\mathfrak{M}$ discussed earlier: a new structure on DQ–modules, closely related to Saito's theory of mixed Hodge modules. This will result in a graded category, which is to $D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}))$ as DQ is to $D^b(\mathsf{Coh}(\mathfrak{M}))$.

4.10 Projectives

As described above, we'll construct projective covers in $\widetilde{\mathsf{dq}}$. As usual, let us first construct these on $\widetilde{\mathfrak{Z}}$.

Consider $A = \mathbb{C}[x, y, \hbar]$ with the usual Moyal star product defined above. There are unique dq–modules $\mathcal{P}^{(k)}_*$ and $\mathcal{P}^{(k)}_!$ over $\mathbb{C}^2$ whose sections are the quotients

$$H^0(\mathbb{C}^2; \mathcal{P}^{(k)}_*) = A/A \star (y \star x)^{\star k} \quad \text{and} \quad H^0(\mathbb{C}^2; \mathcal{P}^{(k)}_!) = A/A \star (x \star y)^{\star k}.$$

Identifying A with the Rees algebra of differential operators D_x on $\mathbb{C}[x]$ (sending $y \mapsto \hbar\, \partial/\partial x$), these modules become the Rees modules of D–modules $\mathscr{P}^{(k)}_*$ and $\mathscr{P}^{(k)}_!$ on $\mathbb{A}^1$ with coordinate x. We can identify these with the $*$– and $!$–pushforwards of the D–module $L^{(k)}$ on $\mathbb{C}^* = \mathrm{Spec}(\mathbb{C}[x, x^{-1}])$ defined by the connection $\nabla = d - N/x$ on the trivial bundle with fiber $\mathbb{C}^k$, where N is the regular nilpotent matrix

$$N = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Both $\mathscr{P}_*^{(k)}$ and $\mathscr{P}_!^{(k)}$ are projective in the category of D–modules on $\mathbb{A}^1$ which are smooth away from the origin, whose monodromy around the origin has nilpotent part of length $\leq k$. The D–module $\mathscr{P}_!^{(k)}$ is the projective cover of the D–module of polynomials on $\mathbb{A}^1$, and $\mathscr{P}_*^{(k)}$ is the projective cover of the delta functions at the origin.

Our presentation of these D–modules induces a good filtration on them; in DQ–module terms, this is an equivariant structure for the cotangent scaling $\mathbb{S}$ which has weight 0 on x and weight 1 on y. In fact, we will want to use shifts of this filtration, corresponding to $\hbar\mathcal{P}_*^{(k)}$ and $\hbar^{1/2}\mathcal{P}_!^{(k)}$ — note that the latter is only equivariant under the squared scaling. In D–module terms, this means that we endow $\mathscr{P}_*^{(k)}$ with the good filtration such that the image of $\mathbb{C}[x] \subset D_x$ spans $F_{-1}\mathscr{P}_*^{(k)}$ and $F_p\mathscr{P}_*^{(k)} = F_{p+1}D_x \cdot F_{-1}\mathscr{P}_*^{(k)}$, and $\mathscr{P}_!^{(k)}$ with the good filtration such that the image of $\mathbb{C}[x] \subset D_x$ spans $F_{-1/2}\mathscr{P}_!^{(k)}$ and $F_{p-1/2}\mathscr{P}_!^{(k)} = F_pD_x \cdot F_{-1/2}\mathscr{P}_!^{(k)}$. These might seem like slightly strange choices: they are deliberately chosen so that in both cases, the unique simple quotient carries a pure Hodge structure of weight 0.

We will need certain morphisms between these DQ–modules:

(1) The linear map N on $\mathbb{C}^k$ induces an endomorphism on $\mathcal{L}^{(k)}$ and hence of $\mathcal{P}_*^{(k)}$ and $\mathcal{P}_!^{(k)}$. This is the same as right multiplication by $y \star x$ or $x \star y$, respectively.

(2) We have a $c^-\colon \mathcal{P}_*^{(k)} \to \mathcal{P}_!^{(k)}$, induced by multiplication on the right by y. Note that this map becomes an isomorphism if we invert y, and consider these as D–modules on $\operatorname{Spec}\mathbb{C}[y, y^{-1}]$.

(3) In the opposite direction we have a map $c^+\colon \mathcal{P}_!^{(k)} \to \mathcal{P}_*^{(k)}$, induced by multiplication on the right by x; this is also induced by the identity on the local system $\mathcal{L}^{(k)}$. Similarly, this map becomes an isomorphism if we invert x.

Note that the morphisms c^- and c^+ shift the good filtration by $\frac{1}{2}$. By [3, Theorem 2.12], we can identify these maps with the logarithm of the monodromy around the origin, the canonical map from nearby to vanishing cycles and the modified variation map discussed in [3, Section 2.7].

Lemma 4.28 *The algebra* $\operatorname{End}(\mathcal{P}_*^{(k)} \oplus \mathcal{P}_!^{(k)})$ *is generated by* $c^\pm$ *subject to the relations*

$$c^+c^- = N, \quad c^-c^+ = N, \quad N^k = 0. \tag{4-10}$$

Proof By construction, $\operatorname{Hom}(\mathcal{P}_!^{(k)}, \mathcal{M})$ is the kernel of the k^{th} power of the logarithm of the monodromy on the stalk of $\mathcal{M}$ at a generic point, given by the image of $(0, \dots, 0, 1)$ in this stalk. In particular, for $\operatorname{Hom}(\mathcal{P}_!^{(k)}, \mathcal{P}_!^{(k)})$, this is $\mathbb{C}^k$ itself, and the map sending $(0, \dots, 0, 1)$ to $(a_1, \dots, a_k)$ is

$$a_k + a_{k-1}N + \dots + a_1N^{k-1}.$$

Similarly, for $\operatorname{Hom}(\mathcal{P}_!^{(k)}, \mathcal{P}_*^{(k)})$, this stalk is the same, but now the map sending $(0, \dots, 0, 1)$ to $(a_1, \dots, a_k)$ is $(a_k + a_{k-1}N + \dots + a_1N^{k-1})c^+$. A symmetric argument holds with $*$ and $!$ reversed. □

As noted before, the map τ induces an isomorphism $\mathbb{C}^2 \setminus \{y = 0\} \cong \mathbb{C}^2 \setminus \{x = 0\}$. We can construct a DQ–module on $\mathfrak{W}_i \cup \mathfrak{W}_{i+1}$ glued using τ, and placing $\mathcal{P}_*^{(k)}$ or $\mathcal{P}_!^{(k)}$ on each $\mathfrak{W}_i$:

- If the two modules are different, ie $\mathcal{P}_*^{(k)}$ on $\mathfrak{W}_i$ and $\mathcal{P}_!^{(k)}$ on $\mathfrak{W}_{i+1}$ or vice versa, then we use the natural isomorphism induced by swapping the roles of x and y.
- If they are the same, ie $\mathcal{P}_*^{(k)}$ or $\mathcal{P}_!^{(k)}$ on both $\mathfrak{W}_i$ and $\mathfrak{W}_{i+1}$, then we use the isomorphisms of multiplication by $y^{\pm 1}$ on $\mathfrak{W}_i$, or equivalently $x^{\pm 1}$ on $\mathfrak{W}_{i+1}$.

Iterating this process, we can construct a DQ–module on $\tilde{\mathfrak{Z}}$ associated to a choice of integer k and a map $\wp\colon \mathbb{Z} \to \{*, !\}$, isomorphic to $\mathcal{P}_{\wp(i)}^{(k)}$ on $\mathfrak{W}_i$. To endow this DQ–module with a global $\mathbb{S}$–action, we will need to shift the natural $\mathbb{S}$–action on the local components $\mathcal{P}_{\wp(i)}^{(k)}$ by a certain amount, determined as follows.

We can associate to $\mathcal{P}_*^{(k)}$ a local system on each of the two components of its singular support, $\{x=0\}$ and $\{y=0\}$, both described in terms of the vector space $\mathbb{V}^{(k)} \cong \mathbb{C} \oplus \mathbb{C}(1) \oplus \cdots \oplus \mathbb{C}(k-1)$; here (p) represents shifting the good filtration/$\mathbb{S}$–action, though when we discuss Hodge modules, we will want to use it to represent Tate twist by the same amount. At a generic point of $\{x=0\}$, the fiber is $\mathbb{V}^{(k)}\big(\frac{1}{2}\big)$ (so we obtain local systems of weights $0, 2, \ldots, 2(k-1)$), and at a generic point of $\{y=0\}$, the fiber is $\mathbb{V}^{(k)}(1)$; for $\mathcal{P}_!^{(k)}$, these swap roles. Thus, in order to have matching $\mathbb{S}$–actions (or equivalently, good filtrations), we need to choose a function $\varsigma : \mathbb{Z} \to \frac{1}{2}\mathbb{Z}$ with the property that

$$\varsigma(m+1) = \begin{cases} \varsigma(m) & \text{if } \wp(m) \neq \wp(m+1), \\ \varsigma(m) - \frac{1}{2} & \text{if } \wp(m) = \wp(m+1) = *, \\ \varsigma(m) + \frac{1}{2} & \text{if } \wp(m) = \wp(m+1) = !, \end{cases} \tag{4-11}$$

and place $\mathcal{P}_{\wp(i)}^{(k)}(\varsigma(i))$ on $\mathfrak{W}_i$. The most important modules constructed this way, denoted by $\mathcal{P}_i^{(k)}$, are given by the functions

$$\wp(i) = \begin{cases} ! & \text{if } m > i, \\ * & \text{if } m \leq i, \end{cases} \quad \text{and} \quad \zeta(i) = \begin{cases} \frac{1}{2}(m-i-1) & \text{if } m > i, \\ \frac{1}{2}(i-m) & \text{if } m \leq i. \end{cases} \tag{4-12}$$

Lemma 4.29 *The DQ–module $\mathcal{P}_i^{(k)}$ is the projective cover of $\mathcal{L}_i$ in the subcategory of $\widetilde{\mathsf{dq}}$ on $\tilde{\mathfrak{Z}}$ where the nilpotent part of the monodromy has length $\leq k$.*

Proof We can reduce to the case where $i=0$ using the $\mathbb{Z}$ action. First, we must prove that $\mathcal{L}_0$ is the unique simple quotient of $\mathcal{P}_i^{(k)}$. On $\mathfrak{W}_0 \cup \mathfrak{W}_1 \cong T^*\mathbb{P}^1$, this module is the pushforward $j_!\mathscr{L}^{(k)}$, where $j : \mathbb{C}^* \hookrightarrow \mathbb{P}^1$ is the inclusion of the complement of the north and south poles. This has unique simple quotient given the intermediate extension of the one-dimensional local system with the standard connection. This matches the simple $\mathcal{L}_0$. Any other simple quotient must be $\mathcal{L}_m$ with $m \neq 0$. If $m < 0$, this would induce a map on $\mathfrak{W}_m$ of $\mathcal{P}_!^{(k)}$ to the delta function D–module; similarly, if $m > 0$, it would induce a map on $\mathfrak{W}_{m+1}$ of $\mathcal{P}_*^{(k)}$ to the function D–module. No such map exists, so indeed $\mathcal{L}_0$ is the unique quotient.

Now we need to show it is projective. Assume that $\mathcal{M}$ is an object in $\widetilde{\mathsf{dq}}$ with nilpotent part of the monodromy of length $\leq k$, and that there is a surjective map $\mathcal{M} \to \mathcal{L}_0$. First, we note that we can restrict $\mathcal{M}$ to $T^*\mathbb{P}^1$ and obtain a D–module on $\mathbb{P}^1$ smooth on $\mathbb{C}^*$. Since $\mathcal{L}_0$ is the only simple in $\widetilde{\mathsf{dq}}$

supported on $\mathbb{P}^1$, the local system we obtain on $\mathbb{C}^*$ is regular, so it is on the trivial vector bundle with fiber $\mathbb{C}^d$ with a connection of the form $d - N'/x$ for $N' \colon \mathbb{C}^d \to \mathbb{C}^d$ a nilpotent map, and the map to $\mathcal{L}_0$ is induced by a map $\phi \colon \mathbb{C}^d \to \mathbb{C}$ whose kernel contains the image of N'. We can lift this up to a map $\mathcal{P}_i^{(k)}|_{T^*\mathbb{P}^1} \to \mathcal{M}|_{T^*\mathbb{P}^1}$ by defining a map $\mathbb{C}^k \to \mathbb{C}^d$ by sending $(0,\dots,0,1)$ to any vector v with nonzero image under ϕ, and then extending by the rule that $N^r(0,\dots,0,1) \mapsto (N')^r v$. By assumption, $(N')^k = 0$, so this sends the standard basis of $\mathbb{C}^k$ to the vectors $(N')^r v$ for $r = 0,\dots,k-1$; there a unique linear map satisfying this property.

Now, we change focus to $\mathfrak{W}_{-1}$; by the projective property of $\mathcal{P}_!^{(k)}$, induced the map of local systems on $\mathfrak{X}_{-1}$ extends to a map of $\mathcal{P}_i^{(k)}|_{\mathfrak{W}_{-1}} \to \mathcal{M}|_{\mathfrak{W}_{-1}}$. Applying the same argument again to $\mathfrak{W}_{-2}$ gives a compatible map $\mathcal{P}_i^{(k)}|_{\mathfrak{W}_{-2}} \to \mathcal{M}|_{\mathfrak{W}_{-2}}$. By induction, we can extend to all $\mathfrak{W}_i$ with $i < 0$. A symmetric argument shows how to extend to $i > 1$. This establishes the result. $\square$

Lemma 4.30 *The stalk of $\sigma_{i'}(\mathcal{P}_i^{(k)})$ at a generic point in $\mathbb{P}^1$ is $\mathbb{V}^{(k)}\big(\frac{1}{2}(|i-i'|+1)\big)$.*

Proof By their identification with the $*$– and $!$–pushforwards, $\mathcal{P}_*^{(k)}$ and $\mathcal{P}_!^{(k)}$ both have stalk $\mathbb{V}^{(k)}$ on $\mathbb{A}^1 - \{0\}$. We need to understand how these correspond to the generic fiber on $y = 0$. This is the same as the vanishing cycles with respect to x at $x = 0$.

In the case $\mathcal{P}_*^{(k)}$, the canonical map induces an isomorphism of these vanishing cycles to $\mathbb{V}^{(k)}\big(-\frac{1}{2}\big)$; in the case $\mathcal{P}_!^{(k)}$, the variation map induces an isomorphism of these vanishing cycles to $\mathbb{V}^{(k)}\big(\frac{1}{2}\big)$. This makes it clear that on each component, we have shift of the $\mathbb{S}$–structure on $\mathbb{V}^{(k)}$, and that this shift is $\frac{1}{2}|i-i'|$. $\square$

This means that $\operatorname{Hom}(\mathcal{P}_i^{(k)}, \mathcal{P}_{i'}^{(k)}) = \mathbb{V}^{(k)}\big(\frac{1}{2}|i-i'|\big)$. The morphisms $c^{\pm}$ and N induce morphisms of DQ–modules:

$$N : \mathcal{P}_i^{(k)} \to \mathcal{P}_i^{(k)}(1), \quad c^- : \mathcal{P}_i^{(k)} \to \mathcal{P}_{i+1}^{(k)}\big(\tfrac{1}{2}\big), \quad c^+ : \mathcal{P}_i^{(k)} \to \mathcal{P}_{i-1}^{(k)}\big(\tfrac{1}{2}\big).$$

By Lemma 4.28, these morphisms generate the endomorphism algebra $\bigoplus_{i,j\in\mathbb{Z}} \operatorname{Hom}(\mathcal{P}_i^{(k)}, \mathcal{P}_j^{(k)})$ subject to the same relations (4-10). That is:

Lemma 4.31 *The endomorphism algebra $\bigoplus_{i,j\in\mathbb{Z}} \operatorname{Hom}(\mathcal{P}_i^{(k)}, \mathcal{P}_j^{(k)})$ is isomorphic to the quotient of the algebra $\widetilde{H}_{\mathbb{C}}$ attached to the usual action of $\mathbb{G}_m$ on $\mathbb{A}^1$, modulo the relations $\mathsf{s}_i{}^k = 0$ for all i.*

Now, we extend this to the general case. Given $\boldsymbol{x}$, we can define a projective by the exterior tensor product $\mathcal{P}_{x_1}^{(k)} \boxtimes \cdots \boxtimes \mathcal{P}_{x_n}^{(k)}$, and consider the action of the torus $T_{\mathbb{Q}}$. We let $\mathcal{Q}_{\boldsymbol{x}}^{(k)}$ be the unique largest quotient of this exterior product where the monodromy around $T_{\mathbb{Q}}$–orbits is trivial. Concretely, the exterior tensor product above carries an action of $\mathbb{C}[N_1,\dots,N_n] = U_{\mathbb{C}}(\mathfrak{d})$, which can be interpreted as the logarithms of monodromy along orbits of the larger torus $D_{\mathbb{Q}}$. The monodromy is trivial along T–orbits if this action factors through the quotient $U_{\mathbb{C}}(\mathfrak{d}) \to S$. We therefore have

$$Q_{\boldsymbol{x}}^{(k)} = (\mathcal{P}_{x_1}^{(k)} \boxtimes \cdots \boxtimes \mathcal{P}_{x_n}^{(k)}) \otimes_{U_{\mathbb{C}}(\mathfrak{d})} S.$$

This quotient has a natural strong T–equivariant structure.

Definition 4.32 Let $\mathcal{P}_{\boldsymbol{x}}^{(k)}$ be the Hamiltonian reduction of $\mathcal{Q}_{\boldsymbol{x}}^{(k)}$ on $\widetilde{\mathfrak{Y}}$; we consider this as a DQ–module.

Lemma 4.33 *The object $\mathcal{P}_{\boldsymbol{x}}^{(k)}$ is the projective cover of $\mathcal{L}_{\boldsymbol{x}}$ in the category of DQ–modules with monodromy around $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$ unipotent of length $\leq k$ for all $\boldsymbol{y}$ with $|\boldsymbol{x}-\boldsymbol{y}|_1 = 1$.*

Proof The desired map from $\mathcal{P}_{\boldsymbol{x}}^{(k)} \to \mathcal{L}_{\boldsymbol{x}}$ is induced by the simple quotient of $\mathcal{P}_{x_i}^{(k)}$ for all i. Thus, we need to show the projective property, and the fact that there are no other simple quotients, which we will do by induction on the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ in the taxicab metric. The restriction of $\mathcal{P}_{\boldsymbol{x}}^{(k)}$ to $T^*\mathfrak{X}_{\boldsymbol{x}}$ is $j_!\mathscr{L}^{(k)}$ where $\mathscr{L}^{(k)}$ is the induced D–module on the complement of the intersection with all other components in $\mathfrak{X}_{\boldsymbol{x}}$. There is only one map to $\mathcal{L}_{\boldsymbol{x}}$ since $\mathscr{L}^{(k)}$ is indecomposable and has unique simple quotient. Since there are no maps of $j_!\mathscr{L}^{(k)}$ to D–modules supported on intersections with other components, we have no maps to $\mathcal{L}_{\boldsymbol{y}}$ for $|\boldsymbol{x}-\boldsymbol{y}|_1 = 1$. As in Lemma 4.29, we can extend this argument to all other $\boldsymbol{y}$, since the map can't be nonzero on $\boldsymbol{y}$ if it is zero on all $\boldsymbol{y}'$ closer to $\boldsymbol{x}$. This shows that $\mathcal{L}_{\boldsymbol{x}}$ is the unique simple quotient of $\mathcal{P}_{\boldsymbol{x}}^{(k)}$.

Now, let us prove the projective property for $\mathcal{P}_{\boldsymbol{x}}^{(k)}$. That is, let $\mathcal{M}$ be an object in $\widetilde{\mathsf{dq}}$ with monodromy around $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{x}'}$ for all $|\boldsymbol{x}-\boldsymbol{x}'|_1 = 1$ unipotent of length $\leq k$, and with a map $\mathcal{M} \to \mathcal{L}_{\boldsymbol{x}}$. We wish to show that we have an induced map $\psi \colon \mathcal{P}_{\boldsymbol{x}}^{(k)} \to \mathcal{M}$ making the usual diagram commute.

Now, let $\mathfrak{D}_{\leq p}$ be the union of the subspaces $T^*\mathfrak{X}_{\boldsymbol{y}}$ for $|\boldsymbol{x}-\boldsymbol{y}|_1 \leq p$. We will show that the map ψ exists by constructing it inductively of $\mathfrak{D}_{\leq p}$.

On $\mathfrak{D}_{\leq 0} \cong T^*\mathfrak{X}_{\boldsymbol{x}}$, we have an induced map from $L^{(k)}$ to the local system given by the restriction of $\mathcal{M}$ to the open orbit in $\mathfrak{X}_{\boldsymbol{x}'}$ by the universal property of $L^{(k)}$, and thus an induced map

$$\psi|_{\mathfrak{D}_{\leq 0}} \colon \sigma_{\boldsymbol{x}}(\mathcal{P}_{\boldsymbol{x}}^{(k)}) \cong j_! L^{(k)} \to \sigma_{\boldsymbol{x}}(\mathcal{M}).$$

Now, assume that we have defined the map ψ on $\mathfrak{D}_{<p}$, and that $|\boldsymbol{y}-\boldsymbol{x}|_1 = p$. If $U = \mathfrak{D}_{<p} \cap \mathfrak{X}_{\boldsymbol{y}}$, then by assumption we have defined a map $\sigma_{\boldsymbol{y}}(\mathcal{P}_{\boldsymbol{x}}^{(k)})|_U \to \sigma_{\boldsymbol{y}}(\mathcal{M})|_U$. By construction, $\sigma_{\boldsymbol{y}}(\mathcal{P}_{\boldsymbol{x}}^{(k)}) = i_!(\sigma_{\boldsymbol{y}}(\mathcal{P}_{\boldsymbol{x}}^{(k)})|_U)$, where $i \colon U \hookrightarrow \mathfrak{X}_{\boldsymbol{y}}$ is the inclusion. Thus, we have a unique induced map $\sigma_{\boldsymbol{y}}(\mathcal{P}_{\boldsymbol{x}}^{(k)}) \to \sigma_{\boldsymbol{y}}(\mathcal{M})$; applying this for each $\boldsymbol{y}$ extends this map to $\mathfrak{D}_{\leq p}$. This shows that we have the projective property, and we have already confirmed it is the indecomposable projective cover of $\mathcal{L}_{\boldsymbol{x}}$. □

4.11 An equivalence of categories

Let $\widetilde{H}_{\mathbb{K}}^{(k)}$ be the quotient of $\widetilde{H}_{\mathbb{K}}^{\lambda}$ by the two-sided ideal generated by s_i^k.

Lemma 4.34 *The endomorphism ring $\bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\widetilde{\Lambda}} \operatorname{Hom}_{\mathsf{dq}}(\mathcal{P}_{\boldsymbol{x}}^{(k)}, \mathcal{P}_{\boldsymbol{y}}^{(k)})$ is isomorphic to $\widetilde{H}_{\mathbb{C}}^{(k)}$.*

Proof This map is induced by sending the morphism $c_{\boldsymbol{x}}^{\pm i}$ to the morphism $c^{\pm}$ in the i^{th} factor of the exterior product, and $\boldsymbol{s}_i$ to the endomorphism N of this tensor factor.

We check that this is well-defined. The linear relations among the variables $\boldsymbol{s}_i$ in S correspond to the triviality of monodromy along T–orbits in $\mathcal{P}_{\boldsymbol{x}}^{(k)}$. The relations (3-6a) are a consequence of Lemma 4.28, while the relations (3-6b) and (3-6c) follow from the fact that these are the tensor product of endomorphisms of two different tensor factors. This shows that we have the desired map.

Now, consider $\mathrm{Hom}(\mathcal{P}_{\boldsymbol{x}}^{(k)}, \mathcal{P}_{\boldsymbol{y}}^{(k)})$; this is a quotient of $\mathbb{C}[\boldsymbol{s}_1, \dots, \boldsymbol{s}_n]/(\boldsymbol{s}_i^k)$, which is the stalk of $\boxtimes_{i=1}^n \mathcal{P}_{y_i}^{(k)}$ on the component $\mathfrak{X}_{y_1} \times \cdots \times \mathfrak{X}_{y_n}$. Killing the monodromy on T gives us the quotient $S/(\boldsymbol{s}_i^k)$. This is generated by the image of $c_{\boldsymbol{x},\boldsymbol{y}}$, so our homomorphism is surjective, and the fact that $\widetilde{H}_{\mathbb{C}}$ is free as an S–module shows it is also injective. □

Assume that M is a finite-dimensional right $\widetilde{H}_{\mathbb{C}}^{\lambda}$–module. Assume k is chosen large enough that $\boldsymbol{s}_i^k$ kills M for all i. We can thus write M as a quotient of $\bigoplus_{p=1}^q 1_{\boldsymbol{x}_p} \widetilde{H}_{\mathbb{Q}}^{(k)}$ for some $\boldsymbol{x}_p$, and in fact as the cokernel of a map

$$f\colon \bigoplus_{r=1}^{s} 1_{\boldsymbol{y}_r} \widetilde{H}_{\mathbb{C}}^{(k)} \to \bigoplus_{p=1}^{q} 1_{\boldsymbol{x}_p} \widetilde{H}_{\mathbb{C}}^{(k)}$$

induced by elements $a_{rp} \in 1_{\boldsymbol{x}_p} \widetilde{H}_{\mathbb{Q}}^{(k)} 1_{\boldsymbol{y}_r}$ of degree $2(\ell_p - \nu_r)$. We can also view f as a morphism of DQ–modules

$$\widetilde{\mathsf{d}}(f)\colon \bigoplus_{r=1}^{s} \mathcal{P}_{\boldsymbol{y}_r}^{(k)} \to \bigoplus_{p=1}^{q} \mathcal{P}_{\boldsymbol{x}_p}^{(k)}.$$

Let $\widetilde{\mathsf{d}}(M)$ denote the cokernel of the map $\widetilde{\mathsf{d}}(f)$. Let $\widetilde{H}_{\mathbb{C}}^{\lambda,\mathrm{op}}$–$\mathrm{mod}_o$ be the category of finite-dimensional right modules of $\widetilde{H}_{\mathbb{C}}^{\lambda}$ on which each s_i acts nilpotently.

Proposition 4.35 *This defines equivalences of categories*

$$\widetilde{\mathsf{d}}\colon \widetilde{H}_{\mathbb{C}}^{\lambda,\mathrm{op}}\text{–}\mathrm{mod}_o \to \widetilde{\mathsf{dq}} \quad \text{and} \quad \mathsf{d}\colon H_{\mathbb{C}}^{\lambda,\mathrm{op}}\text{–}\mathrm{mod}_o \to \mathsf{dq}.$$

Proof By Lemma 3.17, the category of DQ–modules which are quotients of a finite sum of the objects $\mathcal{P}_{\boldsymbol{x}}^{(k)}$ is equivalent to the category $H_{\mathbb{C}}^{(k),\mathrm{op}}$–mod via the functor $\widetilde{\mathsf{d}}$. The dimension of $\widetilde{\mathsf{d}}(M)$ under this equivalence is the same as the composition length of M, so M is in $\widetilde{\mathsf{dq}}$ if and only if its image is finite-dimensional. Thus, we have an equivalence $H_{\mathbb{C}}^{(k),\mathrm{op}}$–$\mathrm{mod}_o \to \widetilde{\mathsf{dq}}^{\leq k}$ between finite-dimensional $H_{\mathbb{C}}^{(k),\mathrm{op}}$–modules and the subcategory $\widetilde{\mathsf{dq}}^{\leq}$ of $\widetilde{\mathsf{dq}}$ where all monodromy has unipotent length $\leq k$. Since $\widetilde{H}_{\mathbb{C}}^{\lambda,\mathrm{op}}$–$\mathrm{mod}_o$ is the union of the modules factoring through the quotients $H_{\mathbb{C}}^{(k),\mathrm{op}}$ for all k, and similarly $\widetilde{\mathsf{dq}} = \cup_k \widetilde{\mathsf{dq}}^{\leq k}$; this induces an equivalence $\widetilde{\mathsf{d}}\colon \widetilde{H}_{\mathbb{C}}^{\xi,\mathrm{op}}$–$\mathrm{mod}_o \to \widetilde{\mathsf{dq}}$, as desired.

The proof for d is word-for-word identical to that for $\widetilde{\mathsf{d}}$, so we leave the details to the reader. □

Since $\widetilde{\mathsf{d}}$ is an exact functor, it extends to a (both left and right) derived functor $\widetilde{\mathsf{d}}\colon D^b(\widetilde{H}_{\mathbb{C}}^{\lambda,\mathrm{op}}\text{–}\mathrm{mod}_o) \to \widetilde{\mathsf{DQ}}$. Combining with Corollary 3.41, we see our version of homological mirror symmetry in this context, as promised in the introduction:

Theorem 4.36 *The functor $\mathscr{M} \mapsto \widetilde{\mathsf{d}}(\mathbb{R}\mathrm{Hom}(\mathscr{T}_{\mathbb{C}}^{\lambda}, \mathscr{M}))$ defines an equivalence of dg–categories*

$$D^b(\mathsf{Coh}_G(\mathfrak{M}_{\mathbb{C}})_o) \to \widetilde{\mathsf{DQ}}.$$

Similarly, d *defines an equivalence* $D^b(\mathsf{Coh}(\mathfrak{M}_{\mathbb{C}})_o) \to \mathsf{DQ}$.

Proof We give the proof for the first equivalence, leaving the second to the reader. We know from Corollary 3.41 that this reduces to showing the derived functor of $\tilde{\mathsf{d}}$ is an equivalence of derived categories $D^b(\tilde{H}_{\mathbb{C}}^{\lambda,\mathrm{op}}\text{–mod}_o) \to \widetilde{\mathsf{DQ}}$. Proposition 4.35 show that this functor is an equivalence of categories on the heart of the usual t–structure. It's enough to additionally check that for a set of generating objects, such as the simples $L_{\boldsymbol{x}}$, the induced map $\mathrm{Ext}^k(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}) \to \mathrm{Ext}^k(\tilde{\mathsf{d}}(L_{\boldsymbol{x}}), \tilde{\mathsf{d}}(L_{\boldsymbol{y}}))$ is an isomorphism for all k, $\boldsymbol{x}$ and $\boldsymbol{y}$; this isomorphism follows for all other objects by a standard long exact sequence argument. Thus, to complete the proof, it is enough to show that $\tilde{\mathsf{d}}$ induces an isomorphism $e_{\boldsymbol{y}}\tilde{H}^!_{\lambda,\mathbb{C}}e_{\boldsymbol{x}} \cong \mathrm{Ext}_{\widetilde{\mathsf{DQ}}}(\tilde{\mathcal{L}}_{\boldsymbol{x}}, \tilde{\mathcal{L}}_{\boldsymbol{y}})$.

Of course, Theorem 4.27 implies that *an* isomorphism between the corresponding Ext–algebras exists. We could carefully confirm that this is (up to sign conventions) the same as that of Theorem 4.27, but this is not strictly necessary. The equivalence of abelian categories of Proposition 4.35 implies this functor induces an isomorphism $\mathrm{Ext}^1(L_{\boldsymbol{x}}, L_{\boldsymbol{y}}) \to \mathrm{Ext}^1(\tilde{\mathsf{d}}(L_{\boldsymbol{x}}), \tilde{\mathsf{d}}(L_{\boldsymbol{y}}))$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$. Since $\tilde{H}^!_{\lambda,\mathbb{C}}$ is generated by elements of degree 1, this implies that the map induced by $\tilde{\mathsf{d}}$ is surjective, and thus an isomorphism since the dimensions of $e_{\boldsymbol{y}}\tilde{H}^!_{\lambda,\mathbb{C}}e_{\boldsymbol{x}}$ and $\mathrm{Ext}_{\widetilde{\mathsf{DQ}}}(\tilde{\mathcal{L}}_{\boldsymbol{x}}, \tilde{\mathcal{L}}_{\boldsymbol{y}})$ in each degree are the same. In fact, since $\mathrm{Ext}^1_{\widetilde{\mathsf{DQ}}}(\tilde{\mathcal{L}}_{\boldsymbol{x}}, \tilde{\mathcal{L}}_{\boldsymbol{y}})$ is at most one-dimensional, we must recover the isomorphism of Theorem 4.27 up to rescaling the image of $d_{\boldsymbol{x},\boldsymbol{y}}$ to be a nonzero scalar multiple of $\boldsymbol{d}_{\boldsymbol{x},\boldsymbol{y}}$. □

This also resolves the concern about formality raised below Theorem 4.27: since $\tilde{H}_{\mathbb{C}}^{\lambda,\mathrm{op}}$ is Koszul, the induced dg–algebra structure on the Ext of simples is formal, and this shows that the same holds in $\widetilde{\mathsf{dq}}$.

5 Hodge structures

5.1 Microlocal mixed Hodge modules

We will need the notion of a unipotent mixed $\mathbb{Q}$–Hodge structure on $\sigma_{\boldsymbol{x}}(\mathcal{M})$; see [34] for a reference. "Unipotent" simply means that the monodromy on every piece of a stratification on which the D–module is smooth is unipotent. Mixed Hodge modules are a very deep subject, but one which we can use in a mostly black-box manner. The important thing for us is that given a holonomic regular D–module $\mathscr{M}$, a mixed Hodge structure can be encoded as real form and a pair of filtrations, a good filtration (often called the Hodge filtration) and the weight filtration (by submodules) on $\mathscr{M}$. As discussed previously, we are allowing good filtrations indexed by $\frac{1}{2}\mathbb{Z}$.

Note that while most references on mixed Hodge modules only consider untwisted D–modules, since a Hodge structure is given by local data, the definition extends to twisted D–modules in an obvious way. We will only be using twists by honest line bundles (as opposed fractional powers), so we have an even easier definition available to us: a mixed/pure Hodge structure on a module $\mathscr{M}$ over differential operators twisted by a line bundle L is the same structure on the untwisted D–module $L^* \otimes \mathscr{M}$. Since we will be working with fixed twists in what follows, we will conceal this choice and simply speak of mixed Hodge modules on $\mathfrak{X}_{\boldsymbol{x}}$ rather than twisted mixed Hodge modules.

Given an $\mathbb{S}G$–equivariant $\widetilde{\mathcal{O}}_{\phi}^{\hbar}$–module $\mathcal{M}$ in dq, a $\widetilde{\mathcal{O}}_{\phi}^{\hbar}(0)$–lattice $\mathcal{M}(0)$ induces a good filtration on $\sigma_{\boldsymbol{x}}(\mathcal{M})$ for each $\boldsymbol{x}$.

A $\mathbb{Q}$–form of $\sigma_{\boldsymbol{x}}(\mathcal{M})$ is a perverse sheaf L on $\mathfrak{X}_{\boldsymbol{x}}$ with coefficients in $\mathbb{Q}$ with a fixed isomorphism $L \otimes_{\mathbb{Q}} \mathbb{C} \cong \mathrm{Sol}(\sigma_{\boldsymbol{x}}\mathcal{M})$. We wish to define a $\mathbb{Q}$–form of $\mathcal{M}$ analogously, but we need to think carefully about compatibility between different $\boldsymbol{x}$.

Definition 5.1 An $\mathbb{Q}$–form for $\mathcal{M} \in \mathcal{O}_{\phi}^{\hbar}\text{–mod}^{\mathbb{S}G}$ is a perverse sheaf $L_{\boldsymbol{x}}$ on $\mathfrak{X}_{\boldsymbol{x}}$ for each $\boldsymbol{x}$ with a fixed isomorphism $L_{\boldsymbol{x}} \otimes_{\mathbb{Q}} \mathbb{C} \cong \mathrm{Sol}(\sigma_{\boldsymbol{x}}\mathcal{M})$ such that the isomorphism (4-6) induces an isomorphism $\mathscr{F}_{\boldsymbol{y},\boldsymbol{x}}(L_{\boldsymbol{x}}|_{N_{\boldsymbol{x},\boldsymbol{y}}}) \cong L_{\boldsymbol{y}}|_{N_{\boldsymbol{y},\boldsymbol{x}}}$, that is, it is compatible with the induced conjugation maps on the solution sheaves $\mathrm{Sol}(\sigma_{\boldsymbol{x}}\mathcal{M})$.

A *mixed Hodge structure* on $\mathcal{M}$ consists of a lattice $\mathcal{M}(0)$, $\mathbb{Q}$–form $L_{\boldsymbol{x}}$ for all $\boldsymbol{x}$ and an increasing weight filtration $W_\bullet$ of $\mathcal{M}$ by submodules such that the induced good filtration, $\mathbb{Q}$–form and weight filtration on $\sigma_{\boldsymbol{x}}(\mathcal{M})$ is a unipotent mixed $\mathbb{Q}$–Hodge structure on this D–module. The real forms are required to be compatible under the isomorphism (4-6).

Remark 5.2 This definition does not provide any hope of giving a general definition of "mixed Hodge DQ–modules". The space $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$ is a union of cotangent bundles of smooth varieties, with the scaling action on the cotangent bundle of each component extending to a global action on $\widetilde{\mathfrak{D}}^{\mathrm{alg}}$. We don't know of any similar situation outside the hypertoric case. Generalizing this definition to other cases is, of course, a quite interesting question, but not one on which we can provide much insight at the moment.

5.2 Hodge structures on projectives

One natural operation on mixed Hodge DQ–modules is that of *Tate twist*, which shifts the filtrations by $F_i\mathcal{M}(k) = F_{i+k}\mathcal{M}$ and $W_i\mathcal{M}(k) = W_{i+2k}\mathcal{M}$ for $k \in \frac{1}{2}\mathbb{Z}$. Note that defining Tate twists for half-integers requires using good filtrations which are indexed by $k \in \frac{1}{2}\mathbb{Z}$, this explains our cryptic introduction of half-integers in earlier sections. We're only interesting in understanding simple modules up to this operation. We can easily check that:

Lemma 5.3 *If $\mathcal{M}$ is supported on the core $\mathfrak{C}$, then the D–module $\sigma_{\boldsymbol{x}}(\mathcal{M})$ is smooth along the orbit stratification of $\mathfrak{X}_{\boldsymbol{x}}$ as a toric variety.*

Lemma 5.4 *The sheaf $\mathcal{L}_{\boldsymbol{x}}$ has a unique mixed Hodge structure whose associated mixed Hodge modules are pure of weight* 0.

Proof The trivial local system on $\mathfrak{X}_{\boldsymbol{x}}$ has the structure of a variation of Hodge structure which is pure of weight 0. This is unique by [17, Proposition 1.13]. Of course, any mixed Hodge structure of weight 0 on $\mathcal{L}_{\boldsymbol{x}}$ must be induced by this VMHS, which shows uniqueness.

Thus, we only need to show that the induced lattice $\mathcal{L}_{\boldsymbol{x}}(0)$, real form, and (trivial) weight filtration induce mixed Hodge structures on the microlocalizations $\sigma_{\boldsymbol{y}}(\mathcal{L}_{\boldsymbol{x}})$ for each $\boldsymbol{y}$. Recall that $\sigma_{\boldsymbol{y}}(\mathcal{L}_{\boldsymbol{x}})$ is the pushforward of the trivial line bundle on $\mathfrak{X}_{\boldsymbol{x}} \cap \mathfrak{X}_{\boldsymbol{y}}$, so the result follows from the compatibility of mixed Hodge structure with pushforward. □

Unfortunately, while the Hodge structure on a simple module is unique up to Tate twist, there are "too many" different Hodge structures on other objects in dq. For example, $\mathcal{L}_{\boldsymbol{x}} \oplus \mathcal{L}_{\boldsymbol{x}}(k)$ has a nontrivial moduli of Hodge structures, induced by the same phenomenon on $\mathbb{Q} \oplus \mathbb{Q}(k)$.

Thus, we need to find a way of avoiding these sort of deformations of Hodge structure. We do this by constructing a natural Hodge structure on the modules $\mathcal{P}_{\boldsymbol{x}}^{(k)}$.

Recall that we started the construction of these projectives by considering modules $\mathcal{P}_*^{(k)}$ and $\mathcal{P}_!^{(k)}$ over $A = \mathbb{C}[x, y, \hbar]$ with the usual Moyal star product. We make these into mixed Hodge modules on $\mathbb{A}^1$ by endowing $\mathcal{P}_*^{(k)}$ with the good filtration such that the image of $\mathbb{C}[x] \subset D_x$ spans $F_{-1}\mathcal{P}_*^{(k)}$ and $F_p\mathcal{P}_*^{(k)} = F_{p+1}D_x \cdot F_{-1}\mathcal{P}_*^{(k)}$, and $\mathcal{P}_!^{(k)}$ with the good filtration such that the image of $\mathbb{C}[x] \subset D_x$ spans $F_{-1/2}\mathcal{P}_!^{(k)}$ and $F_{p+1/2}\mathcal{P}_!^{(k)} = F_pD_x \cdot F_{-1/2}\mathcal{P}_!^{(k)}$. These might seem like slightly strange choices: they are deliberately chosen so that in both cases, the unique simple quotient carries a pure Hodge structure of weight 0.

Now, we consider Hodge structures on these DQ–modules extending the good filtrations defined above on $\mathcal{P}_*^{(k)}$ and $\mathcal{P}_!^{(k)}$. Their real form is the obvious one where x and y are conjugation invariant; this corresponds to the obvious real form of $L^{(k)}$. We define the weight filtration on $\mathcal{P}_*^{(k)}$ by

$$W_p\mathcal{P}_*^{(k)} = \begin{cases} 0 & \text{if } p < -2k+1, \\ D_x(\partial_x x)^{-p/2}/D_x(\partial_x x)^k & \text{if } 0 > p \geq 2k+1, \text{ with } p \text{ even}, \\ D_x x(\partial_x x)^{-(p+1)/2}/D_x(\partial_x x)^k & \text{if } 0 > p \geq 2k+1, \text{ with } p \text{ odd}, \\ \mathcal{P}_*^{(k)} & \text{if } p > 0, \end{cases}$$

and the weight filtration on $\mathcal{P}_!^{(k)}$ analogously, swapping x and y.

Lemma 5.5 *These data define mixed Hodge structures on $\mathcal{P}_*^{(k)}$ and $\mathcal{P}_!^{(k)}$.*

Proof First, let's consider $\mathcal{P}_*^{(k)}$. By the definition above, $W_p\mathcal{P}_*^{(k)}/W_{p-1}\mathcal{P}_*^{(k)} \cong D_x/D_x x$ if p is even and $0 \geq p \geq 2k+1$; this is equipped the good filtration where the image of ∂_x^r for $r < s$ span $F_{s+p/2}$. On the other hand, the V–filtration of this D–module for the function x has V^ℓ spanned by y^r for $r \geq -\ell$. Thus, the vanishing cycles $\Phi = \phi(W_p\mathcal{P}_*^{(k)}/W_{p-1}\mathcal{P}_*^{(k)})$ are spanned by the image of 1, ie they are one-dimensional. Accounting for the shift of good filtration (as in [34, (2.1.7)]) they are equipped with the good filtration

$$F_{s+p/2}(\Phi) = \begin{cases} \Phi & \text{if } s \geq 0, \\ 0 & \text{if } s < 0. \end{cases}$$

This means that $W_p\mathcal{P}_*^{(k)}/W_{p-1}\mathcal{P}_*^{(k)}$ is isomorphic to the usual Tate pure Hodge structure of weight p on $\mathbb{Q}$, pushed forward at the origin $x = 0$. If p is odd, then we have $W_p\mathcal{P}_*^{(k)}/W_{p-1}\mathcal{P}_*^{(k)} \cong D_x/D_x\partial_x$;

exactly as above, the generic fiber of this local system has the Tate Hodge structure of weight $p-1$, and so gives a pure Hodge module of weight p.

For $\mathcal{P}_!^{(k)}$, the calculations are the same, but odd and even cases swap roles. In particular, we see that half-integral filtrations are needed so that we can endow $\mathbb{Q}$ with a Tate Hodge structure of odd weight (ie a half-integral Tate twist). □

We defined above morphisms N and $c^{\pm}$ between these DQ–modules. These morphisms preserve the mixed Hodge structure up to Tate twist and become morphisms of mixed Hodge modules

$$N\colon \mathcal{P}_*^{(k)} \to \mathcal{P}_*^{(k)}(1), \quad c^-\colon \mathcal{P}_*^{(k)} \to \mathcal{P}_!^{(k)}\big(\tfrac{1}{2}\big), \quad c^+\colon \mathcal{P}_!^{(k)} \to \mathcal{P}_*^{(k)}\big(\tfrac{1}{2}\big).$$

This means that they define Tate elements of the endomorphism algebra $\operatorname{End}(\mathcal{P}_*^{(k)} \oplus \mathcal{P}_!^{(k)})$, and since they generate, they show that the induced Hodge structure on this algebra is of Tate type agreeing with the grading $\deg(c^{\pm}) = 1$ and $\deg(N) = 2$.

As noted before, the map τ induces an isomorphism

$$\mathbb{C}^2 \setminus \{y=0\} \;\cong\; \mathbb{C}^2 \setminus \{x=0\}.$$

We can construct a DQ–module on $\mathfrak{W}_i \cup \mathfrak{W}_{i+1}$ glued using τ, and placing $\mathcal{P}_*^{(k)}$ or $\mathcal{P}_!^{(k)}$ on each $\mathfrak{W}_i$.

- If the two modules are different, ie $\mathcal{P}_*^{(k)}$ on $\mathfrak{W}_i$ and $\mathcal{P}_!^{(k)}$ on $\mathfrak{W}_{i+1}$ or vice versa, then we use the natural isomorphism induced by swapping the roles of x and y.
- If they are the same, ie $\mathcal{P}_*^{(k)}$ or $\mathcal{P}_!^{(k)}$ on both $\mathfrak{W}_i$ and $\mathfrak{W}_{i+1}$, then we use the isomorphisms of multiplication by $y^{\pm 1}$ on $\mathfrak{W}_i$, or equivalently $x^{\pm 1}$ on $\mathfrak{W}_{i+1}$.

Of course, if we don't include shifts, this gluing will not respect the Hodge structure, so we need to glue these DQ–modules with Tate twists in them. The functions we Tate twist by have already been constructed in (4-11), based on a choice of which version of the module we will take on each component, expressed by a function $\wp$. This makes the modules $\mathcal{P}_i^{(k)}$ into mixed Hodge modules.

Remark 5.6 These modules are not projective in the category of mixed Hodge modules (even with appropriate monodromy restrictions) since they don't account for non-Tate extensions.

This induces a Hodge structure on the module $\mathcal{P}_{\boldsymbol{x}}^{(k)}$ defined in Definition 4.32, and thus on the endomorphism ring $\bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\widetilde{\Lambda}} \operatorname{Hom}_{\mathsf{dq}}(\mathcal{P}_{\boldsymbol{x}}^{(k)}, \mathcal{P}_{\boldsymbol{y}}^{(k)})$.

Lemma 5.7 *The Hodge structure on the endomorphism ring $\bigoplus_{\boldsymbol{x},\boldsymbol{y}\in\widetilde{\Lambda}} \operatorname{Hom}_{\mathsf{dq}}(\mathcal{P}_{\boldsymbol{x}}^{(k)}, \mathcal{P}_{\boldsymbol{y}}^{(k)})$ is Tate and matches that constructed from the grading on $H_{\mathbb{Q}}^{(k)}$.*

5.3 The category of mixed Hodge modules

Now, we wish to establish a graded version of the equivalence of Theorem 4.36. As discussed above, looking at all mixed Hodge structures on DQ–modules results in "too many" objects; in particular, the graded lift of a projective object will not be projective in the category of all mixed Hodge structures on objects in $\widetilde{\mathsf{dq}}$, which is not the behavior we expect from adding a grading to a ring. In more categorical terms, the functor of forgetting Hodge structure is not a "degrading" functor.

Thus, we will consider objects in $\widetilde{\mathsf{dq}}$ with a more restricted set of Hodge structures, only those which arise as a quotient of the objects $\mathcal{P}_{\boldsymbol{x}}^{(k)}$; it's worth noting that while these objects have a projective property in $\widetilde{\mathsf{dq}}$ (subject to a restriction on monodromy), they are not projective amongst mixed Hodge DQ–modules with this monodromy. The important effect this has is that it forces the local systems on the open part of $\mathfrak{X}_{\boldsymbol{x}}$ to be Tate as mixed Hodge structures; typically, the structures we wish to avoid will not have this property.

Definition 5.8 We let $\mu\mathsf{m}$ and $\widetilde{\mu\mathsf{m}}$ be the categories of mixed Hodge DQ–modules in dq and $\widetilde{\mathsf{dq}}$ which are quotients of a sum of the form $\bigoplus_{p=1}^{q} \mathcal{P}_{\boldsymbol{x}_p}^{(k)}(\ell_p)$ for some $k \geq 0$, $\ell_p \in \frac{1}{2}\mathbb{Z}$ and $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_q\} \subset \widetilde{\Lambda}$.

We let $\mu\mathsf{M}$ and $\widetilde{\mu\mathsf{M}}$ be the standard dg–enhancements of the derived categories $D^b(\mu\mathsf{m})$ and $D^b(\widetilde{\mu\mathsf{m}})$ (the quotient of the dg–category of all complexes modulo that of acyclic complexes).

Now, assume that M is a finite-dimensional graded right $\widetilde{H}_{\mathbb{Q}}^{\zeta}$–module. Recall that $M(\ell)$ denotes M with the grading shifted down by ℓ. Assume k is chosen large enough that $\boldsymbol{s}_i^k$ kills M for all i. We can thus write M as a quotient of $\bigoplus_{p=1}^{q} 1_{\boldsymbol{x}_p} H_{\mathbb{Q}}^{(k)}(2\ell_p)$ with $\boldsymbol{x}_p$ and ℓ_p as above, and in fact as the cokernel of a map

$$f\colon \bigoplus_{r=1}^{s} 1_{\boldsymbol{y}_r} H_{\mathbb{Q}}^{(k)}(2\nu_r) \to \bigoplus_{p=1}^{q} 1_{\boldsymbol{x}_p} H_{\mathbb{Q}}^{(k)}(2\ell_p)$$

induced by elements $a_{rp} \in 1_{\boldsymbol{x}_p} H_{\mathbb{Q}}^{(k)} 1_{\boldsymbol{y}_r}$ of degree $2(\ell_p - \nu_r)$. We can also view A as a morphism of Hodge DQ–modules

$$\widetilde{\mathsf{m}}(f)\colon \bigoplus_{r=1}^{s} \mathcal{P}_{\boldsymbol{y}_r}^{(k)}(\nu_r) \to \bigoplus_{p=1}^{q} \mathcal{P}_{\boldsymbol{x}_p}^{(k)}(\ell_p).$$

Let $\widetilde{\mathsf{m}}(M)$ denote the cokernel of the map $\widetilde{\mathsf{m}}(f)$. Let $\widetilde{H}_{\mathbb{Q}}^{\zeta,\mathrm{op}}$–gmod$_o$ be the category of finite-dimensional graded right modules of $\widetilde{H}_{\mathbb{Q}}^{\zeta}$.

Theorem 5.9 *This defines equivalences of categories*

$$\widetilde{\mathsf{m}}\colon \widetilde{H}_{\mathbb{Q}}^{\zeta,\mathrm{op}}\text{–gmod}_o \to \widetilde{\mu\mathsf{m}} \quad \textit{and} \quad \mathsf{m}\colon H_{\mathbb{Q}}^{\zeta,\mathrm{op}}\text{–gmod}_o \to \mu\mathsf{m},$$

sending grading shift (ℓ) *to the Tate twist* $\left(\frac{1}{2}\ell\right)$.

Proof If $f\colon M \to M'$ is a homogeneous map of modules, the construction of $\widetilde{\mathsf{m}}(f)$ by presenting M and M' as cokernels proceeds exactly as in the proof of Proposition 4.35, as does the proof that this functor is fully faithful.

The only point where we need a bit more care is in the proof of essential surjectivity. By definition, any module $\mathcal{M}$ in $\widetilde{\mu\mathsf{m}}$ is a quotient of $\widetilde{\mathsf{m}}(\mathcal{P}_0)$ for some $\mathcal{P}_0$. Thus, we need to show that the kernel $\mathcal{K}$ is also an object in $\widetilde{\mu\mathsf{m}}$. The object $\mathcal{K}$ has a largest semisimple quotient, ie its cosocle. This is a finite sum of objects of the form $\mathcal{L}_{\boldsymbol{y}_r}(\nu_r)$. This shows that $\mathcal{K}$ is generated by the images of maps (of DQ–modules, ignoring Hodge structure) from $\mathcal{P}_{\boldsymbol{y}_r}^{(k)}$ for $r = 1, \dots, s$. Note that $\mathrm{Hom}(\mathcal{P}_{\boldsymbol{y}_r}^{(k)}, \mathcal{K})$ carries a mixed Hodge structure which is a subobject of $\mathrm{Hom}(\mathcal{P}^{(k)}_{\boldsymbol{y}_r}, \widetilde{\mathsf{m}}(\mathcal{P}_0))$; the former has Tate type since the latter does as well. Thus, there is a module M_1 such that

$$\widetilde{\mathsf{m}}(\mathcal{P}_1) = \bigoplus_{r=1}^{s} \mathrm{Hom}(\mathcal{P}^{(k)}_{\boldsymbol{y}_r}, \mathcal{K}) \otimes_{\mathbb{Q}} \mathcal{P}^{(k)}_{\boldsymbol{y}_r}$$

as mixed Hodge DQ–modules; of course, the image of the induced map $\widetilde{\mathsf{m}}(\mathcal{P}_1) \to \widetilde{\mathsf{m}}(\mathcal{P}_0)$ is exactly $\mathcal{K}$, and so $\mathcal{M} = \widetilde{\mathsf{m}}(M)$ where M is the cokernel of the map $\mathcal{P}_1 \to \mathcal{P}_0$. This completes the proof that $\widetilde{\mathsf{m}}$ is an equivalence. The second equivalence is proven the same way. □

Analogous to the proof of Theorem 4.36, we have the following:

Corollary 5.10 *There are equivalences of categories*

$$D^b(\mathsf{Coh}_{\mathbb{G}_m \times G}(\mathfrak{M}_{\mathbb{Q}})_o) \to \widetilde{\mu\mathsf{M}} \quad \textit{and} \quad D^b(\mathsf{Coh}_{\mathbb{G}_m}(\mathfrak{M}_{\mathbb{Q}})_o) \to \mu\mathsf{M}$$

sending grading shift (ℓ) *to the Tate twist* $\left(\frac{1}{2}\ell\right)$.

We conclude with a few questions raised by this result. Under our equivalence, the $\mathbb{G}_m$–action on $\mathfrak{M}_{\mathbb{C}}$ corresponds to the weight grading on $\mu\mathsf{M}$. This action, which dilates the symplectic form, is key to the enumerative geometry of hypertoric varieties. Indeed, the symplectic structure on $\mathfrak{M}_{\mathbb{C}}$ implies that the nonequivariant quantum connection of $\mathfrak{M}_{\mathbb{C}}$ is essentially trivial. Its $\mathbb{G}_m$–equivariant version, on the other hand, is the hypergeometric system studied in [30]. The same is true for more general symplectic resolutions: for instance, the $\mathbb{G}_m$–equivariant quantum connection of the Springer resolution is the decidedly nontrivial affine KZ connection [12]. Our result thus suggests that the mirror description of these connections can be approached via microlocal Hodge structures.

We also note that whereas the left-hand side of both of our equivalences is a geometrically defined category, the right-hand sides are defined by picking certain generators inside the ambient category of deformation-quantization modules. This is in contrast to the equivalence proven in the sequel to this paper [21], which equates coherent sheaves on $\mathfrak{M}_{\mathbb{C}}$ with the wrapped Fukaya category of its mirror. A more direct geometric definition of $\mu\mathsf{M}$ and its grading, in particular, would be of great interest.

References

[1] **M Abouzaid**, *On homological mirror symmetry for toric varieties*, PhD thesis, University of Chicago (2007) Available at `https://www.proquest.com/docview/304795694`

[2] **R Anno**, **R Bezrukavnikov**, **I Mirković**, *Stability conditions for Slodowy slices and real variations of stability*, Mosc. Math. J. 15 (2015) 187–203 MR Zbl

[3] **D Arapura**, *Mixed Hodge structures associated to geometric variations*, from "Cycles, motives and Shimura varieties" (V Srinivas, editor), Tata Inst. Fund. Res. Stud. Math. 21, Narosa, New Delhi (2010) 1–34 MR Zbl

[4] **A Beilinson**, **V Ginzburg**, **W Soergel**, *Koszul duality patterns in representation theory*, J. Amer. Math. Soc. 9 (1996) 473–527 MR Zbl

[5] **R Bezrukavnikov**, *Noncommutative counterparts of the Springer resolution*, from "International Congress of Mathematicians, II" (M Sanz-Solé, J Soria, J L Varona, J Verdera, editors), Eur. Math. Soc., Zürich (2006) 1119–1144 MR Zbl

[6] **R Bezrukavnikov**, **D Kaledin**, *Fedosov quantization in positive characteristic*, J. Amer. Math. Soc. 21 (2008) 409–438 MR Zbl

[7] **T Braden**, **A Licata**, **N Proudfoot**, **B Webster**, *Gale duality and Koszul duality*, Adv. Math. 225 (2010) 2002–2049 MR Zbl

[8] **T Braden**, **A Licata**, **N Proudfoot**, **B Webster**, *Hypertoric category $\mathscr{O}$*, Adv. Math. 231 (2012) 1487–1545 MR Zbl

[9] **T Braden**, **A Licata**, **N Proudfoot**, **B Webster**, *Quantizations of conical symplectic resolutions, II: Category $\mathscr{O}$ and symplectic duality*, Astérisque 384, Soc. Math. France, Paris (2016) 75–179 MR Zbl

[10] **T Braden**, **N Proudfoot**, **B Webster**, *Quantizations of conical symplectic resolutions, I: Local and global structure*, Astérisque 384, Soc. Math. France, Paris (2016) 1–73 MR Zbl

[11] **A Braverman**, **M Finkelberg**, **H Nakajima**, *Towards a mathematical definition of Coulomb branches of 3–dimensional $\mathcal{N}=4$ gauge theories, II*, Adv. Theor. Math. Phys. 22 (2018) 1071–1147 MR Zbl

[12] **A Braverman**, **D Maulik**, **A Okounkov**, *Quantum cohomology of the Springer resolution*, Adv. Math. 227 (2011) 421–458 MR Zbl

[13] **J Brundan**, **N Davidson**, *Categorical actions and crystals*, from "Categorification and higher representation theory" (A Beliakova, A D Lauda, editors), Contemp. Math. 683, Amer. Math. Soc., Providence, RI (2017) 105–147 MR Zbl

[14] **M Bullimore**, **T Dimofte**, **D Gaiotto**, **J Hilburn**, *Boundaries, mirror symmetry, and symplectic duality in 3D $\mathcal{N}=4$ gauge theory*, J. High Energy Phys. (2016) art. id. 108 MR Zbl

[15] **A Cannas da Silva**, *Lectures on symplectic geometry*, Lecture Notes in Math. 1764, Springer (2001) MR Zbl

[16] **Z Dancso**, **M McBreen**, **V Shende**, *Deletion-contraction triangles for Hausel–Proudfoot varieties*, preprint (2019) arXiv 1910.00979

[17] **P Deligne**, *Un théorème de finitude pour la monodromie*, from "Discrete groups in geometry and analysis" (R Howe, editor), Progr. Math. 67, Birkhäuser, Boston, MA (1987) 1–19 MR Zbl

[18] **P Freyd**, *Representations in abelian categories*, from "Proceedings of the Conference on Categorical Algebra" (S Eilenberg, D K Harrison, H Röhrl, editors), Springer (1966) 95–120 MR Zbl

[19] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Lagrangian intersection Floer theory: anomaly and obstruction, I*, AMS/IP Stud. Adv. Math. 46.1, Amer. Math. Soc., Providence, RI (2009) MR Zbl

[20] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Lagrangian intersection Floer theory: anomaly and obstruction, II*, AMS/IP Stud. Adv. Math. 46.2, Amer. Math. Soc., Providence, RI (2009) MR Zbl

[21] **B Gammage**, **M McBreen**, **B Webster**, *Homological mirror symmetry for hypertoric varieties, II*, preprint (2019) arXiv 1903.07928

[22] **I Ganev**, *Quantizations of multiplicative hypertoric varieties at a root of unity*, J. Algebra 506 (2018) 92–128 MR Zbl

[23] **M Groechenig**, **M McBreen**, *Hypertoric Hitchin systems and Kirchhoff polynomials*, Int. Math. Res. Not. 2022 (2022) 15271–15312 MR Zbl

[24] **M Gross**, **P M H Wilson**, *Large complex structure limits of $K3$ surfaces*, J. Differential Geom. 55 (2000) 475–546 MR Zbl

[25] **D Kaledin**, *Derived equivalences by quantization*, Geom. Funct. Anal. 17 (2008) 1968–2004 MR Zbl

[26] **M Kashiwara**, **R Rouquier**, *Microlocalization of rational Cherednik algebras*, Duke Math. J. 144 (2008) 525–573 MR Zbl

[27] **S-C Lau**, **X Zheng**, *SYZ mirror symmetry for hypertoric varieties*, Comm. Math. Phys. 373 (2020) 1133–1166 MR Zbl

[28] **D Maulik**, **A Okounkov**, *Quantum groups and quantum cohomology*, Astérisque 408, Soc. Math. France, Paris (2019) MR Zbl

[29] **V Mazorchuk**, **S Ovsienko**, **C Stroppel**, *Quadratic duals, Koszul dual functors, and applications*, Trans. Amer. Math. Soc. 361 (2009) 1129–1172 MR Zbl

[30] **M B McBreen**, **D K Shenfeld**, *Quantum cohomology of hypertoric varieties*, Lett. Math. Phys. 103 (2013) 1273–1291 MR Zbl

[31] **I M Musson**, **M Van den Bergh**, *Invariants under tori of rings of differential operators and related topics*, Mem. Amer. Math. Soc. 650, Amer. Math. Soc., Providence, RI (1998) MR Zbl

[32] **H Nakajima**, *Towards a mathematical definition of Coulomb branches of 3–dimensional $\mathcal{N} = 4$ gauge theories, I*, Adv. Theor. Math. Phys. 20 (2016) 595–669 MR Zbl

[33] **N J Proudfoot**, *A survey of hypertoric geometry and topology*, from "Toric topology" (M Harada, Y Karshon, M Masuda, T Panov, editors), Contemp. Math. 460, Amer. Math. Soc., Providence, RI (2008) 323–338 MR Zbl

[34] **M Saito**, *A young person's guide to mixed Hodge modules*, from "Hodge theory and L^2–analysis" (L Ji, editor), Adv. Lect. Math. 39, International, Somerville, MA (2017) 517–553 MR Zbl

[35] **T J Stadnik, Jr**, *Étale splittings of certain Azumaya algebras on toric and hypertoric varieties in positive characteristic*, Adv. Math. 233 (2013) 268–290 MR Zbl

[36] **B Webster**, *Coherent sheaves and quantum Coulomb branches, I: Tilting bundles from integrable systems*, preprint (2019) arXiv 1905.04623

[37] **B Webster**, *Coherent sheaves and quantum Coulomb branches, II: Quiver gauge theories and knot homology*, preprint (2022) arXiv 2211.02099

MM: *Department of Mathematics, Chinese University of Hong Kong*
Hong Kong

BW: *Department of Pure Mathematics, University of Waterloo*
Waterloo, ON, Canada

BW: *Perimeter Institute for Theoretical Physics*
Waterloo, ON, Canada

mcb@math.cuhk.edu.hk, ben.webster@uwaterloo.ca

Proposed: Mark Gross — Received: 9 October 2020
Seconded: Richard P Thomas, Dan Abramovich — Revised: 22 August 2022

Geometry & Topology **28**:3 (2024) 1065–1098
DOI: 10.2140/gt.2024.28.1065
Published: 10 May 2024

Moduli spaces of Ricci positive metrics in dimension five

MCFEELY JACKSON GOODMAN

We use the η invariants of spin^c Dirac operators to distinguish connected components of moduli spaces of Riemannian metrics with positive Ricci curvature. We then find infinitely many nondiffeomorphic five-dimensional manifolds for which these moduli spaces each have infinitely many components. The manifolds are total spaces of principal S^1 bundles over $\#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ and the metrics are lifted from Ricci positive metrics on the bases. Along the way we classify 5–manifolds with fundamental group $\mathbb{Z}_2$ admitting free S^1 actions with simply connected quotients.

53C20, 53C27, 58D19, 58D27, 58J28; 19K56

Many closed manifolds are known to admit Riemannian metrics of positive Ricci curvature — for example, all compact, simply connected homogeneous spaces, biquotients, and cohomogeneity-one manifolds; see Berestovskiĭ [5], Grove and Ziller [23], Schwachhöfer and Tuschmann [35]. Systematic methods for constructing such metrics on certain connected sums and bundles have been explored in Corro and Galaz-García [11], Gilkey, Park and Tuschmann [20], Nash [33], Searle and Wilhelm [36], Sha and Yang [37] and Wraith [44].

Once we know that a manifold admits positive Ricci curvature we ask how many such metrics it admits. The space of geometrically distinct metrics of positive Ricci curvature on a manifold M is the moduli space $\mathfrak{M}_{\mathrm{Ric}>0}(M) = \mathfrak{R}_{\mathrm{Ric}>0}(M)/\mathrm{Diff}(M)$, where $\mathfrak{R}_{\mathrm{Ric}>0}(M)$ is the set of positive Ricci curvature metrics on M and $\mathrm{Diff}(M)$ is the diffeomorphism group, acting by pullbacks. The number of path components of $\mathfrak{M}_{\mathrm{Ric}>0}$ serves as a coarse measure of distinct positive Ricci curvature metrics on M.

We identify an infinite family of 5–manifolds M with $\pi_1(M) = \mathbb{Z}_2$ such that $\mathfrak{M}_{\mathrm{Ric}>0}(M)$ has infinitely many path components.

Theorem A *Let $B^4 = \#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ with $a+b \geq 2$, and let $S^1 \to M^5 \to B^4$ be a principal bundle with first Chern class $2d$, where $d \in H^2(B^4, \mathbb{Z})$ is primitive and $w_2(TB^4) = d \bmod 2$. Then $\mathfrak{M}_{Ric>0}(M^5)$ has infinitely many path components.*

Here w_2 is the second Stiefel–Whitney class, and a primitive class is one that is not a positive integer multiple of any other. We will see that for each 4–manifold B there are 2, 3 or 4 diffeomorphism types of such total spaces M, depending on the value of $|a-b| \bmod 4$, each of which admits infinitely many inequivalent free S^1 actions with quotient B. The only other five-dimensional manifolds for which $\mathfrak{M}_{\mathrm{Ric}>0}$ is known to have infinitely many components are the four homotopy real projective spaces recently described by Dessai and González-Álvaro [14] and five quotients of $S^2 \times S^3$ recently described by Wermelinger [42].

The conditions on the first Chern class in Theorem A are equivalent to the statement that $\pi_1(M^5) = \mathbb{Z}_2$, M^5 is nonspin, and the universal cover of M^5 is spin. M^5 can be constructed by taking five-dimensional homotopy real projective spaces, removing tubular neighborhoods of generators of the fundamental group, and gluing along the boundaries of the tubular neighborhoods. By the classification of Smale [39] and Barden [4], the universal cover $\widetilde{M}^5$ is diffeomorphic to $\#^{a+b-1} S^3 \times S^2$. But we do not know an explicit description of the deck group action by $\mathbb{Z}_2$ on $\widetilde{M}^5$.

Our second theorem identifies conditions under which M^5 admits one, and infinitely many, free S^1 actions. As an application, we will show that the manifolds in Theorem A admit infinitely many free S^1 actions. We construct the metrics used in Theorem A by lifting metrics from the quotients of M^5 by those actions. Here $b_2(M)$ is the second Betti number of M.

Theorem B *Let M^5 be a 5–manifold with $\pi_1 = \mathbb{Z}_2$. Then M admits a free S^1 action with a simply connected quotient if and only if M is orientable, $H_2(M, \mathbb{Z})$ is torsion-free and $\pi_1(M)$ acts trivially on $\pi_2(M)$. Furthermore, if $b_2(M) = 0$, then M is diffeomorphic to $\mathbb{R}P^5$. If $b_2(M) > 0$ and M admits a free S^1 action with simply connected quotient B^4, then M admits infinitely many inequivalent free S^1 actions with quotients diffeomorphic to B^4.*

Note that here B^4 can be any simply connected 4–manifold, and need not be one of the manifolds of Theorem A. Theorem 1.11 provides greater detail about the correspondence between a 5–manifold M^5 and the set $Q(M)$ of possible quotients $B^4 = M^5/S^1$. Given M^5 satisfying the hypotheses of Theorem B, we give conditions on the cohomology ring of a 4–manifold B^4 which are necessary and sufficient for B to be in $Q(M)$. In particular, any smooth manifold homeomorphic to a manifold in $Q(M)$ is in $Q(M)$. In Corollary 1.12 we see that for any such M, $Q(M)$ contains either $\#^c S^2 \times S^2$ or $\#^a \mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ for some $a, b, c \in \mathbb{Z}$. Those manifolds admit metrics with positive Ricci curvature, which can be lifted to M. Thus we have:

Corollary *Let M be a 5–manifold with $\pi_1(M) = \mathbb{Z}_2$ admitting a free S^1 action with a simply connected quotient. Then M admits a metric with positive Ricci curvature.*

Furthermore, it follows from Theorem 1.11 that given a simply connected 4–manifold B^4, the set of diffeomorphism types of total spaces M^5 with $\pi_1(M^5) = \mathbb{Z}_2$ of S^1 bundles over B^4 depends only on the cohomology ring of B^4. In particular, Theorem A would describe the same set of 5–manifolds if we replaced $\#^a \mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ with one of the manifolds homeomorphic to it.

We first review previous work with methods and results relevant to Theorem A. In [30] Kreck and Stolz invented a moduli space invariant $s(M, g) \in \mathbb{Q}$ for a metric g of positive scalar curvature on a closed spin manifold M. The metric is based on the η spectral invariant of the Dirac operator defined in Atiyah, Patodi and Singer [1]. If $s(M, g_1) \neq s(M, g_2)$ then g_1 and g_2 represent elements in different path components of $\mathfrak{M}_{\mathrm{scal}>0}$. Kreck and Stolz use the invariant to prove that for M^{4k+3} with a unique spin structure and vanishing rational Pontryagin classes $\mathfrak{M}_{\mathrm{scal}>0}(M)$ is either empty or has infinitely many components.

Since a path of Riemannian metrics which maintains positive Ricci curvature maintains positive scalar curvature as well, the s invariant can detect connected components of $\mathfrak{M}_{\mathrm{Ric}>0}$. Kreck and Stolz calculated s for the Einstein metrics on S^1 bundles $N^7_{k,l}$ over $\mathbb{C}P^1 \times \mathbb{C}P^2$ described by Wang and Ziller [41]. Kreck and Stolz showed, using the diffeomorphism classification in [28], that when k is even and $\gcd(k,l) = 1$, $N_{k,l}$ is diffeomorphic to infinitely many manifolds in the same family. As the s invariant takes infinitely many values on those metrics, the authors concluded that $\mathfrak{M}_{\mathrm{Ric}>0}(N_{k,l})$ has infinitely many components. Similar results have since been proved for S^1 bundles over $\mathbb{C}P^1 \times \mathbb{C}P^{2n}$ with $n \geq 1$; see Dessai, Klaus and Tuschmann [15].

Wraith showed that for a homotopy sphere σ^{4k-1} bounding a parallelizable manifold, $\mathfrak{M}_{\mathrm{Ric}>0}(\sigma)$ has infinitely many components. The procedure known as plumbing with disc bundles over spheres produces infinitely many parallelizable manifolds with boundaries diffeomorphic to σ. Wraith [43] constructed metrics of positive Ricci curvature on each boundary, and calculated the s invariant of each metric in [45].

Dessai [13] and the author [21] used the s invariant to find several infinite families of seven-dimensional sphere bundles M^7 such that $\mathfrak{M}_{\mathrm{Ric}>0}(M)$ and $\mathfrak{M}_{\sec\geq 0}(M)$ have infinitely many path components. Grove and Ziller [22; 24] constructed metrics of nonnegative sectional curvature on the manifolds in those families, and the diffeomorphism classifications in Crowley and Escher [12] and Escher and Ziller [18] show that each manifold is diffeomorphic to infinitely many other members of the family.

More recently, Dessai and González-Álvaro [14] showed that if M^5 is one of the four closed manifolds homotopy equivalent to $\mathbb{R}P^5$ then $\mathfrak{M}_{\sec\geq 0}(M)$ and $\mathfrak{M}_{\mathrm{Ric}>0}(M)$ have infinitely many path components. López de Medrano [32] showed that each such M^5 admits infinitely many descriptions as a quotient of a Brieskorn variety, and Grove and Ziller [23] showed the each quotient admits a metric of nonnegative sectional curvature. Dessai and González-Álvaro calculated the relative η invariant for those metrics to distinguish the path components. Wermelinger [42] extended their method to prove the same conclusion for five $\mathbb{Z}_2$ quotients of $S^2 \times S^3$.

We now outline the proof of Theorem A. We use Theorem B to show that each manifold M^5 in Theorem A admits infinitely many inequivalent free S^1 actions with quotient $B^4 = \#^a \mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$. We modify a result of Perelman [34] to show that B admits a metric of positive Ricci curvature. That metric can be lifted to a metric of positive Ricci curvature on M by Gilkey, Park and Tuschmann [20]. The lifted metrics depend on the S^1 action, and we get infinitely many distinct metrics on M.

We show that in dimensions $4k+1$, the η invariant of a certain spin^c Dirac operator constructed for a positive Ricci curvature metric g depends only on the connected component of the class of g in $\mathfrak{M}_{\mathrm{Ric}>0}$. To complete the proof we calculate η for each metric on M and show that it obtains infinitely many values. This is the most intricate part of our proof.

The standard method for calculating the η invariant of a spin Dirac operator on a manifold M with positive scalar curvature is to extend the metric over a manifold W with $\partial W = M$ so that the extension has positive scalar curvature as well. When M is not spin but spin^c, both the metric and a unitary connection on the complex line bundle associated to the spin^c structure must be extended. The desired condition then involves the curvatures of both metric and connection. In their work, Dessai and González-Álvaro passed to the universal cover to find a suitable W over which the connection could be extended to a flat connection. They use equivariant η invariants on the cover to compute the η invariant on the quotient.

In this paper we work directly on M and use a manifold with boundary W over which the connection cannot be extended to a flat connection, but the curvature of the extension can be explicitly controlled. To be specific, we extend the metric and connection on M to a metric h and connection ∇ on the disc bundle $W = M \times_{S^1} D^2$ associated to the S^1 bundle. We then use the Atiyah–Patodi–Singer index theorem [1] to obtain a formula for η in terms of the index of the spin^c Dirac operator on W and topological data on W. The index will vanish as long as

$$\mathrm{scal}(h) > 2|F^{\nabla}|_h,$$

where F^{∇} is the curvature form of the connection ∇. We accomplish the extension for a general class of S^1–invariant metrics of positive scalar curvature. This is more general than we need but may be of independent interest. In fact we construct h and ∇ such that

$$\mathrm{scal}(h) > \ell|F^{\nabla}|_h,$$

where ℓ is a positive integer such that the first Chern class of the S^1 bundle is ℓ times the canonical class of a spin^c structure on the quotient.

Sha and Yang [38] constructed metrics of positive Ricci curvature on the 4–manifolds $\#^{a-b}\mathbb{C}P^2 \#^b S^2 \times S^2$ with $a > b$. Those manifolds are diffeomorphic to $\#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$, so a manifold M satisfying the hypotheses of Theorem A also admits a free S^1 action with quotient $\#^{a-b}\mathbb{C}P^2 \#^b S^2 \times S^2$. One can lift the Sha–Yang metric to M, and there is no reason to expect that the resulting metric lies in the same component as the metric lifted from $\#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ in the proof of Theorem A. We will see, however, that the computation of the η invariant involves only the cohomology ring of the quotient, and we cannot distinguish any new components in this way.

In [37] Sha and Yang also found metrics of positive Ricci curvature on $\#^b S^2 \times S^2$. One might expect our methods to yield a similar result in this case. The 5–manifolds, however, would be spin, and the η invariant of the spin Dirac operator in dimension $4k+1$ vanishes, even when twisted with certain complex line bundles; see Botvinnik and Gilkey [7].

We now discuss Theorem B. In [26], Hambleton and Su find a complete diffeomorphism classification of 5–manifolds M with $\pi_1(M) = \mathbb{Z}_2$ when M is orientable, $H_2(M, \mathbb{Z})$ is torsion-free, and $\pi_1(M)$ acts trivially on $\pi_2(M)$. They apply the classification to investigate the diffeomorphism type of the total space of an S^1 bundle over a simply connected 4–manifold. When the total space is nonspin but has a spin universal cover, as is the case in Theorem A, they can only restrict the diffeomorphism type to two possibilities. Furthermore, an error is present in that calculation, which we correct in Lemma 1.7.

To prove Theorem B, we use the data of a principal S^1 bundle, namely the base and the first Chern class, to compute the diffeomorphism invariants used by Hambleton and Su for the total space. One, the second Betti number, is calculated easily. When the total space is nonspin but has a spin universal cover, we show how the other invariant can be computed by applying a map from $\Omega_4^{\mathrm{Spin}^c} \to \Omega_4^{\mathrm{Pin}^+}$ to the base. While a two-fold ambiguity remains in determining which diffeomorphism type corresponds to a specific first Chern class, we are nonetheless able to determine which pairs of invariants are achieved, and achieved infinitely many times, by bundles over a given 4–manifold.

The paper is organized as follows. In Section 1 we examine S^1 actions on 5–manifolds with $\pi_1 = \mathbb{Z}_2$ and prove Theorem B. In Section 2 we discuss the η invariant of a spin^c Dirac operator and show that it can be used to detect connected components of the moduli space in the context of Theorem A. In Section 3 we compute η in the case of certain $(4n+1)$–manifolds admitting free S^1 actions and prove Theorem A. In Section 4 we construct the metrics and connections used in the computations of Section 3.

Acknowledgements I would like to acknowledge my PhD advisor Wolfgang Ziller for all his help, as well as Anand Dessai, David González-Álvaro, Fernando Galaz-García and Diego Corro for helpful discussions. I am further grateful to Yang Su for pointing out how to work around an error in [26] and for other useful insights. This research was partially supported by National Science Foundation grant DMS-2001985.

1 S^1 actions on 5–manifolds with $\pi_1 = \mathbb{Z}_2$

Our methods for constructing metrics with positive Ricci curvature and for calculating η use the structure of a principal S^1 bundle. In this section we prove Theorem 1.11, which classifies 5–manifolds with $\pi_1 = \mathbb{Z}_2$ admitting one, or infinitely many, free S^1 actions with simply connected quotients. Theorem 1.11 also identifies those quotients. In particular, we prove Theorem B and show that a manifold M^5 satisfying the hypotheses of Theorem A admits infinitely many inequivalent S^1 actions with the same quotient. Our proof relies on a diffeomorphism classification of 5–manifolds with fundamental group $\mathbb{Z}_2$ carried out by Hambleton and Su [26].

Given a manifold M with $\pi_1(M) = \mathbb{Z}_2$, a characteristic submanifold $P \subset M$ is defined as follows. For N sufficiently large let $f : M \to \mathbb{R}P^N$ be a classifying map of the universal covering $\widetilde{M} \to M$. We can choose f to be transverse to $\mathbb{R}P^{N-1}$, and hence $P = f^{-1}(\mathbb{R}P^{N-1})$ is a smooth manifold. One checks that any two manifolds defined in this way are cobordant.

Alternatively, assume that $P \subset M$ is a submanifold such that the inverse image $\widetilde{P} \subset \widetilde{M}$ under the universal covering splits $\widetilde{M}$ into two components $\widetilde{M}_1$ and $\widetilde{M}_2$. Furthermore $\partial \widetilde{M}_1 = \partial \widetilde{M}_2 = \widetilde{P}$ and the covering transformation acting on $\widetilde{M}$ switches $\widetilde{M}_1$ and $\widetilde{M}_2$. One can then construct a map $f : M \to \mathbb{R}P^N$ such that $P = f^{-1}(\mathbb{R}P^{N-1})$. For details see [19; 32].

The key invariant of the classification in [26] is the class of P in an appropriate cobordism group. The appropriate structure on P depends on the second Stiefel–Whitney classes w_2 of M and $\widetilde{M}$. Hambleton and Su use the following labels for a manifold M with $\pi_1(M) = \mathbb{Z}_2$ and universal cover $\widetilde{M}$:

- **Type I** $w_2(T\widetilde{M}) \neq 0$.
- **Type II** $w_2(TM) = 0$.
- **Type III** $w_2(TM) \neq 0$ and $w_2(T\widetilde{M}) = 0$.

A characteristic submanifold P of a Type III manifold admits a pin$^+$ structure, and all such P are pin$^+$ cobordant. Here $\mathrm{Pin}^{\pm}(n)$ is the extension of $O(n)$ by $\mathbb{Z}_2$ such that a preimage of a reflection squares to ± 1 and $\Omega_n^{\mathrm{Pin}^{\pm}}$ is the cobordism group of n–manifolds with pin$^{\pm}$ structures. For details, see [26] and [19].

We review the construction of a pin$^+$ structure on P as we will use it later. Let $\mu = \widetilde{M} \times_{\mathbb{Z}_2} \mathbb{R}$ be the unique nontrivial real line bundle over M. Recall that $\widetilde{M} = \widetilde{M}_1 \cup_{\widetilde{P}} \widetilde{M}_2$, and the covering transformation exchanges the components. Thus the normal bundle $N\widetilde{P}$ of $\widetilde{P}$ is trivial and the covering transformation reverses the orientation of the fibers. The normal bundle NP of P satisfies

$$NP = N\widetilde{P}/\mathbb{Z}_2 \cong \widetilde{P} \times_{\mathbb{Z}_2} \mathbb{R} = \mu|_P.$$

Since M is orientable,

$$w_1(NP) = w_1(TP) = w_1(\det(TP)),$$

so $NP \cong \det(TP)$. Thus

$$(TM \oplus 2\mu)|_P = TP \oplus 3NP = TP \oplus 3\det(TP). \tag{1.1}$$

Using [19, Lemma 9; 26, Lemma 2.3], one checks that $w_2(TM \oplus 2\mu) = 0$. We can apply [27, Lemma 1.7] to see that a spin structure on $TP \oplus 3\det(TP)$ induces a pin$^+$ structure on TP. A similar argument on a cobordism shows that any two characteristic submanifolds are pin$^+$ cobordant.

Let $b_2(M)$ denote the second Betti number of a manifold M. The main theorem for Type III manifolds is [26, Theorem 3.1]:

Theorem 1.2 [26] *Let M_1 and M_2 be Type III 5–manifolds such that $\pi_1(M_i) \cong \mathbb{Z}_2$ acts trivially on $\pi_2(M_i)$ and $H_2(M_i, \mathbb{Z})$ is torsion-free for $i = 1, 2$. Then M_1 is diffeomorphic to M_2 if and only if*

$$b_2(M_1) = b_2(M_2) \quad \text{and} \quad [P_1] = \pm[P_2] \in \Omega_4^{\mathrm{Pin}^+},$$

where P_i is a characteristic submanifold of M_i.

We will take the data of a principal S^1 bundle, namely the base and the first Chern class, and identify the diffeomorphism type of the total space. In particular, we will identify when the total space satisfies the hypotheses of Theorem 1.2, and then compute b_2 and $[P]$. That computation combined with the classification of Type I and II total spaces in [26, Theorems 6.5 and 6.8] finishes the proof of Theorem 1.11, which in turn implies Theorem B.

A straightforward computation using the long exact homotopy and Gysin sequences proves the following; see for instance [26, Proposition 6.1].

Lemma 1.3 *Let B^n be a simply connected manifold and let $M^{n+1} \to B^n$ be a nontrivial principal S^1 bundle with first Chern class kd, where d is a primitive element of $H^2(B,\mathbb{Z})$ and $k \neq 0$ is an integer. Then M is orientable, $H_2(M,\mathbb{Z})$ is torsion-free and $b_2(M) = b_2(B) - 1$. The fundamental group $\pi_1(M) \cong \mathbb{Z}_k$ is generated by any S^1 fiber and acts trivially on $\pi_2(M)$. The universal cover of M is the total space of an S^1 bundle over B with first Chern class d. If $k = 2$, M is Type III if and only if and $w_2(TB) = d \bmod 2$.*

The condition $w_2(TB) = d \bmod 2$ implies the existence of a spin^c structure on B. We call d the canonical class of that spin^c structure. On a simply connected manifold a spin^c structure is uniquely determined by its canonical class. Thus in the Type III case, given a simply connected spin^c 4–manifold B^4 with primitive canonical class d, we want to know the diffeomorphism type of the total space M^5 of the S^1 bundle over B^4 with first Chern class $2d$. Since $b_2(M)$ is determined by Lemma 1.3, it remains to find the pin$^+$ cobordism class of a characteristic submanifold $P^4 \subset M^5$. In fact, the spin^c structure on B^4 will naturally induce a pin$^+$ structure on P^4.

To see this let $\rho\colon M \to B$ be the bundle map and let $\lambda \to B$ be a complex line bundle with first Chern class d; then $\rho^* d$ is the unique nontrivial torsion element of $H^2(M,\mathbb{Z})$. Let $\mu \to M$ be the unique nontrivial real line bundle over M. As in the proof that a characteristic submanifold of M will admit a pin$^+$ structure — see [19, Lemma 9; 26, Lemma 2.3] — $w_2(\mu \oplus \mu) = w_1(\mu)^2 \neq 0$. So $\mu \oplus \mu$ with its natural orientation is a nontrivial complex line bundle. Since $\mu \otimes \mu$ is trivial, $c_1(\mu \oplus \mu)$ is torsion, and we conclude that $\rho^*\lambda \cong \mu \oplus \mu$.

The S^1 action on M splits TM into a horizontal bundle isomorphic to $\rho^* TB$ and a vertical bundle, trivialized by an action field, which we call TS^1. The spin^c structure on B is equivalent to a spin structure on $TB \oplus \lambda$. That spin structure induces a spin structure on

$$\rho^*(TB \oplus \lambda) \oplus TS^1 \cong TM \oplus \mu \oplus \mu \tag{1.4}$$

and in turn a pin$^+$ structure on $P \subset M$ using (1.1). Denote by $\beta(B,d) \in \Omega_4^{\mathrm{Pin}^+}$ the cobordism class of P with this pin$^+$ structure. We synthesize the construction with the results of Lemma 1.3 as follows:

Lemma 1.5 *Let B^4 be a simply connected 4–manifold and let M^5 be the total space of a principal S^1 bundle over B with first Chern class $2d \in H^2(B,\mathbb{Z})$ where d is a primitive element such that $w_2(TB) = d \bmod 2$. Then M satisfies the conditions of Theorem 1.2 with $b_2(M) = b_2(B) - 1$ and $[P] = \beta(B,d)$.*

In the next lemma, we will see that β is a spin^c cobordism invariant whenever it is defined.

Lemma 1.6 *Let B_1 and B_2 be spin^c manifolds with primitive canonical classes d_1 and d_2, respectively. Then:*

(a) $\beta(B_1 \amalg B_2, d_1 + d_2) = \beta(B_1, d_1) + \beta(B_2, d_2)$.

(b) *If B_1 is spin^c cobordant to B_2 then $\beta(B_1, d_1) = \beta(B_2, d_2)$.*

Proof Part (a) follows immediately since the total space of the relevant bundle and the characteristic submanifold of that total space will be disjoint unions.

To prove part (b), let W be a simply connected spin^c cobordism between B_1 and B_2 with canonical class d. Then $d|_{B_i} = d_i$ for each $i = 1, 2$, and d must be a primitive class. Let $\pi\colon N \to B$ be the principal S^1 bundle over W with first Chern class $2d$. By Lemma 1.3, $\pi_1(N) = \mathbb{Z}_2$. We have that $\partial N = \pi^{-1}(B_1) \amalg \pi^{-1}(B_2)$ and $M_i = \pi^{-1}(B_i) \to B_i$ is the principal S^1 bundle with first Chern class $2d_i$.

Let $f\colon N \to \mathbb{R}P^N$ be a classifying map for the universal cover of N which is transverse to $\mathbb{R}P^{N-1}$. By Lemma 1.3, $\pi_1(N)$ is generated by any S^1 orbit, so $\pi_1(M_i) \to \pi_1(N)$ is an isomorphism, and $f|_{M_i}$ is a classifying map for the universal cover of M_i. Thus $P_i = f^{-1}(\mathbb{R}P^{N-1}) \cap M_i$ is a characteristic submanifold of M_i and $f^{-1}(\mathbb{R}P^{N-1})$ is a cobordism between P_1 and P_2. The argument before Lemma 1.5 proves that the spin^c structure on W induces a pin$^+$ structure on $f^{-1}(\mathbb{R}P^{N-1})$. That pin$^+$ structure restricts to the pin$^+$ structures induced on P_i by the spin^c structures on B_i. To see this one must simply note that the nontrivial real line bundle over N restricts to the nontrivial real line bundle over M_i. We conclude that

$$\beta(B_1, d_1) = [P_1] = [P_2] = \beta(B_2, d_2). \qquad \square$$

We now see that β defines a map between the spin^c and pin$^+$ cobordism groups. The four-dimensional spin^c cobordism group $\Omega_4^{\mathrm{Spin}^c}$ is isomorphic to $\mathbb{Z}^2$. The isomorphism takes a spin^c manifold B with canonical class d to the characteristic numbers

$$\langle d^2, [B]\rangle \quad \text{and} \quad \tfrac{1}{8}(\langle d^2, [B]\rangle - \operatorname{sign} B).$$

Here $\operatorname{sign}(B)$ is the signature, and the second integer is the index of the spin^c Dirac operator, which we denote by $\operatorname{ind}(B, d)$. See [3; 40] for details. To construct generators of $\Omega_4^{\mathrm{Spin}^c}$ let $x \in H^*(\mathbb{C}P^2, \mathbb{Z})$ be the generator which is the first Chern class of the Hopf bundle. Give $X = \mathbb{C}P^2$ the spin^c structure with canonical class x and $Y = \mathbb{C}P^2 \# \mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$ the spin^c structure with canonical class $d_Y = (3x, x, x) \in H^2(Y, \mathbb{Z}) \cong \oplus^3 H^2(\mathbb{C}P^2, \mathbb{Z})$. Then $[X], [Y] \in \Omega_4^{\mathrm{Spin}^c}$ represent $(1, 0)$ and $(9, 1)$ under the isomorphism with $\mathbb{Z}^2$ and form a minimal generating set of $\Omega_4^{\mathrm{Spin}^c}$. Since X and Y have primitive canonical classes, and their inverses in the cobordism group are given by reversing orientation, we conclude that every class in $\Omega_4^{\mathrm{Spin}^c}$ can be represented by a simply connected manifold B with primitive canonical class d. Lemma 1.6 implies that by mapping the cobordism class of such a pair to $\beta(B, d)$ we can define a homomorphism $\beta\colon \Omega_4^{\mathrm{Spin}^c} \to \Omega_4^{\mathrm{Pin}^+}$.

Using the isomorphism $\Omega_4^{\mathrm{Pin}^+} \cong \mathbb{Z}_{16}$ generated by a pin$^+$ structure on $\mathbb{R}P^4$ we prove the following:

Lemma 1.7 *We have that*

$$\beta(B, d) = \langle d^2, [B]\rangle + 4\epsilon \operatorname{ind}(B, d) \bmod 16$$

for an unknown sign $\epsilon = \pm 1$.

This lemma corrects a mistake in the statement of [26, Theorem 6.7]. Our argument uses ideas from the proof in [26] as well as corrections suggested to the author by Yang Su.

Proof We will see that $\beta(X,x) = 1$ and $\beta(Y,d_Y) = 5$ or 13. The lemma then follows since β is a homomorphism and $\Omega_4^{\mathrm{Spin}^c} \cong \mathbb{Z}^2$.

The principal S^1 bundle $\mathbb{R}P^5 \to \mathbb{C}P^2$ which is a $\mathbb{Z}_2$ quotient of the Hopf bundle has first Chern class $2x$. Since $\mathbb{R}P^4$ is a characteristic submanifold of $\mathbb{R}P^5$, it follows that

$$\beta(X,x) = [\mathbb{R}P^4] = 1 \in \Omega_4^{\mathrm{Pin}^+}.$$

The second calculation is more involved. We use the notation $[z_0,z_1,z_2] \in \mathbb{C}P^2$ and $[z_0,z_1,z_2]_\pm \in \mathbb{R}P^5$ for the respective images of the point $(z_0,z_1,z_2) \in S^5 \subset \mathbb{C}^3$. Let $\rho\colon M \to Y$ be the principal S^1 bundle with first Chern class $2d_Y \in H^2(Y,\mathbb{Z})$ as defined above. By Lemma 1.3, the double cover $\widetilde{M}$ of M is the total space of a principal S^1 bundle $\tilde{\rho}\colon \widetilde{M} \to Y$ with first Chern class d_Y. Let $g\colon Y \to \mathbb{C}P^2$ be a classifying map for $\tilde{\rho}$ which is transverse to $\mathbb{C}P^1 \subset \mathbb{C}P^2$ and has a regular value $[1,0,0] \in \mathbb{C}P^1$. Then $g^*x = d_Y$ and the pullback of $\pi\colon \mathbb{R}P^5 \to \mathbb{C}P^2$ by f has first Chern class $2d_Y$. There is a map of principal S^1 bundles $f\colon M \to \mathbb{R}P^5$ covering g, that is, an S^1 equivariant map making the following diagram commute:

$$\begin{tikzcd} M \arrow[r, "f"] \arrow[d, "\rho"] & \mathbb{R}P^5 \arrow[d, "\pi"] \\ Y \arrow[r, "g"] & \mathbb{C}P^2 \end{tikzcd}$$

Since the fundamental groups of M and $\mathbb{R}P^5$ are generated by S^1 orbits (see Lemma 1.3), the homomorphism $f_*\colon \pi_1(M) \to \pi_1(\mathbb{R}P^5)$ is an isomorphism and f is a classifying map for the double cover $\widetilde{M} \to M$. Thus if we show that f is transverse to $\mathbb{R}P^4 \subset \mathbb{R}P^5$, we can conclude that $P = f^{-1}(\mathbb{R}P^4)$ is a characteristic submanifold of M. Then given the correct pin$^+$ structure on P, $\beta(Y,d_Y) = [P] \in \Omega_4^{\mathrm{Pin}^+}$.

To see that f is transverse to $\mathbb{R}P^4 = \{[z_0,z_1,r]_\pm \in \mathbb{R}P^5 \mid r \in \mathbb{R}\}$ note that at points in $\pi^{-1}(\mathbb{C}P^2 \backslash \mathbb{C}P^1)$, $\mathbb{R}P^4$ is transverse to the S^1 orbits, which are contained in the image of the equivariant map f. At points in $\pi^{-1}(\mathbb{C}P^1)$, we associate the horizontal space of the S^1 action with $T\mathbb{C}P^2$. By assumption on g, f is transverse to $T\mathbb{C}P^1$, and $T\mathbb{C}P^1 \subset T\mathbb{R}P^4$.

For later, we also note that f is transverse to $\mathbb{R}P^2 = \{[z_0,r,0] \in \mathbb{R}P^5 \mid r \in \mathbb{R}\}$ since $T\mathbb{C}P^1 \subset T\mathbb{R}P^2$ except at $[1,0,0]$, which is a regular value of f by assumption on g.

There is a short exact sequence

(1.8) $$0 \to \mathbb{Z}_2 \to \Omega_4^{\mathrm{Pin}^+} \xrightarrow{\phi} \Omega_2^{\mathrm{Pin}^-} \to 0,$$

where ϕ is given by taking the cobordism class of a submanifold dual to w_1^2; see [26, page 172] and [27, page 217] for details. Thus $\Omega_2^{\mathrm{Pin}^-}$ is isomorphic to $\mathbb{Z}_8$ with generator $[\mathbb{R}P^2]$. We now compute $\phi([P]) = 5$, which restricts the possible values of $\beta(Y,d_Y) = 5$ or 13, as desired.

We need to find a submanifold of P dual to $w_1^2(TP)$. Denote by $N\mathbb{R}P^4$ the normal bundle of $\mathbb{R}P^4$ in $\mathbb{R}P^5$ and by NP the normal bundle of P in M. Then $f^*N\mathbb{R}P^4 = NP$. Since $\mathbb{R}P^5$ and M are orientable,

$$w_1(TP) = w_1(NP) = f^*w_1(N\mathbb{R}P^4) = f^*w_1(T\mathbb{R}P^4).$$

Since $w_1(T\mathbb{R}P^4)^2$ is dual to $\mathbb{R}P^2 \subset \mathbb{R}P^4$, as long as the mod 2 degree of $f\colon f^{-1}(\mathbb{R}P^2) \to \mathbb{R}P^2$ is 1, it follows that $f^{-1}(\mathbb{R}P^2)$ is dual to $w_1(TP)^2$. For convenience let $\Sigma = f^{-1}(\mathbb{R}P^2)$. Since $[1,0,0]$ is a regular point of g, $[1,0,0]_\pm$ is a regular point of f, and the degree of f is the same as the degree of $f|_\Sigma$. The degree of f is the same as the degree of g. The degree of g is given by

$$\langle g^*x^2, Y\rangle = \langle d_Y^2, [Y]\rangle = 9.$$

Thus the mod 2 degree of $f|_\Sigma$ is 1 and $\phi([P]) = [\Sigma] \in \Omega_2^{\mathrm{Pin}^-}$.

Let U be a tubular neighborhood of the S^1 orbit of $[1,0,0]_\pm$ and $V = \mathbb{R}P^2 \backslash U$. Since $[1,0,0]$ is a regular value of g we can choose U to be made up of regular values of f. Then $f|_{f^{-1}(U)}$ is a covering map. Since f maps S^1 fibers to S^1 fibers, $f_*\colon \pi_1(f^{-1}(U)) \to \pi_1(U)$ is surjective and the covering is trivial. Thus $f^{-1}(U)$ is the disjoint union of $\deg(f) = 9$ copies of U and $f^{-1}(U \cap \mathbb{R}P^2)$ is 9 copies of $U \cap \mathbb{R}P^2$. The S^1 orbit of $[1,0,0]_\pm$ is a nontrivial loop in $\mathbb{R}P^2$, and $U \cap \mathbb{R}P^2$ is a tubular neighborhood of that loop, diffeomorphic to $\mathbb{R}P^2 \backslash D^2$ (the Möbius band). The local inverses to $f|_{f^{-1}(U)}$ are equivariant embeddings of the oriented tubular neighborhood U and are all isotopic. It follows that the 9 embedding of $\mathbb{R}P^2 \backslash D^2$ making up $f^{-1}(U \cap \mathbb{R}P^2)$ are all isotopic. Thus the process by which TM induces a pin$^+$ structure on P, which in turn induces a pin$^-$ structure on Σ, will induce the same pin$^-$ structure on each of the 9 copies of $\mathbb{R}P^2 \backslash D^2$.

Since $\pi(\mathbb{R}P^2) = \mathbb{C}P^1$ and $\pi(U) \cap \mathbb{C}P^1$ is diffeomorphic to a disc D^2 around $[1,0,0]$ made up of regular values of g, $g^{-1}(\pi(U) \cap \mathbb{C}P^1)$ is 9 copies of D^2 and $\pi(V) = \mathbb{C}P^1 \backslash D^2$. $\pi|_{\mathbb{R}P^2}$ is injective away from the orbit of $[1,0,0]_\pm$, and thus is injective on V. It follows that ρ maps $f^{-1}(V)$ injectively onto $g^{-1}(\pi(V))$. Thus $f^{-1}(V)$ is diffeomorphic to $g^{-1}(\mathbb{C}P^2)$ with 9 discs removed while $f^{-1}(U \cap \mathbb{R}P^2)$ is 9 copies of $\mathbb{R}P^2 \backslash D^2$. In other words,

$$\Sigma \cong g^{-1}(\mathbb{C}P^1) \# \mathbb{R}P^2 \# \cdots \# \mathbb{R}P^2, \tag{1.9}$$

and the nine summands of $\mathbb{R}P^2$ all have the same pin$^-$ structure. $\Omega_2^{\mathrm{Pin}^-}$ is generated by $[\mathbb{R}P^2]$, and so it remains to compute the value of $[g^{-1}(\mathbb{C}P^1)]$.

Let $\chi = g^{-1}(\mathbb{C}P^2)$. We will use a general method to define a pin$^-$ structure called r_χ on χ and compute $[\chi] \in \Omega_2^{\mathrm{Pin}^-}$ with this structure. We will then show that r_χ is the correct pin$^-$ structure to use, that is, r_χ is compatible under (1.9) with the pin$^-$ structure used to identify $[\Sigma]$ with $\phi([P])$, which we will call r.

Consider a simply connected spin^c 4–manifold B with canonical class d and ν the complex line bundle with $c_1(\nu) = d$. Let $N \subset B$ be a smooth submanifold dual to d. Then $\nu|_N$ is isomorphic to the normal bundle of N. The spin^c structure on B is equivalent to a spin structure, called s, on $TB \oplus \nu$. Restricted to N, this is a spin structure on $TN \oplus 2\nu$. The transition functions for 2ν admit a canonical lift from SO(4) to Spin(4); simply multiply two copies of any lift for the transition functions of ν, and the sign ambiguities cancel. Note that the identity lifts to the identity in this way. Using this lift, s induces a spin structure s_N on N.

The spin cobordism class of N depends only on the spin^c cobordism class of B. To see this, note that the dual to the canonical class of a spin^c cobordism will be a spin cobordism between the two relevant submanifolds. Thus we have a homomorphism

$$\psi : \Omega_4^{\mathrm{Spin}^c} \to \Omega_2^{\mathrm{Spin}} \cong \mathbb{Z}_2$$

defined by $\psi([B]) = [N]$. Indeed, there is a long exact sequence

$$\to \Omega_4^{\mathrm{Spin}} \to \Omega_4^{\mathrm{Spin}^c} \to \Omega_2^{\mathrm{Spin}}(\mathrm{BU}(1)) \to \Omega_3^{\mathrm{Spin}} = 0$$

as in [26, page 154; 25, page 654]. We see that ψ is surjective by noting that ψ is the composition of $\Omega_4^{\mathrm{Spin}^c} \to \Omega_2^{\mathrm{Spin}}(\mathrm{BU}(1))$ with the surjective map $\Omega_2^{\mathrm{Spin}}(\mathrm{BU}(1)) \to \Omega_2^{\mathrm{Spin}}$, which ignores the map to BU(1).

Recall that X, Y generate $\Omega_4^{\mathrm{Spin}^c}$. The canonical class of X is dual to $\mathbb{C}P^1 \subset \mathbb{C}P^2$, which is nullcobordant, so $\psi([X]) = 0$. Since ψ is surjective, $\psi([Y])$ generates Ω_2^{Spin}. Since $\mathbb{C}P^1$ contains a regular value of g, the degree of $g|_\chi$ equals the degree of g and χ is dual to $g^*x = d_Y$. Giving χ the spin structure s_χ used to define ψ, we have $\psi([Y]) = [\chi] \neq 0$.

Spin(n) embeds naturally into both Pin$^\pm(n)$, so a spin structure induces a natural pin$^-$ structure. Kirby and Taylor show that in dimension 2, the corresponding map

$$\Omega_2^{\mathrm{Spin}} \cong \mathbb{Z}_2 \to \Omega_2^{\mathrm{Pin}^-} \cong \mathbb{Z}_8$$

is injective; see [27, Proposition 3.8]. Let r_χ be the Pin$^-$ structure on χ induced by s_χ. Using that structure, we have $[\chi] = 4 \in \Omega_2^{\mathrm{Pin}^-}$. Once we confirm that r_χ is the correct structure, we conclude with (1.9) that $\phi([P]) = 5$, completing the proof of Lemma 1.7.

Let r be the pin$^-$ structure on Σ used to define $\phi([P])$. Recall that ρ is a diffeomorphism between the open set $O = f^{-1}(V) \subset \Sigma$ and $\rho(O)$, which is χ with 9 discs removed. It remains only to check that $r = \rho^* r_\chi$ on O.

We first recall the definition of r. Let μ be the nontrivial real line bundle over M and let $E = TM \oplus 2\mu$. Let λ be the complex line bundle over Y with $c_1(\lambda) = d_Y$ and let s be spin structure on $TY \oplus \lambda$ used in the definition of ψ. With the isomorphism (1.4), s induces a spin structure on E called s_E. Then (1.1) shows

$$E|_P = TP \oplus 3\det(TP)$$

and we induce a pin$^+$ structure on TP using a canonical lift of the transition functions of $3\det(TP)$ from $O(3)$ to Pin$^-(3)$. In turn,

$$TP|_\Sigma = T\Sigma \oplus 2\det(T\Sigma),$$

and using a canonical lift of the transition functions of $2\det(T\Sigma)$ from $O(2)$ to $\mathrm{Pin}^+(2)$ we induce the pin^- structure r on Σ. Note that the normal bundle of Σ in P is orientable and thus

$$w_1(\det(T\Sigma)) = w_1(\det(TP)|_\Sigma).$$

In this way we can combine the two steps and see that s_E induces r on $T\Sigma$ using the isomorphism

$$E|_\Sigma = T\Sigma \oplus 5\det(T\Sigma) \tag{1.10}$$

and a canonical lift of the transition functions of $5\det(T\Sigma)$ from $O(5)$ to $\mathrm{Pin}^+(5)$. The details of the canonical lifts involved can be found in [27, Lemma 1.7]; the salient fact is that each lifts the identity to the identity.

Next, we note that $\det(T\Sigma)$ and $\rho^*\lambda$ are trivial over O. The former follows because O is an open set in Σ, but is orientable since it is diffeomorphic to an open set in χ. As for the latter, we have seen that $\rho^*\lambda \cong 2\mu$, $\mu|_P = \det(TP)$, and $\det(TP)|_\Sigma = \det(T\Sigma)$. Since ρ is a diffeomorphism on O and $\rho^*\lambda$ is trivial, λ is trivial on $\rho(O)$.

Let t_{ij} be transition functions with values in $\mathrm{SO}(2)$ for $T\chi$. As we saw in the definition of ψ, for points in χ,

$$TY \oplus \lambda \cong T\chi \oplus 2\lambda.$$

Thus on $\rho(O)$ the transition functions for λ can be chosen to be the identity and the transition functions for $(TY \oplus \lambda)|_\chi$ can be chosen to be t_{ij}. The spin structure s gives a lift of t_{ij} to $\tilde{t}_{ij}$ in $\mathrm{Spin}(2)$. Since the canonical lift of the transition functions for 2λ will also be the identity, $\tilde{t}_{ij}$ is also the lift given by s_χ and r_χ.

Furthermore, using (1.4), $t_{ij} \circ \rho$ are transition functions for E on O. By definition, s_E gives the lift $\tilde{t}_{ij} \circ \rho$. Using (1.10), $t_{ij} \circ \rho$ are transition functions for both $E|_O$ and $T\Sigma$, compatible by picking trivial transition functions for $5\det(T\Sigma)$. The canonical lift of the transition functions for $5\det(T\Sigma)$ will also be trivial, and the lift given by r will simply be the inclusion of $\tilde{t}_{ij} \circ \rho$ into $\mathrm{Pin}^-(2)$. Thus $r = \rho^* r_\chi$ on O. This completes the proof of Lemma 1.7. □

We can now prove Theorem B. In fact, we prove the following more detailed theorem, which includes the statement of Theorem B. Here we use the notation of Hambleton and Su, where $\#_{S^1}$ is gluing along the boundary of a tubular neighborhood of a generator of π_1. The $X(q)$ for $q = 1, 3, 5, 7$ are the four closed manifolds homotopy equivalent to $\mathbb{R}P^5$, with $X(1) = \mathbb{R}P^5$, and the $X(q)$ for $q = 0, 2, 4, 6, 8$ are constructed from pairs of homotopy $\mathbb{R}P^5$'s using the operation $\#_{S^1}$. The labeling is such that a characteristic submanifold $P \subset X(q)$ has class $q \in \Omega_4^{\mathrm{Pin}^+}/\pm = \{0, \dots, 8\}$. See the discussion before [26, Theorem 3.7] for details.

Theorem 1.11 *Let M be a 5–manifold with $\pi_1 = \mathbb{Z}_2$. Let $P \subset M$ be a characteristic submanifold.*

(1) *M admits a free S^1 action with a simply connected quotient if and only if M is orientable, $H_2(M, \mathbb{Z})$ is torsion-free, and $\pi_1(M)$ acts trivially on $\pi_2(M)$. Furthermore if $b_2(M) = 0$ then M is diffeomorphic to $\mathbb{R}P^5$.*

M^5	$Q(M^5) =$ simply connected 4–manifolds B^4 such that:
Type II	B is spin and $b_2(B) = b_2(M) + 1$.
Type III	B is nonspin, $b_2(B) = b_2(M) + 1$ and $\operatorname{sign}(B) = \pm[P] \bmod 4$.
Type I and $M \notin S$	B is nonspin and $b_2 = b_2(M) + 1$.
$X(q) \#_{S^1} (\mathbb{C}P^2 \times S^1)$ with $q = 0, 4$	B is nonspin, $b_2 = 3$ and $\lvert\operatorname{sign} B\rvert = 1$.
$X(q) \#_{S^1} (S^2 \times \mathbb{R}P^3)$ with $q = 0, 4$	B is nonspin, $b_2 = 4$ and $\lvert\operatorname{sign} B\rvert < 4$.

Table 1

(2) *Suppose M^5 satisfies the conditions in* (1). *Let $Q(M)$ be the set of quotients of M by free S^1 actions. Table 1 gives necessary and sufficient conditions for a 4–manifold to be in $Q(M)$. S is a set of four exceptional 5–manifolds of Type I described in the final two rows. If $b_2(M) > 0$ then for each $B \in Q(M)$, M admits infinitely many inequivalent S^1 actions with quotients diffeomorphic to B.*

Thus given M^5 satisfying the hypotheses of (1) and matching the description of one of the rows in the left column, a 4–manifold B^4 is diffeomorphic to a quotient of M^5 by a free S^1 action if and only if it satisfies the conditions given in the corresponding row of the right column.

Proof We prove (2) first. Let M be an orientable 5–manifold with $\pi_1(M) = \mathbb{Z}_2$ acting trivially on $\pi_2(M)$, $H_2(M, \mathbb{Z})$ torsion-free, and $b_2(M) > 0$ unless $M \cong \mathbb{R}P^2$. Let $P \subset M$ be a characteristic submanifold.

If $M \to B$ is a principal S^1 bundle, the long exact homotopy sequence implies that $\pi_1(M) \to \pi_1(B)$ is surjective. If $\pi_1(B) = \mathbb{Z}_2$, then the Gysin sequence implies that $H^3(B) \to H^3(M)$ is injective. Since M, and thus B, is orientable, $H^3(B) = \mathbb{Z}_2$ and $H_2(M)$ would not be torsion-free. Thus any quotient of M by a free S^1 action is simply connected.

***M* is Type II** First, suppose $M \to B$ is a principal S^1 bundle. By Lemma 1.3, $b_2(B) = b_2(M) + 1$ and by [26, Proposition 6.1], B is spin.

Conversely, let B be a simply connected spin 4–manifold with $b_2(B) = b_2(M) + 1$. It follows from [26, Proposition 6.1] that all of the total spaces of principal S^1 bundle over B with $\pi_1 = \mathbb{Z}_2$ are Type II and have second Betti number $b_2(B) - 1$. By [26, Theorem 3.1] all such total spaces are diffeomorphic to M. If $b_2(M) \geq 1$ there are infinitely many primitive elements of $H^2(B, \mathbb{Z}) = \mathbb{Z}^{b_2(M)+1}$ and thus infinitely many nonisomorphic such bundles.

***M* is Type III** Suppose $M \to B$ is a principal S^1 bundle. By Lemma 1.3, $b_1(B) = b_1(M) + 1$ and the first Chern class of the bundle is $2d$, where d is a primitive element of $H^2(B, \mathbb{Z})$ such that $w_2(TB) = d \bmod 2$. It follows that B is nonspin, and by [31, Corollary II.2.12] the intersection form of B is odd. By the classification of integral forms and Donaldson's theorem [16, page 5 and Theorem 1.3.1], the intersection form of B is diagonal, and so $H^*(B,\mathbb{Z}) = H^*(\#^a\mathbb{C}P^2\,\#^b\,\overline{\mathbb{C}P^2},\mathbb{Z})$ for some integers a and b. Then using [31, Corollary II.2.12] again we see that

$$w_2(B) = (1,1,\dots,1) \in H^2(B,\mathbb{Z}_2) \cong \mathbb{Z}_2^{a+b}.$$

Thus $d = (d_1,\dots,d_{a+b}) \in H^2(B,\mathbb{Z}) \cong \mathbb{Z}^{a+b}$, where each d_i is an odd integer. This completes the proof of one direction of (2) since

$$[P] = \beta(B,d) = \langle d^2,[B]\rangle = \sum_{i=1}^{a} d_i^2 - \sum_{j=a+1}^{b} d_j^2 = \operatorname{sign} B \mod 4.$$

Conversely, Let B be a nonspin simply connected 4–manifold with $b_2(B) = b_2(M)+1$. Assume further that $\operatorname{sign}(B) = [P] \in \mathbb{Z}_4/\pm$. Again, $H^*(B,\mathbb{Z}) = H^*(\#^a\mathbb{C}P^2\,\#^b\,\overline{\mathbb{C}P^2},\mathbb{Z})$, where $b_2(B) = a+b$ and $\operatorname{sign}(B) = a-b$. Choose $c \in \{0,1,2,3\}$ such that $\pm[P] = a-b+4c \mod 16$. If $b_2(M) > 0$, choose k such that

$$(4+2\epsilon)k(k+1) = 4c \mod 16,$$

where $\epsilon = \pm 1$ is the sign from Lemma 1.7. If $b_2(M) = 0$ then choose $k = 0$. Set

$$d_k = (1+2k,1,\dots,1) \in H^2(B,\mathbb{Z}) \cong \mathbb{Z}^{a+b}.$$

Then d is primitive and as above, we see that $w_2(TB) = d \mod 2$. Using Lemma 1.7 we have

$$\beta(B,d_k) = \operatorname{sign} B + (4+2\epsilon)k(k+1) = \pm[P] \mod 16.$$

Hence, by Lemma 1.5 and Theorem 1.2, M is diffeomorphic to the total space of an S^1 bundle over B with first Chern class $2d_k$. In the case where $b_2(M) > 1$, there are infinitely many choices of k yielding distinct classes d_k, and M is diffeomorphic to infinitely many total spaces of nonisomorphic S^1 bundles over B.

***M* is Type I** Suppose $M \to B$ is a principal S^1 bundle. By Lemma 1.3, $b_1(B) = b_1(M)+1$ and by [26, Proposition 6.1] B is nonspin and the first Chern class of the bundle is $2d$, where d is a primitive element of $H^2(B,\mathbb{Z})$ such that $w_2(TB) \neq d \mod 2$.

If $M = X(q)\,\#_{S^1}(\mathbb{C}P^2\times S^1)$ with $q = 0,4$, then $b_2(B) = 3$ and by [26, Theorem 6.8] $\langle d^2,[B]\rangle = \pm q \mod 8$. If $\operatorname{sign}(B) = \pm 3$, then up to orientation as above $H^*(B,\mathbb{Z}) = H^*(\#^3\mathbb{C}P^2,\mathbb{Z})$ and $w_2(TB) = (1,1,1)$. Thus

$$d = (d_1,d_2,d_3) \in H^2(B,\mathbb{Z}) \cong \mathbb{Z}^3,$$

and some d_i must be even. Since d is primitive, some d_i must be odd. One easily checks that under these conditions, $\langle d^2,[B]\rangle \neq 0,4 \mod 8$. So $\operatorname{sign}(B) = \pm 1$.

If $M = X(q)\#_{S^1}(S^2\times\mathbb{R}P^3)$ with $q = 0,4$, then $b_2(B) = 4$ and $\langle d^2,[B]\rangle = \pm q \mod 8$. If $\operatorname{sign}(B) = \pm 4$, then up to orientation by the argument in the Type III case, $H^*(B,\mathbb{Z}) = H^*(\#^4\mathbb{C}P^2,\mathbb{Z})$ and

$$d = (d_1,d_2,d_3,d_4) \in H^2(B,\mathbb{Z}) \cong \mathbb{Z}^3$$

with at least one d_i even and at least one d_i odd. Again $\langle d^2,[B]\rangle \neq 0,4 \mod 8$, so $|\operatorname{sign}(B)| < 4$.

Conversely, Let B be a simply connected nonspin 4–manifold satisfying the conditions given by the table for $Q(M)$. Then $H^*(B,\mathbb{Z}) = H^*(\#^a \mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}, \mathbb{Z})$ for some integers a, b such that $a+b = b_1(M)+1$. Let $(q,s) \in \mathbb{Z}_8 \oplus \mathbb{Z}_2$ represent the cobordism class of $P \subset M$ in the pin^c cobordism group $\Omega_4^{\mathrm{Pin}^c} \cong \mathbb{Z}_8 \oplus \mathbb{Z}_2$; see [26, page 154]. By [26, Theorem 3.6], $q+s = b_2(M)+1 \bmod 2$.

If $q = 0, 4$ then [26, Theorem 3.7] implies that $a+b \geq 3$, so we can assume that up to orientation $a \geq 2$ and using Table 1 either $a+b \geq 5$ or $|\mathrm{sign}(B)| < b_2(B)$, which implies $b > 0$. Define the following elements $d_k \in H^2(B,\mathbb{Z}) \cong \mathbb{Z}^a \oplus \mathbb{Z}^b$ for each $k \in \mathbb{Z}$:

$$q = 0\colon \quad d_k = \begin{cases} (1+8k, 0, \dots, 0, 1) & \text{if } b > 0, \\ (2+8k, 1, 1, 1, 1, 0, \dots, 0) & \text{if } b = 0, \end{cases}$$

$$q = 4\colon \quad d_k = \begin{cases} (2+8k, 1, 0, \dots, 0, 1) & \text{if } b > 0, \\ (1+8k, 1, 1, 1, 0, \dots, 0) & \text{if } b = 0. \end{cases}$$

If $q = 2$, [26, Theorem 3.7] implies that $a+b \geq 3$ and we can assume $a \geq 2$ and define

$$q = 2\colon \quad d_k = (1+8k, 1, 0, \dots, 0).$$

If q is odd, by [26, Theorem 3.7] $a+b \geq 2$, and we can assume $a \geq 1$. Define

$$q = 1\colon \quad d_k = (1+8k, 4, 0, \dots, 0),$$

$$q = 3\colon \quad d_k = (1+8k, 2, 0, \dots, 0).$$

In each case d_k is primitive, $w_2(TB) \neq d_k \bmod 2$, and $q = \pm\langle d_k^2, [B]\rangle \bmod 8$. By [26, Theorem 6.8] the S^1 bundle over B with first Chern class $2d_k$ is diffeomorphic to M. Again, infinitely many k yield distinct classes d_k and thus nonisomorphic bundles.

To prove (1), first assume M is a 5–manifold with $\pi_1(M) = \mathbb{Z}_2$ admitting a free S^1 action with simply connected quotient B. By Lemma 1.3, M is orientable, $\pi_1(M)$ acts trivially on $\pi_2(M)$ and $H_2(M,\mathbb{Z})$ is torsion-free. If $b_2(M) = 0$, then $b_2(B) = 1$ and up to orientation $H^*(B,\mathbb{Z}) \cong H^*(\mathbb{C}P^2,\mathbb{Z})$ and $w_2(TB)$ is nonzero. There are only two primitive classes $\pm d \in H^2(B,\mathbb{Z}) \cong \mathbb{Z}$, each restricting to $w_2(B)$ mod 2. Thus B is of Type III and $\beta([B,d]) = \pm 1$. By Theorem 1.2, M is diffeomorphic to $\mathbb{R}P^5$.

To prove the converse, suppose M is an orientable 5–manifold with $\pi_1(M) = \mathbb{Z}_2$ acting trivially on $\pi_2(M)$ and $H_2(M,\mathbb{Z})$ torsion-free. Let $P \subset M$ be a characteristic submanifold. Since $\mathbb{R}P^5$ admits a free S^1 action induced by the Hopf action we assume $b_2(M) > 0$. We must show the set $Q(M)$ described in Table 1 is nonempty.

If M is Type II, by [26, Theorem 3.6] $b_2(M)$ is odd. Then $B = \#^{(b_2(M)+1)/2} S^2 \times S^2 \in Q(M)$. If M is Type I then $B = \#^{b_2(M)} \mathbb{C}P^2 \# \overline{\mathbb{C}P^2} \in Q(M)$. If M is Type III, let $0 \leq c < 16$ be an integer such that $[P] = c \bmod 16$. By [26, Theorem 3.6] we see that $c = b_2(M)+1 \bmod 2$. Choose l such that

$$0 \leq c - 4l < 4.$$

Then

$$a = \tfrac{1}{2}(b_2(M)+1+c-4l) \quad \text{and} \quad b = \tfrac{1}{2}(b_2(M)+1-c+4l)$$

are nonnegative integers. Let $B = \#^a \mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$. Then $b_2(B) = b_2(M) + 1$ and $\operatorname{sign}(B) = [P] \in \mathbb{Z}_4/\pm$. So $B \in Q(M)$. □

We note that the final paragraph of the proof above in fact shows the following, which we will make use of later.

Corollary 1.12 *Let M be a 5–manifold with $\pi_1 = \mathbb{Z}_2$ admitting a free S^1 action with a simply connected quotient. Then M admits a free S^1 action with quotient diffeomorphic to either $\#^c S^2 \times S^2$ or $\#^a \mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ for some $a, b, c \in \mathbb{Z}$.*

Combining Theorem 1.11 with [26, Theorem 3.7], we can characterize the manifolds satisfying Theorem A.

Corollary 1.13 *Let M^5 be a 5–manifold. The following are equivalent*:

(1) *M^5 is Type III and admits a free S^1 action with a simply connected quotient.*

(2) *There exists $B^4 = \#^a \mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ with $a, b \in \mathbb{Z}_{\geq 0}$ such that M^5 is the total space of a principal bundle $S^1 \to M^5 \to B^4$ with first Chern class $2d$, where $d \in H^2(B^4, \mathbb{Z})$ is primitive and $w_2(TB^4) = d \bmod 2$.*

(3) *There exist $k \in \mathbb{Z}_{\geq 0}$ and $q \in \{0, 1, \dots, 8\}$, with $k > 0$ if q is 3, 5 or 7, such that M^5 is diffeomorphic to*
$$X(q) \#_{S^1} (\#^k (S^2 \times S^2) \times S^1).$$

If those conditions are satisfied then $\mathfrak{M}_{\mathrm{Ric}>0}(M^5)$ has infinitely many path components.

Proof (1) implies (2) by Lemma 1.3 and Corollary 1.12. If we assume (2), Lemma 1.3 implies that M is a Type III manifold with π_1 acting trivially on π_2 and $H_2(M, \mathbb{Z})$ torsion-free. [26, Theorem 3.7] shows that every such manifold is diffeomorphic to $X(q) \#_{S^1} (\#^k (S^2 \times S^2) \times S^1)$ for some $q \in \{0, \dots, 8\}$ and some $k \in \mathbb{Z}_{\geq 0}$. If $k = 0$ and q is odd, $b_2(X(q)) = 0$ and using Theorem 1.11, M must be diffeomorphic to $\mathbb{R}P^5 = X(1)$.

By [26, Theorem 3.7], $M = X(q) \#_{S^1} (\#^k (S^2 \times S^2) \times S^1)$ is an orientable Type III manifold with $\pi_1(M)$ acting trivially on $\pi_2(M)$, $H_2(M, \mathbb{Z})$ torsion-free, and $b_2(M) = 2k + (1 + (-1)^q)/2$. Thus by Theorem 1.11, (3) implies (1).

Now assume M satisfies the conditions. If $b_2(M) > 0$, then by Lemma 1.3 the integers a, b in (2) must satisfy $a + b \geq 2$. Then Theorem A implies that $\mathfrak{M}_{\mathrm{Ric}>0}(M)$ has infinitely many path components. By Theorem 1.11, if $b_2(M) = 0$, then (1) implies that $M \cong \mathbb{R}P^5$. $\mathfrak{M}_{\mathrm{Ric}>0}(\mathbb{R}P^5)$ is shown to have infinitely many path components in [14]. □

Remark 1.14

- By the discussion preceding [26, Theorem 3.7] the manifolds described in Corollary 1.13 can also be constructed by applying $\#_{S^1}$ to the homotopy $\mathbb{R}P^5$'s. For instance,
$$X(q) \#_{S^1} (\#^k (S^2 \times S^2) \times S^1) \cong X(q) \#_{S^1} X(0) \#_{S^1} \cdots \#_{S^1} X(0).$$

- It is shown in [14] that $\mathfrak{M}_{\mathrm{Ric}>0}$ also has infinitely many components for the homotopy $\mathbb{R}P^5$'s $X(3)$, $X(5)$ and $X(7)$.
- A characteristic submanifold $P \subset X(q)\#_{S^1}(\#^k(S^2\times S^2)\times S^1)$ has class $q \in \Omega_4^{\mathrm{Pin}^+}/\pm = \{0,\dots,8\}$. If we fix a nonspin simply connected 4–manifold B^4, then a Type III total space of a principal S^1 bundle over B will be diffeomorphic to $X(q)\#_{S^1}(\#^k(S^2\times S^2)\times S^1)$. Using Table 1 we see that q must satisfy $q = \pm\,\mathrm{sign}\,B \bmod 4$. It follows that there are 2, 3 or 4 choices of q, and the same number of diffeomorphism types of Type III total spaces, if $\mathrm{sign}(B)$ is 2, 0 or $\pm 1 \bmod 4$, respectively. The value of q can be determined, up to two possibilities, using Lemma 1.7. The set of diffeomorphism types of Type I total spaces is more complicated, but can be computed using Theorem 1.11 and [26, Theorem 3.7]. If B^4 is a simply connected spin 4–manifold, there exists a unique diffeomorphism type of total spaces with $\pi_1 = \mathbb{Z}_2$, represented by $(S^2\times\mathbb{R}P^3)\#_{S^1}(\#^{(b_2(B)-2)/2}(S^2\times S^2)\times S^1)$.

Using a result of Gilkey, Park and Tuschmann, we can lift metrics from the quotients described by Corollary 1.12 to prove the following:

Corollary 1.15 *Let M be a 5–manifold with $\pi_1(M) = \mathbb{Z}_2$ admitting a free S^1 action with a simply connected quotient. Then M admits a metric with positive Ricci curvature.*

Proof In [37] Sha and Yang put a metric of positive Ricci curvature on $\#^c S^2\times S^2$. A modification of Perelman's construction in [34] puts such a metric on $\#^a\mathbb{C}P^2\,\#^b\,\overline{\mathbb{C}P^2}$; see Lemma 3.10. Corollary 1.12 shows that M^5 admits a free S^1 action with quotient B^4 diffeomorphic to one of those manifolds. Gilkey, Park and Tuschmann [20] showed that if B^4 admits $\mathrm{Ric}>0$, M^5 is the total space of a principal bundle over B^4 with compact connected structure group G, and $\pi_1(M^5)$ is finite, then M admits a G–invariant metric with $\mathrm{Ric}>0$. In this case $G = S^1$, $\pi_1(M) = \mathbb{Z}_2$ and the corollary follows. □

The corresponding result in the simply connected case was proved by Corro and Galaz-Garcia in [11]. By Lichnerowicz's theorem, many simply connected 4–manifolds, such as a K3 surface, do not admit even positive scalar curvature. It is interesting to note that Corollary 1.15 and the results of [11] imply that total spaces with $\pi_1 = 0$ or $\mathbb{Z}_2$ of principal S^1 bundles over such manifolds nonetheless admit metrics of positive Ricci curvature.

2 The η invariant

We use the η invariant of the spin^c Dirac operator, which we define in this section, to distinguish components of geometric moduli spaces. A manifold M is spin^c if there exists a complex line bundle λ over M such that the frame bundle of $TM\oplus\lambda$, a principal $\mathrm{SO}(n)\times U(1)$ bundle, lifts to a principal $\mathrm{Spin}^c(n) = \mathrm{Spin}(n)\times_{\mathbb{Z}_2} U(1)$ bundle. A manifold is spin^c if and only if the second Stiefel–Whitney class $w_2(TM)$ is the image of an integral class $c \in H^2(M,\mathbb{Z})$ under the map $H^2(M,\mathbb{Z}) \to H^2(M,\mathbb{Z}_2)$. In this case c, which we call the canonical class of the spin^c structure, is the first Chern class of λ, which we call the canonical bundle.

Using complex representations of $\mathrm{Spin}^c(n)$ we form spin^c spinor bundles and equip them with actions of the complex Clifford algebra bundle $\mathbb{C}\ell(TM)$. When the dimension of M is even there is a unique irreducible such bundle S with a natural grading $S = S^+ \oplus S^-$. Given a metric g on M and a unitary connection ∇ on λ, we can construct a spinor connection ∇^s on S, compatible with Clifford multiplication, and a spin^c Dirac operator $D^c_{g,\nabla}$ acting on sections of S. See [31, Appendix D] for details. The Bochner–Lichnerowicz identity for this operator is

$$(D^c_{g,\lambda})^2 = (\nabla^s)^*\nabla^s + \tfrac{1}{4}\,\mathrm{scal}(g) + \tfrac{i}{2}F^{\nabla}, \tag{2.1}$$

where the complex two-form F^∇ is the curvature of ∇. This form acts on the spinor bundle S by way of the vector bundle isomorphism $\Lambda T^*M \to \Lambda TM \to \mathbb{C}\ell(TM)$ given by g. The operator $(\nabla^s)^*\nabla^s$ is nonnegative definite with respect to the L^2 inner product on a closed manifold or a compact manifold with boundary on which the Atiyah–Patodi–Singer boundary conditions have been applied. See [2, Theorem 3.9] for details. The remaining term $\frac{1}{4}\,\mathrm{scal}(g) + \frac{i}{2}F^\nabla$ is positive definite if

$$\mathrm{scal}(g) > 2|F^\nabla|_g, \tag{2.2}$$

where the norm $|\cdot|_g$ is the operator norm on $\mathbb{C}\ell(TM)$ acting on S. In particular, $\ker(D^c_{g,\nabla}) = 0$ if (2.2) is satisfied. For a later purpose we note that for $\omega \in \Omega^2(M,\mathbb{C})$ and an orthonormal basis $\{e_i\}$ of TM with respect to g, we have

$$|\omega|_g \le \sum_{i<j} |\omega(e_i,e_j)|. \tag{2.3}$$

Suppose W is a spin^c manifold with boundary $\partial W = M$, with λ and c defined on W as above. W induces a spin^c structure on M with canonical class $c|_{\partial W}$ and canonical bundle $\lambda|_{\partial W}$. Choose a metric h on W and a connection ∇ on λ which are product-like near ∂W, ie

$$h = h|_{\partial W} + dr^2 \quad\text{and}\quad \nabla = \mathrm{proj}_M^*(\nabla|_{\partial W})$$

on a collar neighborhood $U \cong M \times I$, where I is an interval with coordinate r. Applying the Atiyah–Patodi–Singer boundary conditions, the Atiyah–Patodi–Singer index theorem [1] states that

$$\mathrm{ind}(D^c_{h,\nabla}|_{S^+}) = \int_W e^{c_1(\nabla)/2}\hat{A}(p(g)) - \tfrac{1}{2}\big(\dim(\ker(D^c_{h|_{\partial W},\nabla|_{\partial W}})) + \eta(D^c_{h|_{\partial W},\nabla|_{\partial W}})\big). \tag{2.4}$$

Here $c_1(\nabla)$ and $p(g)$ are the Chern–Weil Chern and Pontryagin forms constructed from the curvature tensors of the connection and metric, respectively. $\hat{A}$ is the polynomial in the Pontryagin forms and $D^c_{h|_{\partial W},\nabla|_{\partial W}}$ is the spin^c Dirac operator on M constructed using the induced metric and connection.

The η invariant is an analytic invariant of the spectrum of an elliptic operator defined in [1]. Given an elliptic differential operator D with spectrum $\{\lambda_i\}$, we define a complex function

$$\eta(D,s) = \sum_{\lambda_i \neq 0} \mathrm{sign}(\lambda_i)|\lambda_i|^{-s}.$$

One shows that the function is analytic when the real part of s is large, and Atiyah, Patodi and Singer showed that it can be analytically continued to a meromorphic function which is analytic at 0. Thus we define $\eta(D) = \eta(D, 0)$. If a diffeomorphism ϕ preserves the spin^c structure, then $D^c_{\phi^* g, \phi^* \nabla}$ is conjugate to $D^c_{g,\nabla}$, and hence they have the same spectrum and the same values of η. We will use (2.4) to calculate η for an operator $D_{g,\overline{\nabla}}$ on a manifold M by finding a suitable W with $\partial W = M$ and extending g and $\overline{\nabla}$ to product-like h and ∇ on W.

Kreck and Stolz combined the η invariant with information about the Chern–Weil forms of the metric to get an invariant for metrics on $(4n+3)$–dimensional spin manifolds. We prove that the η invariant alone provides the desired invariant for certain $(4n+1)$–dimensional spin^c manifolds.

Theorem 2.5 *Let M^{4n+1} be a closed* spin^c *manifold with canonical class $c \in H^2(M, \mathbb{Z})$ and canonical bundle λ. Suppose c and the Pontryagin classes $p_i(TM)$ are torsion and g_t, where $t \in [0, 1]$, is a smooth path of metrics on M with* $\mathrm{scal}(g_t) > 0$. *If ∇_0 and ∇_1 are flat unitary connections on λ, then*

$$\eta(D^c_{g_0,\nabla_0}) = \eta(D^c_{g_1,\nabla_1}).$$

Proof Modifying g_t if necessary, we assume it is a constant path for t near 0 and 1. Given $L \in \mathbb{R}_{>0}$, define a smooth metric g on $M \times [0, 1]$ by

$$g = g_t + L^2\, dt^2.$$

Then g is product-like near $M \times \{0, 1\}$. One sees that $\mathrm{scal}(g)$ differs from $\mathrm{scal}(g_t)$ by terms depending on the second fundamental form of each slice $M \times \{t\}$, but the second fundamental form tends to 0 as $L \to \infty$, so for large L we have $\mathrm{scal}(g) > 0$.

The difference of unitary connections on a complex line bundle is an imaginary one-form. Define $\alpha \in \Omega(M)$ such that

$$i\alpha = \nabla_1 - \nabla_0.$$

Since both connections are flat, $d\alpha = 0$. Let $\pi : M \times [0, 1] \to M$ be the projection and let $f : M \times [0, 1] \to [0, 1]$ be the projection onto $[0, 1]$ followed by a smooth function which is 0 in a neighborhood of 0 and 1 in a neighborhood of 1. Define a connection on $\pi^*\lambda$ by

$$\nabla = \pi^*\nabla_0 + i f \pi^*\alpha.$$

Then, since ∇_0 is flat,

$$F^\nabla = i\, df \wedge \pi^*\alpha.$$

Let e_i be an orthonormal frame for g at a point (p, t), such that $e_1 = (1/L)\partial_t$. Then

$$2\sum_{i<j} |(df \wedge \alpha)(e_i, e_j)| = \frac{2\partial_t f}{L} \sum_{i>1} \alpha(e_i).$$

Since e_i for $i > 2$ is tangent to $M \times \{t\}$, it does not depend on L. Using (2.3), for large L we have

$$\mathrm{scal}(g) > 2|F^\nabla|_g.$$

The definition of f ensures that ∇ is product-like near $\partial(M\times I)$. Then by (2.1), $D^c_{g,\nabla}$ has trivial kernel and $\operatorname{ind}(D^c_{g,\nabla}|_{S^+})=0$.

Since $F^{\nabla_i}=0$ for $i=1,2$,

$$\operatorname{scal}(g_i)>0=2|F^{\nabla_i}|_{g_i}$$

and hence (2.1) implies $\ker D^c_{g_i,\nabla_i}=\{0\}$. We now apply the Atiyah–Patodi–Singer index theorem (2.4). The boundary of $M\times I$ is two copies of M with opposite orientations. The spectrum of the Dirac operator on $M\times\{0,1\}$ is the union of the spectra on $M\times\{0\}$ and $M\times\{1\}$, and the η invariant is the sum of the two η invariants. When we change the orientation of an odd-dimensional manifold, the Dirac operator changes by a sign. Thus the Atiyah–Patodi–Singer theorem yields

$$\begin{aligned}&\operatorname{ind}(D^c_{g,\nabla}|_{S^+})\\ &\quad=\int_{M\times[0,1]}e^{c_1(\nabla)/2}\widehat{A}(p(g))-\tfrac{1}{2}\big(\dim(\ker(D^c_{g_0,\nabla_0}))+\dim(\ker(D^c_{g_1,\nabla_1}))+\eta(D^c_{g_0,\nabla_0})-\eta(D^c_{g_1,\nabla_1})\big),\end{aligned}$$

and hence

$$\eta(D^c_{g_1,\nabla_1})-\eta(D^c_{g_0,\nabla_0})=2\int_{M\times[0,1]}e^{c_1(\nabla)}\widehat{A}(p(g)).$$

Since π_1^*c is torsion, $c_1(\nabla)$ is exact. Because ∇ is flat near the boundary, $c_1(\nabla)|_{\partial(M\times I)}=0$. Furthermore, g is product-like near the boundary so $p(g)|_{M\times\{i\}}=p(g_i)$. Since the real Pontryagin classes of M vanish, $p_j(g_i)$ is exact for $j>0$. By Stokes' theorem, and since the dimension of M is $4n+1$, the integral vanishes. □

As a corollary we show how to use the η invariant to detect path components of moduli spaces of metrics with curvature conditions no weaker than positive scalar curvature.

Corollary 2.6 *Let M be as in Theorem 2.5. Let (g_i,∇_i) be a sequence of Riemannian metrics g_i with $\mathrm{Ric}(g_i)>0$, and flat connections ∇_i on λ such that $\{\eta(D^c_{g_i,\nabla_i})\}_i$ is infinite. Then $\mathfrak{M}_{\mathrm{Ric}>0}(M)$ and $\mathfrak{M}_{\mathrm{scal}>0}(M)$ have infinitely many path components.*

Proof Let $\mathrm{Diff}^c(M)$ be the set of diffeomorphisms of M which fix the spin^c structure. For $g\in\mathfrak{R}_{\mathrm{scal}>0}$ let $[g]$ represent the image in $\mathfrak{M}_{\mathrm{scal}>0}$ and $[g]^c$ the image in $\mathfrak{R}_{\mathrm{scal}>0}/\mathrm{Diff}^c(M)$. It follows from Ebin's slice theorem [17; 9] that if $[g_i]$ and $[g_j]$ are in the same connected component of $\mathfrak{R}_{\mathrm{scal}>0}/\mathrm{Diff}^c(M)$ then g_i and ϕ^*g_j are in the same path component of $\mathfrak{R}_{\mathrm{scal}>0}$ for some $\phi\in\mathrm{Diff}^c(M)$. Then there is a path between them maintaining positive scalar curvature, and by Theorem 2.5 and the spin^c diffeomorphism invariance of η we have $\eta(D^c_{g_i,\nabla_i})=\eta(D^c_{\phi^*g_j,\phi^*\nabla_j})=\eta(D^c_{g_j,\nabla_j})$. Since $\{\eta(D^c_{g_i,\nabla_i})\}$ is infinite, $\mathfrak{R}_{\mathrm{scal}>0}/\mathrm{Diff}^c(M)$ has infinitely many components.

Any diffeomorphism ϕ pulls back the spin^c structure to another one with canonical class ϕ^*c, a torsion class in $H^2(M,\mathbb{Z})$. There are finitely many such classes. The finite group $H^1(M,\mathbb{Z}_2)$ indexes the spin^c structures associated to each class. Thus the orbit of the spin^c structure under $\mathrm{Diff}(M)$ and the set $\mathrm{Diff}(M)/\mathrm{Diff}^c(M)$ are finite. The fibers of $\mathfrak{R}_{\mathrm{scal}>0}/\mathrm{Diff}^c(M) \to \mathfrak{M}_{\mathrm{scal}>0}$ are no larger than $\mathrm{Diff}(M)/\mathrm{Diff}^c(M)$, implying that $\mathfrak{M}_{\mathrm{scal}>0}$ has infinitely many components.

The proof is identical for $\mathfrak{M}_{\mathrm{Ric}>0}$ since $\mathrm{Ric} > 0$ implies $\mathrm{scal} > 0$. □

3 The η invariant in dimension $4n+1$ with free S^1 actions

In this section we prove Theorem A. We want to use the Atiyah–Patodi–Singer index theorem to calculate the η invariant of a metric on M. Many authors have computed η and related invariants on spin manifolds M by extending metrics to manifolds W with boundary diffeomorphic to M. If the extension has positive scalar curvature, the index of the Dirac operator will vanish. In the spin^c case, we must also extend an auxiliary connection. A difficulty arises when the extended connection cannot be flat because the canonical class of the spin^c structure on W is not torsion. Then the metric and connection must satisfy (2.2). The following theorem, which we prove in Section 4, illustrates how to use certain free S^1 actions on M to accomplish this.

Theorem 3.1 *Let S^1 act freely on M by isometries of a Riemannian metric g_M with $\mathrm{scal}(g_M) > 0$ and assume $\pi_1(M)$ is finite. Let $B = M/S^1$ be the quotient and $\rho\colon W = M \times_{S^1} D^2 \to B$ the associated disc bundle. Suppose the first Chern class of the principal S^1 bundle $\pi: M \to B$ is ℓd for $d \in H^2(B,\mathbb{Z})$ and $\ell \in \mathbb{Z}$. If λ is the complex line bundle over W with first Chern class $\rho^* d$, then there exists a metric g_W on W and a connection ∇ on λ such that*

(3.2) $$\mathrm{scal}(g_W) > \ell |F^\nabla|_{g_W}.$$

Furthermore there is a collar neighborhood $V \cong M \times [0,N]$ of $\partial W \cong M$ such that for $t \in [0,N]$ near 0, g_W is a product metric

(3.3) $$g_W \cong g_M + dt^2,$$

and

(3.4) $$\nabla \cong \mathrm{proj}^*_{V,M}\, \bar{\nabla},$$

where $\bar{\nabla}$ is any flat unitary connection on $\lambda|_{\partial W}$.

Notice that here there are no restrictions on the dimension or Pontryagin classes of M, d need not be primitive, and no spin^c structure is required. We next use Theorem 3.1 and (2.4) to calculate η for S^1–invariant metrics on certain spin^c manifolds in dimensions $4n+1$.

Theorem 3.5 *Let S^1 act freely on a $4n+1$ manifold M by isometries of a Riemannian metric g with $\mathrm{scal}(g) > 0$. Assume $\pi_1(M)$ is finite and let $B = M/S^1$ be the quotient. Suppose the first Chern class of the principal bundle $S^1 \to M \xrightarrow{\pi} B$ is ℓd, where ℓ is a positive even integer and $w_2(TB) = d \bmod 2$.*

Finally assume the real Pontryagin classes of M vanish. Then M admits a spin^c *structure with canonical class π^*d. If $\bar{\bar{\nabla}}$ is a flat connection on the canonical bundle of this* spin^c *structure and $D^c_{g,\bar{\bar{\nabla}}}$ is the* spin^c *Dirac operator, then*

$$\eta(D^c_{g,\bar{\bar{\nabla}}}) = \left\langle \frac{\sinh\left(\frac{1}{2}d\right)\widehat{A}(TB)}{\sinh\left(\frac{1}{2}\ell d\right)}, [B] \right\rangle.$$

When $n = 1$,

$$\eta(D^c_{g,\bar{\bar{\nabla}}}) = \left\langle -\frac{(\ell^2-1)d^2 + p_1(TB)}{24\ell}, [B] \right\rangle. \tag{3.6}$$

Proof Since TM is the direct sum of π^*TB and a trivial bundle generated by the action field of the S^1 action,

$$w_2(TM) = \pi^* w_2(TB) = \pi^* d \bmod 2.$$

Let μ be the complex line bundle over B associated to $\pi\colon M \to B$. Let $W = M \times_{S^1} D^2$ and let $\rho\colon W \to B$ be the disc bundle associated to $\pi\colon M \to B$. Then $TW = \rho^*(TB \oplus \mu)$ and, since ℓ is even,

$$w_2(TW) = \rho^*(d + \ell d) \bmod 2 = \rho^* d \bmod 2.$$

It follows that W admits a spin^c structure with canonical class ρ^*d. We call the canonical bundle λ. The spin^c structure on W induces one on M with canonical class π^*d.

Then M, W and λ satisfy the hypotheses of Theorem 3.1. We construct the metric g_W on W and connection ∇ on λ as in the theorem such that $g_W|_M = g_M$ and $\nabla|_M = \bar{\bar{\nabla}}$. Define the spin^c Dirac operator $D^c_{g_W,\nabla}$ on W and $D^c_{g_M,\bar{\bar{\nabla}}}$ as in Section 2. Given that g_W and ∇ are product-like near ∂W, we can apply (2.4). Since $\bar{\bar{\nabla}}$ is flat,

$$\operatorname{scal}(g_M) > 2|F^{\bar{\bar{\nabla}}}|_{g_M} = 0,$$

and, by (3.2),

$$\operatorname{scal}(g_W) > \ell |F^{\nabla}|_{g_W} \geq 2|F^{\nabla}|_{g_W}.$$

Then (2.1) implies that $\operatorname{ind}(D^c_{g_W,\nabla}) = 0$ and $\ker(D^c_{g_M,\bar{\bar{\nabla}}}) = \{0\}$. It follows from (2.4) that

$$\eta(D^c_{g_M,\bar{\bar{\nabla}}}) = 2\int_W e^{c_1(\nabla)/2}\,\widehat{A}(p(g_W)). \tag{3.7}$$

To evaluate that integral, we use [29, Lemma 2.7]:

Lemma 3.8 [30] *Let W be a manifold with boundary, and let α and β be closed forms on W such that $\alpha|_{\partial W} = d\hat{\alpha}$ and $\beta|_{\partial W} = d\hat{\beta}$. Then*

$$\int_W \alpha \wedge \beta = \int_{\partial W} \hat{\alpha} \wedge \beta + \langle j^{-1}(\alpha) \cup j^{-1}(\beta), [W, \partial W] \rangle,$$

where j^{-1} represents any preimage under the long exact sequence map

$$j\colon H^*(W, \partial W; \mathbb{Q}) \to H^*(W, \mathbb{Q}).$$

To apply Lemma 3.8 to (3.7), let $\alpha = e^{c_1(\nabla)/2}$ and $\beta = \widehat{A}(p(g_W))$. Since g_W is product-like near the boundary, $p_i(g_W)|_{\partial W} = p_i(g_M)$. For $i > 0$, $p_i(g_M)$ is exact by the assumption on the Pontryagin classes of M. Since $c_1(\nabla)|_{\partial W} = c_1(\overline{\nabla})$ and $\overline{\nabla}$ is flat, we can choose $\widehat{\alpha} = 0$. The form $c_1(\nabla)$ represents the cohomology class $c_1(\lambda) = \rho^* d$. Thus

$$\eta(D^c_{g,\overline{\nabla}}) = 2\langle j^{-1}[e^{\rho^* d/2}] \cup j^{-1}[\widehat{A}(TW)], [W, \partial W]\rangle.$$

The following cup product diagram commutes:

$$\begin{array}{ccc} H^s(W,\partial W) \oplus H^t(W,\partial W) & \xrightarrow{\cup} & H^{s+t}(W,\partial W) \\ \big\downarrow{\scriptstyle (\mathrm{Id},j)} & & \big\downarrow \\ H^s(W,\partial W) \oplus H^t(W) & \xrightarrow{\cup} & H^{s+t}(W,\partial W) \end{array}$$

Thus,

$$\eta(D^c_{g,\overline{\nabla}}) = 2\langle j^{-1}[e^{\rho^* d/2}] \cup [\widehat{A}(TW)], [W, \partial W]\rangle. \tag{3.9}$$

Since the terms of $\widehat{A}(TW)$ have degree $4k$, with $k \in \mathbb{Z}$, and the dimension of W is $4n+2$, only terms of degree $4k+2$ in $e^{\rho^* d/2}$ will contribute. In those degrees, $e^{\rho^* d/2} = \sinh(\rho^* d/2)$ as power series.

Since $TW = \rho^*(TB \oplus \mu)$, we have $\widehat{A}(TW) = \rho^*(\widehat{A}(TB)\widehat{A}(\mu))$. For the complex line bundle μ, we have

$$\widehat{A}(\mu) = \frac{\frac{1}{2}c_1(\mu)}{\sinh\bigl(\frac{1}{2}c_1(\mu)\bigr)} = \frac{\ell d}{2\sinh\bigl(\frac{1}{2}\ell d\bigr)}$$

as a formal power series. The series $\sinh\bigl(\frac{1}{2}d\bigr)$ is divisible by d, so

$$\rho^*\left(\frac{\sinh\bigl(\frac{1}{2}d\bigr)}{\ell d}\right) \in H^*(W, \mathbb{Q}).$$

Let $\Phi \in H^2(W, \partial W, \mathbb{Z})$ be the Thom class of $\rho\colon W \to B$. Then $j(\Phi) = \rho^* c_1(\mu) = \rho^*(\ell d)$. By means of another commutative diagram

$$\begin{array}{ccc} H^*(W,\partial W) \oplus H^*(W) & \xrightarrow{\cup} & H^*(W,\partial W) \\ \big\downarrow{\scriptstyle (j,\mathrm{Id})} & & \big\downarrow{\scriptstyle j} \\ H^*(W) \oplus H^*(W) & \xrightarrow{\cup} & H^*(W) \end{array}$$

we see

$$j\left(\Phi \cup \rho^*\left(\frac{\sinh\bigl(\frac{1}{2}d\bigr)}{\ell d}\right)\right) = \rho^*\left(\ell d \cup \frac{\sinh\bigl(\frac{1}{2}d\bigr)}{\ell d}\right) = \rho^* \sinh\bigl(\tfrac{1}{2}d\bigr).$$

Substituting into (3.9),

$$\begin{aligned} \eta(D^c_{g,\overline{\nabla}}) &= 2\left\langle \Phi \cup \rho^*\left(\frac{\sinh\bigl(\frac{1}{2}d\bigr)}{\ell d}\right) \cup \rho^*\left(\frac{\widehat{A}(TB)\cdot \ell d}{2\sinh\bigl(\frac{1}{2}\ell d\bigr)}\right), [W,\partial W]\right\rangle \\ &= \left\langle \Phi \cup \rho^*\left(\frac{\sinh\bigl(\frac{1}{2}d\bigr)\widehat{A}(TB)}{\sinh\bigl(\frac{1}{2}\ell d\bigr)}\right), [W,\partial W]\right\rangle. \end{aligned}$$

The Thom isomorphism yields

$$\eta(D^c_{g,\overline{\nabla}}) = \left\langle \frac{\sinh\left(\frac{1}{2}d\right)\widehat{A}(TB)}{\sinh\left(\frac{1}{2}\ell d\right)}, [B] \right\rangle.$$

When $n = 1$ the dimension of B is four and we have, as series in $H^*(B, \mathbb{Z})$,

$$\widehat{A}(TB) = 1 - \tfrac{1}{24}p_1(TB) \quad \text{and} \quad \frac{\sinh\left(\frac{1}{2}d\right)}{\sinh\left(\frac{1}{2}\ell d\right)} = \frac{1}{\ell}\left(1 - \tfrac{1}{24}(\ell^2 - 1)d^2\right).$$

Multiplying and isolating terms of degree four yields (3.6). □

We are now ready to prove Theorem A. We first construct metrics of Ric > 0 on $\#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$. Perelman [34] constructed a metric with Ric > 0 on arbitrary connected sums of $\mathbb{C}P^2$ with its standard orientation. More details on Perelman's proof can be found in [10; 8]. With a slight adjustment to the construction one can reverse the orientation on some of the copies of $\mathbb{C}P^2$, proving the following.

Lemma 3.10 *$\#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$ admits a metric with positive Ricci curvature for all a and b.*

Proof In [34], Perelman puts a metric on $\#^c\mathbb{C}P^2$ for all values of c. The construction involves c copies of $\mathbb{C}P^2$ attached to a central S^4 by "necks" $S^3 \times I$. The metric on the necks is of the form

$$ds^2 = dt^2 + A^2(t,x)\,dx^2 + B^2(t,x)\,d\sigma^2,$$

where t is the coordinate on the interval I; see [34, page 159]. Furthermore, S^3 is represented as the product of S^2 and an interval with the top and bottom each identified to a point, and x is the coordinate on that interval, while $d\sigma^2$ is the standard metric on S^2.

An orientation-reversing isometry of $d\sigma^2$, such as the antipodal map, extends naturally to a diffeomorphism of $\phi\colon S^3 \to S^3$, which induces an isometry of ds^2. Let $c = a + b$, and take Perelman's metric on $\#^c\mathbb{C}P^2$. For b of the necks, we cut along a copy of S^3 and reglue with ϕ rather than the identity. Because ϕ reverses orientation, the resulting manifold is $\#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$. Because ϕ induces an isometry on $S^3 \times I$, the same metrics on the pieces extend smoothly over the gluing, completing the proof of the lemma. □

Let M^5 satisfy the hypotheses of Theorem A. By Lemma 1.3, M is Type III and by Theorem 1.11 M is the total space of infinitely many nonisomorphic principal S^1 bundles $\pi_k\colon M^5 \to B^4 = \#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}$. From the proof of Theorem 1.11 we see that the first Chern class of π_k is $2d_k$, where

$$d_k = (1 + 2k, 1, \dots, 1) \in H^2(\#^a\mathbb{C}P^2 \#^b \overline{\mathbb{C}P^2}, \mathbb{Z}) \cong \mathbb{Z}^{a+b}$$

for a certain infinite set of integers k.

Using the result of [20] (see Corollary 1.15) we see that since B admits a metric of positive Ricci curvature by Lemma 3.10, $\pi_k\colon M \to B$ is a principal S^1 bundle, and $\pi_1(M)$ is finite, then for each k M admits a metric g_k with $\mathrm{Ric}(g_k) > 0$ such that the S^1 action corresponding to the principal bundle $\pi_k\colon M \to B$ acts by isometries of g_k.

Using the Gysin sequence it follows that $H^4(M,\mathbb{R})=0$ and M, g_k and B satisfy the hypotheses of Theorem 3.5 with $g_M=g_k$, $d=d_k$, $\ell=2$ and $\overline{\nabla}$ any flat connection on the canonical bundle of the spin^c structure. By (3.6) we have

$$\eta(D^c_{g_k,\overline{\nabla}})=-\tfrac{1}{16}(\langle c_k^2,[B]\rangle+\operatorname{sign}B)=-\tfrac{1}{16}(\pm 4k^2\pm 4k+2\operatorname{sign}B),$$

using the fact that $\left\langle \frac{1}{3}p_1(TB),[B]\right\rangle=\langle L(TB),[B]\rangle$ is equal to the signature of B.

Thus $\eta(D^c_{g_k,\overline{\nabla}})$ is a nontrivial polynomial in k and takes on infinitely many values for the infinite set of integers k. Corollary 2.6 implies that $\mathfrak{M}_{\mathrm{Ric}>0}(M)$ has infinitely many components, completing the proof of Theorem A. □

Note that Corollary 2.6 also implies that $\mathfrak{M}_{\mathrm{scal}>0}(M)$ has infinitely many components.

4 Metric and connection

In this section we prove Theorem 3.1. We first set up notation for the tangent space to W. We consider D^2 to be the unit disc in $\mathbb{C}$. Let $\sigma\colon M\times D^2\to W$ be the quotient map so $\sigma(p,x)=[p,x]$. Then $\rho([p,x])=\pi(p)$. The metric g_M and the S^1 action induce an orthogonal splitting $T_pM=\overline{H}_p\oplus\overline{V}_p$ into horizontal space $\overline{H}_p$ and vertical space $\overline{V}_p$ of π. Define horizontal and vertical spaces of ρ to be

$$H_{[p,x]}=\sigma_*(\overline{H}_p\oplus\{0\})\quad\text{and}\quad V_{[p,x]}=\sigma_*(\{0\}\oplus T_xD^2)$$

for $p\in M$ and $x\in D^2$.

These are well defined since for $z\in S^1$, $\overline{H}_{zp}=z_*\overline{H}_p$ and $T_{zx}D^2=z_*T_xD^2$. One can use a local section of σ to see that $H_{[p,x]}$ and $V_{[p,x]}$ are smooth distributions on W. Note that $V_{[p,x]}$ is the tangent space to the fiber $\rho^{-1}(\pi(p))=\sigma(\{p\}\times D^2)$ and $T_{[p,x]}W=H_{[p,x]}\oplus V_{[p,x]}$. Away from the zero section of ρ, $V_{[p,x]}$ is spanned by

$$W_r=\sigma_*(0,\partial_r)\quad\text{and}\quad W_\theta=\sigma_*(0,\partial_\theta).$$

These are well-defined smooth vector fields since ∂_θ and ∂_r are S^1–invariant vector fields on D^2.

Fix $0<L<1$ and define a diffeomorphism

$$\tau\colon M\times[L,1]\hookrightarrow M\times D^2\xrightarrow{\sigma}W$$

of $M\times[L,1]$ to a collar neighborhood U of ∂W. Let t be the coordinate on $[L,1]$ and, in a slight abuse of notation, let $\mathrm{proj}_{U,M}\colon M\times[L,1]\to M$ be the projection. Thus,

$$\rho\circ\tau=\pi\circ\mathrm{proj}_{U,M},\quad \tau_*(\overline{H}_p\oplus\{0\})=H_{[p,x]},\quad \tau_*(0,\partial_t)=W_r.$$

Let

$$X^*(p)=\frac{d}{dt}\bigg|_{t=0}e^{it}\cdot p$$

be the action field of the S^1 action on M, which spans $\bar{V}_p$. Then, since $\sigma_*(X^*, \partial_\theta) = 0$,

$$\tau_*(X^*, 0) = -W_\theta.$$

Furthermore, $\tau|_{M\times\{1\}}$ identifies M and ∂W, sending $\bar{H}_p$ to $H_{[p,1]}$ and X^* to $-W_\theta$.

We keep track of the maps in the following diagram:

$$\begin{tikzcd}
 & M \times D^2 \arrow[d, "\sigma"] \\
M \times I \arrow[ur, hook] \arrow[r, "\tau"] \arrow[d, "\mathrm{proj}_{U,M}"'] & W \arrow[d, "\rho"] \\
M \arrow[r, "\pi"] & B
\end{tikzcd}$$

To construct g_W and ∇ we will use two smooth functions on the interval $[0, 1]$. Let $f_1 : [0, 1] \to [0, 1]$ be a smooth monotone function which is 0 in a neighborhood of 0, and 1 in a neighborhood of $[L, 1]$.

For a constant $\epsilon > 0$, let

$$f_2(r) = -\frac{1}{2}\int_0^r f_1(t)\,dt - \epsilon r^3 + r.$$

One easily sees that $f_2 > 0$ on $(0, 1]$ for small ϵ.

4.1 Metric

We define a Riemannian metric at a point $(p, (r, \theta)) \in M \times D^2$, where r and θ are polar coordinates on D^2, by

$$g_{M\times D^2}(p, (r, \theta)) = g_M(p) + \epsilon^2 |X^*(p)|^2_{g_M}\left(dr^2 + \frac{f_2(r)^2}{1 - \epsilon^2 f_2(r)^2}\, d\theta^2\right).$$

By converting to Cartesian coordinates on D^2, one sees that $g_{M\times D^2}$ is smooth as long as

$$\frac{1}{r^4}\left(\frac{f_2^2}{1 - \epsilon^2 f_2^2} - r^2\right)$$

is a smooth function of $r \in [0, 1]$. This is easily seen to hold since for r near 0, $f_2(r) = r - \epsilon r^3$. Since $g_{M\times D^2}$ is invariant under the diagonal action of S^1 on $M \times D^2$, it induces a metric g_W on W such that $g_{M\times D^2}$ and g_W make σ into a Riemannian submersion. Similarly, let g_B be the metric on B such that g_M and g_B make π into a Riemannian submersion.

Lemma 4.1 *The metrics g_W and g_B make ρ into a Riemannian submersion.*

Proof With respect to $g_{M\times D^2}$, $\bar{H}_p \oplus \{0\}$ is orthogonal to X^* and TD^2. Thus $\bar{H}_p \oplus \{0\}$ is orthogonal to the vertical space of σ, which is spanned by (X^*, ∂_θ), and to the horizontal projection of TD^2 as well. It follows that with respect to g_W, $H_{[p,x]}$ is orthogonal to $V_{[p,x]}$ and is the horizontal space of ρ. Finally, we have

$$g_W|_{H_{[p,x]}} \cong g_{M\times D^2}|_{\bar{H}_p \oplus \{0\}} \cong g_M|_{\bar{H}_p} \cong g_B|_{T_{\pi(p)}B}. \qquad \square$$

We first describe the induced metric on the D^2 fibers of ρ.

Lemma 4.2
$$g_W|_{\rho^{-1}(\pi(p))} \cong \epsilon^2 |X^*(p)|_{g_M}(dr^2 + f_2(r)^2\, d\theta^2).$$

Proof The restriction $\sigma|_{\{p\}\times D^2}\colon D^2 \to \rho^{-1}(\pi(p))$ is a diffeomorphism such that ∂_r and ∂_θ are mapped to W_r and W_θ, respectively. Since σ is a Riemannian submersion with vertical space generated by (X^*, ∂_θ), we calculate:

$$\begin{aligned}
|W_r|^2_{g_W} &= |(0,\partial_r)|^2_{g_{M\times D^2}} = \epsilon^2 |X^*|^2_{g_M},\\
|W_\theta|^2_{g_W} &= |(0,\partial_\theta)|^2_{g_{M\times D^2}} - \frac{\langle (0,\partial_\theta),(X^*,\partial_\theta)\rangle^2_{g_{M\times D^2}}}{\langle (X^*,\partial_\theta),(X^*,\partial_\theta)\rangle_{g_{M\times D^2}}}\\
&= \epsilon^2 |X^*|^2_{g_M}\left(\frac{f_2(r)^2}{1-\epsilon^2 f_2(r)^2}\right)\\
&\qquad - \epsilon^4 |X^*|^4_{g_M}\left(\frac{f_2(r)^2}{1-\epsilon^2 f_2(r)^2}\right)^2 \left(\frac{1}{|X^*|^2_{g_M} + \epsilon^2 |X^*|^2_{g_M}\big(f_2(r)^2/(1-\epsilon^2 f_2(r)^2)\big)}\right)\\
&= \epsilon^2 |X^*|^2_{g_M} f_2(r)^2,\\
\langle W_r, W_\theta\rangle_{g_W} &= \langle (0,\partial_r),(0,\partial_\theta)\rangle_{g_{M\times D^2}} = 0.
\end{aligned}$$
□

We next modify g_W to have the desired product structure near ∂W. We use a technique of Wraith, which allows deformations of metrics with positive mean curvature at the boundary.

Lemma 4.3 *∂W has positive mean curvature with respect to an inward normal vector.*

Proof Let $\bar{X}_i$ be local S^1–invariant vector fields extending an orthonormal frame of $\bar{H}_p$, and define $X_i = \sigma_*(\bar{X}_i, 0)$. At a point $[p,1]$,
$$\left\{X_i, \frac{1}{\epsilon |X^*|_{g_M} f_2} W_\theta\right\}$$
is an orthonormal basis of $T\partial W$ and
$$-\frac{1}{\epsilon |X^*|_{g_M}} W_r$$
is an inward-pointing unit normal vector. Since
$$[X_i, W_r] = [\sigma_*(\bar{X}_i,0), \sigma_*(0,\partial_r)] = \sigma_*[(\bar{X}_i,0),(0,\partial_r)] = 0$$
and $|X_i| = 1$,
$$\frac{1}{\epsilon |X^*|_{g_M}} \langle \nabla_{X_i} X_i, -W_r\rangle = \frac{1}{\epsilon |X^*|_{g_M}} \langle X_i, \nabla_{X_i} W_r\rangle = \frac{1}{\epsilon |X^*|_{g_M}} \langle X_i, \nabla_{W_r} X_i\rangle = 0.$$
Thus
$$\frac{1}{\epsilon^3 |X^*|^3_{g_M} f_2(1)^2} \langle \nabla_{W_\theta} W_\theta, -W_r\rangle = \frac{1}{2\epsilon^3 |X^*|^3_{g_M} f_2(1)^2} W_r(|W_\theta|^2) = \frac{f_2'(1)}{\epsilon |X^*|_{g_M} f_2(1)}.$$

Evaluating that quantity at $r = 1$ we see that the mean curvature is

$$\frac{\frac{1}{2} - 3\epsilon}{\epsilon |X^*|_{g_M} f_2(1)} > 0$$

for sufficiently small ϵ. $\square$

We see that $g_W|_{\partial W}$ is obtained from g_M by shrinking the S^1 fibers of π, a process which preserves positive scalar curvature.

Lemma 4.4 *There exists a smooth path of metrics $g_M(s)$ on M, with $s \in [\epsilon^2 f_2(1)^2, 1]$, such that $g_M(\epsilon^2 f_2(1)^2) = g_W|_{\partial W}$, $g_M(1) = g_M$ and $\mathrm{scal}(g_M(s)) > 0$ for all s.*

Proof We recall that $\tau|_{M\times\{1\}} \colon M \to \partial W$ is a diffeomorphism. We see that

$$((\tau|_{M\times\{1\}})^* g_W)|_{\bar{H}_p} = g_W|_{H_{[p,1]}} = g_M|_{\bar{H}_p}$$

and

$$|X^*(p)|^2_{(\tau|_{M\times\{1\}})^* g_W} = |W_\theta([p,1])|^2_{g_W} = \epsilon^2 f_2(1)^2 |X^*(p)|^2_{g_M}.$$

Thus, defining

$$g_M(s) = g_M|_{\bar{H}_p} + s g_M|_{\bar{V}_p},$$

we have, for ϵ small enough, that

$$\epsilon^2 f_2(1)^2 < 1, \quad g_M(\epsilon^2 f_2(1)^2) = (\tau|_{M\times\{1\}})^* g_W \quad \text{and} \quad g_M(1) = g_M.$$

Since the metric is not changing on the horizontal space of π, each $g_M(s)$ makes π into a Riemannian submersion with g_B. The O'Neil formula [6] then implies

$$\mathrm{scal}(g_M(s)) = \mathrm{scal}(g_B) - s|A_\pi|^2 - |T_\pi|^2 - |N_\pi|^2 - 2\delta N_\pi \geq \mathrm{scal}(g_M) > 0,$$

where A_π, T_π and N_π are the tensors defined for the Riemannian submersion π with respect to g_M. $\square$

Use the normal exponential map from ∂W to define a collar neighborhood $V \cong M \times [0, N]$, where $t \in [0, N]$ is the distance to ∂W. We choose N small such that $V \subset U$. Using this identification, g_W has the form

$$g_W = g(t) + dt^2,$$

where $g(t) = g_W|_{M\times t}$ is a smooth path of metrics on M. Since $g(0) = g_W|_{\partial W}$ has positive scalar curvature, we can choose N small such that $\mathrm{scal}(g(t)) > 0$ for all $t \in [0, N]$.

Lemma 4.5 *We can alter g_W inside of V such that it is product-like near ∂W with $g_W|_{\partial W} = g_M$ and $\mathrm{scal}(g_W|_V) > 0$.*

Proof We use the paths $g_M(s)$ and $g(s)$ and the following lemma from [45] to replace g_W near the boundary with a product metric restricting to g_M at the boundary.

Lemma 4.6 [45] *Let $g(t) + dt^2$ be a metric of positive scalar curvature on $M \times [0, N]$ such that* $\mathrm{scal}(g(t)) > 0$ *and $M \times \{0\}$ has positive mean curvature with respect to the inward normal vector ∂_t. Let $\overline{g}(t)$ be a smooth path of metrics on M such that* $\mathrm{scal}(\overline{g}(t)) > 0$ *for $t \in [0, N]$ and $\overline{g}(t) = g(t)$ for t in a neighborhood of N. Then there exists a function $\beta\colon [0, N] \to \mathbb{R}_+$ such that $\beta = 1$ for t in a neighborhood of N, $\beta = \beta(0)$ is constant for t in a neighborhood of 0, and $\overline{g}(t) + \beta(t)\, dt^2$ has positive scalar curvature.*

To define our replacement path $\overline{g}$, we define two smooth functions:

$\chi_1\colon \left[0, \tfrac{1}{2}N\right] \to [\epsilon^2 f_2(1)^2, 1]$ such that $\chi_1(t) = 1$ for t near 0 and $\chi_1(t) = \epsilon^2 f_2(1)^2$ for t near $\tfrac{1}{2}N$,

$\chi_2\colon \left[\tfrac{1}{2}N, N\right] \to [0, 1]$ such that $\chi_2(t) = 0$ for t near $\tfrac{1}{2}N$ and $\chi_2(t) = t$ for t near N.

We then define a smooth path of metrics

$$\overline{g}(t) = \begin{cases} g_M \circ \chi_1(t) & \text{if } t \in \left[0, \tfrac{1}{2}N\right], \\ g \circ \chi_2(t) & \text{if } t \in \left[\tfrac{1}{2}N, N\right]. \end{cases}$$

By Lemma 4.4 and the definition of g, $\mathrm{scal}(\overline{g}(t)) > 0$ for all t. Then Lemmas 4.3 and 4.6 imply that $\overline{g}(t) + \beta(t)\, dt^2$ has positive scalar curvature for the function $\beta(t)$ given by Lemma 4.6. For t near N, $\overline{g}(t) = g(t)$ and $\beta(t) = 1$, so $\overline{g}(t) + \beta(t)\, dt^2 = g_W$. Thus replacing $g_W|_V$ with this metric results in a new smooth metric, for which we reuse the notation g_W. Since $\overline{g}(t) = g$ and $\beta(t)$ is constant for t near 0, $\overline{g}(t) + \beta(t)\, dt^2$ has the desired product structure (3.3). This proves Lemma 4.5. □

4.2 Connection

Let $\beta \in \Omega^2(B)$ represent the image of ℓd in $H^2(B, \mathbb{R})$. The Gysin sequence for an S^1 bundle shows that $\pi^* \ell d = 0$, so we can choose $\alpha \in \Omega^1(M)$ such that $\pi^*\beta = d\alpha$. Since $\pi^*\beta$ is S^1–invariant, we can choose α to be S^1–invariant by averaging.

Lemma 4.7
$$\alpha(X^*) = -\frac{1}{2\pi}.$$

Proof Let $\Phi \in \Omega^2(W)$ be a Thom form of the disc bundle $\rho\colon W \to B$. Since

$$[\Phi] \mapsto \rho^* \ell d$$

under the long exact sequence map $H^2(W, \partial W) \to H^2(W)$, we have

$$\rho^*\beta - \Phi = d\overline{\alpha}$$

for some $\overline{\alpha} \in \Omega^1(W)$. Since Φ vanishes near ∂W,

$$d\overline{\alpha}|_M = \rho^*\beta|_M = \pi^*\beta = d\alpha.$$

Since $\pi_1(M)$ is finite, $\overline{\alpha}|_M - \alpha$ is exact. By the defining property of the Thom form, for any point $q \in B$, $\int_{\rho^{-1}(q)} \Phi = 1$. We use Stokes' theorem to compute

$$-1 = \int_{\rho^{-1}(q)} \rho^* \beta - \Phi = \int_{\rho^{-1}(q)} d\overline{\alpha} = \int_{\pi^{-1}(q)} \overline{\alpha} = \int_{\pi^{-1}(q)} \alpha = 2\pi\alpha(X^*). \qquad \square$$

We next construct a form $\gamma \in \Omega^1(W)$ extending $2\pi\alpha/\ell$. We first define a form $\overline{\gamma} \in \Omega(M \times D^2)$. At $(p, x) \in M \times D^2$, $x \neq 0$, set

$$\overline{\gamma}|_{\bar{H}_p \times \{0\}} = \frac{2\pi}{\ell}\alpha_{\bar{H}_p}, \quad \overline{\gamma}(X^*, 0) = -\frac{f_1(r)}{\ell}, \quad \overline{\gamma}(0, \partial_r) = 0, \quad \overline{\gamma}(0, \partial_\theta) = \frac{f_1(r)}{\ell},$$

where r is the radial coordinate on D^2. This form extends smoothly to the origin of D^2 since f_1 is zero in a neighborhood of $r = 0$. Since r, $\bar{H}_p \oplus \{0\}$, α, ∂_r, ∂_θ and X^* are all preserved by the S^1 action, $\overline{\gamma}$ is S^1–invariant. The vertical space of σ is generated by (X^*, ∂_θ), and so $\overline{\gamma}$ vanishes on the vertical space. It follows that there is a unique form $\gamma \in \Omega(W)$ such that $\sigma^*\gamma = \overline{\gamma}$.

Lemma 4.8
$$\tau^*\gamma = \frac{2\pi}{\ell} \operatorname{proj}^*_{U,M} \alpha.$$

Proof Recall that $f_1(r) = 1$ for r in the image of τ and note that $\tau^*\gamma = (\sigma^*\gamma)|_{M \times [L,1]} = \overline{\gamma}|_{M \times [L,1]}$. Thus:

$$\begin{aligned}
\tau^*\gamma|_{\bar{H}_p \oplus \{0\}} &= \overline{\gamma}|_{\bar{H}_p \oplus \{0\}} = \frac{2\pi}{\ell}\alpha_{\bar{H}_p}, \\
\tau^*\gamma(X^*, 0) &= \overline{\gamma}(X^*, 0) = -\frac{f_1(r)}{\ell} = \frac{2\pi}{\ell}\alpha(X^*), \\
\tau^*\gamma(0, \partial_t) &= \ \overline{\gamma}(0, \partial_r) \ = 0 = \frac{2\pi}{\ell}\alpha(\operatorname{proj}_{M*}(0, \partial_t)).
\end{aligned} \qquad \square$$

Let λ_B be the complex line bundle with $c_1(\lambda_B) = d$. Given a differential form in the de Rahm cohomology class of $2\pi i$ times the first Chern class of a complex line bundle, there is a unitary connection on the line bundle whose curvature is that differential form. Thus, since β represents ℓd, let ∇_B be a unitary connection on λ_B with curvature

$$F^{\nabla_B} = \frac{2\pi i}{\ell}\beta.$$

We now define a connection on λ:

$$\nabla = \rho^*\nabla_B - i\gamma.$$

Lemma 4.9 *∇ is flat on U.*

Proof We need to show that $F^{\tau^*\nabla} = 0$. Using Lemma 4.8 it follows that

$$\tau^*\nabla = \tau^*\rho^*\nabla_B - i\tau^*\gamma = \operatorname{proj}^*_{U,M}\Big(\pi^*\nabla_B - \frac{2\pi i}{\ell}\alpha\Big),$$

and hence the curvature of the term in the parentheses is

$$\frac{2\pi i}{\ell}\pi^*\beta - \frac{2\pi i}{\ell}\, d\alpha = 0. \qquad \square$$

We finish the construction of ∇ by modifying it so that it is product-like near ∂W and restricts to $\overline{\nabla}$ at ∂W. Let $\mathrm{proj}^*_{V,M}\colon V \to M$ be the projection defined by the identification $V \cong M \times [0,N]$ from Section 4.1. Note that while $V \subset U$, $\mathrm{proj}_{V,M}$ and $\mathrm{proj}_{U,M}$ will not in general agree (the latter was defined independently of h, and the former using h). Since $V \subset U$, ∇ is flat on V. Since $\mathrm{proj}_{V,M}$ and the inclusion of $\partial W \cong M \times \{0\}$ are homotopy inverses, $\mathrm{proj}^*_{M,V}(\lambda|_M) = \lambda|_V$. Thus $\nabla|_V$ and $\mathrm{proj}^*_{V,M}\,\overline{\nabla}$ are both flat unitary connections on $\lambda|_V$ and

$$\mathrm{proj}^*_{V,M}(\overline{\nabla}) - \nabla|_V = i\delta$$

for some closed form $\delta \in \Omega^1(V)$. Since $\pi_1(V) = \pi_1(M)$ is finite, $\delta = df$ for a smooth function f on V. We modify f to a function $\bar{f}$ which is equal to f near $\partial W \cong M \times \{0\}$ and equal to 0 near $M \times \{N\}$. We then replace ∇ with $\nabla + i\,d\bar{f}$ on V. We see that

∇ is still smooth, flat on V, and near ∂W, $\nabla = \mathrm{proj}^*_{V,M}\,\overline{\nabla}$, satisfying (3.4).

4.3 Curvature

We complete the proof of Theorem 3.1 by showing that (3.2) holds. On V, ∇ is flat and by Lemma 4.5 $\mathrm{scal}(g_W) > 0$, so the inequality is satisfied. For the remainder of the proof we consider $W \backslash V$. Then $\mathrm{scal}(g_W)$ is given by Lemma 4.1 and the O'Neil formula for the scalar curvature of Riemannian submersion

$$\mathrm{scal}(g_W) = \mathrm{scal}(g_W|_{\rho^{-1}(\pi(p))}) + \mathrm{scal}(g_B) - |A_\rho|^2 - |T_\rho|^2 - |N_\rho|^2 - 2\delta N_\rho.$$

As $\epsilon \to 0$, $|A_\rho| \to 0$, while the final three terms remain constant. By Lemma 4.2,

$$\mathrm{scal}(g_W|_{\rho^{-1}(\pi(p))}) = -\frac{2}{\epsilon^2 |X^*|^2_{g_M}} \left(\frac{f_2''}{f_2}\right).$$

Therefore, as $\epsilon \to 0$,

$$\mathrm{scal}(g_W) = -\frac{2}{\epsilon^2 |X^*|^2_{g_M}} \left(\frac{f_2''}{f_2}\right) + O(1).$$

Let $\bar{X}_i$ be an orthonormal basis of $\bar{H}_p$ with respect to g_M. Let $X_i = \sigma_*(\bar{X}_i, 0)$. Then $\{X_i\}$ is an orthonormal basis of $H_{[p,x]}$ with respect to g_W outside of V. Away from the zero section,

$$\left\{\frac{1}{\epsilon |X^*|_{g_M}} W_r, \frac{1}{\epsilon |X^*|_{g_M} f_2} W_\theta\right\}$$

is an orthonormal basis of $V_{[p,x]}$. Neither the $\bar{X}_i$ nor ∇ depend on ϵ. Then as $\epsilon \to 0$, using (2.3),

$$|F^\nabla|_{g_M} \le \frac{1}{\epsilon^2 |X^*|^2_{g_M} f_2} |F^\nabla(W_r, W_\theta)| + \sum_i \frac{1}{\epsilon |X^*|_{g_M}} |F^\nabla(W_r, X_i)| + \frac{1}{\epsilon |X^*|_{g_M} f_2} |F^\nabla(W_\theta, X_i)| + O(1).$$

Lemma 4.10 *$F^\nabla(W_r, W_\theta) = -i f_1'(r)/\ell$ and $F^\nabla(W_r, X_i) = F^\nabla(W_\theta, X_i) = 0$.*

Proof Since $\rho_* W_r = \rho_* W_\theta = 0$,

$$F^\nabla(W_r, W_\theta) = -i\, d\gamma(W_r, W_\theta) = -i\, d\gamma(\sigma_*(0,\partial_r), \sigma_*(0,\partial_\theta))$$
$$= -i\sigma^* d\gamma((0,\partial_r),(0,\partial_\theta)) = -i\, d\bar{\gamma}((0,\partial_r),(0,\partial_\theta)) = -i\partial_r \bar{\gamma}(0,\partial_\theta) = -i\frac{f_1'(r)}{\ell}.$$

Similarly,

$$F^\nabla(W_r, X_i) = -i\, d\bar{\gamma}((0,\partial_r),(\bar{X}_i,0)) = -i\left(\partial_r\left(\frac{2\pi}{\ell}\alpha(\bar{X}_i)\right) - \bar{X}_i\left(\frac{f_1(r)}{\ell}\right)\right) = 0,$$
$$F^\nabla(W_\theta, X_i) = -i\, d\bar{\gamma}((0,\partial_\theta),(\bar{X}_i,0)) = -i\left(\partial_\theta\left(\frac{2\pi}{\ell}\alpha(\bar{X}_i)\right)\right) = 0. \qquad \square$$

Lemma 4.10 implies that as $\epsilon \to 0$,

$$\operatorname{scal}(g_W) - \ell|F^\nabla|_{g_M} = \frac{1}{\epsilon^2|X^*|^2_{g_M}}\left(\frac{-2f_2'' - f_1'}{f_2}\right) + O(1) = \frac{12}{\epsilon|X^*|^2_{g_M}}\left(\frac{r}{f_2}\right) + O(1).$$

From the definition of f_2 one sees that $r/f_2 \to 1$ as $r \to 0$. It follows that we can choose ϵ small enough that (3.2) holds, completing the proof of Theorem 3.1. $\square$

In [30, Lemma 4.2], Kreck and Stolz constructed positive scalar curvature metrics on associated disc bundles in order to calculate their invariant for spin manifolds with free S^1 actions. In their proof, they needed to assume that the S^1 orbits were geodesics. The metric g_W constructed in Theorem 3.1 generalizes their method to a free isometric S^1 action without the geodesic condition.

References

[1] **M F Atiyah**, **V K Patodi**, **I M Singer**, *Spectral asymmetry and Riemannian geometry, I*, Math. Proc. Cambridge Philos. Soc. 77 (1975) 43–69 MR Zbl

[2] **M F Atiyah**, **V K Patodi**, **I M Singer**, *Spectral asymmetry and Riemannian geometry, II*, Math. Proc. Cambridge Philos. Soc. 78 (1975) 405–432 MR Zbl

[3] **A Bahri**, **P Gilkey**, *The eta invariant,* Pin^c *bordism, and equivariant* Spin^c *bordism for cyclic 2–groups*, Pacific J. Math. 128 (1987) 1–24 MR Zbl

[4] **D Barden**, *Simply connected five-manifolds*, Ann. of Math. 82 (1965) 365–385 MR Zbl

[5] **V N Berestovskiĭ**, *Homogeneous Riemannian manifolds of positive Ricci curvature*, Mat. Zametki 58 (1995) 334–340 MR Zbl In Russian; translated in Math. Notes 58 (1995) 905–909

[6] **A L Besse**, *Einstein manifolds*, Ergebnisse der Math. (3) 10, Springer (1987) MR Zbl

[7] **B Botvinnik**, **P B Gilkey**, *The eta invariant and metrics of positive scalar curvature*, Math. Ann. 302 (1995) 507–517 MR Zbl

[8] **B Botvinnik**, **M G Walsh**, **D J Wraith**, *Homotopy groups of the observer moduli space of Ricci positive metrics*, Geom. Topol. 23 (2019) 3003–3040 MR Zbl

[9] **J-P Bourguignon**, *Une stratification de l'espace des structures riemanniennes*, Compositio Math. 30 (1975) 1–41 MR Zbl

[10] **B L Burdick**, *Ricci-positive metrics on connected sums of projective spaces*, Differential Geom. Appl. 62 (2019) 212–233 MR Zbl

[11] **D Corro**, **F Galaz-García**, *Positive Ricci curvature on simply-connected manifolds with cohomogeneity-two torus actions*, Proc. Amer. Math. Soc. 148 (2020) 3087–3097 MR Zbl

[12] **D Crowley**, **C M Escher**, *A classification of S^3–bundles over S^4*, Differential Geom. Appl. 18 (2003) 363–380 MR Zbl

[13] **A Dessai**, *On the moduli space of nonnegatively curved metrics on Milnor spheres*, preprint (2018) arXiv 1712.08821v2

[14] **A Dessai**, **D González-Álvaro**, *Moduli space of metrics of nonnegative sectional or positive Ricci curvature on homotopy real projective spaces*, Trans. Amer. Math. Soc. 374 (2021) 1–33 MR Zbl

[15] **A Dessai**, **S Klaus**, **W Tuschmann**, *Nonconnected moduli spaces of nonnegative sectional curvature metrics on simply connected manifolds*, Bull. Lond. Math. Soc. 50 (2018) 96–107 MR Zbl

[16] **S K Donaldson**, **P B Kronheimer**, *The geometry of four-manifolds*, Oxford Univ. Press (1990) MR Zbl

[17] **D G Ebin**, *The manifold of Riemannian metrics*, from "Global analysis" (S-S Chern, S Smale, editors), Proc. Sympos. Pure Math. XIV–XVI, Amer. Math. Soc., Providence, RI (1970) 11–40 MR Zbl

[18] **C Escher**, **W Ziller**, *Topology of non-negatively curved manifolds*, Ann. Global Anal. Geom. 46 (2014) 23–55 MR Zbl

[19] **H Geiges**, **C B Thomas**, *Contact topology and the structure of 5–manifolds with $\pi_1 = Z_2$*, Ann. Inst. Fourier (Grenoble) 48 (1998) 1167–1188 MR Zbl

[20] **P B Gilkey**, **J Park**, **W Tuschmann**, *Invariant metrics of positive Ricci curvature on principal bundles*, Math. Z. 227 (1998) 455–463 MR Zbl

[21] **M J Goodman**, *On the moduli spaces of metrics with nonnegative sectional curvature*, Ann. Global Anal. Geom. 57 (2020) 305–320 MR Zbl

[22] **K Grove**, **W Ziller**, *Curvature and symmetry of Milnor spheres*, Ann. of Math. 152 (2000) 331–367 MR Zbl

[23] **K Grove**, **W Ziller**, *Cohomogeneity one manifolds with positive Ricci curvature*, Invent. Math. 149 (2002) 619–646 MR Zbl

[24] **K Grove**, **W Ziller**, *Lifting group actions and nonnegative curvature*, Trans. Amer. Math. Soc. 363 (2011) 2865–2890 MR Zbl

[25] **I Hambleton**, **M Kreck**, **P Teichner**, *Nonorientable 4–manifolds with fundamental group of order 2*, Trans. Amer. Math. Soc. 344 (1994) 649–665 MR Zbl

[26] **I Hambleton**, **Y Su**, *On certain 5–manifolds with fundamental group of order 2*, Q. J. Math. 64 (2013) 149–175 MR Zbl

[27] **R C Kirby**, **L R Taylor**, Pin *structures on low-dimensional manifolds*, from "Geometry of low-dimensional manifolds, II" (S K Donaldson, C B Thomas, editors), London Math. Soc. Lecture Note Ser. 151, Cambridge Univ. Press (1990) 177–242 MR Zbl

[28] **M Kreck**, **S Stolz**, *A diffeomorphism classification of 7–dimensional homogeneous Einstein manifolds with $\mathrm{SU}(3) \times \mathrm{SU}(2) \times U(1)$–symmetry*, Ann. of Math. 127 (1988) 373–388 MR Zbl

[29] **M Kreck**, **S Stolz**, *Some nondiffeomorphic homeomorphic homogeneous 7–manifolds with positive sectional curvature*, J. Differential Geom. 33 (1991) 465–486 MR Zbl

[30] **M Kreck**, **S Stolz**, *Nonconnected moduli spaces of positive sectional curvature metrics*, J. Amer. Math. Soc. 6 (1993) 825–850 MR Zbl

[31] **H B Lawson, Jr**, **M-L Michelsohn**, *Spin geometry*, Princeton Mathematical Series 38, Princeton Univ. Press (1989) MR Zbl

[32] **S López de Medrano**, *Involutions on manifolds*, Ergebnisse der Math. (3) 59, Springer (1971) MR Zbl

[33] **J C Nash**, *Positive Ricci curvature on fibre bundles*, J. Differential Geometry 14 (1979) 241–254 MR Zbl

[34] **G Perelman**, *Construction of manifolds of positive Ricci curvature with big volume and large Betti numbers*, from "Comparison geometry" (K Grove, P Petersen, editors), Math. Sci. Res. Inst. Publ. 30, Cambridge Univ. Press (1997) 157–163 MR Zbl

[35] **L J Schwachhöfer**, **W Tuschmann**, *Metrics of positive Ricci curvature on quotient spaces*, Math. Ann. 330 (2004) 59–91 MR Zbl

[36] **C Searle**, **F Wilhelm**, *How to lift positive Ricci curvature*, Geom. Topol. 19 (2015) 1409–1475 MR Zbl

[37] **J-P Sha**, **D Yang**, *Positive Ricci curvature on the connected sums of $S^n \times S^m$*, J. Differential Geom. 33 (1991) 127–137 MR Zbl

[38] **J-P Sha**, **D Yang**, *Positive Ricci curvature on compact simply connected 4–manifolds*, from "Differential geometry: Riemannian geometry, III" (R Greene, S T Yau, editors), Proc. Sympos. Pure Math. 54, Amer. Math. Soc., Providence, RI (1993) 529–538 MR Zbl

[39] **S Smale**, *On the structure of 5–manifolds*, Ann. of Math. 75 (1962) 38–46 MR Zbl

[40] **R E Stong**, *Relations among characteristic numbers, I*, Topology 4 (1965) 267–281 MR Zbl

[41] **M Y Wang**, **W Ziller**, *Einstein metrics on principal torus bundles*, J. Differential Geom. 31 (1990) 215–248 MR Zbl

[42] **J Wermelinger**, *Moduli space of metrics of nonnegative sectional or positive Ricci curvature on Brieskorn quotients in dimension* 5, preprint (2020) arXiv 2006.13697

[43] **D Wraith**, *Exotic spheres with positive Ricci curvature*, J. Differential Geom. 45 (1997) 638–649 MR Zbl

[44] **D J Wraith**, *New connected sums with positive Ricci curvature*, Ann. Global Anal. Geom. 32 (2007) 343–360 MR Zbl

[45] **D J Wraith**, *On the moduli space of positive Ricci curvature metrics on homotopy spheres*, Geom. Topol. 15 (2011) 1983–2015 MR Zbl

Department of Mathematics, Colby College
Waterville, ME, United States

mjgoodma@colby.edu

Proposed: Tomasz S Mrowka
Seconded: Tobias H Colding, David Fisher

Received: 4 December 2020
Revised: 1 April 2022

msp

Geometry & Topology **28**:3 (2024) 1099–1112
DOI: 10.2140/gt.2024.28.1099
Published: 10 May 2024

Riemannian manifolds with entire Grauert tube are rationally elliptic

XIAOYANG CHEN

It was conjectured by Bott, Grove and Halperin that a compact simply connected Riemannian manifold M with nonnegative sectional curvature is rationally elliptic. We confirm this conjecture under the stronger assumption that M has entire Grauert tube, ie M is a real-analytic Riemannian manifold that has a unique adapted complex structure defined on the whole tangent bundle TM. Our result also provides a strong topological obstruction to the existence of an entire Grauert tube.

53C20

1 Introduction

The following conjecture is a central problem in the study of Riemannian manifolds with nonnegative sectional curvature; see Berger and Bott [2] and Grove and Halperin [7].

Conjecture (Bott–Grove–Halperin) A compact simply connected Riemannian manifold M with nonnegative sectional curvature is rationally elliptic.

Here M is said to be rationally elliptic if and only if it has finite-dimensional rational homotopy groups, ie all but finitely many homotopy groups of M are finite; otherwise M is said to be rationally hyperbolic. It is a well-known simple consequence of Sullivan's minimal model theory [15] that M being rationally elliptic is equivalent to polynomial growth of the sequence of Betti numbers of its based loop space ΩM relative to rational coefficient. If M is rationally elliptic, then there are severe topological restrictions of M. For example, M has nonnegative Euler characteristic number and $\dim H_*(M,\mathbb{Q}) \le 2^n$; see Félix, Halperin and Thomas [5] and Grove and Halperin [7].

It is known that compact simply connected homogeneous spaces and cohomogeneity one-manifolds are rationally elliptic; see Grove and Halperin [8]. Grove, Wilking and Yeager [9] confirmed the Bott–Grove–Halperin conjecture under the additional assumption that M supports an isometric action with principal orbits of codimension two.

In this paper we confirm the Bott–Grove–Halperin conjecture under the stronger assumption that M has entire Grauert tube:

Theorem 1.1 *Let (M, g) be an n–dimensional compact simply connected real-analytic Riemannian manifold that has entire Grauert tube, then M is rationally elliptic.*

Remark 1.2 In fact, our proof shows that M is topologically elliptic, ie the Betti numbers of its loop space relative to any field of coefficients grow at most polynomially.

Here (M, g) is said to be real-analytic if M is a real-analytic manifold with a real-analytic Riemannian metric g. Then there is a unique adapted complex structure defined on $T^R M = \{v \in TM \mid g(v,v) < R^2\}$ for some $R > 0$; see Guillemin and Stenzel [10], Lempert and Szőke [12] and Szőke [16]. When $R = \infty$, then M is said to have entire Grauert tube. It was shown in [12] that a Riemannian manifold with entire Grauert tube has nonnegative sectional curvature. Hence Theorem 1.1 gives a partial answer to the Bott–Grove–Halperin conjecture. On the other hand, it also provides a strong topological obstruction to the existence of an entire Grauert tube.

Aguilar [1] showed that the quotient of a Riemannian manifold with entire Grauert tube by a group of isometries acting freely also has entire Grauert tube. All known manifolds with entire Grauert tube are obtained by Aguilar's construction: starting with a compact Lie group with a bi-invariant metric, or the product of such a group with Euclidean space, one takes the quotient by some group of isometries acting freely. Such quotient manifolds include almost all closed manifolds which are known to have Riemannian metrics with nonnegative sectional curvature.

It was conjectured by Hopf that the Euler characteristic number of a compact Riemannian manifold with nonnegative sectional curvature is nonnegative. The following corollary settles this conjecture under the stronger assumption that M has entire Grauert tube.

Corollary 1.3 *Let M be a compact Riemannian manifold with entire Grauert tube. Then M has nonnegative Euler characteristic number.*

Proof If M has finite fundamental group, then its universal cover $\widetilde{M}$ with the induced Riemannian metric also has entire Grauert tube. By Theorem 1.1, the Euler characteristic number of $\widetilde{M}$ is nonnegative. Hence M has nonnegative Euler characteristic number. If M has infinite fundamental group, as M has nonnegative sectional curvature, then the Euler characteristic number of M is zero; see Cheeger and Gromoll [4]. □

A related conjecture proposed by Totaro [17] predicts that a compact Riemannian manifold M with nonnegative sectional curvature has a good complexification, ie M is diffeomorphic to a smooth affine algebraic variety U over the real numbers such that the inclusion $U(\mathbb{R}) \to U(\mathbb{C})$ is a homotopy equivalence. The Euler characteristic number of a compact manifold which has a good complexification is also nonnegative. Also, a conjecture by Burns [3] predicts that for every compact Riemannian manifold M with entire Grauert tube, the complex manifold TM is an affine algebraic variety in a natural way. If this is correct, the complex manifold TM would be a good complexification of M in the above sense. Both conjectures of Totaro and Burns are still open.

The proof of Theorem 1.1 is based on the counting function introduced in Berger and Bott [2], Gromov [6] and Paternain [14]. For $x \in M$ and each $T > 0$, let

$$D_T := \{v \in T_x M \mid g(v,v) \le T^2\}$$

be the disk of radius T in $T_x M$. Define the counting function $n_T(x,y)$ by

$$n_T(x,y) := \sharp((\exp_x)^{-1}(y) \cap D_T).$$

In other words, $n_T(x,y)$ counts the number of geodesic arcs joining x to y with length $\le T$. When M is simply connected, then we have the crucial inequality

$$\sum_{j=0}^{k-1} \dim H_j(\Omega M, F) \le \frac{1}{\mathrm{Vol}_g(M)} \int_M n_{Ck}(x,y)\, dy, \tag{1-1}$$

where C is a positive constant independent of k and F is any field of coefficients; see [6; 14].

For any $x \in M$, Berger and Bott [2] proved that $\int_M n_T(x,y)\, dy$ can be computed by Jacobi fields on M; see also Paternain [14]. Precisely, they showed that

$$\int_M n_T(x,y)\, dy = \int_0^T d\sigma \int_{\mathbb{S}} \sqrt{\det\bigl(g(J_j(\sigma), J_k(\sigma))\bigr)_{j,k=1,2,\dots,n-1}}\, d\theta, \tag{1-2}$$

where $\mathbb{S}$ is the unit sphere of $T_x M$. Moreover, the J_j for $j = 1, 2, \dots, n-1$ are Jacobi fields along the unique geodesic γ determined by $\theta \in \mathbb{S}$ (ie $\gamma(0) = x$, $\gamma'(0) = \theta$) with initial conditions

$$J_j(0) = 0, \quad J_j'(0) = v_j,$$

where the v_j for $j = 1, 2, \dots, n-1$ form an orthonormal basis of $T_\theta \mathbb{S}$.

If (M, g) has entire Grauert tube, the right-hand side in (1-2) can be further described by a matrix valued holomorphic function on the upper half plane. Applying Fatou's representation theorem to this function, we will show that $\int_M n_T(x,y)\, dy$ is a polynomial function of T. When M is simply connected, it follows that $\sum_{j=0}^{k-1} \dim H_j(\Omega M, F)$ has polynomial growth for any field of coefficients. Hence M is topologically elliptic.

We finally mention that based on an iterated use of the Rauch comparison theorem for Jacobi fields, an estimate for the Betti numbers of ΩM for manifolds with $0 < \delta \le \sec M \le 1$ was derived in [2]. Although the estimate is given in terms of the pinching constant δ, its growth rate is exponential.

Acknowledgements The author is partially supported by NSFC 12171364 and 23JC1403600 (project title *On the topology of almost nonnegatively curved manifolds*). He would like to thank the referee for helpful suggestions.

2 Vertical and horizontal subbundles

In this section we recall some basic facts on the geometry of the tangent bundle TM. For more details, see [14]. Let $\pi: TM \to M$ be the canonical projection, ie if $\theta = (x, v) \in TM$, then $\pi(\theta) = x$. There exists a canonical subbundle of TTM, called the vertical subbundle, whose fiber at θ is given by the tangent vectors of curves $\sigma: (-\epsilon, \epsilon) \to TM$ of the form: $\sigma(t) = (x, v + t\omega)$, where $\omega \in T_x M$. In other words,

$$V(\theta) = \ker((\pi_*)_\theta).$$

Suppose that M is endowed with a Riemannian metric g. We shall define the connection map

$$K: TTM \to TM$$

as follows: let $\xi \in T_\theta TM$ and $z: (-\epsilon, \epsilon) \to TM$ be an adapted curve to ξ, that is, with initial conditions

$$z(0) = \theta, \quad z'(0) = \xi.$$

Such a curve gives rise to a curve $\alpha: (-\epsilon, \epsilon) \to M$ given by $\alpha := \pi \circ z$, and a vector field Z along α; equivalently, $z(t) = (\alpha(t), Z(t))$. Define

$$K_\theta(\xi) := (\nabla_\alpha Z)(0) = \lim_{t \to 0} \frac{(P_t)^{-1} Z(t) - Z(0)}{t},$$

where $P_t: T_x M \to T_{\alpha(t)} M$ is the linear isomorphism defined by the parallel transport along α. The horizontal subbundle is the subbundle of TTM whose fiber at θ is given by

$$H(\theta) = \ker K_\theta.$$

Another equivalent way of constructing the horizontal subbundle is by means of the horizontal lift

$$L_\theta: T_x M \to T_\theta TM,$$

which is defined as follows. Let $\theta = (x, v)$. Given $\omega \in T_x M$ and $\alpha: (-\epsilon, \epsilon) \to M$ an adapted curve of ω, ie $\alpha(0) = x$, $\alpha'(0) = \omega$, let $Z(t)$ be the parallel transport of v along α and $\sigma: (-\epsilon, \epsilon) \to TM$ be the curve $\sigma(t) = (\alpha(t), Z(t))$. Then

$$L_\theta(w) = \sigma'(0) \in T_\theta TM.$$

Proposition 2.1 *K_θ and L_θ have the following properties*:

$$(\pi_*)_\theta \circ L_\theta = \mathrm{Id} \quad \text{and} \quad K_\theta \circ i_* = \mathrm{Id},$$

where $i: T_x M \to TM$ is the inclusion map. Moreover,

$$T_\theta TM = H(\theta) \oplus V(\theta)$$

and the map $j_\theta: T_\theta TM \to T_x M \times T_x M$ given by

$$j_\theta(\xi) = ((\pi_*)_\theta(\xi), K_\theta(\xi))$$

is a linear isomorphism.

For each $\theta \in TM$, there is a unique geodesic γ_θ in M with initial condition θ. Let $\xi \in T_\theta TM$ and $z\colon (-\epsilon, \epsilon) \to TM$ be an adapted curve to ξ, that is, with initial conditions

$$z(0) = \theta, \quad z'(0) = \xi.$$

Then the map $(s,t) \mapsto \pi \circ \phi_t(z(s))$ gives rise to a variation of γ_θ. Here $\pi\colon TM \to M$ is the projection map and ϕ_t is the geodesic flow of TM. The curves $t \mapsto \pi \circ \phi_t(z(s))$ are geodesics and therefore the corresponding variational vector fields $J_\xi := (\partial/\partial s)|_{s=0}\, \pi \circ \phi_t(z(s))$ is a Jacobi field with initial conditions

$$J_\xi(0) = (\pi_*)_\theta(\xi), \quad J'_\xi(0) = K_\theta(\xi).$$

3 Adapted complex structure on the tangent bundle

In this section we describe the adapted complex structure on the tangent bundle. Let (M, g) be a compact smooth Riemannian manifold. Then $TM \setminus M$ carries a natural foliation by Riemannian surfaces defined as follows. For $\tau \in \mathbb{R}$ denote by $N_\tau\colon TM \to TM$ the smooth mapping defined by multiplication by τ in the fibers. If $\gamma\colon \mathbb{R} \to M$ is a geodesic, define an immersion $\phi_\gamma\colon \mathbb{C} \to TM$ by

$$\phi_\gamma(\sigma + i\tau) = N_\tau \gamma'(\sigma).$$

If for two geodesics γ and δ, it holds that $\phi_\gamma(\mathbb{C} \setminus \mathbb{R})$ and $\phi_\delta(\mathbb{C} \setminus \mathbb{R})$ intersect each other, then γ and δ are the same geodesic traversed with different velocities, hence $\phi_\gamma(\mathbb{C}) = \phi_\delta(\mathbb{C})$. Therefore the images of $\mathbb{C} \setminus \mathbb{R}$ under the mapping ϕ_γ defines a smooth foliation of $TM \setminus M$ by surfaces. Moreover, each leaf has complex structure that it inherits from $\mathbb{C}$ via ϕ_γ. The leaves along with their complex structure extend across M, but of course, on M the foliation $\mathcal{F}$ becomes singular.

Given $R > 0$, put

$$T^R M = \{v \in TM \mid g(v,v) < R^2\}.$$

A smooth complex structure on $T^R M$ will be called adapted if the leaves of the foliation $\mathcal{F}$ with the complex structure inherited from $\mathbb{C}$ are complex submanifolds of $T^R M$.

Theorem 3.1 [10; 12; 16] *Let M be a compact real-analytic manifold equipped with a real-analytic metric g. Then there exists some $R > 0$ such that $T^R M$ carries a unique adapted complex structure.*

When the adapted complex structure is defined on the whole tangent bundle, ie $R = \infty$, then M is said to have entire Grauert tube. It was shown in [12] that a Riemannian manifold with entire Grauert tube has nonnegative sectional curvature. The adapted complex structure on $T^R M$ can be described as follows. For this purpose let $\theta \in T^R M \setminus M$ and $x = \pi(\theta)$, where $\pi : TM \to M$ is the projection map. Let γ be a geodesic determined by θ. Choose tangent vectors $v_1, v_2, \dots, v_{n-1}$ such that $v_1, v_2, \dots, v_{n-1}, v_n := \gamma'(0)/|\gamma'(0)|$ form an orthonormal basis of $T_x M$.

Denote by L_θ the leaf of the foliation $\mathscr{F}$ passing through θ. A vector $\overline{\xi} \in T_\theta TM$ determines a vector field ξ (we call it the parallel vector field) along L_θ by defining it to be invariant under two semigroup actions. Namely, ξ is invariant under N_τ and the geodesic flow. For this parallel field ξ, we get that $\xi|_{\mathbb{R}}$ is a Jacobi field along γ.

Now choose a set of vectors $\overline{\xi}_1, \overline{\xi}_2, \dots, \overline{\xi}_n, \overline{\eta}_1, \overline{\eta}_2, \dots, \overline{\eta}_n \in T_\theta TM$ satisfying

$$(\pi_*)_\theta(\overline{\xi}_j) = v_j, \quad K_\theta(\overline{\xi}_j) = 0,$$
$$(\pi_*)_\theta(\overline{\eta}_j) = 0, \quad K_\theta(\overline{\eta}_j) = v_j.$$

Here $K : TTM \to TM$ is the connection map described in Section 2. Extend $\overline{\xi}_j$ and $\overline{\eta}_j$ to get parallel vector fields $\xi_1, \xi_2, \dots, \xi_n, \eta_1, \eta_2, \dots, \eta_n$ along L_θ. Then the Jacobi fields $\xi_1|_{\mathbb{R}}, \xi_2|_{\mathbb{R}}, \dots, \xi_n|_{\mathbb{R}}$ are linearly independent except on a discrete subset S_1 of $\mathbb{R}$. Hence there are smooth real-valued functions ϕ_{jk} defined on $\mathbb{R} \setminus S_1$ such that

$$\eta_k|_{\mathbb{R}} = \sum_{j=1}^{n} \phi_{jk} \xi_j|_{\mathbb{R}}.$$

From the presence of the adapted complex structure it follows that the functions ϕ_{jk} have meromorphic extension f_{jk} over the domain

$$D = \left\{ \sigma + i\tau \in \mathbb{C} \;\middle|\; |\tau| < \frac{R}{\sqrt{g(\theta,\theta)}} \right\}$$

such that for each pair j,k, the poles of f_{jk} lie on $\mathbb{R}$ and the matrix $\mathrm{Im}(f_{jk})|_{D\setminus\mathbb{R}}$ is invertible. Let $(e_{jk}) = (\mathrm{Im}\, f_{jk}(i))^{-1}$. Then the complex structure J satisfies

$$J\overline{\xi}_h = \sum_{k=1}^{n} e_{kh} \times \left[\overline{\eta}_k - \sum_{j=1}^{n} \mathrm{Re}\, f_{jk}(i) \overline{\xi}_j \right].$$

Remark 1 Because $\xi_1|_{\mathbb{R}}, \xi_2|_{\mathbb{R}}, \dots, \xi_{n-1}|_{\mathbb{R}}, \eta_1|_{\mathbb{R}}, \eta_2|_{\mathbb{R}}, \dots, \eta_{n-1}|_{\mathbb{R}}$ are normal Jacobi fields, while $\xi_n|_{\mathbb{R}}$ and $\eta_n|_{\mathbb{R}}$ are tangential Jacobi fields, for $1 \le j,k \le n-1$ we have

$$\phi_{nk} = \phi_{jn} \equiv 0, \quad f_{nk} = f_{jn} \equiv 0, \quad e_{nk} = e_{jn} \equiv 0.$$

Consider the n–tuples

$$\Xi = (\xi_1, \xi_2, \dots, \xi_n) \quad \text{and} \quad H = (\eta_1, \eta_2, \dots, \eta_n),$$

and holomorphic n–tuples

$$\Xi^{1,0} = (\xi_1^{1,0}, \xi_2^{1,0}, \dots, \xi_n^{1,0}) \quad \text{and} \quad H^{1,0} = (\eta_1^{1,0}, \eta_2^{1,0}, \dots, \eta_n^{1,0}),$$

where $\xi_j^{1,0} = \frac{1}{2}(\xi_j - iJ\xi_j)$ and J is the adapted complex structure. Then we have

$$H(\sigma) = \Xi(\sigma) f(\sigma) \quad \text{and} \quad H^{1,0}(\sigma + i\tau) = \Xi^{1,0}(\sigma + i\tau) f(\sigma + i\tau)$$

where

$$f(\sigma + i\tau) = (f_{jk}(\sigma + i\tau)) \quad \text{for } \sigma \in \mathbb{R} \setminus S_1 \text{ and } |\tau| < \frac{R}{\sqrt{g(\theta,\theta)}}.$$

The following facts are proved in [12; 16].

Proposition 3.2 (1) *The vectors $\xi_1^{1,0}, \xi_2^{1,0}, \dots \xi_n^{1,0}$ are linearly independent over $\mathbb{C}$ on $D \setminus \mathbb{R}$. The same is true for the vectors $\eta_1^{1,0}, \eta_2^{1,0}, \dots, \eta_n^{1,0}$.*

(2) *The $2n$ vectors ξ_j, η_k are linearly independent at points $\sigma + i\tau \in D \setminus \mathbb{R}$.*

Theorem 3.3 *The matrix-valued meromorphic function $f(\sigma + i\tau)$ is symmetric (as a matrix) and satisfies*

$$f(0) = 0, \quad f'(0) = \mathrm{Id}.$$

Moreover, if $\sigma + i\tau \in D$ with $\tau > 0$, then $\operatorname{Im} f(\sigma + i\tau)$ is a symmetric, positive definite matrix.

4 Growth rate of counting functions

In this section we prove Theorem 1.1.

Let M be an n–dimensional compact manifold endowed with a Riemannian metric g. For $x \in M$ and each $T > 0$, let

$$D_T := \{v \in T_x M \mid g(v, v) \le T^2\}$$

be the disk of radius T in $T_x M$. Define the counting function $n_T(x, y)$ by

$$n_T(x, y) := \sharp\big((\exp_x)^{-1}(y) \cap D_T\big).$$

In other words, $n_T(x, y)$ counts the number of geodesic arcs joining x to y with length $\le T$.

The following theorems proved in [2; 6; 14] will be crucial for us.

Theorem 4.1 *We have*

$$\int_M n_T(x, y)\, dy = \int_0^T d\sigma \int_{\mathbb{S}} \sqrt{\det\big(g(J_j(\sigma), J_k(\sigma))\big)_{j,k=1,2,\dots,n-1}}\, d\theta, \tag{4-1}$$

where $\mathbb{S}$ is the unit sphere of $T_x M$. Moreover, J_j for $j = 1, 2, \dots, n-1$ are Jacobi fields along the unique geodesic γ determined by $\theta \in \mathbb{S}$ (ie $\gamma(0) = x$ and $\gamma'(0) = \theta$) *with initial conditions*

$$J_j(0) = 0, \quad J_j'(0) = v_j,$$

where the v_j with $j = 1, 2, \dots, n-1$ form an orthonormal basis of $T_\theta \mathbb{S}$.

Theorem 4.2 *Suppose that M is an n–dimensional compact simply connected manifold endowed with a Riemannian metric g. Then*

$$\sum_{j=0}^{k-1} \dim H_j(\Omega M, F) \le \frac{1}{\mathrm{Vol}_g(M)} \int_M n_{Ck}(x, y)\, dy, \tag{4-2}$$

where C is a positive constant independent of k and F is any field of coefficients.

Remark 4.3 The assumption that M is simply connected in Theorem 4.2 is essential.

When M has entire Grauert tube, we will see that the right-hand side in Theorem 4.1 can be further described by a matrix-valued holomorphic function on the upper half-plane. Applying Fatou's representation theorem to this function, we will derive that $\int_M n_T(x,y)\,dy$ has polynomial growth and hence M is topologically elliptic.

Now we give the details of the proof. Let $\mathbb{S}$ be the unit sphere of T_xM and γ the unique geodesic determined by $\theta \in \mathbb{S}$, ie $\gamma(0)=x$, $\gamma'(0)=\theta$. Let $v_1, v_2, \dots, v_n := \gamma'(0)$ be an orthonormal basis of T_xM.

As in Section 3, choose a set of vectors $\overline{\xi}_1, \overline{\xi}_2, \dots, \overline{\xi}_n, \overline{\eta}_1, \overline{\eta}_2, \dots, \overline{\eta}_n \in T_\theta TM$ satisfying

$$\pi_*(\overline{\xi}_j) = v_j, \quad K\overline{\xi}_j = 0,$$
$$\pi_*(\overline{\eta}_j) = 0, \quad K\overline{\eta}_j = v_j.$$

Here $K\colon TTM \to TM$ is the connection map described in Section 2. Extend $\overline{\xi}_j$ and $\overline{\eta}_j$ to get parallel vector fields $\xi_1, \xi_2, \dots, \xi_n, \eta_1, \eta_2, \dots, \eta_n$. Then $J_j := \eta_j|\mathbb{R}$ for $j = 1, 2, \dots, n-1$ are normal Jacobi fields along γ with initial conditions

$$J_j(0) = 0, \quad J_j'(0) = v_j.$$

Moreover, $\xi_1|_\mathbb{R}, \xi_2|_\mathbb{R}, \dots, \xi_n|_\mathbb{R}$ are linearly independent except on a discrete subset S_1 of $\mathbb{R}$. Hence there are smooth real-valued functions ϕ_{jk} defined on $\mathbb{R} \setminus S_1$ such that

$$\eta_k|_\mathbb{R} = \sum_{j=1}^n \phi_{jk}\xi_j|_\mathbb{R}.$$

As M has entire Grauert tube, it follows that the functions ϕ_{jk} have a meromorphic extension f_{jk} over the whole complex plane such that for each pair j,k, the poles of f_{jk} lie on $\mathbb{R}$ and the matrix $\operatorname{Im}(f_{jk})|_{\mathbb{C}\setminus\mathbb{R}}$ is invertible.

Consider the n–tuples

$$\Xi = (\xi_1, \xi_2, \dots, \xi_n) \quad \text{and} \quad H = (\eta_1, \eta_2, \dots, \eta_n),$$

and holomorphic n–tuples

$$\Xi^{1,0} = (\xi_1^{1,0}, \xi_2^{1,0}, \dots, \xi_n^{1,0}) \quad \text{and} \quad H^{1,0} = (\eta_1^{1,0}, \eta_2^{1,0}, \dots, \eta_n^{1,0}),$$

where $\xi_j^{1,0} = \frac{1}{2}(\xi_j - iJ\xi_j)$ and J is the adapted complex structure.

Then we have

$$H(\sigma) = \Xi(\sigma)f(\sigma) \quad \text{and} \quad H^{1,0}(\sigma+i\tau) = \Xi^{1,0}(\sigma+i\tau)f(\sigma+i\tau),$$

where

$$f(\sigma+i\tau) = (f_{jk}(\sigma+i\tau)) \quad \text{for } \sigma \in \mathbb{R}\setminus S_1.$$

Lemma 4.4 *If $\sigma+i\tau \in \mathbb{C}\setminus\mathbb{R}$, then $\operatorname{Im} f^{-1}(\sigma+i\tau)$ is invertible.*

Proof The proof is almost identical to the proof of Proposition 6.8 in [12]. Suppose there is a nonzero column vector $v=(v_j)\in\mathbb{R}^n$ such that $\operatorname{Im} f^{-1}(\sigma+i\tau)v=0$, $\tau\neq 0$, ie $\omega=(\omega_k)=f^{-1}(\sigma+i\tau)v\in\mathbb{R}^n$. By Proposition 3.2, f^{-1} exists on $\mathbb{C}\setminus\mathbb{R}$. Then we have

$$\Xi^{1,0}=H^{1,0}f^{-1}$$

at the point $\sigma+i\tau$. Hence

$$\sum \xi_j^{1,0}v_j=\Xi^{1,0}v=H^{1,0}f^{-1}v=H^{1,0}\omega=\sum \eta_k^{1,0}\omega_k.$$

Taking real parts, we get

$$\sum \xi_j v_j=\sum \eta_k\omega_k,$$

in contradiction with Proposition 3.2. □

Lemma 4.5 *$G(\zeta):=-f^{-1}(\zeta)$ is a matrix-valued meromorphic function on $\mathbb{C}$ whose pole lies in a discrete subset of $\mathbb{R}$ and $\operatorname{Im} G(\zeta)$ is positive definite for $\zeta=\sigma+i\tau\in\mathbb{C}^+$, where $\mathbb{C}^+$ is the upper half-plane.*

Proof Since $H^{1,0}$ and $\Xi^{1,0}$ are invertible on $\mathbb{C}$ except on a discrete subset, combined with $H^{1,0}=\Xi^{1,0}f$ we get that $G(\zeta)$ is a matrix-valued meromorphic function on $\mathbb{C}$ whose pole lies in a discrete subset of $\mathbb{R}$. By Theorem 3.3, we have

$$f(0)=0,\qquad f'(0)=\mathrm{Id}.$$

Then for small positive τ, we get

$$\begin{aligned}\operatorname{Im} G(i\tau)=\operatorname{Im}(-f^{-1}(i\tau))&=\operatorname{Im}(-(f(0)+i\tau f'(0)+O(\tau^2))^{-1})\\&=\operatorname{Im}(-i\tau\,\mathrm{Id}+O(\tau^2))^{-1}=\operatorname{Im}\Big(\frac{i}{\tau}(\mathrm{Id}+O(\tau))^{-1}\Big).\end{aligned}$$

Hence $\operatorname{Im} G(i\tau)$ is positive definite for small positive τ. As $\operatorname{Im} G(\zeta)$ is nondegenerate on $\mathbb{C}^+$ by Lemma 4.4, $\operatorname{Im} G(\zeta)$ is positive definite for $\zeta=\sigma+i\tau\in\mathbb{C}^+$. □

Let $f_1=(f_{jk})$ with $j,k=1,2,\dots,n-1$. Then we have:

Lemma 4.6 *There exists a discrete subset $S_2\subset\mathbb{R}$ such that for $\sigma\in\mathbb{R}\setminus S_2$, we have*

$$\det(g(J_j(\sigma),J_k(\sigma))_{j,k=1,2,\dots,n-1}=\frac{1}{\det((-f_1^{-1})'(\sigma))},\tag{4-3}$$

where J_j for $j=1,2,\dots,n$ are normal Jacobi fields along γ with initial conditions

$$J_j(0)=0,\qquad J_j'(0)=v_j,$$

and $v_1,v_2,\dots,v_n:=\gamma'(0)$ is an orthonormal basis of T_xM.

Proof We can view $\Xi(\sigma)$ and $H(\sigma)$ as linear mappings $\mathbb{R}^n \to T_{\gamma(\sigma)}M$, given by

$$(\omega_j) = \omega \mapsto \Xi(\sigma)\omega = \sum_{j=1}^{n} \omega_j \xi_j(\sigma)$$

and similarly for $H(\sigma)$. Denote by $\Xi^*(\sigma)$ and $H^*(\sigma)$ the adjoints of $\Xi(\sigma)$ and $H(\sigma)$, respectively (adjoint defined using the Euclidean scalar product on $\mathbb{R}^n$ and the Riemannian metric on $T_{\gamma(\sigma)}M$). Let $\{e_j\}$ be the standard orthonormal basis of $\mathbb{R}^n$. Then we have

$$g(J_j(\sigma), J_k(\sigma)) = g(H(\sigma)e_j, H(\sigma)e_k) = \langle H^*(\sigma)H(\sigma)e_j, e_k\rangle.$$

By the proof of Proposition 6.11 in [12], we get

$$\Xi^*(\sigma)\Xi(\sigma)f'(\sigma) = \mathrm{Id} \quad \text{for } \sigma \in (0, c)$$

for some small positive constant c. On the other hand, we have

$$\Xi(\sigma)e_j = \xi_j(\sigma) \quad \text{and} \quad \Xi^*(\sigma)\Xi(\sigma)e_j = g(\xi_j(\sigma), \xi_k(\sigma))e_k.$$

Hence $\Xi^*(\sigma)\Xi(\sigma)$ is real-analytic over $\mathbb{R}$ and so it has a holomorphic extension to a small open set in $\mathbb{C}$ containing $\mathbb{R}$. As M has entire Grauert tube, it follows that $f(\sigma)$ has a meromorphic extension over the whole complex plane such that its poles lie on a discrete subset $S_1 \subset \mathbb{R}$. Then we have

$$\Xi^*(\sigma)\Xi(\sigma)f'(\sigma) = \mathrm{Id} \quad \text{for } \sigma \in \mathbb{R} \setminus S_1.$$

On the other hand, $f^{-1}(\sigma)$ exists on $\sigma \in \mathbb{R} \setminus S_1'$ for some discrete subset S_1'. Moreover, $f(\sigma)$ is symmetric, by Proposition 6.11 in [12] and analytic continuation. Let $S_2 = S_1 \cup S_1'$. For $\sigma \in \mathbb{R} \setminus S_2$, we have

$$\begin{aligned} H^*(\sigma)H(\sigma) &= (\Xi(\sigma)f(\sigma))^*\Xi(\sigma)f(\sigma) = f(\sigma)\Xi^*(\sigma)\Xi(\sigma)f(\sigma) \\ &= f(\sigma)(f'(\sigma))^{-1}f(\sigma) = ((-f^{-1})'(\sigma))^{-1}. \end{aligned}$$

Since $f_{jn} = f_{nk} = 0$ for $j, k = 1, 2, \dots, n-1$, we see that

$$\det(g(J_j(\sigma), J_k(\sigma))_{j,k=1,2,\dots,n-1} = \frac{1}{\det((-f_1^{-1})'(\sigma))} \quad \text{for } \sigma \in \mathbb{R} \setminus S_2. \qquad \square$$

The following version of Fatou's representation theorem will be crucial for us.

Proposition 4.7 *Suppose that F is an $n \times n$ matrix-valued holomorphic function on the upper half-plane $\mathbb{C}^+ = \{\xi \in \mathbb{C} \mid \operatorname{Im} \zeta > 0\} \cup (\mathbb{R} \setminus P)$, where P is a discrete subset of $\mathbb{R}$ consisting of poles of F. Suppose that for every $\zeta \in \mathbb{C}^+$, $\operatorname{Im} F(\zeta)$ is a symmetric, positive definite matrix, whereas for $\zeta \in \mathbb{R} \setminus P$, $\operatorname{Im} F(\zeta) = 0$. Then there is an $n \times n$ symmetric matrix $\mu = (\mu_{jk})$ whose entries are real-valued, signed Borel measures on $\mathbb{R}$ such that*:

(1°) *μ_{jk} does not have mass on any interval which does not contain a pole of F.*

(2°) $\displaystyle\int_{-\infty}^{+\infty} \frac{|d\mu_{jk}(t)|}{1+t^2} < \infty.$

(3°) *μ is positive semidefinite in the sense that for any $(\omega_j) \in \mathbb{R}^n$, the measure $\sum \omega_j \omega_k \mu_{jk}$ is nonnegative.*

(4°) *For $\zeta \in \mathbb{C}^+$,*

$$F'(\zeta) = A + \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{d\mu(t)}{(\zeta - t)^2},$$

where A is a symmetric, positive semidefinite constant matrix. In fact, we have

$$A = \lim_{\tau \to +\infty} \frac{\operatorname{Im} F(i\tau)}{\tau},$$

and $d\mu(\sigma)$ is the weak limit of $\operatorname{Im} F(\sigma + i\tau)$ as $\tau \to 0^+$.

Proof See [11] and Proposition 7.4 in [12]. The only difference is that we require F has a holomorphic extension to $\mathbb{R} \setminus P$, hence we get that μ_{jk} does not have mass on any interval which does not contain a pole of F. □

Now we are going to finish the proof of Theorem 1.1. Applying Proposition 4.7 to the matrix-valued holomorphic function $(-f_1^{-1})$ on the upper half-plane, we get

$$(-f_1^{-1})'(\zeta) = A + \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{d\mu(t)}{(\zeta - t)^2} \quad \text{for } \zeta \in \mathbb{C}^+, \tag{4-4}$$

where $A = (a_{jk})$ is a symmetric, positive semidefinite constant matrix and μ is an $n \times n$ positive semidefinite symmetric matrix whose entries are real-valued, signed Borel measures on $\mathbb{R}$. By analytic continuation, equation (4-4) also holds on $\mathbb{R}$ except a discrete subset. Moreover, μ does not have mass on any interval which does not contain a pole of $-f_1^{-1}$. This yields that

$$(-f_1^{-1})'(\sigma) = A + \frac{1}{\pi} \sum_j \frac{\mu(t_j)}{(\sigma - t_j)^2} \quad \text{for } \sigma \in \mathbb{R} \setminus \{t_1, t_2, \dots\},$$

where $\{t_1, t_2, \dots\}$ are poles of $-f_1^{-1}$. As $f(0) = 0$, we see that 0 is pole of $-f_1^{-1}$.

Lemma 4.8
$$\mu(0) = \pi \operatorname{Id}.$$

Proof By Proposition 4.7, we get

$$\begin{aligned}
\mu(0) &= \lim_{\delta \to 0^+} \mu(-\delta, \delta) = \lim_{\delta \to 0^+} \lim_{\tau \to 0^+} \int_{-\delta}^{\delta} \operatorname{Im}(-f_1^{-1}(\sigma + i\tau))\, d\sigma \\
&= \lim_{\delta \to 0^+} \lim_{\tau \to 0^+} \int_{-\delta}^{\delta} \operatorname{Im}\Big(-\big(f_{jk}(0) + f'_{jk}(0)(\sigma + i\tau) + O(\sigma + i\tau)^2\big)^{-1}_{1 \le j,k \le n-1}\Big)\, d\sigma \\
&= \lim_{\delta \to 0^+} \lim_{\tau \to 0^+} \int_{-\delta}^{\delta} \operatorname{Im}\Big(-\big((\sigma + i\tau) \operatorname{Id} + O(\sigma + i\tau)^2\big)^{-1}\Big)\, d\sigma \\
&= \lim_{\delta \to 0^+} \lim_{\tau \to 0^+} \int_{-\delta}^{\delta} \operatorname{Im}\Big(-\frac{1}{\sigma + i\tau} (\operatorname{Id} + O(\sigma + i\tau))^{-1}\Big)\, d\sigma \\
&= \lim_{\delta \to 0^+} \lim_{\tau \to 0^+} \int_{-\delta}^{\delta} \operatorname{Im}\Big(-\frac{1}{\sigma + i\tau} \operatorname{Id} + O(1)\Big)\, d\sigma = \lim_{\delta \to 0^+} \lim_{\tau \to 0^+} \int_{-\delta}^{\delta} \frac{\tau}{\sigma^2 + \tau^2}\, d\sigma \operatorname{Id} = \pi \operatorname{Id}.
\end{aligned}$$
□

Given Lemma 4.8, then we have

$$(-f_1^{-1})'(\sigma) = \frac{1}{\sigma^2}\,\mathrm{Id} + B,$$

where

$$B = A + \frac{1}{\pi}\sum_{t_j \neq 0} \frac{\mu(t_j)}{(\sigma - t_j)^2}$$

is positive semidefinite.

Lemma 4.9 *Let A_1 and A_2 be two $k \times k$ Hermitian positive semidefinite complex matrix, then*

$$\det(A_1 + A_2) \geq \det A_1 + \det A_2.$$

Proof It follows from the Minkowski determinant theorem [13, page 115] that

$$\big(\det(A_1 + A_2)\big)^{1/k} \geq (\det A_1)^{1/k} + (\det A_2)^{1/k}. \qquad \square$$

By Theorem 3.3, we get that $f(\sigma + i\tau)$ is a symmetric matrix, so is $-f_1^{-1}(\sigma + i\tau)$. By Proposition 4.7, we see that A and $\mu(t_j)$ are real-valued symmetric positive semidefinite matrix. By Lemma 4.9, we get

$$\frac{1}{\det\big((-f_1^{-1})'(\sigma)\big)} \leq \sigma^{2n-2}.$$

By Theorem 4.1 and Lemma 4.6, we see

$$\int_M n_T(x, y)\,dy \leq p(T),$$

where $p(T)$ is a polynomial of degree at most n. By Theorem 4.2, $\sum_{j=0}^{k-1} \dim H_j(\Omega M, F)$ has polynomial growth for any field of coefficients. It follows that M is topologically elliptic.

To illustrate the idea of the above proof, we give two examples here. Let M be an n–dimensional compact manifold of constant sectional curvature c. From the proof of Theorem 2.5 in [16], we have

$$f_1(\sigma + i\tau) = \begin{cases} (\sigma + i\tau)\,\mathrm{Id} & \text{if } c = 0, \\ (tg(\sigma + i\tau))\,\mathrm{Id} & \text{if } c = 1. \end{cases}$$

Case 1 When $c = 0$, then $-f_1^{-1}(\sigma + i\tau) = (-1/(\sigma + i\tau))\,\mathrm{Id}$. Hence

$$(-f_1^{-1})'(\sigma) = \frac{1}{\sigma^2}\,\mathrm{Id}.$$

Let $F(\sigma + i\tau) := -f_1^{-1}(\sigma + i\tau)$. In this case, the matrix A and measure μ in Proposition 4.7 can be computed by

$$A = \lim_{\tau \to +\infty} \frac{\mathrm{Im}\, F(i\tau)}{\tau} = 0,$$

$$\mu(0) = \lim_{\delta \to 0+} \mu(-\delta, \delta) = \lim_{\delta \to 0+} \lim_{\tau \to 0+} \int_{-\delta}^{\delta} \mathrm{Im}\, F(\sigma + i\tau)\,d\sigma = \pi\,\mathrm{Id}.$$

Then $\int_M n_T(x, y)\,dy$ has polynomial growth of degree n.

Case 2 When $c = 1$, then $-f_1^{-1}(\sigma + i\tau) = (-\cot(\sigma + i\tau))\,\mathrm{Id}$. Hence

$$(-f_1^{-1})'(\sigma) = \frac{1}{\sin^2(\sigma)}\,\mathrm{Id}.$$

Let $F(\sigma + i\tau) := -f_1^{-1}(\sigma + i\tau)$. In this case, the matrix A and measure μ in Proposition 4.7 can be computed by

$$A = \lim_{\tau\to+\infty} \frac{\mathrm{Im}\,F(i\tau)}{\tau} = 0,$$

$$\mu(j\pi) \equiv \mu(0) = \lim_{\delta\to0+} \mu(-\delta,\delta) = \lim_{\delta\to0+}\lim_{\tau\to0+} \int_{-\delta}^{\delta} \mathrm{Im}\,F(\sigma+i\tau)\,d\sigma = \pi\,\mathrm{Id} \quad \text{for } j \in \mathbb{Z}.$$

Then $\int_M n_T(x,y)\,dy$ has linear growth.

References

[1] **R M Aguilar**, *Symplectic reduction and the homogeneous complex Monge–Ampère equation*, Ann. Global Anal. Geom. 19 (2001) 327–353 MR Zbl

[2] **M Berger**, **R Bott**, *Sur les variétés à courbure strictement positive*, Topology 1 (1962) 301–311 MR Zbl

[3] **D M Burns, Jr**, **K K Leung**, *The complex Monge–Ampère equation, Zoll metrics and algebraization*, Math. Ann. 371 (2018) 1–40 MR Zbl

[4] **J Cheeger**, **D Gromoll**, *On the structure of complete manifolds of nonnegative curvature*, Ann. of Math. 96 (1972) 413–443 MR Zbl

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

[6] **M Gromov**, *Homotopical effects of dilatation*, J. Differential Geom. 13 (1978) 303–310 MR Zbl

[7] **K Grove**, **S Halperin**, *Contributions of rational homotopy theory to global problems in geometry*, Inst. Hautes Études Sci. Publ. Math. 56 (1982) 171–177 MR Zbl

[8] **K Grove**, **S Halperin**, *Dupin hypersurfaces, group actions and the double mapping cylinder*, J. Differential Geom. 26 (1987) 429–459 MR Zbl

[9] **K Grove**, **B Wilking**, **J Yeager**, *Almost non-negative curvature and rational ellipticity in cohomogeneity two*, Ann. Inst. Fourier (Grenoble) 69 (2019) 2921–2939 MR Zbl

[10] **V Guillemin**, **M Stenzel**, *Grauert tubes and the homogeneous Monge–Ampère equation*, J. Differential Geom. 34 (1991) 561–570 MR Zbl

[11] **P Koosis**, *Introduction to H_p spaces*, Lond. Math. Soc. Lect. Note Ser. 40, Cambridge Univ. Press (1980) MR Zbl

[12] **L Lempert**, **R Szőke**, *Global solutions of the homogeneous complex Monge–Ampère equation and complex structures on the tangent bundle of Riemannian manifolds*, Math. Ann. 290 (1991) 689–712 MR Zbl

[13] **M Marcus**, **H Minc**, *A survey of matrix theory and matrix inequalities*, revised edition, Dover, New York (2010)

[14] **G P Paternain**, *Geodesic flows*, Progr. Math. 180, Birkhäuser, Boston, MA (1999) MR Zbl

[15] **D Sullivan**, *Infinitesimal computations in topology*, Inst. Hautes Études Sci. Publ. Math. 47 (1977) 269–331 MR Zbl

[16] **R Szőke**, *Complex structures on tangent bundles of Riemannian manifolds*, Math. Ann. 291 (1991) 409–428 MR Zbl

[17] **B Totaro**, *Complexifications of nonnegatively curved manifolds*, J. Eur. Math. Soc. 5 (2003) 69–94 MR Zbl

School of Mathematical Sciences
Key Laboratory of Intelligent Computing and Applications (Ministry of Education)
Institute for Advanced Study, Tongji University
Shanghai, China

xychen100@tongji.edu.cn

Proposed: Gang Tian
Seconded: John Lott, Dmitri Burago

Received: 3 January 2021
Revised: 5 May 2022

Geometry & Topology **28**:3 (2024) 1113–1152
DOI: 10.2140/gt.2024.28.1113
Published: 10 May 2024

On certain quantifications of Gromov's nonsqueezing theorem

Kevin Sackel
Antoine Song
Umut Varolgunes
Jonathan J Zhu
Appendix by Joé Brendel

Let $R > 1$ and let B be the Euclidean 4–ball of radius R with a closed subset E removed. Suppose that B embeds symplectically into the unit cylinder $\mathbb{D}^2 \times \mathbb{R}^2$. By Gromov's nonsqueezing theorem, E must be nonempty. We prove that the Minkowski dimension of E is at least 2, and we exhibit an explicit example showing that this result is optimal at least for $R \le \sqrt{2}$. In the appendix by Joé Brendel, it is shown that the lower bound is optimal for $R < \sqrt{3}$. We also discuss the minimum volume of E in the case that the symplectic embedding extends, with bounded Lipschitz constant, to the entire ball.

53D05, 53D35

1 Introduction

Consider $\mathbb{R}^{2n}$ with its standard symplectic structure $\omega = \sum dx_i \wedge dy_i$. A prototypical question in symplectic geometry is to ask whether one domain of $\mathbb{R}^{2n}$ symplectically embeds into another (ie via an embedding Φ with $\Phi^*\omega = \omega$). At the very least, a symplectic embedding preserves the standard volume form $(1/n!)\omega^n$. However, there is more rigidity in symplectic geometry than just volume. We recall the most famous result certifying this bold claim.

Let $B^{2n}(\pi R^2) \subset \mathbb{R}^{2n}$ be the open ball of radius R, and let $Z^{2n}(\pi r^2) = B^2(\pi r^2) \times \mathbb{R}^{2n-2} \subset \mathbb{R}^{2n}$ be the open cylinder of radius r.

Theorem 1.1 (Gromov's nonsqueezing theorem [11]) *A symplectic embedding of $B^{2n}(\pi R^2)$ into $Z^{2n}(\pi r^2)$ exists if and only if $R \le r$.*

Our goal in this paper is to try to quantify the failure of $B^{2n}(\pi R^2)$ to symplectically embed into $Z^{2n}(\pi r^2)$, when $R > r$, via the following motivating question:

Motivating question How much do we need to remove from $B^{2n}(\pi R^2)$ so that it embeds symplectically into $Z^{2n}(\pi r^2)$?

As a first attempt, one may try to use volume to answer this question. Over all possible symplectic embeddings $B^{2n}(\pi R^2) \hookrightarrow \mathbb{R}^{2n}$, what is the minimal possible volume excluded from $Z^{2n}(\pi r^2)$? This question has a straightforward answer, as the following result of Katok reveals.

Theorem 1.2 (Katok [16]) *Given a compact set X in $(\mathbb{R}^{2n}, \omega_{\mathrm{std}})$, for every $\epsilon > 0$, there exists a Hamiltonian diffeomorphism $\phi\colon (\mathbb{R}^{2n}, \omega_{\mathrm{std}}) \to (\mathbb{R}^{2n}, \omega_{\mathrm{std}})$ such that*

$$\mathrm{Vol}(\phi(X) \setminus Z(\pi)) \leq \epsilon.$$

We therefore modify our attempt to produce a more meaningful version of our question. We discuss two specific instances, though we focus predominantly on the first:

(1) What is the smallest Minkowski dimension of a subset $E \subset B^{2n}(\pi R^2)$ with the property that there is a symplectic embedding $B^{2n}(\pi R^2) \setminus E \to Z^{2n}(\pi r^2)$?

(2) Over all possible symplectic embeddings $B^{2n}(\pi R^2) \hookrightarrow \mathbb{R}^{2n}$ with Lipschitz constant at most $L > 0$, what is the minimal possible volume excluded from $Z^{2n}(\pi r^2)$?

Here Minkowski dimension stands for the lower Minkowski dimension, which is defined for any subset of $B^{2n}(\pi R^2)$. Heuristically, E having Minkowski dimension $d \in \mathbb{R}$ means that as $\epsilon \to 0$, the volume of the ϵ–neighborhood of E behaves as $c\epsilon^{2n-d}$, for some constant $c > 0$.

For each of these two questions, there is some quantity we are trying to minimize: either the Minkowski dimension of E, or the volume excluded from the cylinder under an L–Lipschitz symplectic embedding. In either case, there are two key aspects to discuss.

- **Constructive** We find an explicit symplectic embedding which provides an upper bound on the quantity that we are trying to minimize.
- **Obstructive** For a purported symplectic embedding, we find a lower bound on the quantity that we are trying to minimize.

A full answer to these two questions would require that the obstructive and constructive bounds match. Although these questions are interesting in general dimensions, in this paper we will restrict our attention only to the case of dimension $2n = 4$, save for a few open questions posed in Section 6. There are also a plethora of further questions that will be posed in the final section of the paper.

We now discuss our results for each of these two questions.

1.1 The Minkowski dimension problem

Recall that here we are asking for the smallest lower Minkowski dimension of a subset $E \subset B(\pi R^2)$ with the property that there is a symplectic embedding $B(\pi R^2) \setminus E \to Z(\pi r^2)$, assuming $R > r$. (Here and henceforth we drop the superscripts from $B^4(\,.\,)$ and $Z^4(\,.\,)$ since the dimension is always 4.) By a scaling argument, it suffices to consider the case $r = 1$.

Let us start by discussing the constructive side. Observe that if we remove a union of codimension-one affine hyperplanes along a sufficiently fine grid, we end up with many connected components, each of which embeds into $Z(\pi)$ by translations. Interestingly, at least in a certain range of R, one can do better and find a two-dimensional submanifold E whose complement embeds into $Z(\pi)$.

To explain this result we need to introduce some notation. Let us consider the Lagrangian disk

$$L := B(2\pi) \cap \{y_1 = y_2 = 0\}.$$

Let us also define $\mathscr{E}(\pi, 4\pi) \subset \mathbb{R}^4$ to be the open ellipsoid

$$\left\{(x_1, y_1, x_2, y_2) \mid x_1^2 + y_1^2 + \tfrac{1}{4}(x_2^2 + y_2^2) < 1\right\},$$

and let

$$\mathscr{C} := \mathscr{E}(\pi, 4\pi) \cap \{x_2 = y_2 = 0\}.$$

Theorem 1.3 *$B(2\pi) \setminus L$ is symplectomorphic to $\mathscr{E}(\pi, 4\pi) \setminus \mathscr{C}$. Consequently, $B(2\pi) \setminus L$ symplectically embeds into $Z(\pi)$.*

In particular, removing a Lagrangian plane from $B(2\pi)$ halves its Gromov capacity. Our proof of Theorem 1.3 has a great deal in common with Section 3 of Oakley and Usher's beautiful paper [25], where the same geometries are used for a different purpose. In fact, we show in Section 4 how the projective space $\mathbb{CP}^2$ is symplectomorphic to the boundary reduction of the unit cotangent bundle $D^*\mathbb{RP}^2$ by using the explicit map of [25, Lemma 3.1]. Theorem 1.3 can also be derived from the proof of Biran's general decomposition theorem [2, Theorem 1.A; Example 3.1.2], and Opshtein [26, Lemma 3.1]. We use the latter in our argument as well.

On the obstructive side, we show that removing a two-dimensional subset as in Theorem 1.3 is the best one can do in general:

Theorem 1.4 *Let E be a closed subset of $\mathbb{R}^4$ and let $R > 1$. Suppose that $B(\pi R^2) \setminus E$ symplectically embeds into the cylinder $Z(\pi)$. Then the lower Minkowski dimension of E is at least 2.*

For the proof of Theorem 1.4, we build on Gromov's original nonsqueezing argument by adding a key new ingredient: the waist inequality, which was also introduced by Gromov [12]; see also Memarian [23]. Crucially, we require the sharp version due to Akopyan and Karasev [1], as well as the Heintze–Karcher [15] bound on the volumes of tubes around minimal surfaces, in place of the monotonicity inequality for minimal surfaces.

Remark 1.5 Let $R_{\sup} \in (1, \infty]$ be the supremum of the radii R such that there is a codimension-two subset of $B(\pi R^2)$ whose complement can symplectically embed into $Z(\pi)$. In the first version of this article we had conjectured that $R_{\sup}$ should be equal to $\sqrt{2}$. However shortly after its appearance, Joé Brendel informed us that using a construction inspired by Hacking and Prokhorov [13], he can prove $R_{\sup} \geq \sqrt{3}$. As a consequence, we changed our conjecture to a question; see Section 6.3. His construction appears in the appendix.

We further remark that in Theorem 1.3, we remove a Lagrangian plane. In the construction of Joé Brendel in the appendix realizing $R_{\text{sup}} \geq \sqrt{3}$, he removes a union of Lagrangians together with a symplectic divisor. In higher dimensions (see eg Section 6.1), this distinction could be interesting.

1.2 The Lipschitz problem

Recall that for fixed $L > 1$, we are asking for the smallest volume of the region

$$E(\Phi) := B(\pi R^2) \setminus \Phi^{-1}(Z(\pi)) = \Phi^{-1}(\mathbb{R}^4 \setminus Z(\pi))$$

over all symplectic embeddings $\Phi\colon B(\pi R^2) \hookrightarrow \mathbb{R}^4$ of Lipschitz constant bounded above by L. (We note that although we use the letter L for both the Lipschitz constant as well as for the Lagrangian disk of Theorem 1.3, there will be no confusion given the context.)

On the obstructive side, we obtain the following as a corollary of the proof of the obstructive bound for the Minkowski question (Theorem 1.4):

Theorem 1.6 *Let $R > 1$. Then there exists a constant $c = c(R) > 0$ such that for all constants L and all symplectic embeddings $\Phi\colon B(\pi R^2) \hookrightarrow \mathbb{R}^4$ with Lipschitz constant at most L, we have*

$$\mathrm{Vol}_4(E(\Phi)) \geq \frac{c}{L^2}.$$

It is worth noting that one may use the standard nonsqueezing theorem alone to find a weaker quantitative obstructive bound of c/L^3 as follows. Suppose we had an L–Lipschitz symplectic embedding $\phi\colon B(\pi R^2) \hookrightarrow \mathbb{R}^4$ for $R > 2$. Then by Gromov's nonsqueezing theorem and the Lipschitz condition, one can check that there is a ball of radius of order $1/L$ embedded inside $E(\phi)$. Hence,

$$\mathrm{Vol}(E(\phi)) \gtrsim \frac{1}{L^4}.$$

With a little more effort, one may find order L many disjoint such balls inside $E(\phi)$, yielding the obstructive bound c/L^3. However, jumping from c/L^3 to our obstructive bound of c/L^2 appears to require a new tool, which in our case is Gromov's waist inequality.

On the constructive side, we adapt Katok's ideas in [16] to prove the following:

Theorem 1.7 *Let $R > 1$. Then there exists a constant $C = C(R) > 0$ such that for all constants L, there exists a symplectic embedding $\Phi\colon B^4(R) \hookrightarrow \mathbb{R}^4$ with Lipschitz constant at most L such that*

$$\mathrm{Vol}_4\big(E(\Phi)\big) \leq \frac{C}{L}.$$

Remark 1.8 As was pointed out to us by Felix Schlenk, our construction is a simplified version of multiple symplectic folding; see Schlenk [28, Sections 3 and 4].

One would obviously like to push the obstructive and constructive bounds together.

Organization of the paper We start by recalling some definitions and known theorems in Section 2, which are then applied in Section 3 to prove an obstructive bound implying Theorem 1.4. In Section 4 we construct the symplectomorphism of Theorem 1.3. The Lipschitz problem, including Theorems 1.6 and 1.7, are discussed in Section 5. We also list several related questions in the final Section 6. In the appendix, written by Joé Brendel, the construction mentioned in Remark 1.5 appears.

Acknowledgements We thank Michael Usher for a useful e-mail correspondence. We also thank Felix Schlenk and Joé Brendel for their interest and helpful comments our paper. Sackel thanks Larry Guth for originally suggesting the version of this problem involving Lipschitz constants which provided the initial impetus for this project. Varolgunes thanks Grigory Mikhalkin for very useful discussions on Theorem 1.3 and also sketching a proof of Corollary 6.5 that we ended up not using; and Kyler Siegel for a discussion regarding Section 6.2.

Sackel was partially supported by the National Science Foundation under grant DMS-1547145. This research was conducted during the period Song served as a Clay Research Fellow. Zhu was supported in part by the National Science Foundation under grant DMS-1802984 and the Australian Research Council under grant FL150100126.

2 Preliminaries

We use the usual asymptotic notation $f \in O(g)$ to mean $|f| \le Cg$ for some constant C, and $f \in o(g)$ to mean that $\lim_{t\to 0} f(t)/g(t) = 0$. We write $f \in \Theta(g)$ if $f \in O(g)$ and $g \in O(f)$.

Let $S \subset \mathbb{R}^n$ be any bounded subset. Let $N_t(S)$ denote the open t–neighborhood of S with respect to the standard metric. If Σ is a compact submanifold (possibly with boundary), let $V_t(\Sigma)$ be the exponential t–tube of Σ, ie the image by the normal exponential map of the open t–neighborhood of the zero-section in the normal bundle of Σ (which is endowed with the natural metric).

We denote by Vol_n the Euclidean n–volume of a set and set

$$\alpha_l = \frac{\pi^{l/2}}{\Gamma\left(\frac{1}{2}l + 1\right)}.$$

Note that when n is a natural number, $\alpha_n = \mathrm{Vol}_n(B^n)$ is precisely the Euclidean volume of the unit n–dimensional ball $B^n \subset \mathbb{R}^n$. For $s \ge 0$, the s–dimensional lower Minkowski content of S is defined as

$$\underline{\mathcal{M}}_s(S) := \liminf_{t\to 0^+} \frac{\mathrm{Vol}_n(N_t(S))}{\alpha_{n-s} t^{n-s}}.$$

Note that the normalization is chosen so that if $\Sigma^k \subset \mathbb{R}^n$ is a closed k–dimensional submanifold, then $\underline{\mathcal{M}}_k(\Sigma) = \mathrm{Vol}_k(\Sigma)$ coincides with the Euclidean k–volume of Σ.

The lower Minkowski dimension of S is defined as

$$\underline{\dim}_{\mathcal{M}}(S) := \inf_{s>0}\{\underline{\mathcal{M}}_s(S) = 0\} = \sup_{s\geq 0}\{\underline{\mathcal{M}}_s(S) > 0\}.$$

There are similar notions of upper Minkowski dimension and upper Minkowski content, which we will not need in this paper since a lower bound of the lower Minkowski content implies by definition the same lower bound for the upper Minkowski content. There are also equivalent definitions using ball packings. Replacing S by its closure $\overline{S}$ does not change the Minkowski upper/lower dimensions.

2.1 Waist inequalities

The waist inequality for round spheres proved by Gromov [12] and with more details by Memarian [23] was extended to the case of maps from Euclidean balls by Akopyan and Karasev [1, Theorem 1]. The proof of the latter immediately implies the following:

Theorem 2.1 (waist inequality) *For any positive integers n and k, there exists a continuous function $h_{n,k} : (0,\infty) \to \mathbb{R}$ such that $h_{n,k} \in o(t^k)$ and the following holds: for any continuous map $f : B^n \to \mathbb{R}^k$, there exists $y \in \mathbb{R}^k$ such that*

$$\mathrm{Vol}_n\big(N_t(f^{-1}(y)) \cap B^n\big) \geq \alpha_{n-k}\alpha_k t^k - h_{n,k}(t). \tag{2-1}$$

It will be useful for our application that the above estimate is uniform in f. This uniform estimate is indeed implied by the proof of [1, Theorem 1] as they compare the t–neighborhood of a fiber to the t–neighborhood of an equatorial unit sphere $S^{n+1-k} \subset S^{n+1}$ (cf the second-last equation of their proof, with correct normalization). The latter is independent of f, and by explicit calculation one may verify that actually $h_{n,k} \in O(t^{k+2})$.

We remark that the waist inequality for spheres [12; 23] describes a stronger property than the above statement, since it gives optimal bounds on all (not just small) neighborhoods of the big fiber.

2.2 Tubes around minimal submanifolds

The Heintze–Karcher inequality [15] estimates the volume of tubes around compact submanifolds. We need the case of minimal submanifolds in Euclidean space (covered by [15, Theorem 2.3] with $\delta = 0$ and Remark 2 on page 453 in [15]), which may be stated as follows:

Theorem 2.2 (Heintze–Karcher inequality) *For any positive integers n and k, and any smooth compact k–dimensional minimal submanifold $\Sigma^k \subset \mathbb{R}^n$ with boundary, for $t > 0$ we have*

$$\mathrm{Vol}_n(V_t(\Sigma)) \leq \mathrm{Vol}_k(\Sigma)\alpha_{n-k}t^{n-k}. \tag{2-2}$$

Theorem 2.2 is again a uniform estimate on the exponential t–neighborhood; one may compare it to the statement that the (upper) Minkowski k–content of a closed submanifold $\Sigma^k \subset \mathbb{R}^n$ is $\mathrm{Vol}_k(\Sigma)$. The point of Theorem 2.2 is that the constant in this estimate does not depend on the minimal k–submanifold Σ or t. The main ingredient for its proof is an estimate for the Jacobian determinant of the normal exponential, which in Euclidean space is 1 to lowest order. The following term is controlled by the mean curvature $\boldsymbol{H}$, which one may expect from the interpretation of mean curvature as the first variation of (k–)area. Consequently, for a general compact submanifold Σ, (2-2) will have an error term of order $c(\max_\Sigma |\boldsymbol{H}|)t^{n-k+1}$.

2.3 Gromov foliation and maps into the cylinder

For $r < R$, recall that $B(\pi R^2)$ and $Z(\pi r^2)$ denote the open ball and open cylinder of radius R and r, respectively, in $\mathbb{R}^4$. In this subsection, we give a slight modification of the holomorphic foliation argument of Gromov [11] in dimension $n = 4$.

Proposition 2.3 *Let $R, r > 0$. Let E be a compact subset of $\mathbb{R}^4$ and let $\phi\colon B(\pi R^2) \setminus E \to \mathbb{R}^4$ be a smooth symplectic embedding into the cylinder $Z(\pi r^2)$. Let U be the closure of an open neighborhood of $\partial B(\pi R^2) \cup E$ in $\mathbb{R}^4$. Then there exists a smooth map $f\colon B(\pi R^2) \setminus U \to \mathbb{R}^2$ such that*

- *f has no critical points on $B(\pi R^2) \setminus U$, and*
- *for all $y \in \mathbb{R}^2$, if $f^{-1}(y) \cap B(\pi R^2) \setminus U$ is nonempty, then it is a two-dimensional complex submanifold of Euclidean area less than πr^2.*

Proof The following argument is standard in the symplectic community. As mentioned before the statement of the proposition, the ideas are due to Gromov [11], though more thorough analytic details may be found elsewhere; see eg [22].

We define $A := \pi r^2$ for brevity. Since $B(\pi R^2) \setminus U$ has compact closure in $B(\pi R^2) \setminus E$, the image of $B(\pi R^2) \setminus U$ under ϕ lands in $B^2(\pi r_0^2) \times [-K, K]^2$ for some large constant K and $0 < r_0 < r$ (possibly depending upon U). Let $S^2(A)$ denote the 2–sphere with standard symplectic form scaled to have total area A, and let $T_K^2 = (\mathbb{R}/4K\mathbb{Z})^2$ be the 2–torus with symplectic form induced by the standard form on $\mathbb{R}^2$. Then we have a symplectic embedding $B^2(\pi r_0^2) \times [-K, K]^2 \subset S^2(A) \times T_K^2$, and upon composing with ϕ, we arrive at a symplectic embedding, also denoted by ϕ (by abuse of notation),

$$\phi\colon B(\pi R^2) \setminus U \hookrightarrow S^2(A) \times T_K^2.$$

Let J_0 denote the standard complex structure on $B(\pi R^2)$, and let J_1 denote the standard (split) complex structure on $S^2(A) \times T_K^2$. We pick a special almost complex structure J_ϕ on $S^2(A) \times T_K^2$ which incorporates ϕ by requiring that it satisfies the following three properties:

- On the image of ϕ, $J_\phi = \phi_*(J_0)$.
- $J_\phi = J_1$ in a neighborhood of $\{\infty\} \times T_K^2$.
- Everywhere, J_ϕ is compatible with the symplectic form.

By [22, Proposition 9.4.4], which encompasses standard 4–dimensional techniques, the evaluation map

$$\mathrm{ev}\colon \mathcal{M}_{0,1}(\beta, J_\phi) \to S^2(A) \times T^2_K$$

is a diffeomorphism, where $\mathcal{M}_{0,1}(\beta, J_\phi)$ is the moduli space of J_ϕ–holomorphic spheres with one marked point and in the class β. Meanwhile, the map which forgets the marked point $\tau\colon \mathcal{M}_{0,1}(\beta, J_\phi) \to \mathcal{M}_{0,0}(\beta, J_\phi)$ is a smooth fibration, with fibers diffeomorphic to S^2. By positivity of intersections (see [22, Theorem 2.6.3]), each fiber of τ, which is a J_ϕ–holomorphic sphere, intersects $\{\infty\} \times T^2_K$ once and transversely, so by the implicit function theorem we have a canonical diffeomorphism g of T^2_K with $\mathcal{M}_{0,0}(\beta, J_\phi)$. We therefore obtain a map $h\colon S^2(A) \times T^2_K \to T^2_K$ by setting $h(p) = x$ when the unique sphere through p passes through (∞, x). It is clear by construction that the diagram

$$\begin{tikzcd} \mathcal{M}_{0,1}(\beta, J_\phi) \arrow[r, "\mathrm{ev}"] \arrow[d, "\mathrm{for}"] & S^2(A) \times T^2_K \arrow[d, "h"] \\ \mathcal{M}_{0,0}(\beta, J_\phi) \arrow[r, "g"] & T^2_K \end{tikzcd}$$

commutes. In particular since the forgetful map is a smooth S^2 fiber bundle, h is smooth and has no critical points.

Notice that since the image of $B(\pi R^2) \setminus U$ under ϕ is contained in a contractible subset of $S^2(A) \times T^2_K$, we have that the composition $h \circ \phi\colon B(\pi R^2) \setminus U \to T^2_K$ lifts to a map $f\colon B(\pi R^2) \setminus U \to \mathbb{R}^2$. This is the function f we desired in the statement of the proposition, and we must now check it satisfies both of the desired properties.

The fact that $f^{-1}(y)$ is a complex submanifold is simply because it is by definition a subset of a J_ϕ–holomorphic sphere, and J_ϕ is chosen to equal $\phi_*(J_0)$ on the image of ϕ. Finally, the area bound comes from the fact that the area of $f^{-1}(y)$ is at most the symplectic area of the corresponding sphere (the one passing through $(\infty, [y])$), which is just A since symplectic area is purely homological. □

3 A quantitative obstruction to partial symplectic embeddings

As usual, for $r < R$, $B(\pi R^2)$ and $Z(\pi r^2)$ refer to the open ball and open cylinder of radius R and r, respectively, in $\mathbb{R}^4$. The main estimate of this section is the following obstructive bound:

Theorem 3.1 *Let E be a compact subset of $\mathbb{R}^4$ and suppose that $B(\pi R^2) \setminus E$ symplectically embeds into the cylinder $Z(\pi r^2) \subset \mathbb{R}^4$. Then there is a function $k_R \in o(t^2)$ such that for any $t > 0$,*

$$\mathrm{Vol}_4(N_t(E)) \geq \pi^2(R^2 - r^2)t^2 - k_R(t).$$

Proof Let $\Phi\colon B(\pi R^2) \setminus E \to Z(\pi r^2)$ be the symplectic embedding of the statement. Consider $0 < t < \frac{1}{2}(R - r)$ and take $0 < \delta < t$. Later, we will send $\delta \to 0$.

Let U_δ be the closure of $N_\delta(\partial B(\pi R^2) \cup E)$ in $\mathbb{R}^4$ and let $\widetilde{f}_\delta\colon B(\pi R^2) \setminus U_\delta \to \mathbb{R}^2$ be the map given by Proposition 2.3. Note that $B(\pi R^2) \setminus U_\delta = B(\pi(R-\delta)^2) \setminus \overline{N_\delta(E)}$. We then take $f_\delta\colon B(\pi(R-\delta)^2) \to \mathbb{R}^2$ to be any continuous extension of $\widetilde{f}_\delta$. Since f_δ agrees with $\widetilde{f}_\delta$ on $B(\pi(R-\delta)^2) \setminus U_\delta$, by the conclusions of Proposition 2.3 we have for any $y \in \mathbb{R}^2$ that

$$\Sigma_\delta := f_\delta^{-1}(y) \cap \bigl(B(\pi(R-\delta)^2) \setminus U_\delta\bigr)$$

is a minimal submanifold with area less than $A := \pi r^2$.

The main idea of our proof is that the waist inequality guarantees a fiber with large-volume neighborhoods, but by the area bound (and the structure of tubes) this can only happen if the fiber accumulates near the exceptional set. Accordingly, a key component is the following covering claim.

Claim 1 *Let $f_{\delta,t}$ be the restriction of f_δ to the ball $B(\pi(R-2t)^2)$. Then*

$$N_t(f_{\delta,t}^{-1}(y)) \cap B(\pi(R-2t)^2) \subset V_t(\Sigma_\delta) \cup N_{\delta+t}(E). \tag{3-1}$$

Indeed, by definition of Σ_δ and the supposition $\delta < t$ we have that

$$f_{\delta,t}^{-1}(y) \setminus \overline{N_\delta(E)} = \Sigma_\delta \cap B(\pi(R-2t)^2) \subset \Sigma_\delta.$$

Now given any submanifold Σ, its t–neighborhood $N_t(\Sigma)$ is always contained in the union of the tube $V_t(\Sigma)$ and the t–neighborhood $N_t(\partial\Sigma)$ of its boundary. So since $\partial\Sigma_\delta \subset \partial U_\delta$, we have that

$$N_t\bigl(f_{\delta,t}^{-1}(y) \setminus \overline{N_\delta(E)}\bigr) \subset N_t(\Sigma_\delta) \subset V_t(\Sigma_\delta) \cup N_t(\partial U_\delta).$$

By the triangle inequality it follows that

$$N_t(f_{\delta,t}^{-1}(y)) \subset V_t(\Sigma_\delta) \cup N_t(\partial U_\delta) \cup N_{\delta+t}(E). \tag{3-2}$$

But by definition of U_δ, we have $N_t(\partial U_\delta) \subset N_t(\partial B(\pi(R-\delta)^2)) \cup N_{\delta+t}(E)$, and since $\delta < t$, we note that $N_t(\partial B(\pi(R-\delta)^2)) \cap B(\pi(R-2t)^2)) = \varnothing$. Taking the intersection of (3-2) with $B(\pi(R-2t)^2)$ then yields the claim.

Having established the claim, we now estimate the volume of each set in (3-1). First, let $h_{4,2} \in o(t^2)$ be as in Theorem 2.1. By rescaling to the ball $B(\pi(R-2t)^2)$, the waist inequality Theorem 2.1 applied to $f_{\delta,t}\colon B(\pi(R-2t)^2) \to \mathbb{R}^2$ gives that there is some $y \in \mathbb{R}^2$ for which

$$\mathrm{Vol}_4\bigl(N_t(f_{\delta,t}^{-1}(y)) \cap B(\pi(R-2t)^2)\bigr) \geq \pi^2 t^2 (R-2t)^2 - (R-2t)^4 h_{4,2}\left(\frac{t}{R-2t}\right). \tag{3-3}$$

On the other hand, since Σ_δ is minimal with area at most A, the Heintze–Karcher inequality Theorem 2.2 yields

$$\mathrm{Vol}_4(V_t(\Sigma_\delta)) \leq \mathrm{Vol}_2(\Sigma_\delta)\pi t^2 \leq A\pi t^2. \tag{3-4}$$

Combining the covering (3-1) with the estimates (3-3) and (3-4) yields

$$\mathrm{Vol}_4(N_{\delta+t}(E)) \geq \big(\pi^2(R-2t)^2 - A\pi\big)t^2 - (R-2t)^4 h_{4,2}\left(\frac{t}{R-2t}\right).$$

The volume of $N_t(E)$ is nondecreasing with respect to t, and is continuous almost everywhere. Therefore, sending $\delta \to 0$ and recalling that $A = \pi r^2$ in the inequality above, we obtain for any $t > 0$ that

$$\mathrm{Vol}_4(N_t(E)) \geq \pi^2\big((R-2t)^2 - r^2\big)t^2 - (R-2t)^4 h_{4,2}\left(\frac{t}{R-2t}\right).$$

Thus, taking

$$k_R(t) = 4R\pi^2 t^3 + R^4 h_{4,2}\left(\frac{t}{R-2t}\right),$$

for instance, concludes the proof. □

An immediate corollary is a lower bound for the lower Minkowski dimension. We will see that this bound is sharp in the next section, at least for radii R which are not too large.

Corollary 3.2 (Minkowski dimension) *Suppose that $B(\pi R^2) \setminus E$ symplectically embeds into the cylinder $Z(\pi r^2) \subset \mathbb{R}^4$. Then the two-dimensional lower Minkowski content of E satisfies*

$$\underline{\mathcal{M}}_2(E) \geq \pi(R^2 - r^2).$$

In particular, the lower Minkowski dimension $\underline{\dim}_{\mathcal{M}}(E)$ is at least 2.

4 Squeezing the complement of a Lagrangian plane

In this section it will be more convenient to use complex coordinates for the standard symplectic $\mathbb{R}^4$. Therefore we consider $\mathbb{C}^2$ with its standard Kähler structure, ie if x and y are the complex coordinates, then the symplectic form is

$$\frac{i}{2}(dx \wedge d\bar{x} + dy \wedge d\bar{y}).$$

Let us recall the main objects in the statement of Theorem 1.3 in complex notation for convenience. Let $B(2\pi) \subset \mathbb{C}^2$ be the open ball of radius $\sqrt{2}$ centered at the origin. Let $\mathbb{R}^2 \subset \mathbb{C}^2$ be the real part and define

$$L := B(2\pi) \cap \mathbb{R}^2.$$

We also define $\mathscr{E}(\pi, 4\pi) \subset \mathbb{C}^2$ to be the open ellipsoid

$$\big\{(x, y) \mid |x|^2 + \tfrac{1}{4}|y|^2 < 1\big\},$$

and let

$$\mathscr{C} := \mathscr{E}(\pi, 4\pi) \cap \{y = 0\}.$$

Let us introduce the main actors in the proof. Let $\mathbb{CP}^2(2\pi)$ be the symplectic manifold obtained from coisotropic reduction of the sphere S^5 of radius $\sqrt{2}$ in $\mathbb{C}^3$. Denoting the complex coordinates on $\mathbb{C}^3$ by z_1, z_2 and z_3, we are using the symplectic structure $\frac{i}{2}(dz_1 \wedge d\bar{z}_1 + dz_2 \wedge d\bar{z}_2 + dz_3 \wedge d\bar{z}_3)$ on $\mathbb{C}^3$. There is a canonical identification of $\mathbb{CP}^2(2\pi)$ with the complex manifold $\mathbb{CP}^2 := \mathrm{Gr}_{\mathbb{C}}(1,3)$, whose homogeneous coordinates we will denote by $[z_1 : z_2 : z_3]$. Of course, $\mathbb{CP}^2(2\pi)$ is nothing but $\mathbb{CP}^2$ equipped with the Fubini–Study symplectic form scaled so that a complex line (eg $\{z_1 = 0\} \subset \mathbb{CP}^2$) has area 2π.

Let us also specify some submanifolds of $\mathbb{CP}^2(2\pi)$ using its canonical identification with $\mathbb{CP}^2$.

- $L_{\mathbb{RP}}$ is the real part
$$\{[z_1 : z_2 : z_3] \mid \mathrm{Im}(z_1) = \mathrm{Im}(z_2) = \mathrm{Im}(z_3) = 0\}.$$
- For $t = [t_1 : t_2 : t_3] \in \mathbb{RP} := \mathrm{Gr}_{\mathbb{R}}(1,3)$, we define the complex lines
$$S_t := \{[z_1 : z_2 : z_3] \mid t_1 z_1 + t_2 z_2 + t_3 z_3 = 0\}.$$
- FQ is the Fermat quadric
$$\{[z_1 : z_2 : z_3] \mid z_1^2 + z_2^2 + z_3^2 = 0\}.$$

It is well-known that $\mathbb{CP}^2(2\pi) \setminus (S_{[0:0:1]} \cup L_{\mathbb{RP}})$ is symplectomorphic to $B(2\pi) \setminus L$; see Exercise 9.4.11 in [22].

Consider $\mathbb{RP}$ as a smooth manifold in the standard way. Let λ be the tautological one-form on $T^*\mathbb{RP}$ and V be the Liouville vector field, which is a vertical vector field equal to the Euler vector field in each fiber (which is defined on any vector space independently of a basis). We have
$$\omega(V, \cdot) = \lambda,$$
where $\omega = d\lambda$. We denote the zero-section submanifold on $T^*\mathbb{RP}$ by $Z_{\mathbb{RP}}$.

The Riemannian metric on S^2 obtained from its embedding as the round sphere of radius 1 in $\mathbb{R}^3$ induces a metric on its quotient by the antipodal map, which is canonically diffeomorphic to $\mathbb{RP}$. We call this the round metric on $\mathbb{RP}$ and denote it by $g_{\mathbb{RP}}$. We have the diffeomorphism $g^\sharp_{\mathbb{RP}} : T\,\mathbb{RP} \to T^*\mathbb{RP}$, which is in particular linear on the fibers. We can transport the function $K \colon T\,\mathbb{RP} \to \mathbb{R}$, given by lengths of tangent vectors to

$$K^\sharp \colon T^*\mathbb{RP} \to \mathbb{R}. \tag{4-1}$$

On $T\,\mathbb{RP}$ we have the geodesic flow; under the identification by $g^\sharp_{\mathbb{RP}}$ this becomes the Hamiltonian flow of the function $\frac{1}{2}(K^\sharp)^2$. The normalized geodesic flow on $T\,\mathbb{RP} \setminus \mathbb{RP}$ becomes the Hamiltonian flow of $K^\sharp$. Let us call these the geodesic flow and normalized geodesic flow on $T^*\mathbb{RP}$. Note that the normalized geodesic flow on $T^*\mathbb{RP} \setminus Z_{\mathbb{RP}}$ is a π–periodic action of $\mathbb{R}$.

Any unparametrized oriented geodesic circle γ in $\mathbb{RP}$ with its round metric defines a symplectic submanifold C_γ in $T^*\mathbb{RP}$ with boundary on $Z_{\mathbb{RP}}$ by taking points (q,p) such that $q \in \gamma$ and $p = g_q(v,\cdot)$, where v is nonnegatively tangent to γ. Let us denote by $-\gamma$ the geodesic circle with opposite orientation. Clearly, C_γ and $C_{-\gamma}$ intersect along $\gamma \subset Z_{\mathbb{RP}}$ and form a symplectic submanifold of $T^*\mathbb{RP}$, which is diffeomorphic to a cylinder.

Let $D^*\mathbb{RP} \subset T^*\mathbb{RP}$ be the closed unit-disk bundle, which is given by the subset $K^\sharp \le 1$. Also let $U^*\mathbb{RP} = (K^\sharp)^{-1}(1)$ be the unit-sphere bundle, ie the boundary of $D^*\mathbb{RP}$, with its induced contact structure $\theta := \iota^*\lambda$.

The symplectic reduction $U^*\mathbb{RP}/S^1$ is a two-sphere equipped with a canonical symplectic form. Let us call this symplectic manifold (Q,ω_Q). The points of Q are canonically identified with unparametrized oriented geodesic circles in round $\mathbb{RP}$.

Let us denote the boundary reduction symplectic manifold of $D^*\mathbb{RP}$ by $\overline{D^*\mathbb{RP}^2}$ (see Definition 3.9 of [30]; also note the interpretation as one half of a symplectic cut [18]).

Note that Q and $Z_{\mathbb{RP}}$ sit naturally inside $\overline{D^*\mathbb{RP}^2}$. The Poincaré dual of the homology class of Q is $1/\pi$ times the symplectic class. The cylinders $(C_\gamma \cup C_{-\gamma}) \cap D^*\mathbb{RP}$ become symplectic 2–spheres in $\overline{D^*\mathbb{RP}^2}$. They intersect Q positively in two points and $Z_{\mathbb{RP}}$ along the circle γ. Let us call these spheres S_γ, now indexed by unoriented unparametrized geodesic circles on $\mathbb{RP}$. Each S_γ has self-intersection number 1.

Let γ_0 be the oriented unparametrized geodesic on $\mathbb{RP}$ which corresponds to the quotient of the horizontal great circle in $S^2 \subset \mathbb{R}^3$ oriented as the boundary of the lower hemisphere. We define $S := S_{\gamma_0}$.

Proposition 4.1 *There is a symplectomorphism $\overline{D^*\mathbb{RP}^2} \to \mathbb{CP}^2(2\pi)$ with the following properties*:

- *$Z_{\mathbb{RP}}$ is sent to $L_{\mathbb{RP}}$.*
- *S is sent to $S_{[0:0:1]}$.*
- *Q is sent to FQ.*

The proof of this proposition is postponed to Section 4.1. Let us continue with an immediate corollary.

Corollary 4.2 *$\overline{D^*\mathbb{RP}^2} \setminus (S \cup Z_{\mathbb{RP}})$ is symplectomorphic to $\mathbb{CP}^2(2\pi) \setminus (S_{[0:0:1]} \cup L_{\mathbb{RP}})$, and in turn to $B(2\pi) \setminus L$.*

Note that $V \cdot K^\sharp = K^\sharp$ on $T^*\mathbb{RP} \setminus Z_{\mathbb{RP}}$, which means that $K^\sharp$ is an exponentiated Liouville coordinate for V on $T^*\mathbb{RP} \setminus Z_{\mathbb{RP}}$. Hence, we obtain a Liouville isomorphism

$$T^*\mathbb{RP} \setminus Z_{\mathbb{RP}} \simeq (U^*\mathbb{RP} \times (0,\infty)_r, d(r\theta)),$$

where $K^\sharp$ is matched with the function r. The Hamiltonian vector field X_r gives the r–translation invariant Reeb vector field (using $r=1$) on the contact levels. Finally, observe that $C_\gamma \cap (T^*\mathbb{RP} \setminus Z_{\mathbb{RP}})$'s are obtained as the traces of the Reeb orbits on $U^*\mathbb{RP}$ under the Liouville flow.

In particular, we have a foliation of $\overline{D^*\mathbb{RP}^2} \setminus Z_{\mathbb{RP}}$ by open disks which are the reductions of the $C_\gamma \cap (D^*\mathbb{RP} \setminus Z_{\mathbb{RP}})$'s. Let us denote this Q–family of submanifolds by $\mathbb{D}_\gamma$, where $\gamma \in Q$.

Proposition 4.3 *$\overline{D^*\mathbb{RP}^2} \setminus Z_{\mathbb{RP}}$ is symplectomorphic to an area-π standard symplectic disk bundle of (Q, ω_Q) in the sense of Biran* [2, Section 2.1] *in such a way that $\mathbb{D}_\gamma$ is sent to the fiber over γ for every $\gamma \in Q$.*

Remark 4.4 The cohomology class of ω_Q/π is integral and it admits a unique lift to $H_2(Q, \mathbb{Z})$. What we mean by an area-π standard symplectic disk bundle of (Q, ω_Q) is an area-1 standard symplectic disk bundle of $(Q, \omega_Q/\pi)$ with its symplectic form multiplied by π.

Proof The symplectomorphism $D^*\mathbb{RP} \setminus Z_{\mathbb{RP}} \simeq U^*\mathbb{RP} \times (0,1]_r$ induces a symplectomorphism of the boundary reductions of both sides. We note that

$$d(r\widetilde{\theta}) = \underline{\mathrm{pr}}^*\omega_Q - d((1-r)\widetilde{\theta}) \tag{4-2}$$

on $U^*\mathbb{RP} \times (0,1)$, where we define $\widetilde{\theta} = \mathrm{pr}^*\theta$ for clarity.

Here we use the maps in the following commutative diagram, where Pr and pr are the obvious projections, and $U^*\mathbb{RP}^2 \to Q$ is the symplectic reduction map:

$$\begin{tikzcd}
U^*\mathbb{RP} \times [0,\infty)_\rho \arrow[dr, "\mathrm{Pr}"] \arrow[ddr, "\underline{\mathrm{Pr}}"'] & & U^*\mathbb{RP} \times (0,1)_r \arrow[dl, "\mathrm{pr}"'] \arrow[ddl, "\underline{\mathrm{pr}}"] \\
& U^*\mathbb{RP}^2 \arrow[d] & \\
& Q &
\end{tikzcd}$$

Notice that the map $U^*\mathbb{RP} \to Q$ has the structure of a principal $U(1) = \mathbb{R}/\mathbb{Z}$ bundle structure using the Reeb flow of the contact form θ/π, and the associated complex line bundle $\mathscr{L}$ is precisely the fiberwise blow-down of $U^*\mathbb{RP} \times [0,\infty)_\rho$ with respect to its canonical projection $\underline{\mathrm{Pr}}$ to Q. The integral Chern class of this complex line bundle is ω_Q/π (using that $H_*(Q,\mathbb{Z})$ has no torsion) and a transgression 1–form is given by the pull-back of $-\theta/\pi$ by Pr. By definition, the open unit disk bundle $\rho < 1$ inside $\mathscr{L}$, endowed with the symplectic form

$$\underline{\mathrm{Pr}}^*\left(\frac{\omega_Q}{\pi}\right) + d\left(\rho^2 \mathrm{Pr}^*\left(-\frac{\theta}{\pi}\right)\right), \tag{4-3}$$

is an area-1 standard symplectic disk bundle of $(Q, \omega_Q/\pi)$. Therefore, if we multiply this form by π, we obtain an area-π standard symplectic disk bundle of (Q, ω_Q).

Now consider the following commutative diagram:

$$\begin{tikzcd} U^*\mathbb{RP}\times[0,1)_\rho \arrow[r, "r=1-\rho^2"] \arrow[d] & U^*\mathbb{RP}\times(0,1]_r \arrow[d] \\ \{\rho<1\}\subset\mathscr{L} \arrow[r, "F"] & \overline{U^*\mathbb{RP}\times(0,1]} \end{tikzcd}$$

Here the left map is the restriction of the fiberwise blowdown map and the right map is the boundary reduction map. The upper map sends $(x,\rho)\in U^*\mathbb{RP}\times[0,1)_\rho$ to $(x,1-\rho^2)$, and it is a homeomorphism overall as well as a diffeomorphism of the interiors. By construction of symplectic boundary reduction we deduce that there is a canonical diffeomorphism F making this diagram commutative.

It automatically follows from comparing equations (4-2) and (4-3) that F is a symplectomorphism. Composing F with the symplectomorphism from the very beginning of this proof yields the desired symplectomorphism. □

Combined with Corollary 4.2, the following finishes the proof of Theorem 1.3.

Proposition 4.5 *$\overline{D^*\mathbb{RP}^2}\setminus(S\cup Z_{\mathbb{RP}})$ is symplectomorphic to $\mathscr{E}(\pi,4\pi)\setminus\mathscr{C}$.*

Proof We use our Proposition 4.3 and [26, Lemma 2.1] to find an explicit symplectomorphism from the complement of $\mathbb{D}_{\gamma_0}$ in $\overline{D^*\mathbb{RP}^2}\setminus Z_{\mathbb{RP}}$ to $\mathscr{E}(\pi,4\pi)$. Here we use a symplectomorphism between $Q\setminus\{\gamma_0\}$ and the two-dimensional open ellipsoid of area 4π which sends $-\gamma_0$ to the origin, so that $\mathbb{D}_{-\gamma_0}$ is sent to $\mathscr{C}$ by this symplectomorphism. □

4.1 Proof of Proposition 4.1

We will freely use the canonical identification of $\mathbb{CP}^n(2\pi)$ with $(\mathbb{C}^{n+1}\setminus\{0\})/\mathbb{C}^*$. Note that the homogenous coordinates $[z_1:\cdots:z_{n+1}]$ denote the class $[z_1e_1+\cdots+z_{n+1}e_{n+1}]$, where e_i is the standard basis of $\mathbb{R}^{n+1}$ and its complexification $\mathbb{C}^{n+1}$. We also realize $T^*\mathbb{RP}^n$ as

$$\{(q,p)\in S^n\times\mathbb{R}^{n+1}\mid\langle q,p\rangle=0\}/\{\pm1\},$$

where $S^n\subset\mathbb{R}^{n+1}$ is the unit sphere. It is a straightforward computation that the standard symplectic form on $T^*\mathbb{RP}^n$ descends from the restriction of $\sum_{i=1}^{n+1}dp_i\wedge dq_i$ on $\mathbb{R}^{2n+2}$ under this identification. Note also that $K^\sharp([q,p])=|p|$ away from the zero-section.

In [25, Lemma 3.1], Oakley and Usher considered the map

$$\Phi\colon D^*\mathbb{RP}^n\to\mathbb{CP}^n(2\pi)$$

defined by

$$\Phi([q,p]):=\Big[\sqrt{f(|p|)}\,p+\frac{i}{\sqrt{f(|p|)}}q\Big], \tag{4-4}$$

where $f(x) = (1-\sqrt{1-x^2})/x^2$ on $(0,1]$ and $f(0) = \frac{1}{2}$. They proved that $\Phi|_{\mathrm{int}(D^*\mathbb{RP}^n)}$ is a symplectomorphism onto its image $\mathbb{CP}^n(2\pi) \setminus FQ_n$, where

$$FQ_n = \Bigl\{[z_0 : \cdots : z_n] \in \mathbb{CP}^n(2\pi) \Bigm| \sum z_k^2 = 0\Bigr\}$$

is the Fermat quadric. As before, we will denote by $\overline{D^*\mathbb{RP}^n}$ the boundary reduction of $D^*\mathbb{RP}^n$, and by $Z_{\mathbb{RP}^n}$ the zero-section. We have the following, which implies Proposition 4.1:

Proposition 4.6 *The Oakley–Usher map $\Phi\colon D^*\mathbb{RP}^n \to \mathbb{CP}^n(2\pi)$ descends to a symplectomorphism*

$$\bar{\Phi}\colon \overline{D^*\mathbb{RP}^n} \to \mathbb{CP}^n(2\pi).$$

Proof Note that $\overline{D^*\mathbb{RP}^n}$ is canonically homeomorphic to the quotient $D^*\mathbb{RP}^n/\sim$, where $x \sim y$ if $x, y \in U^*\mathbb{RP}^n$ and they are in the same orbit of the geodesic flow. Therefore, it is easy to see from the computations of Oakley–Usher that Φ descends to a bijective continuous map $\bar{\Phi}\colon \overline{D^*\mathbb{RP}^n} \to \mathbb{CP}^n(2\pi)$. We will show that this map is a symplectomorphism.

To do so, it suffices to show that $\bar{\Phi}$ is smooth. Indeed, if $\bar{\Phi}$ is smooth, then by continuity it follows that $\bar{\Phi}$ preserves the symplectic form. This in particular shows that $\bar{\Phi}$ is an immersion, and hence a diffeomorphism that preserves the symplectic forms.

The following point is crucial. The canonical (linear) action of the group $G = \mathrm{SO}(n+1)$ on $\mathbb{R}^{n+1}$ induces an action on $D^*\mathbb{RP}^n$ (and in turn on $\overline{D^*\mathbb{RP}^n}$) and $\mathbb{C}^{n+1}$ (and in turn on $\mathbb{CP}^n(2\pi)$). It is clear from the definition (4-4) that $\bar{\Phi}$ is G–equivariant with respect to these actions.

We first prove smoothness in the case $n = 1$. Equip $\mathbb{CP}^1(2\pi)$ with the induced Riemannian metric (the so called Fubini–Study metric), which makes it isometric to a round S^2. We will use the fact that the image of a linear Lagrangian subspace in $\mathbb{C}^2 \setminus \{0\}$ (with standard Kähler structure) under the canonical projection to $\mathbb{CP}^1(2\pi)$ is a geodesic circle. We will denote $\mathbb{CP}^1(2\pi)$ by $\mathbb{CP}^1$ for brevity.

Note that $Z_{\mathbb{RP}^1}$ is sent to the real part $L_{\mathbb{RP}^1} \subset \mathbb{CP}^1$ under $\bar{\Phi}$. Moreover, $\overline{D^*\mathbb{RP}^1} \setminus \mathrm{int}(D^*\mathbb{RP}^1)$ consists of two points which map to $[1 : \pm i]$. Finally, notice that the images of cotangent fibers (that is, line segments of constant q) are sent to geodesic segments connecting $[1 : i]$ and $[1 : -i]$. It is easy to see that these geodesics are orthogonal to the geodesic circle $L_{\mathbb{RP}^1}$. Also recall that $\bar{\Phi}$ is $\mathrm{SO}(2)$–equivariant as explained above.

Let $(D^*\mathbb{RP}^1)^+ := \{[q, p] \mid p_1q_2 - p_2q_1 \le 0\} \subset D^*\mathbb{RP}^1$. Note that $K^\sharp(q, p) = |p|$ is a smooth function when restricted to $(D^*\mathbb{RP}^1)^+$. On $(D^*\mathbb{RP}^1)^+$ the symplectic form is easily computed to be $dK^\sharp \wedge d\theta$, where $\theta\colon (D^*\mathbb{RP}^1)^+ \to \mathbb{R}/\pi\mathbb{Z}$ is defined by the relation $[q, p] = [\cos(\theta), \sin(\theta), -\sin(\theta)|p|, \cos(\theta)|p|]$.

Let $\mathbb{D} \subset \mathbb{R}^2$ be the unit disk $|x|^2 + |y|^2 \le 1$ with the symplectic form $dx \wedge dy$. It is well-known that there is a symplectic embedding $\mathbb{D} \hookrightarrow \mathbb{CP}^1$ which sends

- the origin to $[1 : i]$,
- the unit circle to $L_{\mathbb{RP}^1}$,

- radial rays to the geodesic segments that connect $L_{\mathbb{RP}^1}$ and $[1:i]$, in a way that preserves angles at the intersections, and
- finally, the disks centered at the origin to balls around $[1:i]$ (in the Fubini–Study metric).

We now consider the induced continuous map $\Pi\colon (D^*\mathbb{RP}^1)^+ \to \mathbb{D}$ defined by the commutative diagram

$$\begin{tikzcd}
(D^*\mathbb{RP}^1)^+ \arrow[r, "\Phi"] \arrow[dr, "\Pi"'] & \mathbb{CP}^1 \\
& \mathbb{D} \arrow[u, hook]
\end{tikzcd}$$

Let (ρ, ϕ) denote polar coordinates on $\mathbb{D}\setminus\{0\}$. We deduce, from our discussion thus far, the following facts:

- There exists a constant c such that $\Pi^*\phi = -2\theta + c$.
- There exists a function $h\colon [0,1) \to (0,1]$ such that
$$\Pi^*\rho = h \circ K^\sharp.$$
- π restricts to a symplectomorphism $(D^*\mathbb{RP}^1)^+ \setminus U^*\mathbb{RP}^1 \to \mathbb{D}\setminus\{0\}$.

These facts imply that h satisfies the differential equation $h'(x)h(x) = -\frac{1}{2}$, which, with the initial condition $h(0) = 1$, has the unique solution $h(x) = \sqrt{1-x}$.[1] We thus observe that the map $\Pi : (D^*\mathbb{RP}^1)^+ \to \mathbb{D}$ is a model for the boundary reduction of $(D^*\mathbb{RP}^1)^+$ at $U^*\mathbb{RP}^1$; see [30, equation (3.1)]. This proves that the map $\overline{\Phi}$ is smooth at both points of $\overline{D^*\mathbb{RP}^1} \setminus \mathrm{int}(D^*\mathbb{RP}^1)$ as we can repeat the same argument on the other half of $\overline{D^*\mathbb{RP}^1}$.

We now move on to the case $n > 1$. We start with a preliminary lemma.

Lemma 4.7 *Let G be a Lie group acting smoothly on M and N. Suppose that S is a smooth submanifold of M and that the multiplication map $G \times S \to M$ is a surjective submersion. If $\phi\colon M \to N$ is a G–equivariant map and $\phi|_S$ is smooth, then ϕ is also smooth.*

Proof We have the following commutative diagram in which each map is known to be smooth except the bottom map:

$$\begin{tikzcd}
G \times S \arrow[r, "\mathrm{id}\times\phi"] \arrow[d, "\mu_M"] & G \times N \arrow[d, "\mu_N"] \\
M \arrow[r, "\phi"] & N
\end{tikzcd}$$

Since the left map is smooth surjective submersion, it has local smooth sections $M \supset U \to G \times S$. The commutativity then implies that ϕ is smooth. $\square$

[1] This form of h can also be recovered from the explicit forms of Φ and the embedding $\mathbb{D} \hookrightarrow \mathbb{CP}^1$.

Note that the orbits of $G = \mathrm{SO}(n+1)$ on $\overline{D^*\mathbb{RP}^n}$ are the submanifolds of constant $|p|$. Fix any unoriented geodesic circle γ in $\mathbb{RP}^n$ with its round metric g. We obtain an embedding of $T^*\mathbb{RP}^1$ in $T^*\mathbb{RP}^n$ by taking points (q, p) such that $q \in \gamma$ and $p = g_q(v, \cdot)$, where v is tangent to γ. This restricts to a smooth embedding of $D^*\mathbb{RP}^1$ into $D^*\mathbb{RP}^n$, and of $\overline{D^*\mathbb{RP}^1}$ into $\overline{D^*\mathbb{RP}^n}$ (the last point is particularly clear in the description of the boundary reductions at hand as in the proof of Proposition 4.3, which we keep in mind for the next point as well.)

It is easy to see that the multiplication map $G \times \overline{D^*\mathbb{RP}^1} \to \overline{D^*\mathbb{RP}^n}$ is indeed a surjective submersion using that $G \times T^*\mathbb{RP}^1 \to T^*\mathbb{RP}^n$ is one. Applying the lemma with $S = \overline{D^*\mathbb{RP}^1}$, smoothness of $\overline{\Phi}$ follows from the smoothness of the $n = 1$ case $\overline{\Phi}_1 \colon D^*\mathbb{RP}^1 \to \mathbb{CP}^1$. $\square$

Remark 4.8 In [29, Chapter V], Seade gives a description of $\mathbb{CP}^n$ as a double mapping cylinder via the natural $\mathrm{SO}(n+1)$ action. One may follow this discussion to obtain the corresponding description of $\overline{D^*\mathbb{RP}^n}$, and that the map Φ factors as the normal exponential map of $\mathbb{RP}^n \hookrightarrow \mathbb{CP}^n$ (with respect to the Fubini–Study metric) composed with the map $D^*\mathbb{RP}^n \to T^*\mathbb{RP}^n \simeq N\mathbb{RP}^n$ induced by $|p| \mapsto \frac{1}{2}\sin^{-1}|p|$; note that $xf(x) = \tan\left(\frac{1}{2}\sin^{-1} x\right)$ and that the focal set of $\mathbb{RP}^n$ is precisely FQ_n. This yields an alternative construction of Φ as well as its extension $\overline{\Phi}$.

5 Lipschitz symplectic embeddings of balls

In this section, we aim to prove Theorems 1.6 and 1.7 from the introduction. To begin, we introduce some slightly more general notation, in which we also vary the radius of the cylinder. Suppose that $\Phi \colon B(\pi R^2) \to \mathbb{R}^4$ is a symplectic embedding with Lipschitz constant $L > 0$. Then we may set

$$E(\Phi, r) := \Phi^{-1}(\mathbb{R}^4 \setminus Z(\pi r^2)).$$

Recall now that Theorem 1.6 is the statement that $\mathrm{Vol}_4(E(\Phi, 1)) \geq c/L^2$ for some constant $c = c(R) > 0$.

Proof of Theorem 1.6 Let δ be any number strictly between 0 and $R-1$. Observe that by the Lipschitz bound, we have

$$N_{\delta/L}(E(\Phi, 1+\delta)) \subset E(\Phi, 1).$$

Applying Theorem 3.1 with

$$E = E(\Phi, 1+\delta)$$

to the symplectic embedding

$$\Phi|_{B(\pi R^2)\setminus E(\Phi,1+\delta)} \colon B(\pi R^2) \setminus E(\Phi, 1+\delta) \to Z(1+\delta),$$

we obtain

$$\mathrm{Vol}_4(E(\Phi, 1)) \geq \pi^2(R^2 - (1+\delta)^2)\frac{\delta^2}{L^2} - o\left(\frac{\delta^2}{L^2}\right),$$

which implies the desired bound after fixing some value for δ, for instance $\delta = \frac{1}{2}(R-1)$. $\square$

Remark 5.1 We may more generally ask about the volume of the region $E(\Phi, r)$ for $R > r$. By a scaling argument, we find that

$$\mathrm{Vol}_4(E(\Phi, r)) \geq \frac{r^4 c(R/r)}{L^2}.$$

Hence, we lose no information by restricting to the case $r = 1$.

Remark 5.2 The proof demonstrates that we may take $c(R) \sim R^4$ as R grows large.

On the other hand, we wish to prove the constructive bound, in which we must find an embedding $\Phi\colon B^4(R) \to \mathbb{R}^4$ of Lipschitz constant bounded by L such that

$$\mathrm{Vol}_4\left(E(\Phi)\right) \leq \frac{C}{L}$$

for some $C = C(R) > 0$. (Recall here that $E(\Phi) = E(\Phi, 1)$.)

Proof of Theorem 1.7 The fact that $\mathrm{Vol}_4(E(\Phi))$ can be made arbitrarily small if there is no restriction on the Lipschitz constant is exemplified by ideas of Katok [16]. Our proof is a quantitative refinement obtained using constructions which appear in symplectic folding. The basic idea is to break the ball $B(\pi R^2)$ into a number of cubes and Hamiltonian isotope each of these cubes into $Z(\pi)$, where the cubes are separated by walls of width $1/L$. To begin, we make three simplifications. First, we replace the domain $B(\pi R^2)$ with the cube $K(R) = [-R, R]^4$, as the volume defect can only increase in size. Second, we replace $K(R)$ with the rectangular prism

$$K'(R) = ([0, 4R^2] \times [0, 1]) \times ([0, 4R^2] \times [0, 1]),$$

where the parentheses indicate a symplectic splitting. Explicitly, the factors refer to the coordinates x_1, y_1, x_2, y_2, with symplectic form $\omega = dx_1 \wedge dy_1 + dx_2 \wedge dy_2$. Notice that the natural symplectomorphism between $K(R)$ and $K'(R)$ has Lipschitz constant $2R$, and in particular, the effect of replacing $K(R)$ with $K'(R)$ only affects our proposition by a factor of R which gets absorbed into the constant C. And finally, we allow the Lipschitz constant to be $O_R(L)$, by which we mean it is bounded by AL, where A is a constant which again depends upon R; we arrive at the proposition as stated by absorbing this constant in C.

Consider now each symplectic factor $[0, 4R^2] \times [0, 1] \subset \mathbb{R}^2$ of $K'(R)$. For $i \in \mathbb{Z}$ with $0 \leq i \leq 4R^2$, let X_i be the region in this rectangle with

$$i \leq x \leq i + 1 - \frac{1}{L}.$$

Our goal is to fit each $X_i \times X_j \subset K'(R)$ into $Z(\pi)$ — indeed, the complement of the union of these regions has volume $\Theta(R^4/L)$; the R^4 factor gets absorbed by the constant C.

To begin, there is an area-preserving map of the rectangle $[0, 4R^2] \times [0, 1]$ into $\mathbb{R}^2$ of Lipschitz constant at most $O(L)$ which translates X_i in the x–direction by i. That is, X_0 stays fixed, X_1 gets shifted to the right

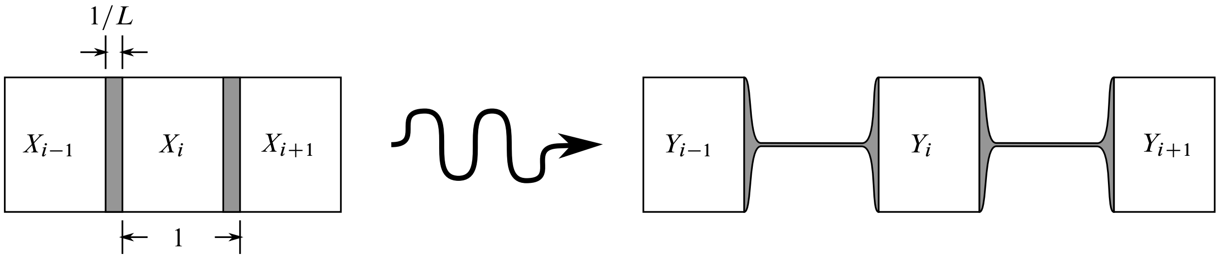


Figure 1: We stretch the region between X_i and X_{i+1} in an area-preserving way so that the images Y_i and Y_{i+1} are separated by distance just over 1.

by 1, and the region between them gets stretched like taffy in an area-preserving way to accommodate for this shift. See Figure 1. Let Y_i be the image of X_i under this map. Then Y_i and Y_{i+1} are separated by distance $1 + 1/L$, with

$$Y_i = \left[2i, 2i + 1 - \frac{1}{L}\right] \times [0, 1].$$

Explicitly, a model for the taffy-stretching map is given as $\phi\colon [0, 1/L] \times [0, 1] \to [0, 1 + 1/L] \times [0, 1]$ of the form

$$\phi(x, y) = \left(f(x), \tfrac{1}{2} + \frac{y - \frac{1}{2}}{f'(x)}\right)$$

(which is automatically area-preserving; see $D\phi$ computed below) with $f\colon [0, 1/L] \to [0, 1+1/L]$ a family of functions, depending upon L, such that

- $f'(x) = 1$ for x in an open neighborhood of the endpoints 0 and $1/L$,
- $f(0) = 0$ and $f(1/L) = 1 + 1/L$,
- $1 \le f'(x) \in O(L)$,
- $|f''(x)/(f'(x))^2| \in O(L)$.

The first two conditions imply that the constructed stretching map ϕ glues to the rigid translations of the X_i in a C^∞ manner. The latter two conditions imply the desired Lipschitz constant bound of $O(L)$, since

$$D\phi = \begin{pmatrix} f'(x) & 0 \\ -\left(y - \frac{1}{2}\right)\dfrac{f''(x)}{(f'(x))^2} & \dfrac{1}{f'(x)} \end{pmatrix},$$

with each entry in this matrix of order $O(L)$.

An example of such a desired function f may be given in the form[2]

$$f(x) = \int_0^x \frac{1}{1 - Cg(y)}\, dy,$$

[2] There are of course many such choices for function f, and we do not claim to make an optimal choice. In an earlier draft, we claimed that the Lipschitz constant of the stretching map could be made less than $2L$. We only need that the Lipschitz constant is $O(L)$ for our purposes, and it is unclear whether $2L$ can be achieved. We thank the reviewer for pointing out this lack of clarity.

where C is a constant dependent upon L and g (as we will explain soon), and where $g:[0,1/L]\to[0,1/4L]$ is a smooth function which is C^0–close to the continuous function

$$g_0(x)=\begin{cases} x & \text{if } 0\le x<1/4L,\\ 1/4L & \text{if } 1/4L\le x<3/4L,\\ 1/L-x & \text{if } 3/4L\le x\le 1/L.\end{cases}$$

We may take g so that $g(x)\equiv 0$ identically near the endpoints $x=0$ and $x=1/L$, and such that $|g'(x)|\le 1+\epsilon$ for any chosen $\epsilon>0$. The constant C is chosen so that $f(1/L)=1+1/L$, ie such that

$$I(C):=\int_0^{1/L}\frac{1}{1-Cg(y)}\,dy=1+\frac{1}{L}.$$

We claim that such a constant C exists. Notice that the value of the integral $I(C)$, as a function of C, is continuous and monotonically nondecreasing on the interval $[0,1/\sup\{g(x)\})$, with

$$I(0)=\frac{1}{L}<1+\frac{1}{L}<\infty=\lim_{C\to 1/\sup\{g(x)\}} I(C),$$

where the limit on the right follows because g attains its maximum value on the interior of the interval $[0,1/L]$, and because g is smooth, we have $g'=0$ at this maximum value. The existence of C now follows from the intermediate value theorem, and monotonicity implies uniqueness. We notice that because $\sup\{g(x)\}\approx 1/4L$ and $C<1/\sup\{g(x)\}$, we have that $C\in O(L)$.

With these choices, all of the conditions on f are now met, so long as we take a close enough approximation g of g_0 (where the closeness depends upon L). The first bullet point follows because $g(x)\equiv 0$ near the endpoints. The second follows because we chose C accordingly. The fourth is guaranteed since

$$\frac{f''(x)}{(f'(x))^2}=Cg'(x),$$

and we have constructed g so that $|g'(x)|\le 1+\epsilon$ and $C\in O(L)$.

That leaves the third bullet point, which we now verify. The fact that $f'(x)\ge 1$ is clear, so it suffices to check $f'(x)\in O(L)$. Solving for C_0 using g_0, we find

$$1+\frac{1}{L}=\int_0^{1/L}\frac{1}{1-C_0g_0(x)}\,dx>\int_{1/4L}^{3/4L}\frac{1}{1-C_0/4L}\,dx=\frac{1}{2L}\cdot\frac{1}{1-C_0/4L},$$

so that $C_0<4L(1-1/(2(L+1)))$. Notice that the value of C depends continuously on the function g (in the C^0–topology), so for any $\epsilon>0$ there is a choice of approximation g so that

$$\sup\{f'(x)\}=\sup\left\{\frac{1}{1-Cg(x)}\right\}<\sup\left\{\frac{1}{1-C_0g_0(x)}\right\}+\epsilon=\frac{1}{1-C_0/4L}+\epsilon<2(L+1)+\epsilon.$$

Hence, $f'(x)\in O(L)$, as required.

Applying the stretching map to each symplectic factor, each region $X_i\times X_j$ is sent to $Y_i\times Y_j$. It suffices now to find a symplectomorphism of $\mathbb{R}^4$ of Lipschitz constant $O_R(1)$ so that each $Y_i\times Y_j$ has image in the cylinder $Z(\pi)$, since in such a case, we compose with our stretching map, and each $X_i \times X_j$ under this composition lands in $Z(\pi)$, where the composition map has total Lipschitz constant $O_R(L)$. We construct this by sliding each $Y_i \times Y_j$ in two steps. We begin by separating the y_2–coordinates of the various blocks based on their x_1–coordinates. That is, we translate $Y_i \times Y_j$ in the y_2–direction by $2i$ units, in other words so that it gets translated to

$$\left[2i, 2i+1-\frac{1}{L}\right] \times [0,1] \times \left[2j, 2j+1-\frac{1}{L}\right] \times [2i, 2i+1].$$

Explicitly, we use a Hamiltonian of the form $H = -\rho(x_1)x_2$, where ρ is a step function with the properties

- $\rho(x_1) = 2i$ when $2i \le x_1 \le 2i+1$ and $0 \le i \le 4R^2$,
- $\|\rho'\|_\infty \in O(1)$,
- $\|\rho''\|_\infty \in O(1)$.

The corresponding Hamiltonian vector field X_H is of the form $2i\,\partial_{y_2}$ when $2i \le x_1 \le 2i+1$, and hence translates $Y_i \times Y_j$ as desired. Explicitly, the time-1 Hamiltonian flow is

$$\phi_H^1(x_1, y_1; x_2, y_2) = (x_1, y_1 + \rho'(x_1)x_2; x_2, y_2 + \rho(x_1)),$$

with derivative

$$D\phi_H^1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \rho''(x_1)x_2 & 1 & \rho'(x_1) & 0 \\ 0 & 0 & 1 & 0 \\ \rho'(x_1) & 0 & 0 & 1 \end{pmatrix}.$$

All terms are $O(1)$ except for $\rho''(x_1)x_2$, because x_2 can grow large. But on the image of our cube after the taffy-stretching step, x_2 is at most $4R^2$, and so the relevant Lipschitz constant of this sliding step is $O(R^2)$.

A similar construction, using a Hamiltonian of the form $-\rho(y_2)y_1$ for the same function ρ, allows us to then take each of these new blocks and translate them in the x_1–direction. After we complete both of these steps, the image of $Y_i \times Y_j$ is

$$\left[0, 1-\frac{1}{L}\right] \times [0,1] \times \left[2j, 2j+1-\frac{1}{L}\right] \times [2i, 2i+1].$$

A final translation simultaneously in the x_1 and y_1 coordinates by $-\frac{1}{2}$ lands each of these blocks in the cylinder $Z(\pi)$, concluding the proof. □

Remark 5.3 As in Remark 5.1, we may also vary the radius of the cylinder r, but where the new constant is $r^4 C(R/r)$.

Remark 5.4 The construction as presented has $C(R) \sim R^9$. Indeed, our volume defect came with a factor of R^4, but the Lipschitz constant has an extra factor of R^5. The first factor of R appearing in the Lipschitz constant came from replacing $K(R)$ with $K'(R)$. An extra two factors of R^2 came from our slide moves. One can optimize a little by performing a single diagonal slide move instead of two separate orthogonal slide moves. Hence, in the end, we may take $C(R) \sim R^7$. We suspect this is far from optimal, though decreasing the exponent appears to require a new idea.

6 Further questions

6.1 Minkowski dimension problem in higher dimensions

In this section we pose the simplest Minkowski dimension question that one could ask in dimensions higher than four, and make a couple of remarks about it. Let us assume $n > 2$ throughout this section.

Question 1 What is the smallest $d \in \mathbb{R}$ such that for some $A > \pi$, there exists a closed subset $E \subset B^{2n}(A)$ of Minkowski dimension d such that $B^{2n}(A) \setminus E$ symplectically embeds into $Z^{2n}(\pi)$?

Assume that for some $A > \pi$ and a closed subset $E \subset B^{2n}(A)$ of Minkowski dimension d, $B^{2n}(A) \setminus E$ symplectically embeds into $Z^{2n}(\pi)$. We find it plausible that a version of our obstructive argument would still give $d \geq 2$, even though we do not have a proof of this.

The argument in Proposition 2.3 suggests that the problem is related to the question of how the 2–width of a round ball changes after the removal of a closed subset. One can explicitly see that for E being the intersection of $B^{2n}(A)$ with a linear Lagrangian subspace L of $\mathbb{R}^{2n}$, an n–dimensional submanifold, there is a (holomorphic!) sweepout of $B^{2n}(A) \setminus E$ with width $\frac{1}{2}A$. Namely, we take the foliation by half-disks that are the connected components of the intersections with $B^{2n}(A) \setminus E$ of affine complex planes that intersect L nontransversely.

6.2 Capacity after removing a linear plane

Throughout this section let $B := B^4(2\pi) \subset \mathbb{C}^2$, where $\mathbb{C}^2$ is equipped with its standard Kähler structure. Let us denote the complex coordinates by x and y.

Let us denote by c_{Gr} the Gromov width, which is a capacity defined on any symplectic manifold Y^{2n} as the supremum of πr^2, where $B^{2n}(\pi r^2)$ symplectically embeds into Y.

Definition 6.1 Let $V \subset \mathbb{C}^2$ be a real subspace of dimension 2. We define the symplecticity of V as $|\omega_{\mathrm{st}}(e_1, e_2)|$, where e_1, e_2 is any orthonormal basis of $T_0 V$.

Notice that V is a Lagrangian plane if and only if its symplecticity is 0. On the other extreme, V is a complex plane if and only if its symplecticity is 1.

The symplecticity defines a surjective continuous function

$$\mathrm{Gr}_{\mathbb{R}}(2,4) \to [0,1].$$

Lemma 6.2 *Two elements of* $\mathrm{Gr}_{\mathbb{R}}(2,4)$ *have the same symplecticity if and only if there is an element of* $U(2)$ *sending one to the other.*

We can therefore define the function (symplecticity to capacity)

$$\mathrm{stc}\colon [0,1] \to (0,\infty)$$

by $s \mapsto c_{\mathrm{Gr}}(B \setminus V_s)$, where $V_s \in \mathrm{Gr}_{\mathbb{R}}(2,4)$ with symplecticity s.

The following remarkable statement is due to Traynor.

Proposition 6.3 (Traynor [31, Proposition 5.2]) *B is symplectomorphic to*

$$B \setminus \bigl(\{xy=0\} \cup \{(e^{it}x, y) \cup (x, e^{it}y) \mid \mathrm{Im}(x) = \mathrm{Im}(y) = 0,\ \mathrm{Re}(x) \ge 0,\ \mathrm{Re}(y) \ge 0,\ t \in [0,2\pi]\}\bigr).$$

Hence, we have that $\mathrm{stc}(1) = 2\pi$. On the other hand, it follows immediately from our Theorem 1.3 that $\mathrm{stc}(0) = \pi$. We finish with the obvious question.

Question 2 What is the function stc? Is it continuous?

6.3 Minkowski dimension problem for large R

Let $R_{\mathrm{sup}} \in (1,\infty]$ be the supremum of the radii R such that there is a closed subset E of Minkowski dimension 2 inside $B(\pi R^2)$ whose complement symplectically embeds into $Z(\pi)$. In Section 4, we showed that $R_{\mathrm{sup}} \le \sqrt{2}$. This inequality will be improved by Joé Brendel to $R_{\mathrm{sup}} \le \sqrt{3}$ using a different construction; see Remark 1.5. An intriguing aspect of both of the squeezing constructions is that they fail for large radii R. This motivates the following:

Question 3 Is R_{sup} a finite number?

6.4 Minkowski dimension problem for extendable embeddings

Here is a variant of our Minkowski dimension question, which is also more directly related to the Lipschitz number question. Assume $R > r$. What is the smallest Minkowski dimension of a subset $E \subset B^4(\pi R^2)$ with the property that for any neighborhood U of E, there is a symplectic embedding $B^4(\pi R^2) \to \mathbb{C}^2$ such that $B^4(\pi R^2) \setminus U$ maps inside $Z^4(\pi r^2)$?

Our obstructive Theorem 1.4 still gives a bound, but our construction in Section 4 does not apply. Recall $B := B^4(2\pi)$.

Proposition 6.4 *Consider the embedding $B \setminus L \hookrightarrow Z(\pi)$ that we constructed in Theorem 1.3. There exists an embedded circle η in $iL \cap B$ that is disjoint from L, and which maps into $\{x_1 = y_1 = 0\}$.*

Proof Recall that our symplectomorphism first sends $B \setminus L$ to $\mathbb{CP}^2(2\pi) \setminus (L_{\mathbb{RP}} \cup S_{[0:0:1]})$ via the restriction of a symplectomorphism $b\colon B \to \mathbb{CP}^2(2\pi) \setminus S_{[0:0:1]}$. The image of $L \cap B$ under b is $L_{\mathbb{RP}} \setminus S_{[0:0:1]}$, whereas the image of $iL \cap B$ is $iL_{\mathbb{RP}} \setminus S_{[0:0:1]}$, where we define

$$iL_{\mathbb{RP}} := \{[z_1 : z_2 : z_3] \mid \mathrm{Re}(z_1) = \mathrm{Re}(z_2) = \mathrm{Im}(z_3) = 0\}.$$

FQ and $iL_{\mathbb{RP}}$ intersect along the circle

$$\widehat{C} = \{[i\cos\theta : i\sin\theta : 1] \mid \theta \in [0, 2\pi]\}.$$

Note that $\widehat{C}$ is disjoint from $L_{\mathbb{RP}} \cup S_{[0:0:1]}$.

Now recall that to complete our symplectomorphism from $B \setminus L$ to $\mathscr{E}\underline{(\pi, 4\pi)} \setminus \mathscr{C}$ we use the Oakley–Usher symplectomorphism between $\mathbb{CP}^2(2\pi) \setminus (L_{\mathbb{RP}} \cup S_{[0:0:1]})$ and $\overline{D^*\mathbb{RP}^2} \setminus (Z_{\mathbb{RP}} \cup S)$, and then the Opshtein symplectomorphism.

From Opshtein's formula we see that all the points on $Q \setminus (Z_{\mathbb{RP}} \cup S)$ and therefore on $FQ \setminus (L_{\mathbb{RP}} \cup S_{[0:0:1]})$ are sent to points in $\{x_1 = y_1 = 0\}$ in $\mathscr{E}(\pi, 4\pi)$. This means that $b^{-1}(\widehat{C})$ satisfies the condition in the statement. $\square$

Corollary 6.5 *Let U be a neighborhood of L that is disjoint from η. Then, we cannot extend the restriction to $B \setminus U$ of our symplectic embedding $B \setminus L$ into $Z(\pi)$ to a symplectic embedding $B \to \mathbb{C}^2$.*

Proof If there were such an embedding, the action of the image of η would have to be simultaneously zero and nonzero, which is absurd. $\square$

6.5 Bounds on Minkowski content of the defect region

We have shown in Corollary 3.2 that the lower Minkowski dimension bound $\underline{\dim}_{\mathcal{M}}(E) \geq 2$ is optimal in the range $R \in (1, \sqrt{2}]$. Is the estimate on the 2–content $\underline{\mathcal{M}}_2(E) \geq \pi(R^2 - 1)$ also sharp? That is, does there exist $R \in (1, \sqrt{2}]$ and E with $\underline{\mathcal{M}}_2(E) = \pi(R^2 - 1)$ such that $B(\pi R^2) \setminus E$ symplectically embeds into $Z(\pi)$?

6.6 Speculations on the Lipschitz question

Consider the volume loss function

$$\mathrm{VL}(L, R) := \inf\{\mathrm{Vol}(E(\Phi))\},$$

where the infimum is taken over all symplectic embeddings $\Phi\colon B^4(\pi R^2) \to \mathbb{R}^4$ with Lipschitz constant at most L, and

$$E(\Phi) = \Phi^{-1}(\mathbb{R}^4 \setminus Z^4(\pi)).$$

In Section 5, we proved Theorems 1.6 and 1.7, which may be summarized by the statement that

$$\frac{c(R)}{L^2} \leq \mathrm{VL}(L, R) \leq \frac{C(R)}{L},$$

where we tacitly assume $R > 1$. Even more, we noted in Remarks 5.2 and 5.4 that our methods show that we may take $c(R) \sim R^4$ and $C(R) \sim R^7$.

One natural question is whether we can bring the two bounds closer together. In terms of factors of R, we suspect that the asymptotics should indeed be $\Theta(R^4)$, ie growing like the total volume, though we could not find constructions which remain under this upper bound for large L. As for the factors of L, the jury is very much out. We nonetheless formulate a precise conjecture.

Conjecture 1 *For any $\ell, r > 0$, the limit*

$$\lim_{R\to\infty}\lim_{L\to\infty}\frac{\mathrm{VL}(\ell L, rR)}{\mathrm{VL}(L, R)}$$

exists and is positive.

The second question is to what extent our methods work in higher dimensions. On the constructive side, we may simply take our constructed symplectic embedding in 4 dimensions and extend it to $2n$ dimensions by acting by the identity on the symplectically complementary $2n-4$ dimensions. This yields a constructive bound of $\Theta(R^{2n+3}/L)$. As for the obstructive bound, a modification of the techniques presented in this paper, in which we obtain a sweepout instead of a foliation if we follow Gromov's nonsqueezing argument, should probably yield a bound of $O(R^{2n}/L^{2n-2})$.

Finally, we describe a quantity which we believe could be interesting. Although we are working with balls $B^4(\pi R^2)$, we could in principle replace these with other subdomains in $\mathbb{R}^4$. To be precise, suppose we fix $X \subset \mathbb{R}^4$ a bounded domain. Consider the generalized volume loss function

$$\mathrm{VL}_X(L, R) := \inf\{\mathrm{Vol}_4(E(\Phi))\},$$

where the infimum is taken over all symplectic embeddings $\Phi\colon RX \to \mathbb{R}^4$ of Lipschitz embedding at most L, and we take

$$E(\Phi) := \Phi^{-1}(\mathbb{R}^4 \setminus Z^4(\pi)).$$

In the case $X = B^4(\pi)$, we recover the usual volume loss function above.

We offer the following reasonable-looking conjecture.

Conjecture 2 *For all bounded domains X, the limit*

$$s_X := \lim_{R\to\infty}\lim_{L\to\infty}\frac{\mathrm{VL}_X(L, R)}{\mathrm{VL}(L, R)}$$

exists and is strictly positive.

Notice that if there exists a symplectic embedding $rX \hookrightarrow Y$ of Lipschitz constant ℓ, then

$$\mathrm{VL}_X(\ell L, rR) \le \mathrm{VL}_Y(L, R).$$

Should Conjecture 2 hold, one might hope to use this inequality to compare the values of s_X and s_Y for bounded domains X and Y.

Appendix Squeezing and degenerations of the complex projective plane by Joé Brendel

A.1 Introduction and main theorem

Our goal is to show the following.

Theorem A.1 *For every $\alpha < 3$, there is a set $\Sigma \subset B^4(\alpha)$ of Minkowski dimension 2 such that $B^4(\alpha) \setminus \Sigma$ symplectically embeds into $Z^4(1) = \mathbb{R}^2 \times D^2(1) \subset \mathbb{R}^4$.*

Our notation corresponds to that of the main body of the text by setting $\alpha = \pi R^2$. The idea of the proof is to view $B^4(\alpha)$ as $\mathbb{CP}^2(\alpha) \setminus \mathbb{CP}^1$ and to use almost toric fibrations of $\mathbb{CP}^2$. As observed in [32], for every *Markov triple* (a, b, c), there is a triangle $\Delta_{a,b,c}(\alpha) \subset \mathbb{R}^2$ and an almost toric fibration on $\mathbb{CP}^2(\alpha)$ with a base diagram whose underlying polytope is $\Delta_{a,b,c}(\alpha)$. Now note that the toric moment map image of $Z^4(1)$ is the half-strip $\mathcal{S} = \mathbb{R}_{\geqslant 0} \times [0, 1)$. We shall show that if the triangle $\Delta_{a,b,c}(\alpha)$ fits into $\mathcal{S}$ (after applying an integral affine transformation), then there is a symplectic embedding of $\mathbb{CP}^2(\alpha)$ into $Z^4(1)$ at the cost of removing a certain subset Σ' from $\mathbb{CP}^2(\alpha)$. The point here is that one can get a good understanding of the subset one needs to remove. Indeed, we show that Σ' is a union of three Lagrangian pinwheels (defined as in [7]) and a symplectic torus. In particular, this set has Minkowski dimension 2. A combinatorial argument shows that for every $\alpha < 3$, there is a Markov triple (a, b, c) and an inclusion $\Delta_{a,b,c}(\alpha) \subset \mathcal{S}$; see Lemma A.5.

Remark A.2 As was pointed out to us by Leonid Polterovich, our results can be combined with Gromov's nonsqueezing to show that any symplectic ball $B^4(1+\varepsilon) \subset \mathbb{CP}^2(\alpha)$ intersects the set $\Sigma' \subset \mathbb{CP}^2$ discussed above. See Corollary A.9 for more details.

Remark A.3 The same strategy may work to produce symplectic embeddings

$$D^2(\alpha) \times D^2(\alpha) \setminus \Sigma \hookrightarrow Z^4(1)$$

of the polydisk of capacity $\alpha < 2$ minus a union of some two-dimensional manifolds into the cylinder. Indeed, one can view the polydisk as the affine part of $S^2 \times S^2$ and use almost toric fibrations of the latter space to carry out the same argument.

The relationship between Markov triples and the complex and symplectic geometry of $\mathbb{CP}^2$ has generated a lot of interest in recent years. It first appeared in the work of Galkin and Usnich [8], who conjectured that for every Markov triple there is an exotic Lagrangian torus in $\mathbb{CP}^2$. This conjecture was proved and generalized by Vianna [32; 33] by the use of almost toric fibrations; see also Symington [30]. On the algebrogeometric side, Hacking and Prokhorov [13] showed that a complex surface X with quotient singularities admits a $\mathbb{Q}$–Gorenstein smoothing to $\mathbb{CP}^2$ if and only if X is a weighted projective space $\mathbb{CP}(a^2, b^2, c^2)$ and (a, b, c) forms a Markov triple. In [7], Evans and Smith studied embeddings of *Lagrangian pinwheels* into $\mathbb{CP}^2$. This is directly related to [13], since Lagrangian pinwheels appear naturally as vanishing cycles of the smoothings of $\mathbb{CP}(a^2, b^2, c^2)$ to $\mathbb{CP}^2$. See also the recent work by Casals and Vianna [5] and the forthcoming paper joint with Mikhalkin and Schlenk [4] for other applications of almost toric fibrations to symplectic embedding problems.

Acknowledgements We heartily thank Jonny Evans, Grisha Mikhalkin, Leonid Polterovich, Kevin Sackel and Umut Varolgunes for valuable discussions and careful comments on a previous draft. We are grateful to the attentive referee for pointing out a gap in the proof. We thank Felix Schlenk for constant encouragement and generous help.

A.2 Some geometry of Markov triangles

Let us recall some facts about Markov numbers and their associated triangles.

Definition A.4 A triple of natural numbers $a, b, c \in \mathbb{N}_{>0}$ is called a *Markov triple* if it solves the *Markov equation*

$$a^2 + b^2 + c^2 = 3abc. \tag{A-1}$$

If (a, b, c) is a Markov triple, then so is $(a, b, 3ab - c)$. Starting from the solution $(1, 1, 1)$, we obtain the so-called *Markov tree* by mutations $(a, b, c) \to (a, b, 3ab - c)$. The first few Markov triples are

$$(1,1,1), \quad (1,1,2), \quad (1,2,5), \quad (1,5,13), \quad (2,5,29), \quad (1,13,34), \quad \dots. \tag{A-2}$$

Given $\alpha > 0$, let $\mathbb{CP}^2$ be equipped with the Fubini–Study symplectic form ω normalized so that $\int_{\mathbb{CP}^1} \omega = \alpha$. For every Markov triple (a, b, c), there is an almost toric fibration of $\mathbb{CP}^2$ with almost toric base diagram a rational triangle $\Delta_{a,b,c}(\alpha) \subset \mathbb{R}^2$, which we call the *Markov triangle* associated to the Markov triple (a, b, c). The first Markov triangle is the (honest) toric moment map image of $\mathbb{CP}^2$ in our normalization,

$$\Delta_{1,1,1}(\alpha) = \{(x, y) \in \mathbb{R}^2_{\geqslant 0} \mid x + y \leqslant \alpha\}. \tag{A-3}$$

In fact, we will slightly abuse notation and at times think of $\Delta_{a,b,c}(\alpha)$ as a triangle in $\mathbb{R}^2$ and at other times as the equivalence class of triangles $\mathbb{R}^2$ under the group of toric symmetries given by *integral affine transformations*, ie elements in $\mathrm{Aff}(2; \mathbb{Z}) = \mathbb{R}^2 \rtimes \mathrm{GL}(2; \mathbb{Z})$. For every mutation of Markov triples $(a, b, c) \to (a, b, 3ab - c)$, there is a corresponding mutation of triangles $\Delta_{a,b,c}(\alpha) \to \Delta_{a,b,3ab-c}(\alpha)$, defined by cutting the triangle in two halves and applying a shear map to one of the halves and gluing it back to the other half. This is called a *branch move* and we refer to Symington [30, Sections 5.3 and 6] and Vianna [32, Section 2] for details. For a concrete description of the triangles $\Delta_{a,b,c}(\alpha)$, see (A-23) and the surrounding discussion. The area of the Markov triangles is well-defined, since it is invariant under $\mathrm{Aff}(2;\mathbb{Z})$. Furthermore, the area is invariant under the mutation of triangles, and hence we obtain

$$\operatorname{area}(\Delta_{a,b,c}(\alpha)) = \operatorname{area}(\Delta_{1,1,1}(\alpha)) = \tfrac{1}{2}\alpha^2. \tag{A-4}$$

Recall that we are interested in *embedding Markov triangles into the half-strip* $\mathcal{S} = \mathbb{R}_{\geqslant 0} \times [0,1)$. This means that, given $\alpha > 0$, we look for a Markov triple (a,b,c) such that $\Delta_{a,b,c}(\alpha) \subset \mathcal{S}$ up to applying an element in $\mathrm{Aff}(2;\mathbb{Z})$. We will prove the following.

Lemma A.5 *For every $0 < \alpha < 3$, there is a Markov triple (a,b,c) such that the Markov triangle $\Delta_{a,b,c}(\alpha) \subset \mathbb{R}^2$ is in $\mathcal{S} = \mathbb{R}_{\geqslant 0} \times [0,1)$ up to applying an element in $\mathrm{Aff}(2;\mathbb{Z})$. This result is sharp in the sense that for $\alpha \geqslant 3$, there is no such Markov triple.*

Let us introduce some definitions from integral affine geometry. A vector $v \in \mathbb{Z}^2$ is called *primitive* if $\beta v \notin \mathbb{Z}^2$ for all $0 < \beta < 1$. Note that to every vector $w \in \mathbb{R}^2$ with rational slope, we can associate a unique primitive vector v such that $w = \gamma v$ for $\gamma > 0$. We call γ the *affine length* of w and denote it by $\ell_{\mathrm{aff}}(w)$. Let $l \subset \mathbb{R}^2$ be a rational affine line (or line segment) with primitive directional vector $v \in \mathbb{Z}^2$ and $p \in \mathbb{R}^2$ be a point. Then the *affine distance* is defined as $d_{\mathrm{aff}}(p,l) = |\det(v,u)|$, where u is any vector such that $p + u \in l$. This does not depend on any of the choices we have made and these quantities are $\mathrm{Aff}(2;\mathbb{Z})$–invariant. See McDuff [21] for more details.

For a given rational (not necessarily Markov) triangle Δ, we denote by E_1, E_2, E_3 its edges and by v_1, v_2, v_3 its vertices such that v_i lies opposite to the edge E_i. We call $\ell_{\mathrm{aff}}(E_1) + \ell_{\mathrm{aff}}(E_2) + \ell_{\mathrm{aff}}(E_3)$ the *affine perimeter* of Δ. Note that the affine perimeter of a Markov triangle $\Delta_{a,b,c}(\alpha)$ is equal to 3α. We have the following formula for the area of Δ,

$$\operatorname{area}(\Delta) = \tfrac{1}{2}\ell_{\mathrm{aff}}(E_i)\, d_{\mathrm{aff}}(v_i, E_i). \tag{A-5}$$

This follows from the definition of affine distance and affine length.

Proof of Lemma A.5 For the first part of the proof we only need one branch of the Markov tree, namely the one where the maximal entry grows the fastest. More precisely, let Δ_n be the Markov triangle associated to the triple (m_{n+2}, m_{n+1}, m_n), where m_k is recursively defined as

$$m_0 = m_1 = m_2 = 1, \quad m_{k+2} = 3m_{k+1}m_k - m_{k-1}. \tag{A-6}$$

The first few terms of this sequence are given by $\{m_k\}_{k\in\mathbb{N}} = \{1,1,1,2,5,29,433,\dots\}$. We clearly have $m_k \to \infty$. Recall from [32] that the affine side lengths of a Markov triangle $\Delta_{a,b,c}(\alpha)$ are given by $\lambda a^2, \lambda b^2, \lambda c^2$ for a proportionality constant $\lambda > 0$. Since the affine perimeter of the Markov triangle is 3α, we obtain $\lambda(a^2+b^2+c^2) = 3\alpha$. Together with the Markov equation (A-1) this yields that the longest edge E_n in $\Delta_n(\alpha)$ has affine length

$$\ell_{\mathrm{aff}}(E_n) = \frac{\alpha m_{n+2}}{m_{n+1}m_n} \overset{\text{(A-6)}}{=} \alpha\left(3 - \frac{m_{n-1}}{m_{n+1}m_n}\right). \tag{A-7}$$

Since the second summand goes to 0 for large n, we obtain $\ell_{\mathrm{aff}}(E_n) \to 3\alpha$. Let v_n be the vertex opposite to E_n. By (A-4) and the area formula (A-5) we obtain

$$\ell_{\mathrm{aff}}(E_n)\, d_{\mathrm{aff}}(v_n, E_n) = \alpha^2. \tag{A-8}$$

This implies that $d_{\mathrm{aff}}(v_n, E_n) \to \frac{1}{3}\alpha$. Now note that we can assume, up to $\mathrm{Aff}(2;\mathbb{Z})$, that the maximal edge E_n lies in $\mathbb{R}_{\geqslant 0} \times \{0\}$ and that $\Delta_n(\alpha)$ lies in the upper half-plane. Then the (Euclidean) height of $\Delta_n(\alpha)$ is equal to $d_{\mathrm{aff}}(v_n, E_n)$. This implies that for all $\alpha < 3$, we have $\Delta_n(\alpha) \subset \mathcal{S}$ for large enough $n \in \mathbb{N}$.

Now let $\alpha \geqslant 3$ and suppose that there is a Markov triangle $\Delta_{a,b,c}(\alpha) \subset \mathcal{S}$. Let E be the longest edge of $\Delta_{a,b,c}(\alpha)$ and v the opposite vertex. We first prove that E is parallel to $e_1 = (1,0)$. Indeed, suppose it is not. Then we can write $E = \ell_{\mathrm{aff}}(E)v$ for a primitive vector $v = (v_1, v_2) \in \mathbb{Z}^2$ with $v_2 \geqslant 1$. But since the affine perimeter of $\Delta_{a,b,c}(\alpha)$ is 3α and E is the longest edge, we obtain $\ell_{\mathrm{aff}}(E) \geqslant \alpha$ and thus $\Delta_{a,b,c}(\alpha)$ is not contained in $\mathcal{S}$. Now if E is parallel to e_1, then the (Euclidean) height of $\Delta_{a,b,c}(\alpha)$ is $d_{\mathrm{aff}}(v, E)$. The affine perimeter 3α is strictly larger than $\ell_{\mathrm{aff}}(E)$ from which we deduce $d_{\mathrm{aff}}(v, E) > \frac{1}{3}\alpha \geqslant 1$ by the area formula (A-5). □

A.3 Proof of the main theorem

Definition A.6 Let p be a positive integer. The topological space obtained from the unit disk D by quotienting out the action of the group of the p^{th} roots of unity on ∂D is called a *p–pinwheel*.

For example, the 2–pinwheel is $\mathbb{RP}^2$. The image of ∂D in the quotient is called the *core circle*. For all $p > 2$ the p–pinwheels are not smooth at points of the core circle. A *Lagrangian pinwheel* in a symplectic manifold M is a Lagrangian *embedding* of a p–pinwheel into M; see [7, Definition 2.3] for the meaning of *embedding* in this context. As it turns out, for every Lagrangian p–pinwheel, there is an additional extrinsic parameter $q \in \{1, \dots, p-1\}$ measuring the twisting of the pinwheel around its core circle. We call such an object (p,q)–pinwheel and denote it by $L_{p,q}$ when this causes no confusion. For us, the following result is key.

Proposition A.7 *Let (a,b,c) be a Markov triple and let $s_a, s_b, s_c \in \mathbb{CP}(a^2, b^2, c^2)$ be the orbifold singular points of the corresponding weighted projective space. Then there is a surjective map*

$$\tilde{\phi}\colon \mathbb{CP}^2 \to \mathbb{CP}(a^2, b^2, c^2)$$

and there are (mutually disjoint) Lagrangian pinwheels $L_{a,q_a}, L_{b,q_b}, L_{c,q_c} \subset \mathbb{CP}^2$ with

$$\tilde{\phi}(L_{a,q_a}) = s_a, \quad \tilde{\phi}(L_{b,q_b}) = s_b, \quad \tilde{\phi}(L_{c,q_c}) = s_c \tag{A-9}$$

such that $\tilde{\phi}$ restricts to a symplectomorphism

$$\phi\colon \mathbb{CP}^2 \setminus (L_{a,q_a} \sqcup L_{b,q_b} \sqcup L_{c,q_c}) \to \mathbb{CP}(a^2, b^2, c^2) \setminus \{s_a, s_b, s_c\}. \tag{A-10}$$

Furthermore, the preimage of the set $\mathfrak{D} \subset \mathbb{CP}(a^2, b^2, c^2)$ — see (A-11) — consists of the union of three Lagrangian pinwheels and a symplectic two-torus intersecting the pinwheels in their respective core circles.

Remark A.8 For the construction of this symplectomorphism, it seems plausible that one can use the existence of a $\mathbb{Q}$–Gorenstein smoothing of $\mathbb{CP}(a^2,b^2,c^2)$ to $\mathbb{CP}^2$ with vanishing locus consisting of a union of three pinwheels. Such smoothings were constructed in [13], by showing that there are no obstructions to piecing together local smoothings of the cyclic quotient singularities holomorphically. Our proof follows the same strategy in the symplectic set-up; see also [7, Examples 2.5–2.6] for a discussion of this, and [17, Section 3; 6, Section 1] for the local smoothings.

We now turn to the proof of Theorem A.1 using Proposition A.7, the proof of which we postpone to Section A.4.

Proof of Theorem A.1 **Step 1** Let $\alpha < 3$ and choose α' so that $\alpha < \alpha' < 3$. Pick a Markov triple (a,b,c) and an associated Markov triangle $\Delta_{a,b,c} = \Delta_{a,b,c}(\alpha')$ which lies in $\mathcal{S} = \mathbb{R}_{\geqslant 0} \times [0,1)$. This is possible by Lemma A.5. Note that $\Delta_{a,b,c}$ is the image of the toric orbifold moment map $\mu\colon \mathbb{CP}(a^2,b^2,c^2) \to \Delta_{a,b,c}$, provided we normalize the orbifold symplectic form appropriately. See the discussion surrounding (A-26) for details on the toric structure on weighted projective space. Let

$$\mathfrak{D} = \mu^{-1}(\partial\Delta_{a,b,c}) \subset \mathbb{CP}(a^2,b^2,c^2) \tag{A-11}$$

be the preimage of the boundary. The set $\mathfrak{D}$ is a union of complex suborbifolds

$$\mathbb{CP}(a^2,b^2) \cup \mathbb{CP}(b^2,c^2) \cup \mathbb{CP}(a^2,c^2)$$

such that each of these suborbifolds projects to one edge of the triangle $\Delta_{a,b,c}$. The complement of $\mathfrak{D}$ admits a symplectic embedding

$$\psi\colon \mathbb{CP}(a^2,b^2,c^2) \setminus \mathfrak{D} \hookrightarrow Z^4(1). \tag{A-12}$$

Indeed, this follows from the inclusion $\operatorname{int}(\Delta_{a,b,c}) \subset \operatorname{int}(\mathcal{S})$ and the fact that inclusions of toric moment map images which respect the boundary stratifications yield (equivariant) symplectic embeddings; see for example [30].

Step 2 By Proposition A.7, there is a symplectomorphism ϕ from the complement of Lagrangian pinwheels L_{a,q_a}, L_{b,q_b}, L_{c,q_c} to the complement of the orbifold points of $\mathbb{CP}(a^2,b^2,c^2)$. By restricting ϕ, we obtain the symplectomorphism

$$\phi'\colon \mathbb{CP}^2 \setminus \Sigma' \to \mathbb{CP}(a^2,b^2,c^2) \setminus \mathfrak{D}. \tag{A-13}$$

Again by Proposition A.7, the set Σ' consists of the union of three pinwheels and a symplectic two-torus.

Step 3 The standard embedding $B^4(\alpha) \subset \mathbb{CP}^2(\alpha')$ together with ϕ' and ψ yields an embedding $B^4(\alpha) \setminus \Sigma \hookrightarrow Z^4(1)$. Here Σ denotes $\Sigma = \Sigma' \cap B^4(\alpha)$. The set Σ has Minkowski dimension two. Indeed, the embedding $B^4(\alpha) \subset \mathbb{CP}^2(\alpha')$ is bilipschitz (its image being contained in a closed ball) and volume-preserving and the set Σ' consists of the union of three pinwheels and a symplectic two-torus. □

As was pointed out to us by Leonid Polterovich, one can combine Theorem A.1 with Gromov's non-squeezing theorem to get certain rigidity results, reminiscent of [2, Theorem 1.B].

Corollary A.9 *Let $\Sigma' \subset \mathbb{CP}^2(\alpha)$ be one of the above sets such that $\mathbb{CP}^2(\alpha) \setminus \Sigma'$ embeds into $Z^4(1)$ for some $1 < \alpha < 3$. Then every symplectic ball $B^4(1+\varepsilon) \subset \mathbb{CP}^2$ for $\varepsilon > 0$ intersects Σ'.*

Proof Assume $B^4(1+\varepsilon) \subset \mathbb{CP}^2(\alpha)$ does not intersect Σ'. The embedding $\mathbb{CP}^2(\alpha) \setminus \Sigma' \hookrightarrow Z^4(1)$ yields a symplectic embedding $B^4(1+\varepsilon) \hookrightarrow Z^4(1)$, contradicting nonsqueezing. □

Note that for a fixed $1 < \alpha < 3$ we get infinitely many sets $\Sigma' \subset \mathbb{CP}^2$ to which Corollary A.9 applies and all of these consist of a union of a symplectic torus and Lagrangian pinwheels.

A.4 Proof of Proposition A.7

Following the exposition in [7, Example 2.5] we consider smoothings of certain orbifold quotients of $\mathbb{C}^2$. This yields the local version from Lemma A.10 of the symplectomorphism in Proposition A.7.

Let a and q be coprime integers with $1 \leqslant q < a$ and take the quotient of $\mathbb{C}^2$ by the action of $(a^2)^{\text{th}}$ roots of unity

$$\zeta.(z_1, z_2) = (\zeta z_1, \zeta^{aq-1} z_2), \quad \text{where } \zeta^{a^2} = 1. \tag{A-14}$$

We denote this quotient by $\mathbb{C}^2/\Gamma_{a,q}$. It can be embedded as $\{w_1 w_2 = w_3^a\}$ into the quotient $\mathbb{C}^3/\mathbb{Z}_a$ by the action

$$\eta.(w_1, w_2, w_3) = (\eta w_1, \eta^{-1} w_2, \eta^q w_3), \quad \eta^a = 1. \tag{A-15}$$

The smoothing is given by

$$\mathscr{X} = \{w_1 w_2 = w_3^a + t\} \subset \mathbb{C}^3/\mathbb{Z}_a \times \mathbb{C}_t, \tag{A-16}$$

which we view as a degeneration by projecting to the t–component, $\pi : \mathscr{X} \to \mathbb{C}_t$. We denote the fibers by $X_t = \pi^{-1}(t)$. The smooth fiber X_1 is a rational homology ball and the vanishing cycle of the degeneration is a Lagrangian pinwheel $L_{a,q}$. This follows from the description of $\mathscr{X}$ as $\mathbb{Z}_a$–quotient of an A_{a-1}–Milnor fiber. Let $s \in X_0$ be the unique isolated singularity of X_0, and $\mathscr{X}^{\text{reg}} = \mathscr{X} \setminus \{s\}$ its complement. The restriction of the standard symplectic form ω_0 on $\mathbb{C}^4 = \mathbb{C}^3 \times \mathbb{C}_t$ yields a symplectic manifold $(\mathscr{X}^{\text{reg}}, \Omega)$. Note that the smooth loci of the fibers $X_{t \neq 0}$ and $X_0 \setminus \{s\}$ are symplectic submanifolds. Let us now construct a symplectomorphism

$$\psi : X_1 \setminus L_{a,q} \to X_0 \setminus \{s\} = \mathbb{C}^2/\Gamma_{a,q} \setminus \{0\}. \tag{A-17}$$

For this, we take the connection on $\mathscr{X}^{\text{reg}}$ defined as the symplectic complement to the vertical distribution,

$$\xi_x = (\ker(\pi_*)_x)^{\Omega} = \{v \in T_x \mathscr{X}^{\text{reg}} \mid \Omega(v, w) = 0 \text{ for all } w \in T_x \pi^{-1}(\pi(x))\}. \tag{A-18}$$

This connection is symplectic in the sense that its parallel transport maps are symplectomorphisms whenever they are defined. In particular, we get a symplectomorphism between any two regular fibers X_{t_1} and X_{t_2} by picking a curve in $\mathbb{C}^\times$ with endpoints t_1 and t_2. Since we are interested in the singular fiber for the construction of (A-17), take the curve $\gamma(r) = 1 - r \in \mathbb{C}$. For every $r < 1$, this yields a symplectomorphism $\psi^r \colon X_1 \to X_{1-r}$. As it turns out, setting $\psi(x) = \lim_{r\to 1} \psi^r(x)$ for $x \in X_1$ yields a well-defined surjective map $\widetilde{\psi}\colon X_1 \to X_0$ with vanishing cycle a Lagrangian pinwheel, $L_{a,q} = \widetilde{\psi}^{-1}(s)$, and which restricts to the desired symplectomorphism (A-17). For the fact that $\widetilde{\psi}$ is well-defined, see the unpublished notes by Evans [6, Lemma 1.2]. We refer to [17, Section 3.1] for more details on the specific degeneration we consider above and to [6, Lemma 1.20] for the fact that the vanishing cycle is a Lagrangian pinwheel.

Lemma A.10 (Evans [6]) *Let $\mathbb{C}^2/\Gamma_{a,q}$ be the quotient by the action (A-14) and $S_{a,q}$ its smoothing as above. Then there is a surjective map $\widetilde{\psi}_a\colon S_{a,q} \to \mathbb{C}^2/\Gamma_{a,q}$ with $\widetilde{\psi}_a(L_{a,q}) = \{0\}$ and which restricts to a symplectomorphism*

$$\psi_a\colon S_{a,q} \setminus L_{a,q} \to (\mathbb{C}^2/\Gamma_{a,q}) \setminus \{0\}. \tag{A-19}$$

Furthermore, the preimage of the set $\mathfrak{D}_{a,q_a} = \{z_1 z_2 = 0\}/\Gamma_{a,q}$ under $\widetilde{\psi}_a$ is the union of $L_{a,q}$ and a symplectic cylinder intersecting $L_{a,q}$ in its core circle.

Proof As explained above, the main statement of the lemma follows from [6]. We only need to identify the preimage of $\mathfrak{D}_{a,q_a} = \{z_1 z_2 = 0\}/\Gamma_{a,q}$, which can be done by keeping track of the parallel transport in the explicit model (A-16). Under the identification of $\mathbb{C}^2/\Gamma_{a,q}$ with X_0 (which we will tacitly use throughout), the set $\{z_1 z_2 = 0\}/\Gamma_{a,q}$ corresponds to $\{w_3 = 0\}/\mathbb{Z}_a \cap X_0$. We claim that the set $\bigcup_t (\{w_3 = 0\}/\mathbb{Z}_a \cap X_t) \subset \mathscr{X}$ is invariant under the symplectic parallel transport induced by (A-18). This proves that $\widetilde{\psi}_a^{-1}(\{w_3 = 0\}/\mathbb{Z}_a \cap X_0) = (\{w_3 = 0\}/\mathbb{Z}_a \cap X_1) \cup L_{a,q}$. Indeed, the preimage of $0 \in \mathbb{C}^2/\Gamma_{a,q}$ is $L_{a,q}$ and on the complement of $L_{a,q}$, the map $\widetilde{\psi}_a$ is a diffeomorphism. To see that $\{w_3 = 0\}/\mathbb{Z}_a \cap X_1$ is a symplectic cylinder intersecting $L_{a,q}$ in its core circle, one can consider the singular fibration structure of the map $X_s \to \mathbb{C}$ given by $(w_1, w_2, w_3) \mapsto w_3^p$; see [6, Section 1.2.3] for more details. Thus it remains to show the invariance of $\bigcup_t (\{w_3 = 0\}/\mathbb{Z}_a \cap X_t)$.

Let $x \in \{w_3 = 0\}/\mathbb{Z}_a \cap X_t$, meaning that x is a $\mathbb{Z}_a$–class of a point $(x_1, x_2, 0) \in \mathbb{C}^3$ with $x_1 x_2 = t$. This gives a natural inclusion $T_{(x,t)}\mathscr{X} = T_x X_t \oplus T_t\mathbb{C} \subset \mathbb{C}^3 \oplus \mathbb{C}$. Since $\Omega = \omega_{\mathbb{C}^3} \oplus \omega_{\mathbb{C}}$, the horizontal lift by the symplectic connection (A-18) of a vector $v \in T_t\mathbb{C} = \mathbb{C}$ is given by $u + v \in \xi_{(x,t)}$, where $u \in (T_x X_t)^{\omega_{\mathbb{C}^3}}$. Note that the subspace $\{u_1 = u_2 = 0\} \subset T_x\mathbb{C}^3 = \mathbb{C}^3$ is contained in $T_x X_t$ and hence its symplectic complement $\{u_3 = 0\} \subset \mathbb{C}^3$ contains the symplectic complement $(T_x X_t)^{\omega_{\mathbb{C}^3}}$. Since u is contained in $(T_x X_t)^{\omega_{\mathbb{C}^3}}$, this proves that u is tangent to the subset $\{w_3 = 0\}$ and hence this subset is preserved under parallel transport. □

We turn to the proof of Proposition A.7. The main idea is to use the fact that $\Delta_{a,b,c}(\alpha)$ is both the almost toric base polytope of $\mathbb{CP}^2(\alpha)$ and the toric base of $\mathbb{CP}(a^2, b^2, c^2)$ with a suitably normalized symplectic form. This shows that there is a symplectomorphism which intertwines the (almost) toric structures on the preimages of a complement of neighborhoods of the vertices. For example, one can choose $W \subset \Delta_{a,b,c}$ as in Figure 2. We use Lemma A.10 to extend this symplectomorphism.

Let us make a few preparations. In particular, we discuss how to use the quotient $\mathbb{C}^2/\Gamma_{a,q}$ as a local toric model for $\mathbb{CP}(a^2,b^2,c^2)$. This is the orbifold version of the toric ball embedding into $\mathbb{CP}^2$ one obtains by the inclusion of the simplex with one edge removed into the standard simplex $\Delta_{1,1,1}$. The toric structure on $\mathbb{C}^2/\Gamma_{a,q}$ is induced by the standard toric structure $(z_1,z_2)\mapsto(\pi|z_1|^2,\pi|z_2|^2)$ on $\mathbb{C}^2$. Indeed, note that the $\Gamma_{a,q}$–action is obtained by restricting the standard $T^2=(\mathbb{R}/\mathbb{Z})^2$–action to a discrete subgroup $\mathbb{Z}_{a^2}$. This implies that we obtain an induced action by $T^2/\Gamma_{a,q}$ on $\mathbb{C}^2/\Gamma_{a,q}$. This action is Hamiltonian and its moment map image, under a suitable identification $T^2\cong T^2/\Gamma_{a,q}$, is given by

$$\angle_{a,q}=\{x_1w_1+x_2w_2\mid x_1,x_2\geqslant 0\},\quad \text{where } w_1=\begin{pmatrix}1\\0\end{pmatrix} \text{ and } w_2=\begin{pmatrix}aq-1\\a^2\end{pmatrix}. \tag{A-20}$$

See for example [7, Remark 2.7] or [30, Section 9]. Note that a ball $B^4(d)=\{\pi(|z_1|^2+|z_2|^2)<d\}\subset\mathbb{C}^2$ quotients to an orbifold ball $B^4(d)/\Gamma_{a,q}\subset\mathbb{C}^2/\Gamma_{a,q}$, which is fibered by the induced toric structure on the quotient. Furthermore, the boundary sphere $S^3(d)\subset\mathbb{C}^2$ quotients to a lens space $\Sigma_{a,q}(d)=S^3(d)/\Gamma_{a,q}$ of type $(aq-1,a^2)$ equipped with its canonical contact structure and which fibers over a segment in $\angle_{a,q}$. We will use this fact in the proof of Proposition A.7.

Let us now show that, for a suitable choice of q, the toric system $\mathbb{C}^2/\Gamma_{a,q}\to\angle_{a,q}$ can be used as a local model around one of the orbifold points of the toric system on $\mathbb{CP}(a^2,b^2,c^2)$. In order to get a concrete description of this toric system on the weighted projective space, recall that the symplectic orbifold $\mathbb{CP}(a^2,b^2,c^2)$ can be defined as a symplectic quotient of $\mathbb{C}^3$,

$$\mathbb{CP}(a^2,b^2,c^2)=H^{-1}(a^2b^2c^2)/S^1,\quad \text{where } H=a^2|z_1|^2+b^2|z_2|^2+c^2|z_3|^2. \tag{A-21}$$

This description (A-21) has the advantage that it is naturally equipped with a Hamiltonian T^2–action inherited from the standard T^3–action on $\mathbb{C}^3$. This induced action is toric and its moment map image is given by the intersection of the plane defined by H and the positive orthant,

$$\widetilde{\Delta}_{a,b,c}=\{a^2y_1+b^2y_2+c^2y_3=a^2b^2c^2\}\cap\mathbb{R}^3_{\geqslant 0}. \tag{A-22}$$

Note that this is a polytope in $\mathbb{R}^3$ and not $\mathbb{R}^2$. We get a Markov triangle $\Delta_{a,b,c}(abc)\subset\mathbb{R}^2$ as in Section A.2 by setting

$$\Delta_{a,b,c}(abc)=\Phi^{-1}(\widetilde{\Delta}_{a,b,c}) \tag{A-23}$$

for Φ an integral affine embedding (see Definition A.11) containing $\widetilde{\Delta}_{a,b,c}$ in its image. Recall that this produces the same triangles (up to integral affine equivalence) as those obtained from almost toric fibrations of $\mathbb{CP}^2$ as discussed in Section A.2; see [32, Section 2]. Hence it makes sense to denote them by $\Delta_{a,b,c}(abc)$. The normalization $\alpha=abc$ of the triangle comes from the choice of level at which we have reduced in (A-21).

Definition A.11 An affine map $\Phi\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $x \mapsto Ax + b$ for $A \in \mathbb{Z}^{3\times 2}$ and $b \in \mathbb{R}^3$ is called an *integral affine embedding* if it is injective and if $A(\mathbb{Z}^2) = A(\mathbb{R}^2) \cap \mathbb{Z}^3$.

By the definition of integral affine embedding, the definition (A-23) makes sense and the polytope it defines is independent of the choice of Φ up to applying an integral affine transformation. We now show that there is a natural number q and an integral affine embedding $\Phi_{a,q}$ such that the triangle $\Phi_{a,q}^{-1}(\widetilde{\Delta}_{a,b,c}) \subset \mathbb{R}^2$ is obtained by intersecting $\angle_{a,q}$ with a half-plane. Indeed, set

$$\Phi_{a,q}\colon \mathbb{R}^2 \to \mathbb{R}^3, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} -b^2 & 1 + \frac{b}{a}(bq - 3c) \\ a^2 & 1 - aq \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b^2c^2 \\ 0 \\ 0 \end{pmatrix}, \tag{A-24}$$

where q satisfies $bq = 3c \mod a$; see also [7, Example 2.6]. The map $\Phi_{a,q}$ has image

$$\{a^2 y_1 + b^2 y_2 + c^2 y_3 = a^2 b^2 c^2\}$$

and it is an integral affine embedding, as can be checked by a computation. Furthermore Φ maps 0 to the vertex $(b^2c^2, 0, 0)$ (corresponding to a^2) of $\widetilde{\Delta}_{a,b,c}$, and v_1 and v_2 to the outgoing edges at $(b^2c^2, 0, 0)$. This means that there is an integral vector (ξ_1, ξ_2) defining a half-plane $K = \{\xi_1 x_1 + \xi_2 x_2 \leqslant k\}$ such that

$$\Phi_{a,q}^{-1}(\widetilde{\Delta}_{a,b,c}) = K \cap \angle_{a,q}. \tag{A-25}$$

From this we deduce the desired toric model. Let $\widetilde{E}_a \subset \widetilde{\Delta}_{a,b,c}$ be the edge opposite the vertex $(b^2c^2, 0, 0)$.

Proposition A.12 *The subset in $\mathbb{CP}(a^2, b^2, c^2)$ fibering over $\widetilde{\Delta}_{a,b,c} \setminus \widetilde{E}_a$ is fibered (orbifold) symplectomorphic to the subset in $\mathbb{C}^2/\Gamma_{a,q}$ fibering over* Int $K \cap \angle_{a,q}$.

Proof We have shown above that $\widetilde{\Delta}_{a,b,c} \setminus \widetilde{E}_a$ and Int $K \cap \angle_{a,q}$ are integral affine equivalent. This implies the claim by the classification of compact toric orbifolds by their moment map images; see [19]. Compactness is not a problem here, since we can compactify the subset fibering over $K \cap \angle_{a,q}$ by performing a symplectic cut at $\{\xi_1 x_1 + \xi_2 x_2 = k\}$. □

Let us now fix a moment map

$$\mu\colon \mathbb{CP}(a^2, b^2, c^2) \to \Delta_{a,b,c} = \Delta_{a,b,c}(abc) \subset \mathbb{R}^2 \tag{A-26}$$

by composing the moment map $\mathbb{CP}(a^2, b^2, c^2) \to \widetilde{\Delta}_{a,b,c}$ with the inverse of a suitable integral affine embedding as described above. Until the end of the proof of Proposition A.7, we simplify notation by writing $\Delta_{a,b,c} = \Delta_{a,b,c}(abc)$.

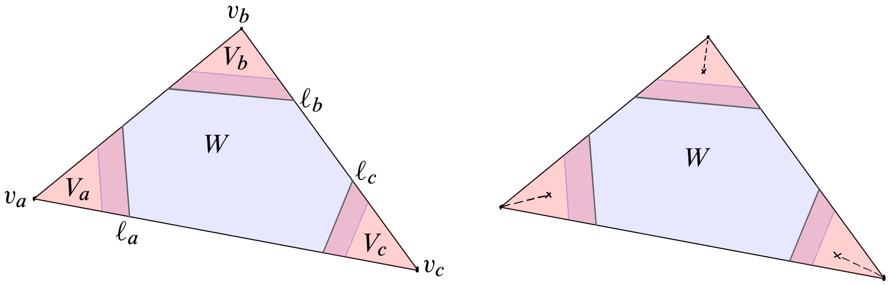


Figure 2: The triangle $\Delta_{a,b,c}$ as union $W \cup V_a \cup V_b \cup V_c$, on the left as the toric moment polytope of $\mathbb{CP}(a^2, b^2, c^2)$ and on the right as almost toric base diagram of $\mathbb{CP}^2$. In both cases the fibration is toric over W and lens spaces fiber over the segments ℓ_a, ℓ_b, ℓ_c.

Proof of Proposition A.7 The main part of the proof will be concerned with proving the existence of the symplectomorphism (A-10) and for readability, we postpone the proof of the existence of the global map $\tilde{\phi}$ and the computation of $\tilde{\phi}^{-1}(\mathfrak{D})$ to Step 5.

Step 1 We start by setting up some notation on the side of the weighted projective space. The orbifold points $s_a, s_b, s_c \in \mathbb{CP}(a^2, b^2, c^2)$ are mapped to the vertices $v_a, v_b, v_c \in \Delta_{a,b,c}$ under the moment map $\mu\colon \mathbb{CP}(a^2, b^2, c^2) \mapsto \Delta_{a,b,c}$. Let us first focus on the orbifold point s_a. Denote the edge opposite to v_a by E_a. By Proposition A.12, there is an orbifold symplectomorphism

$$\rho_a\colon \mu^{-1}(\Delta_{a,b,c} \setminus E_a) \to \mu^{-1}_{\mathbb{C}^2/\Gamma_{a,q}}(\operatorname{Int} K \cap \angle_{a,q}), \tag{A-27}$$

which intertwines the toric structures. Now let $\overline{B}^4(d)/\Gamma_{a,q} \subset \mathbb{C}^2/\Gamma_{a,q}$ be a closed orbifold ball for $\overline{B}^4(d) = \{\pi(|z_1|^2 + |z_2|^2) \leqslant d\}$ and $d > 0$. Its boundary $S^3(d)/\Gamma_{a,q} \subset \mathbb{C}^2/\Gamma_{a,q}$ is a lens space equipped with the standard contact structure. Note that both the orbifold ball and its boundary are fibered by the moment map $\mu_{\mathbb{C}^2/\Gamma_{a,q}}$. Since ρ_a intertwines the toric structures, the image sets

$$B_a^{\mathrm{orb}} = \rho_a^{-1}(\overline{B}^4(d)/\Gamma_{a,q}) \quad \text{and} \quad \Sigma_a = \rho_a^{-1}(S^3(d)/\Gamma_{a,q}) \tag{A-28}$$

are fibered by μ. Then the image of the pair (B_a, Σ_a) under μ is a pair (V_a, ℓ_a) consisting of a segment contained in a triangle around the vertex v_a. Note also that the lens space Σ_a is naturally equipped with its standard contact structure. We do the same procedure around the remaining vertices v_b and v_c, and denote the corresponding sets by $B_b^{\mathrm{orb}}, B_c^{\mathrm{orb}}, \Sigma_b, \Sigma_c \subset \mathbb{CP}(a^2, b^2, c^2)$ and by $V_b, V_c, \ell_b, \ell_c \subset \Delta_{a,b,c}$. We choose the sizes so that B_a^{orb}, B_b^{orb} and B_c^{orb} are mutually disjoint. Furthermore, we choose a set $W \subset \Delta_{a,b,c}$ such that $\Delta_{a,b,c} = W \cup V_a \cup V_b \cup V_c$ and such that the overlap $W \cap V_j$ is a strip around ℓ_j for all $j \in \{a, b, c\}$. Again, see Figure 2.

Step 2 Now consider the almost toric fibration of $\mathbb{CP}^2$ associated to the triangle $\Delta_{a,b,c}$. In the conventions of [30; 32] the triangle $\Delta_{a,b,c}$ is decorated with three dashed line segments of prescribed slope between the vertices and the nodal points. The latter are usually marked by a cross. There is a map $\pi\colon \mathbb{CP}^2 \to \Delta_{a,b,c}$ which is a standard toric fibration away from the dashed lines, but which is only continuous on the preimages of the dashed lines (which encode monodromy of the integral affine structure). By applying *nodal slides* if necessary, we may assume that the dashed lines lie outside of the subset $W \subset \Delta_{a,b,c}$. Since the projection π is standard toric away from the dashed lines, this implies that there is a symplectomorphism

$$\phi_0 \colon \mathbb{CP}^2 \supset \pi^{-1}(W) \to \mu^{-1}(W) \subset \mathbb{CP}(a^2, b^2, c^2), \tag{A-29}$$

which intertwines π and μ. Define the preimages

$$B'_a = \pi^{-1}(V_a) \quad \text{and} \quad \Sigma'_a = \pi^{-1}(\ell_a). \tag{A-30}$$

By [30, Section 9], the set B'_a is a closed rational homology ball and Σ'_a is a lens space of type $(aq-1, a^2)$ equipped with its standard contact structure. In fact, ϕ_0 maps the copy Σ'_a of the lens space to the copy Σ_a. However, contrary to B_a^{orb}, the rational homology ball B'_a is smooth. The same discussion holds for b and c.

Step 3 The key part of the proof is finding extensions

$$\phi_j : B'_j \setminus L_{j,q_j} \to B_j^{\mathrm{orb}} \setminus \{s_j\} \tag{A-31}$$

of the map $\phi_0|_{\Sigma'_j}$, where L_{j,q_j} are Lagrangian pinwheels for $j \in \{a, b, c\}$. For this we use Lemma A.10. Again, restricting our attention to a, let ψ_a be the symplectomorphism from Lemma A.10. Note that we have already established the correspondence between B_a^{orb} and $\mathbb{C}^2/\Gamma_{a,q}$ by ρ_a and that this correspondence is compatible with the toric picture. We now establish a correspondence between the rational homology sphere B'_a and the space $S_{a,q}$ coming from the smoothing in Lemma A.10. Define yet another copy Σ''_a of the lens space by setting $\Sigma''_a = \psi_a^{-1}(\rho_a(\Sigma_{a,q}))$. This lens space is also equipped with the standard contact structure and it bounds a rational homology ball B''_a by [7, Example 2.5]. We now have two pairs (B'_a, Σ'_a) and (B''_a, Σ''_a) consisting of a rational homology ball bounded by a lens space carrying its standard contact structure. By [9, Proposition A.2], which relies on [20], this implies that (B'_a, Σ'_a) and (B''_a, Σ''_a) are equivalent up to symplectic deformation. Let $\chi_a \colon (B'_a, \Sigma'_a) \to (B''_a, \Sigma''_a)$ be the diffeomorphism we obtain from this. Note that we cannot directly use the symplectic deformation to conclude, since the symplectomorphism obtained from a Moser-type argument may not restrict to the desired map on the boundary. More precisely, we obtain a diagram of diffeomorphisms of lens spaces

$$\begin{tikzcd}
B'_a \supset \Sigma'_a \arrow[r, "\phi_0|_{\Sigma'_a}"] \arrow[d, "\chi_a|_{\Sigma'_a}"] & \Sigma_a \subset B_a^{\mathrm{orb}} \arrow[d, "\rho_a|_{\Sigma_a}"] \\
B''_a \supset \Sigma''_a \arrow[r, "\psi_a|_{\Sigma''_a}"] & S^3(d)/\Gamma_{a,q} \subset B^4(d)/\Gamma_{a,q}
\end{tikzcd} \tag{A-32}$$

and this diagram *does not commute*. We may, however, correct the diffeomorphism χ_a so that (A-32) commutes. Recall that $\Sigma_{a,q}$ is a lens space of type $(aq-1, a^2)$. Since $(aq-1)^2 \neq \pm 1 \mod a^2$, it follows from [3, Théorème 3(a)] that the space of diffeomorphisms of Σ''_a has two components, namely the one of the identity and the one of the involution τ induced by the involution $(z_1, z_2) \mapsto (\bar{z}_1, \bar{z}_2)$ of S^3. The diffeomorphism τ extends to a diffeomorphism $\tilde{\tau}$ of $B''_{p,q}$. Up to postcomposing χ_a with $\tilde{\tau}$, we may thus assume that $(\psi_a^{-1}\rho_a\phi_0\chi_a^{-1})|_{\Sigma''_a}$ is isotopic to the identity by an isotopy φ_t. Using this isotopy, we can correct the diffeomorphism χ_a such that the diagram (A-32) commutes. Indeed, the set $\pi^{-1}(V_a \cap W)$ is a collar neighborhood $\Sigma'_a \times [0,2)$ and thus we can use the collar coordinate together with the isotopy φ_t to define a corrected diffeomorphism $\tilde{\chi}_a$, which coincides with the original diffeomorphism χ_a on $\mu^{-1}(V_a \setminus W)$ and with $(\psi_a^{-1}\rho_a\phi_0)|_{\Sigma'_a}$ on Σ'_a. Recall that two collar neighborhoods which agree on the boundary coincide up to applying a smooth isotopy; see Munkres [24, Lemma 6.1]. This means that, after applying an isotopy in (B''_a, Σ''_a), we can assume that the corrected version of χ_a and $\psi_a^{-1}\rho_a\phi_0$ agree on a smaller collar $\Sigma_a \times [0,1)$. Denoting the diffeomorphism we obtain in this way by $\tilde{\chi}_a$, this allows us to define a diffeomorphism

$$\phi_a = \rho_a^{-1} \circ \psi_a \circ \tilde{\chi}_a|_{B'_a \setminus L_{a,q_a}} : B'_a \setminus L_{a,q_a} \to B_a^{\mathrm{orb}} \setminus \{s_a\}, \tag{A-33}$$

which extends ϕ_0 in the sense that it agrees with ϕ_0 on a collar of Σ'_a. Since ψ_a is defined outside of a Lagrangian pinwheel $L_{a,q_a} \subset B''_a$, the diffeomorphism ϕ_a is defined outside of a pinwheel (which we again denote by L_{a,q_a}) in B'_a. We repeat this procedure for b and c to obtain diffeomorphisms ϕ_b and ϕ_c.

Step 4 By construction, the diffeomorphisms ϕ_a, ϕ_b and ϕ_c extend the initial symplectomorphism $\phi_0|_{\pi^{-1}(W \setminus (V_a \cup V_b \cup V_c))}$ and hence we obtain a diffeomorphism

$$\hat{\phi} : \mathbb{CP}^2 \setminus (L_{a,q_a} \sqcup L_{b,q_b} \sqcup L_{c,q_c}) \to \mathbb{CP}(a^2, b^2, c^2) \setminus \{s_a, s_b, s_c\}. \tag{A-34}$$

We now turn to the symplectic forms. On $\mathbb{CP}^2$, we define a symplectic form $\hat{\omega}$ which turns $\hat{\phi}$ into a symplectomorphism as follows. On $\pi^{-1}(W \setminus (V_a \cup V_b \cup V_c))$ we define $\hat{\omega}$ to be the usual Fubini–Study form ω. On V_j we define $\hat{\omega}$ as the pullback form $\tilde{\chi}_j^* \omega_{B''_j}$, where $\tilde{\chi}_j$ is the corrected diffeomorphism constructed at the end of Step 3. This yields a well-defined symplectic form which turns $\hat{\phi}$ into a symplectomorphism. Indeed, this follows from the fact that the maps ϕ_0, ψ_j and ρ_j are symplectomorphisms and ϕ_j is defined as their composition (A-33). This also implies that the symplectic form $\hat{\omega}$ has the same total volume as the Fubini–Study form. By the Gromov–Taubes theorem [22, Remark 9.4.3(ii)], the form $\tilde{\omega}$ is symplectomorphic to the Fubini–Study form and hence postcomposing $\hat{\phi}$ from (A-34) with this symplectomorphism yields the desired symplectomorphism (A-10).

Step 5 The definition of the global map $\tilde{\phi} : \mathbb{CP}^2 \to \mathbb{CP}(a^2, b^2, c^2)$ is obtained by replacing ϕ_a from (A-33) by $\tilde{\phi}_a = \rho_a^{-1} \circ \tilde{\psi}_a \circ \tilde{\chi}_a$ and carrying out the rest of the construction as above. The map $\tilde{\psi}_a$ is given by Lemma A.10. Let us now identify $\tilde{\phi}^{-1}(\mathfrak{D})$, where $\mathfrak{D} = \mu^{-1}(\partial\Delta_{a,b,c})$. Let $\widetilde{W} \subset W$ be the subset of W where $\tilde{\phi}$ coincides with ϕ_0. Since ϕ_0 intertwines the toric structures on $\pi^{-1}(W) \subset \mathbb{CP}^2$ and $\mu^{-1}(W) \subset \mathbb{CP}(a^2, b^2, c^2)$, the set $\tilde{\phi}^{-1}(\mathfrak{D}) \cap \pi^{-1}(\widetilde{W})$ fibers over the three pieces of the boundary given by $\widetilde{W} \cap \partial\Delta_{a,b,c}$ and hence consists of three disjoint symplectic cylinders.

We use Lemma A.10 to prove that the missing pieces $\tilde{\phi}^{-1}(\mathfrak{D}) \cap B'_j$ for $j \in \{a, b, c\}$ are also given by symplectic cylinders and that the union of the six cylinders is given by a torus. We again discuss the case of $j = a$ since the other two are completely analogous. Recall that $\rho_a : B_a^{\mathrm{orb}} \to B^4(d)/\Gamma_{a,q}$ is compatible with the toric structure and thus $\rho_a(\mathfrak{D} \cap B_a^{\mathrm{orb}}) = \mathfrak{D}_{a,q_a} \cap \rho_a(B_a^{\mathrm{orb}})$, where $\mathfrak{D}_{a,q_a} = \{z_1 z_2 = 0\}/\Gamma_{a,q_a}$ as in Lemma A.10. Indeed, this follows from the fact that $\mathfrak{D}_{a,q_a}$ fibers over the boundary of the moment map image $\angle_{a,q_a}$. By Lemma A.10, we deduce that $\widetilde{\chi}_a^{-1}(\widetilde{\phi}_a^{-1}(\rho_a(\mathfrak{D} \cap B_a^{\mathrm{orb}})))$ is given by the union of a pinwheel with a piece of a cylinder. Furthermore, recall that near the lens spaces at the respective boundaries, the map $\widetilde{\rho}_a^{-1}\widetilde{\psi}_a\chi_a$ coincides with ϕ_0. This implies that the cylinder contained in B_a' has two boundary components at $\partial B_a' = \Sigma_a'$, which are smoothly identified in a collar neighborhood with boundary components of the set $\widetilde{\phi}^{-1}(\mathfrak{D}) \cap \pi^{-1}(\widetilde{W})$ discussed in the previous paragraph. This proves the claim. □

Remark A.13 We suspect that there are shorter and more natural proofs of Proposition A.7. In particular, one should be able to avoid Gromov–Taubes. One possibility we have hinted at above is working with a global degeneration and trying to analyze its vanishing cycle. This would completely avoid the use almost toric fibrations. Another possibility, in the spirit of [27; 14], is to equip the explicit local degeneration from [17] with a family of integrable systems avoiding the pinwheel and extending the given toric structure on the boundary. The symplectomorphism from Proposition A.7 then follows from the usual toric arguments and it is automatically equivariant. Although this construction is elementary, it is somewhat outside the scope of this appendix and we hope to carry out the details elsewhere. This is also reminiscent of [10, Section 7], and it is plausible one can apply results from this paper to prove the same result.

References

[1] **A Akopyan**, **R Karasev**, *A tight estimate for the waist of the ball*, Bull. Lond. Math. Soc. 49 (2017) 690–693 MR Zbl

[2] **P Biran**, *Lagrangian barriers and symplectic embeddings*, Geom. Funct. Anal. 11 (2001) 407–464 MR Zbl

[3] **F Bonahon**, *Difféotopies des espaces lenticulaires*, Topology 22 (1983) 305–314 MR Zbl

[4] **J Brendel**, **G Mikhalkin**, **F Schlenk**, *Non-isotopic symplectic embeddings of cubes*, in preparation

[5] **R Casals**, **R Vianna**, *Full ellipsoid embeddings and toric mutations*, Selecta Math. 28 (2022) art. id. 61 MR Zbl

[6] **J D Evans**, *KIAS lectures on symplectic aspects of degenerations*, lecture notes, Korea Inst. Adv. Stud. (2019) Available at `https://github.com/jde27/kias`

[7] **J D Evans**, **I Smith**, *Markov numbers and Lagrangian cell complexes in the complex projective plane*, Geom. Topol. 22 (2018) 1143–1180 MR Zbl

[8] **S Galkin**, **A Usnich**, *Laurent phenomenon for Landau–Ginzburg potential*, preprint 10-0100, Kavli Inst. Phys. Math. Universe (2010) Available at `http://research.ipmu.jp/ipmu/sysimg/ipmu/417.pdf`

[9] **M Golla**, **L Starkston**, *The symplectic isotopy problem for rational cuspidal curves*, Compos. Math. 158 (2022) 1595–1682 MR Zbl

[10] **Y Groman**, **U Varolgunes**, *Locality of relative symplectic cohomology for complete embeddings*, Compos. Math. 159 (2023) 2551–2637 MR Zbl

[11] **M Gromov**, *Pseudo holomorphic curves in symplectic manifolds*, Invent. Math. 82 (1985) 307–347 MR Zbl

[12] **M Gromov**, *Isoperimetry of waists and concentration of maps*, Geom. Funct. Anal. 13 (2003) 178–215 MR Zbl

[13] **P Hacking**, **Y Prokhorov**, *Smoothable del Pezzo surfaces with quotient singularities*, Compos. Math. 146 (2010) 169–192 MR Zbl

[14] **M Harada**, **K Kaveh**, *Integrable systems, toric degenerations and Okounkov bodies*, Invent. Math. 202 (2015) 927–985 MR Zbl

[15] **E Heintze**, **H Karcher**, *A general comparison theorem with applications to volume estimates for submanifolds*, Ann. Sci. École Norm. Sup. 11 (1978) 451–470 MR Zbl

[16] **A B Katok**, *Ergodic perturbations of degenerate integrable Hamiltonian systems*, Izv. Akad. Nauk SSSR Ser. Mat. 37 (1973) 539–576 MR Zbl In Russian; translated in Math. USSR-Izv. 7 (1973) 535–572

[17] **Y Lekili**, **M Maydanskiy**, *The symplectic topology of some rational homology balls*, Comment. Math. Helv. 89 (2014) 571–596 MR Zbl

[18] **E Lerman**, *Symplectic cuts*, Math. Res. Lett. 2 (1995) 247–258 MR Zbl

[19] **E Lerman**, **S Tolman**, *Hamiltonian torus actions on symplectic orbifolds and toric varieties*, Trans. Amer. Math. Soc. 349 (1997) 4201–4230 MR Zbl

[20] **P Lisca**, *On symplectic fillings of lens spaces*, Trans. Amer. Math. Soc. 360 (2008) 765–799 MR Zbl

[21] **D McDuff**, *Displacing Lagrangian toric fibers via probes*, from "Low-dimensional and symplectic topology" (M Usher, editor), Proc. Sympos. Pure Math. 82, Amer. Math. Soc., Providence, RI (2011) 131–160 MR Zbl

[22] **D McDuff**, **D Salamon**, *J–holomorphic curves and symplectic topology*, 2nd edition, Amer. Math. Soc. Colloq. Publ. 52, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[23] **Y Memarian**, *On Gromov's waist of the sphere theorem*, J. Topol. Anal. 3 (2011) 7–36 MR Zbl

[24] **J R Munkres**, *Elementary differential topology*, Ann. of Math. Stud. 54, Princeton Univ. Press (1963) MR Zbl

[25] **J Oakley**, **M Usher**, *On certain Lagrangian submanifolds of $S^2 \times S^2$ and $\mathbb{C}P^n$*, Algebr. Geom. Topol. 16 (2016) 149–209 MR Zbl

[26] **E Opshtein**, *Maximal symplectic packings in $\mathbb{P}^2$*, Compos. Math. 143 (2007) 1558–1575 MR Zbl

[27] **W-D Ruan**, *Lagrangian torus fibration of quintic hypersurfaces, I: Fermat quintic case*, from "Winter school on mirror symmetry, vector bundles and Lagrangian submanifolds" (C Vafa, S-T Yau, editors), AMS/IP Stud. Adv. Math. 23, Amer. Math. Soc., Providence, RI (2001) 297–332 MR Zbl

[28] **F Schlenk**, *Embedding problems in symplectic geometry*, De Gruyter Expos. Math. 40, de Gruyter, Berlin (2005) MR Zbl

[29] **J Seade**, *On the topology of isolated singularities in analytic spaces*, Progr. Math. 241, Birkhäuser, Basel (2006) MR Zbl

[30] **M Symington**, *Four dimensions from two in symplectic topology*, from "Topology and geometry of manifolds", Proc. Sympos. Pure Math. 71, Amer. Math. Soc., Providence, RI (2003) 153–208 MR Zbl

[31] **L Traynor**, *Symplectic packing constructions*, J. Differential Geom. 42 (1995) 411–429 MR Zbl

[32] **R F d V Vianna**, *Infinitely many exotic monotone Lagrangian tori in* $\mathbb{CP}^2$, J. Topol. 9 (2016) 535–551 MR Zbl

[33] **R Vianna**, *Infinitely many monotone Lagrangian tori in del Pezzo surfaces*, Selecta Math. 23 (2017) 1955–1996 MR Zbl

Department of Mathematics and Statistics, University of Massachusetts, Amherst
Amherst, MA, United States

Mathematics Department, California Institute of Technology
Pasadena, CA, United States

Department of Mathematics, Boğaziçi University
İstanbul, Turkey

Department of Mathematics, University of Washington
Seattle, WA, United States

Institut de Mathématiques, Université de Neuchâtel
Neuchâtel, Switzerland

ksackel@umass.edu, aysong@caltech.edu, varolgunesu@gmail.com, jonozhu@uw.edu, joe.brendel@unine.ch

Proposed: Yakov Eliashberg Received: 22 May 2021
Seconded: Leonid Polterovich, Tobias H Colding Revised: 5 April 2022

Geometry & Topology **28**:3 (2024) 1153–1166
DOI: 10.2140/gt.2024.28.1153
Published: 10 May 2024

Zariski dense surface groups in $\mathbf{SL}(2k+1,\mathbb{Z})$

D Darren Long
Morwen B Thistlethwaite

We show that for all k, $\mathrm{SL}(2k+1,\mathbb{Z})$ contains surface groups which are Zariski dense in $\mathrm{SL}(2k+1,\mathbb{R})$.

22E40; 20C15, 20H10

1 Introduction

Let G be a semisimple Lie group, and $\Gamma < G$ a lattice. Following Sarnak [11], a subgroup Δ of Γ is called *thin* if Δ has infinite index in Γ and is Zariski dense. There has been an enormous amount of interest in the nature of thin subgroups of lattices, motivated in part by work on expanders, and in particular the so-called "affine sieve" of Bourgain, Gamburd and Sarnak [3].

Since it is quite standard to exhibit Zariski dense subgroups of lattices that are free products, the case of most interest is when the (finitely generated) thin group Δ does not decompose as a free product. Despite their importance and interest, nonfree thin subgroups in higher rank are extremely difficult to exhibit since the Zariski dense condition makes any given subgroup hard to distinguish from a lattice and freely indecomposable isomorphism types in the higher-rank situation are poorly understood. Our main theorem is the following.

Theorem 1.1 *For every $k \geq 1$, the group $\mathrm{SL}(2k+1,\mathbb{Z})$ contains a faithful representation of a surface group which is Zariski dense in $\mathrm{SL}(2k+1,\mathbb{R})$.*

To the authors' knowledge, this is the first result of this type concerning a freely indecomposable isomorphism type with Zariski closure $\mathrm{SL}(n,\mathbb{R})$ for infinitely many n.

In fact our argument shows that there are infinitely many nonconjugate such representations. We note that previous work of the authors (see Long, Reid and Thistlethwaite [7] and Long and Thistlethwaite [8]) using a totally different approach proved this to be true for $2k+1 = 3, 5$ (and also in fact for $k = \frac{3}{2}$). However, that method seems to have no hope of generalizing for infinitely many n.

Here is an outline of our argument, with careful definitions deferred to the sections below. The starting point is a certain discrete faithful representation of the triangle group $\phi_n : \Delta(3,4,4) \to \mathrm{PSL}(n,\mathbb{R})$, obtained by composing a discrete faithful representation coming from the hyperbolic structure with the irreducible representation

$$\tau_n \colon \mathrm{PSL}(2,\mathbb{R}) \to \mathrm{PSL}(n,\mathbb{R})$$

obtained by the action on homogeneous polynomials of degree $n = 2k+1$ in two variables. Such representations lie on the so-called *Hitchin component* (see Labourie [6] and its generalization to orbifolds coming from Alessandrini, Lee and Schaffhauser [1]), with the key fact being that all representations on the Hitchin component are discrete and faithful [1, Theorem 1.1].

Since n is odd, we can show (Theorem 2.1) that this representation can be conjugated to be integral. Of course, this representation cannot be Zariski dense since it lies inside an algebraic group isomorphic to $\mathrm{PSL}(2,\mathbb{R})$. Indeed, standard theory shows that it has image inside $\mathrm{SO}(J,\mathbb{R})$, where J is a certain quadratic form of signature $(k+1,k)$.

It is this representation we seek to ameliorate using the well-known bending construction, however the triangle group does not lend itself to that, so we pass to a subgroup of index four which is the fundamental group of the orbifold $S^2(3,3,3,3)$. The bending construction is described in detail in Section 3, but briefly: One takes an integral element $\delta \in \mathrm{PSL}(n,\mathbb{R})$ which centralizes the image of an essential simple closed curve γ; this curve splits the surface into two subsurfaces L and R (in the initial case, each is a disc with two orbifold points) and defines a new representation by conjugating ϕ_n by δ on the R surface. This new representation is obviously still inside $\mathrm{SL}(2k+1,\mathbb{Z})$ and it continues to be faithful since we take care to arrange that it lies on the Hitchin component.

If the bent group is not Zariski dense, we conclude that it lies inside $\mathrm{SO}(J)$, and we appeal to a result of Guichard (see Theorem 3.1 and Guichard [5]) to prove that it must have Zariski closure all of $\mathrm{SO}(J)$.

The strategy now is to perform a suitable second bend. This is much more subtle, for example one needs at least to be sure that one can find an element for which all the integral centralizing elements do not lie in $\mathrm{SO}(J)$. However, we can use the extra information that the Zariski closure of the bent group is all of $\mathrm{SO}(J)$ to appeal to the Weisfeiler–Nori version of the strong approximation theorem; see the discussion of Section 3.2. From this follows that for all but finitely many primes p, if one reduces the bent group modulo p it surjects $\Omega(J,p)$, the commutator subgroup of $\mathrm{SO}(J;\mathbb{Z}/p)$. (It is classical that $[\mathrm{SO}(J;\mathbb{Z}/p):\Omega(J,p)] = 2$ for odd p.) Since we can show that group $\Omega(J,p)$ contains elements whose characteristic polynomials are of the form $(Q-1)f(Q)$ where $f(Q)$ is irreducible modulo p, it follows that the original group contains an element η mapping onto such an element, in particular its characteristic polynomial has the form $(Q-1)F(Q)$ where $F(Q)$ is $\mathbb{Z}$ irreducible.

Of course the element η may not be a simple loop, but we show in Section 4 one can ascend a carefully constructed tower of regular coverings so that η lifts to each step of this tower and ultimately becomes a (power of an) essential simple loop on some orbifold surface $S^2(3,3,\dots,3)$, which must still surject $\Omega(J,p)$.

One can then show, using rank considerations, that the characteristic polynomial condition implies that there is an integral element in the centralizer of η inside $SL(2k+1,\mathbb{Z})$ which does not power into $SO(J)$ and from this, it is not hard to see (appealing again to the classification result of Guichard) that bending the group using that element makes the resulting orbifold group Zariski dense. The resulting representation is faithful since it continues to be on the Hitchin component.

We note that the methods of this paper can be used to show that in any dimension, if the Hitchin component contains an integral representation, then it contains an integral Zariski dense representation; see Zshornack [14]. This is because although the finite group situation for n even is slightly more involved, there is no real difficulty extending the observations of Section 3.2 to the symplectic case. However, finding such an integral representation in even dimensions seems to pose significant difficulties.

For the rest of this article, we restrict to the case $n \geq 9$. This is not for any truly essential reason, as the argument present here works if $n > 3$. However, as mentioned above, the cases $n = 3, 5$ are already in the literature and while the case $n = 5$ can be dealt with by the method described here, the case $n = 7$ involves a technical detour which is hardly worth the number of words it would take. This case is resolved explicitly in our computation [9].

Acknowledgements Long was partially supported by the NSF.

2 Integrality of certain representations of the 344–triangle group

This section analyses certain representations of the triangle group

$$\Delta(3,4,4) = \langle a, b, c \mid a^3 = b^4 = c^4 = abc = 1\rangle,$$

which is the group of orientation-preserving symmetries of the tiling of the hyperbolic plane by triangles with angles $\left\{\frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{4}\right\}$.

Let $\phi_2\colon \Delta(3,4,4) \to PSL(2,\mathbb{R})$ be the faithful representation of $\Delta(3,4,4)$ into $\mathrm{Isom}^+(\mathbb{H}^2)$. Hyperbolic triangle groups are rigid and so ϕ_2 is uniquely determined up to conjugacy.

Let $\tau_n\colon PSL(2,\mathbb{R}) \to PSL(n,\mathbb{R})$ for $n \geq 3$ denote the irreducible representation obtained from the standard action on homogeneous polynomials in two variables. We denote the composite representation $\tau_n \circ \phi_2\colon \Delta(3,4,4) \to PSL(n,\mathbb{R})$ by ϕ_n.

Remark For odd n (the situation which primarily concerns us here) the representation ϕ_n lifts to a representation of $\Delta(3,4,4)$ into $SL(n,\mathbb{R})$, but for even n we have to be content with the situation we already see for $n = 2$, namely that we have a representation of a certain pullback group U_{344} into $SL(n,\mathbb{R})$, the pullback being that determined by the representation ϕ_2 together with the natural projection $SL(n,\mathbb{R}) \to PSL(n,\mathbb{R})$.

For ease of notation we denote elements of $\mathrm{PSL}(n,\mathbb{R})$ by representative matrices in $\mathrm{SL}(n,\mathbb{R})$; also, given a subring A of $\mathbb{C}$ we say that a representation into $\mathrm{PSL}(n,\mathbb{R})$ *can be written over A* if the corresponding lifted representation (of $\Delta(3,4,4)$ or U_{344}) can be so written.

Since $\Delta(3,4,4)$ is the fundamental group of a compact orbifold, specifically S^2 with cone points of orders $3,3,4$, it follows that $\phi_2(\Delta(3,4,4))$ contains no parabolic. Therefore all matrices in $\phi_2(\Delta(3,4,4))$ are diagonalizable, and application of τ_n to a diagonal matrix shows that if $A\in\phi_2(\Delta(3,4,4))$ has eigenvalues λ and $1/\lambda$, then the eigenvalues of $\tau_n(A)$ are

$$\lambda^{n-1},\lambda^{n-3},\dots,\lambda^{-(n-3)},\lambda^{-(n-1)}.$$

In this section, we prove the following.

Theorem 2.1 *For odd n, the representation ϕ_n be written over $\mathbb{Z}$.*

For even n, the representation ϕ_n can be written over $\mathbb{Z}[\sqrt{2}]$ but not over $\mathbb{Z}$.

We choose the following faithful representation ϕ_2 of $\Delta(3,4,4)$ into the group $\mathrm{PSL}(2,\mathbb{R})$ of orientation-preserving isometries of $\mathbb{H}^2$, obtained by placing the isosceles $\left(\frac{\pi}{3},\frac{\pi}{4},\frac{\pi}{4}\right)$–triangle symmetrically about the y–axis in the upper half-plane and then putting the matrix for the generator a in rational canonical form:

$$\phi_2(a)=\begin{bmatrix}0&-1\\1&1\end{bmatrix},\quad \phi_2(b)=\begin{bmatrix}0&-1-\sqrt{2}\\-1+\sqrt{2}&\sqrt{2}\end{bmatrix},\quad \phi_2(c)=\begin{bmatrix}1-\sqrt{2}&-\sqrt{2}\\-1+\sqrt{2}&-1\end{bmatrix}.$$

Thus the representation ϕ_2 of $\Delta(3,4,4)$ can be written over the ring $\mathbb{Z}[\sqrt{2}]$, and since application of τ_n to a 2×2 matrix A produces an $n\times n$ matrix whose entries are integer polynomial expressions of those of A, we see that ϕ_n for $n\ge 3$ can also be written over $\mathbb{Z}[\sqrt{2}]$. We shall show that for odd n, ϕ_n can actually be written over $\mathbb{Z}$ and for even n this is not possible.

Our next basic ingredient is the fact that the representation ϕ_3 can be written over the integers. Here is an example of an integral representation ϕ_3' of $\Delta(3,4,4)$, conjugate to ϕ_3:

$$\phi_3'(a)=\begin{bmatrix}1&1&-2\\0&-1&1\\0&-1&0\end{bmatrix},\quad \phi_3'(b)=\begin{bmatrix}1&0&-1\\4&1&-1\\2&0&-1\end{bmatrix},\quad \phi_3'(c)=\begin{bmatrix}1&1&0\\-2&-1&0\\-4&-1&1\end{bmatrix}.$$

Lemma 2.2 *The representation ϕ_n of $\Delta(3,4,4)$ has integral character if and only if n is odd. For even n the character of ϕ_n takes values in $\mathbb{Z}[\sqrt{2}]$.*

Proof Let A be any matrix in $\phi_2(\Delta(3,4,4))$, and let the eigenvalues of A be λ and $1/\lambda$. Since ϕ_3 can be written over the integers, we see that $\lambda^2+1+\lambda^{-2}\in\mathbb{Z}$, hence also $\lambda^2+\lambda^{-2}\in\mathbb{Z}$. For odd n, $\lambda^{n-1}+\lambda^{-(n-1)}=f(\lambda^2+\lambda^{-2})$ for a polynomial $f\in\mathbb{Z}[x]$, so we deduce inductively that the trace of $\tau_n(A)$ is an integer. On the other hand, if n is even, then

$$\lambda^{n-1}+\lambda^{n-3}+\cdots+\lambda^{-(n-3)}+\lambda^{-(n-1)}=(\lambda+\lambda^{-1})(\lambda^{n-2}+\lambda^{n-6}+\cdots+\lambda^{-(n-6)}+\lambda^{-(n-2)}).$$

A similar argument shows that the second factor is an integer, whereas the first factor might not be, eg the trace of $\phi_2(a^{-1}b)$ is $2\sqrt{2}$. We deduce that for even n, the trace of $\phi_n(a^{-1}b)$ is not an integer; however, as ϕ_n can be written over $\mathbb{Z}[\sqrt{2}]$, its character takes values in that ring. □

Lemma 2.3 *For odd n, the representation ϕ_n of $\Delta(3,4,4)$ can be written over the rational numbers.*

Proof We have established that for odd n, ϕ_n has integral character and can be written over $\mathbb{Z}[\sqrt{2}]$. Thus ϕ_n is realizable over a field of degree 2 over $\mathbb{Q}$. Suppose that ϕ_n is not realizable over a field of smaller degree over $\mathbb{Q}$; then the Schur index of the irreducible representation ϕ_n is $[\mathbb{Q}(\sqrt{2}):\mathbb{Q}]=2$. However the Schur index divides the degree n of the representation [2], contradicting the fact that n is odd. □

The conclusion of Theorem 2.1 has already been established for even n; for odd n it follows directly from Lemmas 2.1 and 2.3 together with the proof of Proposition 2.1 of [7].

3 The bending construction

Our construction is reliant upon an orbifold which has somewhat more geometric flexibility than the triangle group $\Delta(3,4,4)$. To this end we note that there is a homomorphism $\Delta(3,4,4)\to\mathbb{Z}/4$ which sends the two elements of order 4 to 1 and -1. This defines an orbifold covering $S^2(3,3,3,3)\to\Delta(3,4,4)$. We may restrict the representation ϕ_n to the subgroup defined by the covering to yield a discrete and faithful representation corresponding to a hyperbolic structure

$$\phi_n\colon \pi_1(S^2(3,3,3,3))\to \mathrm{PSL}(n,\mathbb{Z}).$$

The orbifold $S^2(3,3,3,3)$ has an obvious flexibility (often called *bending* in the setting of $\mathrm{SO}(n,1)$ representations) coming from the following construction: Let $d_1\,.\,d_2=\gamma$ be the simple closed curve on the two-sphere which separates two of the orbifold points from the other two; denote the two sides by L and R. Each contains two cone points of order three. Let δ be any element of $\mathrm{PSL}(n,\mathbb{R})$ which centralizes $\phi_n(\gamma)$. Then we may form a bent representation of $S^2(3,3,3,3)$ by matching $\phi_n(\pi_1(L))$ with $\delta\,.\,\phi_n(\pi_1(R))\,.\,\delta^{-1}$; these representations agree on $\phi_n(\gamma)$ by the choice of δ. We denote this bent representation by ρ^δ (even though this is a bit of an abuse).

With a view to questions concerning faithfulness, we will invariably use bending elements which are in the image of the exponential map (for example they will be diagonalizable and with positive real eigenvalues) that if, for example, $\delta=\exp(\boldsymbol{v})$, then all the elements $\exp(t\boldsymbol{v})$ centralize $\phi_n(\gamma)$. Then there is a path of bendings from ρ to ρ^δ given by $\rho^{\exp(t\boldsymbol{v})}$. It follows that if ρ lies on the Hitchin component, then so does ρ^δ and in particular, this implies that the latter is a faithful representation.

An important ingredient of our approach which gives the requisite control hangs upon the following theorem of Guichard.

Theorem 3.1 (Guichard [5]; see also [1]) *Suppose that $\rho\colon \pi_1(S) \to \mathrm{SL}(m,\mathbb{R})$ is a representation on the Hitchin component and G is the Zariski closure of $\rho(\pi_1(S))$. Then*:

- *If $m = 2n$ is even, then G is conjugate to one of $\tau_m(\mathrm{SL}(2,\mathbb{R}))$, $\mathrm{Sp}(2m,\mathbb{R})$ or $\mathrm{SL}(m,\mathbb{R})$.*
- *If $m = 2n+1 \neq 7$ is odd, then G is conjugate to one of $\tau_m(\mathrm{SL}(2,\mathbb{R}))$, $\mathrm{SO}(n+1,n)$ or $\mathrm{SL}(m,\mathbb{R})$.*
- *If $m = 7$, then G is conjugate to one of $\tau_m(\mathrm{SL}(2,\mathbb{R}))$, $\mathrm{SO}(4,3)$, G_2 or $\mathrm{SL}(7,\mathbb{R})$.*

The relevance of this theorem is that if one can show that a given representation of a surface group leaves no form invariant, then the image is Zariski dense.[1]

3.1 The first bend

The first step is to bend $\phi_n\colon \pi_1(S^2(3,3,3,3)) \to \mathrm{PSL}(n,\mathbb{Z})$. There is a good deal of flexibility in this part of our construction.

To this end, fix any simple closed curve γ which separates two of the orbifold points from the other two and we claim (recall that we are assuming that $n \geq 5$) that there is an element δ in the $\mathrm{PSL}(n,\mathbb{Z})$–centralizer of γ which does not lie in the image $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$. We argue this as follows. As in Lemma 2.2, all the hyperbolic elements in the image $\phi_n(\pi_1(S^2(3,3,3,3))$ have integral character and with eigenvalues 1 and $(n-1)/2$ pairs of the form μ^{2j} and μ^{-2j}. Since $\phi_n(A)$ is integral, it follows that it can be conjugated *over the rationals* into the block form consisting of a single 1 and $(n-1)/2$ block matrices

$$\exp(j\boldsymbol{w}) = \begin{pmatrix} 0 & -1 \\ 1 & K \end{pmatrix}^j \quad \text{for } 1 \leq j \leq \tfrac{1}{2}(n-1).$$

This latter matrix has $\mathbb{Z}$–centralizer isomorphic to $(\pm 1) \oplus \mathbb{Z}^{(n-1)/2}$.

The following is well known.

Proposition 3.2 *Let $M \in M_n(\mathbb{Q})$ be such that $\det(M) = \pm 1$ and the characteristic polynomial of M has integer coefficients. Then some power of M is integral.*

Since the conjugation matrix is rational, it follows from this proposition that the centralizer of $\phi_n(A)$ in $\mathrm{PSL}(n,\mathbb{Z})$ is $\mathbb{Z}^{(n-1)/2}$. Since $n \geq 5$ and the centralizer in $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$ is $\mathbb{Z}$ our claim follows. In fact, with a view to our argument it is important to note a little more is true, namely that this argument shows that we may choose these centralizing elements to be in the image of the exponential map, since we may choose hyperbolic elements with distinct positive eigenvalues.

Accordingly, we may fix some element δ lying in the $\mathrm{PSL}(n,\mathbb{Z})$–centralizer of γ which does not lie in the image $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$; in the interests of being specific, we choose δ as the relevant conjugate of a power of $1 \oplus \mathrm{Id}_2 \oplus \mathrm{Id}_2 \oplus \cdots \oplus \exp(\boldsymbol{w})$. Notice that if we write $\delta = \exp(\boldsymbol{v})$, then the entire path $\exp(t\boldsymbol{v})$ centralizes γ.

[1] Note that $G_2 < \mathrm{SO}(4,3)$.

Theorem 3.3 *The bent representation* ϕ_n^δ (which henceforth we denote by ρ) *is a representation of* $\pi_1(S^2(3,3,3,3))$ *into* $\mathrm{PSL}(n,\mathbb{Z})$ *lying on the Hitchin component. In particular,* ρ *is discrete and faithful. Moreover, the Zariski closure of the image of* ρ *is one of*

- $\mathrm{PSL}(n,\mathbb{R})$, *or*
- $\mathrm{SO}(J)$ *for the form* J *of signature* $(k+1,k)$ *left invariant by* $\phi_n(\mathrm{PSL}(n,\mathbb{R}))$.

Proof If $\delta=\exp(\boldsymbol{v})$, then ρ is the endpoint of the path of representations $\phi_n^{\exp(t\boldsymbol{v})}$ which has one endpoint on the Hitchin component. Since this is a path component of $\mathrm{Hom}(\pi_1(S^2(3,3,3,3)),\mathrm{PSL}(n,\mathbb{R}))$, it follows that ρ is on the Hitchin component and is therefore discrete (which was obvious anyway) and faithful.

The second claim is our first application of Theorem 3.1. The argument is the following. Notice (in the notation of Section 3) that $\pi_1(L)$ (and of course $\pi_1(R)$) is Zariski dense in $\mathrm{PSL}(2,\mathbb{R})$, since this Lie group has no interesting algebraic subgroups. It follows that the algebraic group $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$ is determined by the image $\phi_n(\pi_1(L))$. Referring to the list of Theorem 3.1, we see that the image of the bent representation ρ must be larger than $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$ unless δ normalizes $\phi_n(\pi_1(R))$ and hence $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$.

We claim that this is impossible. For if conjugacy by δ preserved $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$ it would act as an automorphism which commuted with the action by conjugacy of the hyperbolic element γ. However it is well known that $\mathrm{Aut}(\mathrm{PSL}(2,\mathbb{R})))\cong\mathrm{GL}(2,\mathbb{R})$ and we deduce that there would be some nontrivial word $\delta^a\gamma^b$ which centralized the absolutely irreducible representation $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$ and would therefore be trivial, which is impossible by the choice of δ.

Thus we have proved that the Zariski closure of $\rho(\pi_1(S^2(3,3,3,3))$ is strictly larger than $\phi_n(\mathrm{PSL}(2,\mathbb{R}))$.

One now consults the list provided by Theorem 3.1 and we see that (since in particular $n>7$) either we are done or the Zariski closure of $\rho(S^2(3,3,3,3))$ is SO of some form of signature $(k+1,k)$. We claim that this form is necessarily $\mathrm{SO}(J)$, where J is the form mentioned in the introduction. The reason is that on $\pi_1(L)$, the representation ϕ_n and ρ agree and are absolutely irreducible. We now appeal to:

Lemma 3.4 *Suppose that* $\rho\colon G\to\mathrm{SO}(J)\le\mathrm{SL}(n,\mathbb{R})$ *is an absolutely irreducible representation. Then* J *is unique up to a real scaling.*

Proof If J_1 and J_2 are two such forms, then the equations $A^T.J_i.A=J_i$ for $i=1,2$ and any A in the image of ρ imply that $J_2^{-1}.J_1$ centralizes the image of ρ, whence by Schur's lemma is a scalar matrix. □

Returning to the proof of Theorem 3.3, since the group $\phi_n(\pi_1(L))<\mathrm{SO}(J)$, the lemma implies that the Zariski closure of $\rho(S^2(3,3,3,3))$ must be $\mathrm{SO}(J)$, as required. □

Remark In fact, it's useful to note that if ρ is a Hitchin representation, one can weaken the hypothesis in Lemma 3.4 to ask only that ρ be *real* irreducible. The point is that for a Hitchin representation, the infinite-order elements have the property that they have distinct real eigenvalues. Fix such an element and suppose that we have diagonalized it over the reals. Then in the proof above, $J_2^{-1}.J_1$ is real and commutes with a diagonal matrix with distinct real eigenvalues and is therefore diagonal with real eigenvalues. The usual Schur argument using real irreducibility now implies that $J_2^{-1}.J_1 = r.\mathrm{Id}$ for some real r.

Remark Questions concerning *integral* centralizers can be quite delicate because of the rational conjugacy; the need to appeal to Proposition 3.2 results in the loss of any real control.

3.2 The second bend: finding η

To perform the second bend requires a more careful choice of bending curve and this necessitates a discussion of some notation and results from the theory of finite simple groups, in particular from the theory of (special) finite orthogonal groups. This has two steps. We will appeal to a theorem that follows from work of Weisfeiler [13] or Nori [10] to construct an essential curve η on $S^2(3,3,3,3)$ for which we can use purely algebraic considerations to show that $\rho(\eta)$ has a large centralizer in $\mathrm{PSL}(n,\mathbb{Z})$. We cannot use η directly to bend, since it may not be simple on $S^2(3,3,3,3)$, however in Section 4 we show how to improve this situation.

Here is a summary of the algebraic facts that we require. Let J be an m–dimensional quadratic form over the finite field $\mathrm{GF}(p^n)$ of cardinality $q = p^n$. It simplifies the discussion (and this is no loss of generality for us) to assume p is odd. We are interested only in the case that m is also odd; we assume this in the following without further comment. (The situation is slightly more complicated for m even.)

In this case, there is a unique orthogonal group up to isomorphism $O(J,q) = O(2k+1,q)$ which is independent of J; see [12, page 377, Theorem 5.8]. Let $\mathrm{SO}(m,q)$ denote the special orthogonal group and set $\Omega(m,q) = [O(m,q), O(m,q)]$, where $[G,G]$ denotes the commutator subgroup of a group G. We summarize the important fact for us in the following theorem; see [12, pages 383–384] for a discussion.

Theorem 3.5 *When m is odd, $\Omega(J,q)$ is a simple subgroup of $O(J,q)$ of index 4.*

Recall the first bending provided a representation $\rho = \phi_n^\delta\colon \pi_1(S^2(3,3,3,3)) \to \mathrm{SL}(n,\mathbb{Z})$ lying on the Hitchin component and whose image is Zariski dense in $\mathrm{SO}(J,\mathbb{R})$. Given a rational prime p, we may compose with the obvious reduction map modulo p, $\mathrm{SL}(n,\mathbb{Z}) \to \mathrm{SL}(n,\mathbb{Z}/p)$. The following comes from the strong approximation theorem; cf [10; 13].

Theorem 3.6 *In the notation above, when m is odd, for all but a finite number of rational primes p, we have $\Omega(J,p) = \pi_p(\rho(S^2(3,3,3,3))$.*

Proof It follows from strong approximation (cf [10; 13]) that except for finitely many primes we have

$$\Omega(J,p) \le \pi_p\big(\rho(S^2(3,3,3,3))\big) \le \mathrm{SO}(J;p).$$

However, Theorem 3.5, together with the fact that $\pi_1(S^2(3,3,3,3))$ has no subgroup of index 2 implies the result. □

To apply this, we also require the following fact:

Theorem 3.7 *There is an element* $\eta \in S^2(3,3,3,3)$ *with the property that the integer matrix* $\rho(\eta)$ *has characteristic polynomial of the form* $(Q-1)F(Q)$*, where* $F(Q)$ *is irreducible over* $\mathbb{Z}$*.*

The main ingredient in the proof is the following:

Proposition 3.8 *For every prime* p*, there is a matrix in* $\Omega(J,p)$ *with the property that its characteristic polynomial has the form* $(Q-1)f(Q)$*, where* $f(Q)$ *is irreducible modulo* p*.*

This proposition now implies Theorem 3.7. For, by strong approximation, we may choose a prime p so that $\pi_p\rho$ maps $S^2(3,3,3,3)$ onto $\Omega(J,p)$. Pick an element $A(p)\in\Omega(J,p)$ whose characteristic polynomial has the form $(Q-1)f(Q)$, where $f(Q)$ is irreducible modulo p, and choose η so that $\pi_p\rho(\eta)=A(p)$. Then the characteristic polynomial of $\rho(\eta)$ has the form $(Q-1)F(Q)$, and $F(Q)$ is necessarily $\mathbb{Z}$ irreducible, since it is irreducible modulo p.

Proof of Proposition 3.8 We have a fixed $n=2k+1$ and a prime p. We claim that there is an element τ in the algebraic closure of $\mathbb{Z}/p$ which has degree $n-1=2k$ over $\mathbb{Z}/p$ with the property that its minimal polynomial over $\mathbb{Z}/p$, which we denote by $a(Q)$, is symmetric. Further, τ^2 also has degree $n-1$ with (therefore irreducible) symmetric minimal polynomial, which we denote by $a_2(Q)$.

Deferring this claim temporarily, we proceed as follows. Let g be any $\mathbb{Z}/p$ matrix whose characteristic polynomial is $(Q-1)a(Q)$; for example, use the rational canonical form. Let K be a splitting field for $a(Q)$ over $\mathbb{Z}/p$, so that there is a K–matrix m for which $m\,.\,g\,.\,m^{-1}$ is diagonal, with the eigenvalues arranged $1,\lambda_1,\lambda_1^{-1},\dots,\lambda_n,\lambda_n^{-1}$. This matrix is an isometry of the form Σ, whose $(1,1)$ entry is any element of K and then has $(n-1)/2$ blocks which are K multiples of $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$. This gives a $(n+1)/2$–dimensional family of solutions and using the obvious ordered basis, the determinant of such a form Σ is $\pm a_0\cdot a_1\cdots a_{(n-1)/2}$ and is therefore nondegenerate as long as no a_j is zero. Therefore the original matrix g has a K–family of nondegenerate solutions, namely $m^T\,.\,\Sigma\,.\,m$. Now, when we regard the entries of a symmetric matrix σ as indeterminates, the question of whether a form σ satisfies $g^T\,.\,\sigma\,.\,g=\sigma$ is a family of homogeneous linear equations with $\mathbb{Z}/p$ coefficients, and we have just shown that there are nondegenerate solutions over K. It follows that there are nondegenerate $\mathbb{Z}/p$ solutions and such a solution σ gives $g\in\mathrm{SO}(\sigma,p)$ with characteristic polynomial $(Q-1)a(Q)$. As noted above, since n is odd there is only one such orthogonal group up to change of basis, so this characteristic polynomial also occurs in $\mathrm{SO}(J,p)$. After squaring if necessary, one finds the promised element in $\Omega(J,p)$.

It remains to find the polynomials promised in the first paragraph; we sketch an argument. Denote the finite field of degree r over $\mathbb{Z}/p$ by $\mathrm{GF}(p^r)$; it's known there is exactly one such field up to isomorphism for every r. Take an element $x\in\mathrm{GF}(p^k)$ with the property that the polynomial $T^2-x\cdot T+1$ is irreducible over $\mathrm{GF}(p^k)$. Let τ be a root of this polynomial so that $\mathrm{GF}(p^k)(\tau)=\mathrm{GF}(p^{2k})$. By construction, the polynomial for τ over $\mathbb{Z}/p$ has degree $2k=n-1$ and is symmetric and irreducible. Moreover, considering the equation $\tau^2-x\cdot\tau+1=0$, we see that τ^2 cannot lie in $\mathrm{GF}(p^k)$. Therefore the polynomial for τ^2 over $\mathbb{Z}/p$ also has the required properties. □

4 Improving η

At this stage we have a representation $\rho\colon S^2(3,3,3,3) \to \mathrm{PSL}(2k+1,\mathbb{Z})$ so that with finitely many exceptions, the reduction modulo p yields a surjection

$$\pi_p\colon \rho(S^2(3,3,3,3)) \to \Omega(J,p).$$

This was used to find an element η with the property that the characteristic polynomial of $\rho(\eta)$ is of the form $(Q-1)F(Q)$, where it follows from the construction that $F(Q)$ is irreducible over $\mathbb{Z}$. This section is devoted to a proof of the following:

Theorem 4.1 *Denote the orbifold surface which is an S^2 with k–cone points all of order 3 by $F(k)$, where $k \geq 4$. Then given η as above, there is a tower of 3–fold regular coverings*

$$F(4) \leftarrow F(u_1) \leftarrow F(u_2) \leftarrow \cdots \leftarrow F(u_k)$$

with the property that η lifts to each covering and in the covering $F(u_k)$, η is (a power of) a simple loop which encloses at least two cone points on each side.

Proof The construction here is based upon the following simple observation. Fix some surface $F(k)$ and fix two of the cone points, c_1 and c_2, enclosing them with a simple closed curve C. Then C splits the surface into two pieces, one of which is a disc with two cone points of order 3. There is a 3–fold covering of $F(k)$ given by the homomorphism $c_1 \to 1$, $c_2 \to -1$, and all other cone points mapping to zero. It's easy to see that the resulting covering is planar: the disc with two cone points becomes an S^2 with three discs removed, each of which corresponds to a lift of C. The other side of C lifts to three copies, each a trivial covering which is attached to one of these C–lifts. Notice that the number of cone points in the covering strictly increases since it has $3(k-2)$ cone points and this is strictly larger than k for $k \geq 4$.

The coverings being used are all abelian, and we indicate homology class by $[*]$. We begin by observing that it is easy to make coverings where η lifts: for $F(4)$, for example, if $[\eta]$ is not zero, it represents a generator of $H_1(F(4)) \cong (\mathbb{Z}/3)^3$, we can choose two other cone points x and y so that $H_1(F(4)) = \langle \eta, x, y\rangle$ and we form a covering as above with $c_1 = x$ and $c_2 = y$. Clearly η lifts to this covering. If $[\eta]$ is zero in $H_1(F(4))$, we may choose any pair of the cone points. Since the number of cone points only goes up, this procedure can be iterated (increasing the number of supplementary generators if need be).

The proof of Theorem 4.1 is accomplished by showing one can find a tower of coverings where η lifts and has a strictly decreasing number of self-intersections up to the point that we obtain the conclusion of the theorem. For future reference we note that the shape of the characteristic polynomial implies that η has infinite order in the fundamental group of the orbifold so that it cannot encircle just one cone point on either side. In particular, it is a hyperbolic element on the hyperbolic orbifolds in question, so we may make η geodesic.

We construct a planar subsurface X of $F(k)$ as follows. Take a thin regular neighborhood of η and attach discs to all the boundary components which bound discs in the complement of η. By construction, the boundary of the subsurface X consists of simple closed curves all of which are essential, and therefore bound discs which have on them at least one cone point.

Suppose that X has at least three boundary components. Then at least two of these boundary components, $\partial_1 X$ and $\partial_2 X$, say, contain cone points c_1 and c_2, respectively, with the property that

$$\langle c_1, c_2, \eta\rangle \cong (\mathbb{Z}/3)^3 < H_1(F(k)).$$

Define a covering of $F(k)$ as above, mapping $c_1 \to 1$, $c_2 \to -1$ and $\eta \to 0$, and extend to $H_1(F(k))$. This arranges $\partial_1 X \to 1$, $\partial_2 X \to -1$ and $[\eta] \to 0$. The preimage of X in this covering is connected, so that the three preimages $\tilde{\eta}_i$ for $1 \le i \le 3$ of η form a connected graph. A point where $\tilde{\eta}_1$ meets $\tilde{\eta}_2$ corresponds to a self-intersection of η in $F(k)$ which has now disappeared from $\tilde{\eta}_1$. Thus the number of self-intersections of $\tilde{\eta}_1$ is strictly less than the number of self-intersections of η.

We can repeat this process as long as the planar neighborhood X is not an annulus. However in this case, η must be a power of a simple loop, namely the core of the annulus. □

Remark In the latter case, the annulus cannot contain just one cone point on either side, since in this case the loop η would have order dividing 3.

5 Proof of Theorem 1.1

We may now complete the argument of Theorem 1.1.

At the top of the tower provided by Theorem 4.1, either we have a simple loop corresponding to η on a planar surface for which the characteristic polynomial of $\rho(\eta)$ is $(Q-1)F(Q)$, where $F(Q)$ is $\mathbb{Z}$–irreducible, or we have that $\eta = (\eta')^r$, where η' is a simple loop. In this latter case, the characteristic polynomial of $\rho(\eta')$ is also of the form $(Q-1)G(Q)$, where $G(Q)$ is $\mathbb{Z}$–irreducible, since the r^{th} power of the roots of $G(Q)$ give the roots of $F(Q)$, and these all have maximal degree over the rationals. We economize on notation by replacing η' by η and $G(Q)$ by $F(Q)$.

Notice that since ρ was on the Hitchin component for $S^2(3,3,3,3)$, $F(Q)$ is totally real [6, Theorem 1.5]. The simple curve η cannot encircle just one cone point, else it would have order 3 so η splits the surface into two pieces each of which must have at least two cone points in it. By applying the construction of Theorem 4.1 if necessary, we may assume that we have constructed a planar orbifold surface (denote it by Σ) in which η is simple and each of the two sides of η contains a large number of cone points.

We have already observed that the initial representation ρ lies on the Hitchin component for $S^2(3,3,3,3)$ and it follows that ρ restricted to $\pi_1(\Sigma)$ lies on the Hitchin component for Σ; for example, one can use the bending used to construct ρ restricted to the corresponding subgroup of finite index. Moreover, each side of η is a hyperbolic orbifold with totally geodesic boundary, so that the restriction of the given representation to either side gives an element of the Hitchin component of that side in the sense of [1, Theorem 2.28]. In particular, each side is represented irreducibly into $\mathrm{SO}(J;\mathbb{R}) < \mathrm{SL}(n,\mathbb{R})$.

The main claim now is that there is a path of elements $\delta_t = \exp(t\boldsymbol{v})$ all centralizing η and with $\delta = \delta_1$ in the centralizer of η in $\mathrm{SL}(n,\mathbb{Z})$. Moreover, the element δ does not preserve J (even up to scaling).

Once this is established, Theorem 1.1 is proved with mild variations of the arguments we have already used in the first bending: by choice, we have a 1–parameter family of bendings $\rho^{\exp(t\boldsymbol{v})}$ connecting ρ^δ to ρ. The latter representation lies on the Hitchin component and therefore ρ^δ lies in the Hitchin component. It is therefore faithful.

Moreover, after the δ–bending, one side of the orbifold surface Σ represents into $\mathrm{SO}(J;\mathbb{R})$ and the other in $\mathrm{SO}(\delta^T J\delta;\mathbb{R})$; by absolute irreducibility and the property of δ we claimed above, it follows that there is no form left invariant by the whole orbifold surface group.

The proof of Theorem 1.1 is now completed by the following theorem applied to $\xi = \rho^\delta$.

Theorem 5.1 *Suppose that ξ is any representation on the Hitchin component of a hyperbolic orbifold group Γ which leaves no form invariant. Then ξ restricted to any surface subgroup of finite index in Γ is Zariski dense in* $\mathrm{SL}(n,\mathbb{R})$.

Proof It is shown in [6, Theorem 1.5] that for any representation on the Hitchin component, the nonidentity elements are *loxodromic*, that is to say that their eigenvalues are distinct real numbers and moreover, since n is odd, these eigenvalues are all positive. In particular, it follows from [4, Theorem 2] that all the infinite-order elements of $\xi(\Gamma)$ are in a unique one-parameter subgroup of the exponential map $\exp\colon \mathfrak{sl}(n,\mathbb{R}) \to \mathrm{SL}(n,\mathbb{R})$.

Suppose then that some surface subgroup H of finite index in Γ lies inside a proper algebraic subgroup of $\mathrm{SL}(n,\mathbb{R})$; by Guichard's result (Theorem 3.1), it must be contained in $\mathrm{SO}(J)$ for some form J of signature $(k+1,k)$. Take any loxodromic element $\exp(\boldsymbol{v}) = \gamma \in \xi(\Gamma)$ and choose r so that $\exp(r\boldsymbol{v}) = \gamma^r \in H$. The condition that $\gamma^r \in \mathrm{SO}(J)$ is equivalent to $J\boldsymbol{v}J^{-1} = -\boldsymbol{v}^{tr}$, so the entire one-parameter subgroup $\exp(t\boldsymbol{v})$ lies in $\mathrm{SO}(J)$ and in particular, therefore, $\gamma \in \mathrm{SO}(J)$. However, it is clear that Γ is generated by its loxodromic elements and we would deduce that $\Gamma < \mathrm{SO}(J)$, a contradiction. It follows that H must have Zariski closure $\mathrm{SL}(n,\mathbb{R})$. $\square$

In particular, since any subgroup of finite index in $\xi(\Gamma)$ contains a surface group, it shows that the Zariski closure of any subgroup of finite index (in particular, index $= 1$) is all of $\mathrm{SL}(n,\mathbb{R})$.

5.1 The existence of δ

Recall that the characteristic polynomial of the integer matrix $\rho(\eta)$ is $(Q-1)F(Q)$, where $F(Q)$ is symmetric $\mathbb{Z}$–irreducible and with (distinct) real roots, since ρ is on the Hitchin component. One can see (for example by diagonalizing the element $\rho(\eta)$ and considering the possible forms it could leave invariant) that the centralizer of $\rho(\eta)$ in $\mathrm{SO}(J;\mathbb{R})$ has rank $(n-1)/2$. On the other hand, the totally real number field K defined by a root of $f(Q)=0$ has degree $n-1$, so that the unit group of its ring of integers has rank $n-2$, which is $>(n-1)/2$ for $n\geq 5$.

Make a rational change of basis so that $M^{-1}\rho(\eta)\,.\,M=(1)\oplus A$, where A is an integer matrix in rational canonical form. The ring $\mathbb{Z}[A]$ is a matrix representation of the ring of integers $\mathbb{O}_K$, which therefore contains a multiplicative subring of units of rank $n-2$, ie matrices which have determinant ± 1. Since all elements of the form $M\,.\,\big((1)\oplus\sum r_j A^j\big)\,.\,M^{-1}$ clearly commute with $\rho(\eta)$, it follows from the rank considerations described above that we may find a rational matrix with determinant $=1$ and integer characteristic polynomial in the $\mathrm{SL}(n,\mathbb{R})$–centralizer of $\rho(\eta)$ which does not power into $\mathrm{SO}(J;\mathbb{R})$. By Proposition 3.2 there is some power of this matrix which is integral; this is a choice for δ with the required properties. As observed above, δ commutes with an element which has distinct positive real eigenvalues, so that it is diagonalizable, and by squaring if need be, we arrange that δ has positive eigenvalues. Therefore, it is in the image of the exponential map, so that $\exp(\boldsymbol{v})=\delta$. From this it follows that the entire path $\exp(t\boldsymbol{v})$ centralizes $\rho(\eta)$, so that the bent representation lies on the Hitchin component.

References

[1] **D Alessandrini**, **G-S Lee**, **F Schaffhauser**, *Hitchin components for orbifolds*, J. Eur. Math. Soc. 25 (2023) 1285–1347 MR Zbl

[2] **H Bass**, *Groups of integral representation type*, Pacific J. Math. 86 (1980) 15–51 MR Zbl

[3] **J Bourgain**, **A Gamburd**, **P Sarnak**, *Affine linear sieve, expanders, and sum-product*, Invent. Math. 179 (2010) 559–644 MR Zbl

[4] **W J Culver**, *On the existence and uniqueness of the real logarithm of a matrix*, Proc. Amer. Math. Soc. 17 (1966) 1146–1151 MR Zbl

[5] **O Guichard**, *Zariski closure of positive and maximal representations*, in preparation

[6] **F Labourie**, *Anosov flows, surface groups and curves in projective space*, Invent. Math. 165 (2006) 51–114 MR Zbl

[7] **D D Long**, **A W Reid**, **M Thistlethwaite**, *Zariski dense surface subgroups in* $\mathrm{SL}(3,\mathbb{Z})$, Geom. Topol. 15 (2011) 1–9 MR Zbl

[8] **D D Long**, **M B Thistlethwaite**, *Zariski dense surface subgroups in* $\mathrm{SL}(4,\mathbb{Z})$, Exp. Math. 27 (2018) 82–92 MR Zbl

[9] **D D Long**, **M Thistlethwaite**, *Zariski dense surface subgroups in* $\mathrm{SL}(7,\mathbb{Z})$, computational output (2024) Available at `http://tinyurl.com/computational-output`

[10] **M V Nori**, *On subgroups of* $\mathrm{GL}_n(\mathbb{F}_p)$, Invent. Math. 88 (1987) 257–275 MR Zbl

[11] **P Sarnak**, *Notes on thin matrix groups*, from "Thin groups and superstrong approximation" (E Breuillard, H Oh, editors), Math. Sci. Res. Inst. Publ. 61, Cambridge Univ. Press (2014) 343–362 MR Zbl

[12] **M Suzuki**, *Group theory, I*, Grundl. Math. Wissen. 247, Springer (1982) MR Zbl

[13] **B Weisfeiler**, *Strong approximation for Zariski-dense subgroups of semisimple algebraic groups*, Ann. of Math. 120 (1984) 271–315 MR Zbl

[14] **M Zshornack**, *Integral Zariski dense surface groups in* $\mathrm{SL}(n, \mathbb{R})$, preprint (2022) arXiv 2211.08616

Department of Mathematics, University of California, Santa Barbara
Santa Barbara, CA, United States

Department of Mathematics, University of Tennessee
Knoxville, TN, United States

long@math.ucsb.edu, morwenthistlethwaite@gmail.com

Proposed: Anna Wienhard | Received: 21 July 2021
Seconded: Martin R Bridson, Mladen Bestvina | Accepted: 17 August 2022

Geometry & Topology **28**:3 (2024) 1167–1212
DOI: 10.2140/gt.2024.28.1167
Published: 10 May 2024

Scalar and mean curvature comparison via the Dirac operator

Simone Cecchini
Rudolf Zeidler

We use the Dirac operator technique to establish sharp distance estimates for compact spin manifolds under lower bounds on the scalar curvature in the interior and on the mean curvature of the boundary. In the situations we consider, we thereby give refined answers to questions on metric inequalities recently proposed by Gromov. These include optimal estimates for Riemannian bands and for the long neck problem. In the case of bands over manifolds of nonvanishing $\hat{A}$–genus, we establish a rigidity result stating that any band attaining the predicted upper bound is isometric to a particular warped product over some spin manifold admitting a parallel spinor. Furthermore, we establish scalar and mean curvature extremality results for certain log-concave warped products. The latter includes annuli in all simply connected space forms. On a technical level, our proofs are based on new spectral estimates for the Dirac operator augmented by a Lipschitz potential together with local boundary conditions.

53C21, 53C24; 53C23, 53C27, 58J20

1 Introduction

Manifolds of positive scalar curvature have been a central topic in differential geometry and topology in recent decades. On complete spin manifolds, a particularly powerful tool in the study of positive scalar curvature metrics has been the spinor Dirac operator which facilitates a fruitful exchange between geometry and topology. This technique exploits the tension between, on the one hand, the Schrödinger–Lichnerowicz formula

$$\not{D}^2 = \nabla^*\nabla + \tfrac{1}{4}\,\mathrm{scal},$$

which implies invertibility of the spinor Dirac operator $\not{D}$ when the scalar curvature is uniformly positive and, on the other hand, index theory in the sense of Atiyah and Singer, which in various situations yields differential-topological obstructions to invertibility. Until recently, and with the notable exceptions of sharp Dirac eigenvalue estimates (see Friedrich [15] and Hijazi, Montiel and Roldán [27]), sharp K–area estimates (see Goette and Semmelmann [16] and Llarull [31]), and approaches to the positive mass theorem based on an idea of Witten [41], the strongest applications of the Dirac operator technique in positive scalar curvature geometry have been of a fundamentally *qualitative* nature. Indeed, there is a substantial body of celebrated literature addressing existence questions of positive scalar curvature metrics on a given manifold, or more generally, studying the topology of the space of positive scalar curvature metrics via the Dirac method; see Botvinnik, Ebert and Randal-Williams [7], Gromov and Lawson [23; 21], Lichnerowicz [30] and Stolz [39] for a selection. However, Gromov [18; 19] recently directed the focus towards more *quantitative* questions and proposed studying the geometry of scalar curvature via various metric inequalities which have similarities to classical Riemannian comparison geometry. This resulted in a number of conjectures, a few of which we now recall.

Conjecture 1.1 [19, page 103, long neck problem] Let (M, g) be a compact connected n–dimensional Riemannian manifold with boundary whose scalar curvature is bounded below by $n(n-1)$. Suppose that $\Phi\colon M \to S^n$ is a smooth area-nonincreasing map which is locally constant near the boundary. If

$$\mathrm{dist}_g(\mathrm{supp}(\mathrm{d}\Phi), \partial M) \geq \frac{\pi}{n},$$

then the mapping degree of Φ is zero.

Conjecture 1.2 [18, 11.12, Conjecture D′] Let X be a closed manifold of dimension n and such that $X \setminus \{p_0\}$ with $p_0 \in X$ does not admit a complete metric of positive scalar curvature. Let M be the manifold with boundary obtained from X by removing an open ball around p_0. Then for any Riemannian metric of scalar curvature $\geq n(n-1) > 0$ on M, the width of a geodesic collar neighborhood of ∂M is bounded above by π/n.

Conjecture 1.3 [18, 11.12, Conjecture C] Let M be a closed connected manifold of dimension $n-1 \neq 4$ such that M does not admit a metric of positive scalar curvature. Let g be a Riemannian metric on $V = M \times [-1, 1]$ of scalar curvature bounded below by $n(n-1) = \mathrm{scal}_{S^n}$. Then

$$\mathrm{width}(V, g) \leq \frac{2\pi}{n},$$

where $\mathrm{width}(V, g) := \mathrm{dist}_g(\partial_- V, \partial_+ V)$ is the distance between the two boundary components of V with respect to g.

Note that we have strengthened the bounds in Conjectures 1.1 and 1.2 compared to the original sources. All these constants are optimal, as we shall discuss below.

Gromov's first definitive result on these questions [18] was a proof of Conjecture 1.3 for the torus and related manifolds via the geometric measure theory approach to positive scalar curvature going back to the minimal hypersurface method of Schoen and Yau [38]. In fact, it initially appeared that the Dirac operator technique was not suitable to such quantitative questions, in particular because they involve manifolds with boundary. However, in recent articles of the authors [43; 10; 42], we have demonstrated that the Dirac operator method can in principle be used to approach Conjectures 1.1–1.3. A slightly different Dirac operator approach based on quantitative K–theory leading to similar (nonsharp) estimates for bands was subsequently given by Guo, Xie and Yu [24].

In the present article, we advance the spinor Dirac operator method further and put forward a novel point of view towards Conjectures 1.1–1.3, which brings the mean curvature of the boundary into the focus of attention. That is, under the assumption that the scalar curvature is bounded below by $n(n-1)$ and that suitable index-theoretic invariants do not vanish, we establish a precise quantitative relationship between the mean curvature of the boundary and the relevant distance quantity appearing in situations related to Conjectures 1.1–1.3. More precisely, we show that in each case there exist constants $c_n(l) > 0$, depending on a distance parameter $l > 0$ and the dimension of the manifold, such that if the mean curvature is bounded below by $-c_n(l)$, then the relevant distance is at most l. The crucial property of these constants is that $c_n(l) \to \infty$ as l approaches the conjectured distance bound. In other words, as the relevant distance tends to the threshold, the mean curvature tends to $-\infty$ somewhere at the boundary. The geometric intuition behind this behavior is that the metric must collapse as the critical threshold is approached. Moreover, our new point of view allows us to establish rigidity results for certain extremal cases of the predicted quantitative relationship between scalar curvature, mean curvature and distance.

On a technical level, one ingredient is to augment the spinor Dirac operator — similarly to our previous approaches — by a potential defined in terms of a distance function. This procedure modifies the classical Schrödinger–Lichnerowicz formula in a way that allows us to relate distance estimates to spectral properties of the modified operator. However, the crucial new ingredient is that we study a tailor-made boundary value problem associated to the augmented Dirac operator. This enables us to use spinorial techniques to not only quantitatively control the scalar curvature using a differential expression of the potential, but also bring the mean curvature of the boundary into play. The main principle behind our new approach is that we can compare certain spin manifolds to model spaces which are suitable warped products, provided that one can produce a nontrivial solution of a boundary value problem associated to the augmented Dirac operator on the given manifold. Moreover, up to a constant, the potential directly corresponds to the mean curvature of the cross sections in the model warped product space.

We develop this approach in a general setting that allows us to treat the results related to Conjectures 1.1–1.3 as well as further novel results in an essentially unified way. To this end, we introduce a new abstract geometric structure which we call a *relative Dirac bundle*. This is a Dirac bundle $S \to M$ in the sense of Gromov and Lawson [23, Section 1] (see also Lawson and Michelsohn [28]) together with a suitable bundle involution $\sigma \in C^\infty(M \setminus K, \mathrm{End}(S))$, which is defined outside a compact subset $K \subset M^\circ$ of the interior of the manifold; see Section 2 for details. The use of this structure is twofold. Firstly, together with a suitable function $\psi\colon M \to \mathbb{R}$, it allows us to define the potential term necessary for the precise quantitative estimates. This leads to the *Callias*[1] *operator*

$$\mathcal{B}_\psi = \mathcal{D} + \psi\sigma,$$

where $\mathcal{D}$ is the Dirac operator associated to the Dirac bundle S. Secondly, the involution σ can be used to define natural chiral boundary conditions, which are crucial for the development of a suitable index theory for relative Dirac bundles (see Section 3) and allow spectral estimates for $\mathcal{B}_\psi$ (see Section 4) relating the mean curvature with the value of the function ψ along the boundary. These boundary conditions are related to the treatment of the cobordism theorem via a boundary value problem as in [6, Section 21; 4, Section 6.3]. Our chirality also allows for an auxiliary choice of sign for each boundary component, reminiscent of the boundary conditions considered by Freed [14]. This additional choice will be relevant in the proofs of our results related to Conjecture 1.3.

A further notable observation is that our construction has a vague formal similarity to *μ–bubbles* or *generalized soap bubbles*, which have recently led to substantial advances via the geometric measure theory approach to scalar curvature; see Gromov [19, Section 5], Chodosh and Li [12], Gromov [20], Lesourd, Unger and Yau [29] and Zhu [46; 45]. Indeed, the latter can be viewed as an augmentation of the minimal hypersurface method by suitable potentials.

In Sections 1.1–1.3 of the introduction, we present a simplified overview of our main results. In the main body of the article, these are derived by working with a suitable relative Dirac bundle and choosing a potential that is appropriate for the situation at hand.

1.1 Length of the neck

Here we present our main geometric results related to Conjectures 1.1 and 1.2. We improve the upper bound of π/n to an estimate depending on the mean curvature of the boundary.

In our first result we estimate the length of the neck of a Riemannian manifold with boundary. Recall that for a smooth map of Riemannian manifolds $\Phi\colon M \to N$, the *area contraction constant* at $p \in M$ is defined to be the norm of the induced map $\Phi_*\colon \bigwedge^2 \mathrm{T}_p M \to \bigwedge^2 \mathrm{T}_{f(p)} N$ on 2–vectors. We say the map is *area nonincreasing* if the area contraction constant is ≤ 1 at every point. If M is compact and N is closed, both connected and oriented, where $n = \dim M = \dim N \geq 2$, then a smooth map $\Phi\colon M \to N$ that is locally constant near the boundary ∂M has a well-defined mapping degree $\deg(\Phi) \in \mathbb{Z}$.[2] Moreover, given a Riemannian manifold (M, g), we denote the mean curvature of its boundary ∂M by H_g (or simply H if the metric is implicit); see also (2-14) in Section 2 for our sign and normalization conventions.

[1] We use this terminology because the study of Dirac operators with potential was initiated by Callias [9].

[2] This can be defined as usual via counting the preimage of a regular value with signs or, in cohomological terms, using the induced map $\mathbb{Z} \cong \mathrm{H}^n(N;\mathbb{Z}) \cong \mathrm{H}^n(N, \Phi(\partial M);\mathbb{Z}) \xrightarrow{\Phi^*} \mathrm{H}^n(M, \partial M;\mathbb{Z}) \cong \mathbb{Z}$, where we use that $\Phi(\partial M)$ is a finite set of points and $n \geq 2$.

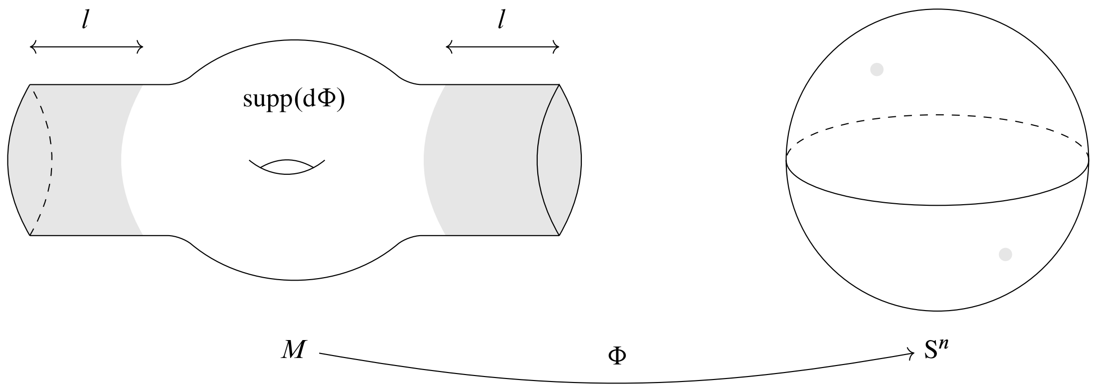


Figure 1: The long neck problem.

Theorem 1.4 (see Section 5) *Let (M, g) be a compact connected Riemannian spin manifold with nonempty boundary, $n = \dim M \geq 2$ even, and let $\Phi\colon M \to \mathrm{S}^n$ be a smooth area-nonincreasing map. Assume that $\mathrm{scal}_g \geq n(n-1)$. Moreover, suppose there exists $l \in (0, \pi/n)$ such that $\mathrm{H}_g \geq -\tan\bigl(\frac{1}{2}nl\bigr)$ and $\mathrm{dist}_g\bigl(\mathrm{supp}(\mathrm{d}\Phi), \partial M\bigr) \geq l$; compare Figure 1. Then $\deg(\Phi) = 0$.*

The statement of Theorem 1.4 is sharp; we discuss this in Section 5, Proposition 5.2. In this context, a subtle point hidden in the statement of the theorem is that we also rule out the equality situation $\mathrm{dist}_g\bigl(\mathrm{supp}(\mathrm{d}\Phi), \partial M\bigr) = l$ under these scalar and mean curvature bounds if $\deg(\Phi) \neq 0$. This is in contrast to the situations of the other conjectures, where the equality situations can be realized; compare Remark 1.13 below. Addressing this detail requires a considerably more precise analysis than in the earlier approach from [10, Theorem A].

Corollary 1.5 *Let (M, g) be a compact connected Riemannian spin manifold with boundary, let $n = \dim M \geq 2$ be even, and let $\Phi\colon M \to \mathrm{S}^n$ be a smooth area-nonincreasing map. Assume that $\mathrm{scal}_g \geq n(n-1)$. If $\mathrm{dist}_g\bigl(\mathrm{supp}(\mathrm{d}\Phi), \partial M\bigr) \geq \pi/n$, then $\deg(\Phi) = 0$.*

This corollary is a direct consequence of Theorem 1.4 because $\tan\bigl(\frac{1}{2}nl\bigr) \to \infty$ as $l \to \pi/n$. Thus this refines the original approach to Conjecture 1.1 from [10, Theorem A]. Proposition 5.2 also shows that the constant π/n is optimal here.

For our second result, we introduce the notion of $\widehat{\mathrm{A}}$–area, which is a generalization of the notion of K–area introduced in [17, Section 4]. For a Hermitian bundle E, denote by R^E the curvature of the connection on E.

Definition 1.6 Let (X, g) be a closed even-dimensional oriented Riemannian manifold. The $\widehat{\mathrm{A}}$*–area* of (X, g) is the supremum of the numbers $\|\mathrm{R}^E\|_\infty^{-1}$, ranging over all Hermitian bundles E with metric connections such that $\int_X \widehat{\mathbf{A}}(X) \wedge \mathrm{ch}(E) \neq 0$, where $\widehat{\mathbf{A}}(X)$ denotes the $\widehat{\mathrm{A}}$–form of X and $\mathrm{ch}(E)$ denotes the Chern character form of E.

The $\widehat{\mathrm{A}}$–*area* of (X, g) depends on the metric g. However, since X is compact, the notion of having *infinite* $\widehat{\mathrm{A}}$–*area* is independent of g. A closed spin manifold of infinite K–area also has infinite $\widehat{\mathrm{A}}$–area. An important class of examples consists of even-dimensional compactly enlargeable manifolds, eg the $2m$–dimensional torus T^{2m}. For the notion of enlargeability and more examples, we refer to [22; 23]. More generally, if X has infinite K–area and Y has nonvanishing $\widehat{\mathrm{A}}$–genus, then $X \times Y$ has infinite $\widehat{\mathrm{A}}$–area. This includes, for example, the Cartesian product $\mathrm{T}^{2m} \times Y$, where Y is the K3–surface.

Now suppose X is a closed n–dimensional enlargeable spin manifold. By [10, Theorem C], $X \setminus \{p_0\}$ with $p_0 \in X$ does not admit a complete metric of positive scalar curvature. Moreover, the double of $M := X \setminus \mathrm{B}^n$, with B^n an open n–ball embedded in X, is enlargeable as well. Therefore, the next theorem in particular refines the upper bound of Conjecture 1.2 in the case of even-dimensional enlargeable spin manifolds, using information from the mean curvature of the boundary.

Theorem 1.7 (see Section 6) *Let (M, g) be a compact connected n–dimensional Riemannian spin manifold with boundary such that the double of M has infinite $\widehat{\mathrm{A}}$–area. Suppose that* $\mathrm{scal}_g > 0$ *and that there exist positive constants κ and l, with $0 < l < \pi/(\sqrt{\kappa} n)$, such that*

$$\mathrm{H}_g \geq -\sqrt{\kappa} \tan\left(\tfrac{1}{2}\sqrt{\kappa} n l\right).$$

Then the boundary ∂M admits no open geodesic collar neighborhood $\mathcal{N} \subseteq M$ of width strictly greater than l such that $\mathrm{scal}_g \geq \kappa n(n-1)$ *on $\mathcal{N}$.*

The estimate in this theorem is also optimal; see Remark 1.13 below.

Remark 1.8 Similarly as before, one deduces from this that the case $l \geq \pi/(n\sqrt{\kappa})$ is ruled out independently of mean curvature restrictions. This also follows from the techniques in [10, Theorem B], and see the discussion in [11].

In particular, Theorem 1.7 implies that a manifold with boundary whose double has infinite $\widehat{\mathrm{A}}$–area cannot carry any metric of positive scalar curvature and mean convex boundary. This also follows from recent results of Bär and Hanke [5].

1.2 Estimates of bands

Here we exhibit our main results related to Conjecture 1.3. Similarly as above, we are able to improve the upper bound of $2\pi/n$ to a bound depending on the mean curvature of the boundary. We will formulate our result for manifolds which are not only cylinders. We say that a *band* is a compact manifold V together with a decomposition $\partial V = \partial_- V \sqcup \partial_+ V$, where $\partial_\pm V$ are unions of components. This notion goes back to Gromov [18], where such manifolds are called compact proper bands. A map $V \to V'$ is a *band map* if it takes $\partial_\pm V$ to $\partial_\pm V'$. The *width* $\mathrm{width}(V, g)$ of a Riemannian band (V, g) is the distance between $\partial_- V$ and $\partial_+ V$ with respect to g.

We now focus on a class of bands to which our results apply and which is simple to describe. An *overtorical band* [18, Section 2] is a band V together with a smooth band map $V \to \mathrm{T}^{n-1} \times [-1, 1]$ of nonzero mapping degree, where T^{n-1} denotes the torus of dimension $n-1$. More generally, we can also consider $\widehat{\mathrm{A}}$–overtorical bands [42], which are defined similarly but replacing the usual mapping degree by the $\widehat{\mathrm{A}}$–degree in the sense of [22, Definition 2.6].

Theorem 1.9 (cf Corollaries 7.7 and 7.8) *Let n be odd and (V, g) be an n–dimensional $\widehat{\mathrm{A}}$–overtorical spin band. Suppose that $\mathrm{scal}_g \geq n(n-1)$. If the mean curvature of the boundary satisfies either*

- $\mathrm{H}_g \geq -\tan\bigl(\frac{1}{4}nl\bigr)$ *for some* $0 < l < 2\pi/n$, *or*
- $\mathrm{H}_g|_{\partial_- V} \geq 0$ *and* $\mathrm{H}_g|_{\partial_+ V} \geq -\tan\bigl(\frac{1}{2}nl\bigr)$ *for some* $0 < l < \pi/n$,

then $\mathrm{width}(V, g) \leq l$.

Again, as the expression $-\tan\bigl(\frac{1}{4}nl\bigr)$ tends to $-\infty$ as $l \to 2\pi/n$, we obtain the strict version of the original estimate desired by Conjecture 1.3. This also follows from [42, Corollary 1.5]. If, in addition, we assume that one of the boundary components is mean convex, then we can even obtain a strict bound of π/n.

Corollary 1.10 (cf Corollary 7.9) *Let n be odd and (V, g) be an n–dimensional $\widehat{\mathrm{A}}$–overtorical spin band. Suppose that $\mathrm{scal}_g \geq n(n-1)$. Then we always have $\mathrm{width}(V, g) < 2\pi/n$. Moreover, if $\partial_- V$ is mean convex, then $\mathrm{width}(V, g) < \pi/n$.*

Remark 1.11 The statement of Theorem 1.9 exhibited here in the introduction is a special case of the more general Theorem 7.6 comparing the mean curvature to arbitrary values of the form $\mp\tan\bigl(\frac{1}{2}nt_\pm\bigr)$, where $-\pi/n < t_- < t_+ < \pi/n$, to get a corresponding width bound $t_+ - t_-$.

Remark 1.12 Our methods apply to a more general class of bands of *infinite vertical* $\widehat{\mathrm{A}}$*–area* which, in particular, includes bands diffeomorphic to $M \times [-1, 1]$ with M being a closed spin manifold of infinite $\widehat{\mathrm{A}}$–area in the sense of Definition 1.6. See Section 7 for details.

Remark 1.13 The band estimates given in Theorem 1.9 are sharp. This follows from the warped product metric $\varphi^2 g_{\mathrm{T}^{n-1}} + \mathrm{d}x \otimes \mathrm{d}x$ on $\mathrm{T}^{n-1} \times [-l, l]$ for any $0 < l < \pi/n$, where $g_{\mathrm{T}^{n-1}}$ is the flat torus metric and $\varphi(t) = \cos\bigl(\frac{1}{2}nt\bigr)^{2/n}$, as indicated by Gromov in [18, page 653]. By rescaling a given arbitrary metric g_M on any manifold M allowed in Conjecture 1.3, the warped product metric $\varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x$ in fact shows optimality of this estimate on any band diffeomorphic to $M \times [-1, 1]$. Incidentally, this construction also shows optimality of the estimate in Theorem 1.7 by forgetting the band structure and simply considering $M = \mathrm{T}^{n-1} \times [-l, l]$ with this warped product metric.

1.3 Extremality and rigidity results

In view of the band width estimates, it is natural to investigate the extremal case. The class of $\widehat{\mathrm{A}}$–overtorical bands discussed above includes the special case of $\widehat{\mathrm{A}}$*–bands*, that is, bands such that $\widehat{\mathrm{A}}(\partial_- V) \neq 0$ (and thus, by bordism invariance also $\widehat{\mathrm{A}}(\partial_+ V) \neq 0$). In this special case, we prove the following rigidity theorem stating that the extremal case can only be achieved by the warped product construction discussed in Remark 1.13 over a Ricci flat manifold.

Theorem 1.14 (cf Corollary 9.2) *Let (V, g) be an n–dimensional band which is a spin manifold and satisfies $\widehat{\mathrm{A}}(\partial_- V) \neq 0$. Suppose that $\mathrm{scal}_g \geq n(n-1)$. Let $0 < d < \pi/n$ and assume furthermore that one of the following conditions holds: either*

- $\mathrm{width}(V, g) \geq 2d$ *and we have* $\mathrm{H}_g|_{\partial V} \geq -\tan\big(\tfrac{1}{2}nd\big)$, *or*
- $\mathrm{width}(V, g) \geq d$ *and we have* $\mathrm{H}_g|_{\partial_- V} \geq 0$ *and* $\mathrm{H}_g|_{\partial_+ V} \geq -\tan\big(\tfrac{1}{2}nd\big)$.

Then (V, g) is isometric to a warped product $(M \times I, \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x)$, where $I = [-d, d]$ or $I = [0, d]$, $\varphi(t) = \cos(nt/2)^{2/n}$ and g_M is some Riemannian metric on M which carries a nontrivial parallel spinor. In particular, g_M is Ricci-flat.

Again, we also have a more general version of this theorem involving arbitrary mean curvature bounds of the form $\mp\tan\big(\tfrac{1}{2}nt_\pm\big)$; see Theorem 9.1.

The study of extremality questions about scalar curvature has a long history initiated by Gromov's K–area inequalities [17]. Llarull [31] proved sharp inequalities using the Dirac operator, which imply that the round metric on the sphere is scalar-curvature extremal, meaning it cannot be enlarged without a decrease in the scalar curvature at some point. Llarull's technique and results were subsequently refined and generalized by Goette and Semmelmann [16] and remain of central importance in contemporary research on scalar curvature; see for instance [18, Section 10; 19, Section 4; 44]. Lott [32] recently extended the technique of Llarull and Goette and Semmelmann to even-dimensional manifolds with boundary using the Dirac operator with (local) boundary conditions. The following results combine the technique of Goette and Semmelmann with our machinery to obtain a new kind of extremality result for a large class of warped product manifolds.

Let M be a manifold with boundary ∂M. We say that a Riemannian metric g_M on M is *scalar-mean extremal* if every metric g on M which satisfies $g \geq g_M$, $\mathrm{scal}_g \geq \mathrm{scal}_{g_M}$ and $\mathrm{H}_g \geq \mathrm{H}_{g_M}$ already satisfies $\mathrm{H}_{g_M} = \mathrm{H}_g$ and $\mathrm{scal}_{g_M} = \mathrm{scal}_g$. Moreover, we say that g_M is *scalar-mean rigid* if any such metric must satisfy $g = g_M$.

Theorem 1.15 (cf Corollary 10.3) *Let n be odd and (M, g_M) be an $(n-1)$–dimensional Riemannian spin manifold of nonvanishing Euler-characteristic whose Riemannian curvature operator is nonnegative. Let $\varphi\colon [t_-, t_+] \to (0, \infty)$ be a smooth strictly logarithmically concave function and consider the warped product metric $g_V = \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x$ on $V := M \times [t_-, t_+]$. Then any metric g on V which satisfies*

(i) $g \geq g_V$,

(ii) $\mathrm{scal}_g \geq \mathrm{scal}_{g_V}$, *and*

(iii) $\mathrm{H}_g \geq \mathrm{H}_{g_V}$

is itself a warped product $g = \varphi^2 \tilde{g}_M + \mathrm{d}x \otimes \mathrm{d}x$ for some metric $\tilde{g}_M$ on M such that $\mathrm{scal}_{\tilde{g}_M} = \mathrm{scal}_{g_M}$. In particular, g_V is scalar-mean extremal.

If, in addition, the metric g_M satisfies $\mathrm{Ric}_{g_M} > 0$, then g_V is scalar-mean rigid.

Note that strictly logarithmically concave means that $\log(\varphi)'' < 0$. Geometrically, this means that the mean curvature of $M \times \{x\}$ in the warped product (with respect to the normal field ∂_x) is strictly increasing.

Remark 1.16 We have a more general version of this theorem involving distance-nonincreasing maps of nonzero degree to these spaces; see Theorem 10.2.

A particularly interesting special case of manifolds which can be written as such log-concave warped products are annuli in spaces of constant curvature. Indeed, as a direct consequence of Theorem 1.15, we obtain the following result.

Corollary 1.17 (cf Corollary 10.5) *Let $n \geq 3$ be odd and (M_κ, g_κ) the n–dimensional simply connected space form of constant sectional curvature $\kappa \in \mathbb{R}$. Let $0 < t_- < t_+ < t_\infty$, where $t_\infty = +\infty$ if $\kappa \leq 0$, and $t_\infty = \pi/\sqrt{\kappa}$ if $\kappa > 0$. Consider the annulus*

$$\mathrm{A}_{t_-,t_+} := \{p \in M_\kappa \mid t_- \leq d_{g_\kappa}(p, p_0) \leq t_+\}$$

around some basepoint $p_0 \in M_\kappa$. Then the metric g_κ is scalar-mean rigid on A_{t_-,t_+}.

Similar statements for log-concave warped products (and, in particular, punctured space forms) have been suggested by Gromov in [19, Section 5.4] based on considerations with μ–bubbles. Moreover, different rigidity statements for hyperbolic space have been studied in the context of the positive mass theorem; see for instance Andersson and Dahl [1], Chruściel and Herzlich [13], Min-Oo [33] and Sakovich and Sormani [37].

Remark 1.18 These results also appear quite similar to Lott's main result in [32] (apart from the fact that they apply in complementary parities of the dimension). They are, however, of an essentially different geometric nature. On the one hand, [32] proves *area* extremality rather than merely (length) extremality as in our results. Our technique here relies on the existence of a suitable Lipschitz function and does not appear to be readily applicable to the study of area-nonincreasing maps. On the other hand, we do not require that the curvature operator of the metric on V itself is nonnegative or that the second fundamental form of the boundary is nonnegative. Indeed, as the examples of hyperbolic annuli show, our results include manifolds of negative sectional curvature. Similarly, suitable spherical annuli are examples which have negative second fundamental form at the boundary. One should note that on a manifold *without boundary*, a metric of negative scalar curvature can never be scalar extremal — simply rescaling using a constant > 1 provides a counterexample. However, in our example of an annulus in hyperbolic space, the outer boundary component has positive mean curvature, which thwarts such a rescaling argument. This shows that the presence of the outer boundary component is crucial for our result to hold in the negative curvature example. More generally, in all the examples covered by Corollary 1.17 it is the case that at least one quantity is positive among the scalar curvature in the interior and the mean curvatures on the two boundary components.

1.4 Higher index theory and future directions

In the present article we are only working with classical Dirac-type operators on finite-dimensional Hermitian vector bundles rather than using *higher index theory*, which would involve bundles with coefficients in infinite-dimensional C*–algebras. This is in contrast to our previous work [43; 10; 42]. The main rationale behind this change of perspective is that it allows a quicker and more accessible exposition of the novel geometric arguments we want to exhibit through this article. At the same time, this does not sacrifice too much of the possible generality of the statements, because many examples which are usually approached via higher index theory can already be dealt with via more classical notions like *enlargeability* or *infinite* K*–area*. It does, however, restrict the parity of the dimension in some of the results.

Note that our central structure of a *relative Dirac bundle* can be straightforwardly generalized to C*–algebra coefficients. In this way it would be possible to reformulate all arguments from [43; 10; 42] in terms of this concept. This will be partly explained in [11]. Since this notion is fairly general, we also expect that it can be applied in many other geometric contexts we have not considered thus far. However, extending the results from this article to higher index theory requires some new analytic work because the present state of the art in the literature on boundary value problems for Dirac-type operators, where we mostly rely on Bär and Ballmann [4; 3], does not cover the case of infinite-dimensional bundles. Since this aspect is orthogonal to the geometric arguments presented here, we will address it separately in future work.

Acknowledgements We thank Misha Gromov for useful comments on an earlier version of this manuscript and pointing us to the examples showing optimality of Theorem 1.4. We also would like to thank Bernd Ammann for a brief helpful correspondence and the referee for valuable suggestions improving the exposition.

Cecchini was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Priority Programme *Geometry at infinity*' (SPP 2026, CE 393/1-1). Zeidler was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 427320536 – SFB 1442, as well as under Germany's Excellence Strategy EXC 2044 390685587, Mathematics Münster: Dynamics–Geometry–Structure, and through the Priority Programme *Geometry at infinity* (SPP 2026, ZE 1123/2-2).

2 Relative Dirac bundles

In this section, we set up the differential geometric preliminaries underlying the rest of this article. In particular, we introduce our new concept of a *relative Dirac bundle*.

We begin by fixing some notation. Let (M, g) be a Riemannian manifold and let $E \to M$ be a Hermitian vector bundle. The space of smooth sections will be denoted by $\mathrm{C}^\infty(M, E)$ and the subspace of compactly supported smooth sections by $\mathrm{C}^\infty_{\mathrm{c}}(M, E)$. We denote fiberwise inner products by $\langle -, - \rangle$ and fiberwise norms by $|-|$. Next we recall the notion of a classic Dirac bundle in the sense of Gromov and Lawson.

Definition 2.1 [23, Section 1], [28] A ($\mathbb{Z}/2$–graded) *Dirac bundle* over M is a Hermitian vector bundle $S \to M$ with a metric connection $\nabla\colon \mathrm{C}^\infty(M,S) \to \mathrm{C}^\infty(M, \mathrm{T}^*M \otimes S)$ (endowed with a parallel and orthogonal $\mathbb{Z}/2$–grading $S = S^+ \oplus S^-$) and a parallel bundle map $\mathrm{c}\colon \mathrm{T}^*M \to \mathrm{End}(S)$, called *Clifford multiplication*, such that $\mathrm{c}(\omega)$ is antiselfadjoint (and odd), and $\mathrm{c}(\omega)^2 = -|\omega|^2$ for all $\omega \in \mathrm{T}^*M$.

For simplicity, we sometimes denote Clifford multiplication by $\omega \cdot u := \mathrm{c}(\omega)u$. Let the associated Dirac operator be $\mathcal{D} = \sum_{i=1}^n \mathrm{c}(e^i)\nabla_{e_i}\colon \mathrm{C}^\infty(M,S) \to \mathrm{C}^\infty(M,S)$, where $e_1,\dots,e_n$ is a local frame and $e^1,\dots,e^n$ the dual coframe. If we have a $\mathbb{Z}/2$–grading $S = S^+ \oplus S^-$, the operator $\mathcal{D}$ is odd, that is, it is of the form

$$\mathcal{D} = \begin{pmatrix} 0 & \mathcal{D}^- \\ \mathcal{D}^+ & 0 \end{pmatrix},$$

where $\mathcal{D}^\pm\colon \mathrm{C}^\infty(M,S^\pm) \to \mathrm{C}^\infty(M,S^\mp)$ are formally adjoint to one another.

We now turn to relative Dirac bundles, an augmentation of a Dirac bundle.

Definition 2.2 Let $K \subset M^\circ$ be compact subset in the interior. A *relative Dirac bundle* with *support* K is a $\mathbb{Z}/2$–graded Dirac bundle $S \to M$ together with an odd, selfadjoint, parallel bundle involution $\sigma \in \mathrm{C}^\infty(M \setminus K, \mathrm{End}(S))$ satisfying $\mathrm{c}(\omega)\sigma = -\sigma\,\mathrm{c}(\omega)$ for every $\omega \in \mathrm{T}^*M|_{M\setminus K}$ and such that σ admits a smooth extension to a bundle map on an open neighborhood of $\overline{M \setminus K}$.

The final technical requirement in particular ensures the existence of a unique continuous extension of σ to the topological boundary of K. We do not consider the further extension to a neighborhood of $\overline{M \setminus K}$ part of the data, we just require its existence.

It follows directly from the definition that the Dirac operator anticommutes with the involution σ where it is defined.

One main use of this structure is that it will allow us to associate local boundary conditions to any choice of a sign for each connected component of ∂M.

Definition 2.3 (boundary chirality for relative Dirac bundles) For a relative Dirac bundle $S \to M$ and a locally constant function $s\colon \partial M \to \{\pm 1\}$, we say that the endomorphism

$$\chi := s\,\mathrm{c}(\nu^\flat)\sigma\colon S|_{\partial M} \to S|_{\partial M} \tag{2-1}$$

is the *boundary chirality* on S associated to the choice of signs s. Here, $\nu \in \mathrm{C}^\infty(\partial M, \mathrm{T}M|_{\partial M})$ is the inward-pointing unit normal field.

Note that χ is a selfadjoint even involution, which anticommutes with $\mathrm{c}(\nu^\flat)$ but commutes with $\mathrm{c}(\omega)$ for all $\omega \in \mathrm{T}^*(\partial M)$. This defines local boundary conditions for sections $u \in \mathrm{C}^\infty(M,S)$ by requiring that

$$\chi(u|_{\partial M}) = u|_{\partial M}. \tag{2-2}$$

In Section 3, we will observe that these boundary conditions are elliptic. The next lemma will be used to show that they are selfadjoint.

Lemma 2.4 *Let $S \to M$ be a relative Dirac bundle and $s\colon \partial M \to \{\pm 1\}$ a choice of signs. Then $L(\chi) := \{u \in S|_{\partial M} \mid \chi(u) = u\}$ is a Lagrangian subbundle of $S|_{\partial M}$ with respect to the fiberwise symplectic form $(u, v) \mapsto \langle u, \mathrm{c}(\nu^\flat)v\rangle$. In other words, $\mathrm{c}(\nu^\flat)L(\chi) = L(\chi)^\perp$.*

Proof Note that $L(\chi)$ is the image of the orthogonal bundle projection $\frac{1}{2}(1+\chi)$. Since $L(\chi)^\perp$ is the image of the complementary projection $\frac{1}{2}(1-\chi)$ and χ anticommutes with $\mathrm{c}(\nu^\flat)$, we obtain $\mathrm{c}(\nu^\flat)L(\chi) = L(\chi)^\perp$. □

In the following, we recall several standard formulas on Dirac bundles for later use.

- **Green's formula** [40, Proposition 9.1] For the Dirac operator,

$$\int_M \langle \mathcal{D}u, v\rangle \,\mathrm{vol}_M = \int_M \langle u, \mathcal{D}v\rangle \,\mathrm{vol}_M + \int_{\partial M} \langle u, \nu^\flat \cdot v\rangle \,\mathrm{vol}_{\partial M}, \tag{2-3}$$

and, for the connection,

$$\int_M \langle \nabla u, \varphi\rangle \,\mathrm{vol}_M = \int_M \langle u, \nabla^* \varphi\rangle \,\mathrm{vol}_M - \int_{\partial M} \langle u, \varphi(\nu)\rangle \,\mathrm{vol}_{\partial M}, \tag{2-4}$$

where $u, v \in \mathrm{C}^\infty_\mathrm{c}(M, S)$, $\varphi \in \mathrm{C}^\infty_\mathrm{c}(M, \mathrm{T}^*M \otimes S)$ and $\nu \in \mathrm{C}^\infty(\partial M, \mathrm{T}M|_{\partial M})$ is the inward-pointing normal field.

- **Bochner–Lichnerowicz–Weitzenböck formula** [23, Proposition 2.5] We have

$$\mathcal{D}^2 = \nabla^*\nabla + \mathcal{R}, \tag{2-5}$$

where $\nabla^*\nabla = -\sum_{i=1}^n \nabla^2_{e_i, e_i}$ is the connection laplacian and

$$\mathcal{R} = \sum_{i<j} \mathrm{c}(e^i)\,\mathrm{c}(e^j)\,\mathrm{R}^\nabla(e_i, e_j)$$

is a bundle endomorphism linearly depending on the curvature tensor R^∇ of ∇.

- **Penrose operator and Friedrich inequality** [8, Section 5.2] We have that

$$|\nabla u|^2 = |\mathcal{P}u|^2 + \frac{1}{n}|\mathcal{D}u|^2 \tag{2-6}$$

for all $u \in \mathrm{C}^\infty(M, S)$, where $\mathcal{P}\colon \mathrm{C}^\infty(M, S) \to \mathrm{C}^\infty(M, \mathrm{T}^*M \otimes S)$ is the *Penrose operator* defined as

$$\mathcal{P}_\xi u := \nabla_\xi u + \frac{1}{n}\xi^\flat \cdot \mathcal{D}u.$$

In particular, we have the *Friedrich inequality*

$$|\nabla u|^2 \geq \frac{1}{n}|\mathcal{D}u|^2,$$

where equality holds if and only if $\mathcal{P}u = 0$, that is, u satisfies the *twistor equation*

$$\nabla_\xi u + \frac{1}{n}\xi^\flat \cdot \mathcal{D}u = 0 \tag{2-7}$$

for all $\xi \in \mathrm{T}M$.

• **Boundary Dirac operator** [4, Appendix 1], [27, Introduction] Let ν be the interior unit normal field. We turn $S^\partial = S|_{\partial M}$ into a Dirac bundle via the Clifford multiplication and connection,

$$\begin{aligned} \mathrm{c}^\partial(\omega) &= \mathrm{c}(\omega)\,\mathrm{c}(\nu^\flat) && \text{for } \omega \in \mathrm{T}^*\partial M, \\ \nabla^\partial_\xi &= \nabla_\xi + \tfrac{1}{2}\,\mathrm{c}^\partial(\nabla_\xi \nu^\flat) && \text{for } \xi \in \mathrm{T}\partial M. \end{aligned} \tag{2-8}$$

Denote the corresponding Dirac operator by

$$\mathcal{A}\colon \mathrm{C}^\infty(\partial M, S^\partial) \to \mathrm{C}^\infty(\partial M, S^\partial), \quad \text{where } \mathcal{A} = \sum_{i=1}^{n-1} \mathrm{c}^\partial(e^i)\nabla^\partial_{e_i}. \tag{2-9}$$

This is the *canonical boundary operator* associated to the Dirac bundle $S \to M$. Note that $\mathcal{A}$ is even with respect to the grading on S^∂ restricted from the grading on S. It satisfies

$$\mathcal{A}\,\mathrm{c}(\nu^\flat) = -\mathrm{c}(\nu^\flat)\mathcal{A} \tag{2-10}$$

and, if $S \to M$ is endowed with the structure of a relative Dirac bundle, then

$$\mathcal{A}\sigma = \sigma\mathcal{A} \quad \text{and} \quad \chi\mathcal{A} = -\mathcal{A}\chi, \tag{2-11}$$

where χ is defined as in (2-1) with respect to any choice of signs. This implies that, if $u \in \mathrm{C}^\infty(M, S)$ satisfies the boundary condition (2-2), then

$$\langle u|_{\partial M}, \mathcal{A}u|_{\partial M}\rangle = 0. \tag{2-12}$$

Moreover,

$$\mathcal{A} = \tfrac{1}{2}(n-1)\,\mathrm{H} - \mathrm{c}(\nu^\flat)\sum_{i=1}^{n-1}\mathrm{c}(e^i)\nabla_{e_i} = \tfrac{1}{2}(n-1)\,\mathrm{H} - \mathrm{c}(\nu^\flat)\mathcal{D} - \nabla_\nu, \tag{2-13}$$

where H is the mean curvature of ∂M with respect to ν. To avoid any confusion about signs and normalization, let us be explicit about our convention for the mean curvature:

$$\mathrm{H} = \frac{1}{n-1}\,\mathrm{tr}(-\nabla\nu) = \frac{1}{n-1}\sum_{i=1}^{n-1}\langle e_i, -\nabla_{e_i}\nu\rangle = \frac{1}{n-1}\sum_{i=1}^{n-1}\mathrm{II}(e_i, e_i). \tag{2-14}$$

Here $-\nabla\nu$ is the shape operator, II denotes the second fundamental form and we use a local orthonormal frame $e_1, \dots, e_{n-1}$ on ∂M.

Finally, in the remainder of this section, we discuss two concrete geometric examples of relative Dirac bundles, which are relevant for our main results.

Example 2.5 Let M be a compact even-dimensional Riemannian spin manifold with boundary and let $\$_M = \$^+_M \oplus \$^-_M$ be the associated $\mathbb{Z}/2$–graded complex spinor bundle endowed with the Levi-Civita connection. Let $E, F \to M$ be a pair of Hermitian bundles equipped with metric connections. Note that the bundle

$$S := \$_M \mathbin{\widehat{\otimes}} (E \oplus F^{\mathrm{op}})$$

is a $\mathbb{Z}/2$–graded Dirac bundle, where the grading $S = S^+ \oplus S^-$ is given by

$$S^+ := (\slashed{S}_M^+ \otimes E) \oplus (\slashed{S}_M^- \otimes F) \quad \text{and} \quad S^- := (\slashed{S}_M^+ \otimes F) \oplus (\slashed{S}_M^- \otimes E).$$

In analogy with Gromov and Lawson [23], we make the following assumption:

(2-15) There exists a compact set $K \subset M^\circ$ and a parallel unitary bundle isomorphism $\mathfrak{t}\colon E|_{M\setminus K} \to F|_{M\setminus K}$ which extends to a smooth bundle map on a neighborhood of $\overline{M \setminus K}$.

In this case, we say that (E, F) is a *GL pair* with *support* K. Note that condition (2-15) implies that S is a relative Dirac bundle with involution

$$\sigma := \mathrm{id}_{\slashed{S}_M} \mathbin{\hat{\otimes}} \begin{pmatrix} 0 & \mathfrak{t}^* \\ \mathfrak{t} & 0 \end{pmatrix} \colon S|_{M\setminus K} \to S|_{M\setminus K},$$

where $\hat{\otimes}$ stands for the graded tensor product of operators; compare [26, Appendix A]. This means $\sigma((u_+ \oplus u_-) \otimes (e \oplus f)) = (u_+ \oplus -u_-) \otimes (\mathfrak{t}^* f \oplus \mathfrak{t} e)$ for sections $u_\pm \in \mathrm{C}^\infty(M, \slashed{S}_M^\pm)$, $e \in \mathrm{C}^\infty(M, E)$ and $f \in \mathrm{C}^\infty(M, F)$. The Dirac operator on S is described as follows. Let $\slashed{D}_E$ and $\slashed{D}_F$ be the operators obtained by twisting the complex spin Dirac operator on M respectively with the bundles E and F. Observe that $\slashed{D}_E$ and $\slashed{D}_F$ are odd operators, that is, they are of the form

$$\slashed{D}_E = \begin{pmatrix} 0 & \slashed{D}_E^- \\ \slashed{D}_E^+ & 0 \end{pmatrix} \quad \text{and} \quad \slashed{D}_F = \begin{pmatrix} 0 & \slashed{D}_F^- \\ \slashed{D}_F^+ & 0 \end{pmatrix},$$

where $\slashed{D}_E^+ \colon \mathrm{C}_\mathrm{c}^\infty(M, \slashed{S}^+ \otimes E) \to \mathrm{C}_\mathrm{c}^\infty(M, \slashed{S}^- \otimes E)$ and $\slashed{D}_F^+ \colon \mathrm{C}^\infty(M, \slashed{S}^+ \otimes F) \to \mathrm{C}^\infty(M, \slashed{S}^- \otimes F)$, and $\slashed{D}_E^-$ and $\slashed{D}_F^-$ are formally adjoint to $\slashed{D}_E^+$ and $\slashed{D}_F^+$, respectively. The Dirac operator on S is given by

$$\mathcal{D} = \begin{pmatrix} 0 & \mathcal{D}^- \\ \mathcal{D}^+ & 0 \end{pmatrix} \colon \mathrm{C}^\infty(M, S) \to \mathrm{C}^\infty(M, S), \tag{2-16}$$

where $\mathcal{D}^\pm \colon \mathrm{C}^\infty(M, S^\pm) \to \mathrm{C}^\infty(M, S^\mp)$ are the operators defined as

$$\mathcal{D}^+ := \begin{pmatrix} 0 & \slashed{D}_F^- \\ \slashed{D}_E^+ & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{D}^- := \begin{pmatrix} 0 & \slashed{D}_E^- \\ \slashed{D}_F^+ & 0 \end{pmatrix}. \tag{2-17}$$

Moreover, the curvature endomorphism $\mathcal{R}$ from (2-5) is given by

$$\mathcal{R} = \tfrac{1}{4}\,\mathrm{scal}_g + \mathcal{R}^{E\oplus F}, \tag{2-18}$$

where

$$\mathcal{R}^{E\oplus F} = \sum_{i<j} \mathrm{c}(e^i)\,\mathrm{c}(e^j)(\mathrm{id}_{\slashed{S}_M} \otimes \mathrm{R}^{\nabla^{E\oplus F}}_{e_i, e_j})$$

is an even endomorphism of the bundle S which depends linearly on the curvature of the connection on $E \oplus F$; compare [28, Theorem 8.17].

Example 2.6 Let (V, g) be an odd-dimensional Riemannian spin band and let $\slashed{S}_V \to V$ be the associated complex spinor bundle, endowed with the connection induced by the Levi-Civita connection. Let $E \to M$

be a Hermitian bundle equipped with a metric connection. Then $S := (\$_V \otimes E) \oplus (\$_V \otimes E)$ is a $\mathbb{Z}/2$–graded Dirac bundle with Clifford multiplication

$$\mathrm{c} := \begin{pmatrix} 0 & \mathrm{c}_{\$} \otimes \mathrm{id}_E \\ \mathrm{c}_{\$} \otimes \mathrm{id}_E & 0 \end{pmatrix},$$

where $\mathrm{c}_{\$}$ is the Clifford multiplication on $\$_V$. Moreover, S turns into a relative Dirac bundle with involution

$$\sigma := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \tag{2-19}$$

globally defined on V (that is, the support is empty). The Dirac operator on S is given by

$$\mathcal{D} = \begin{pmatrix} 0 & \not{D}_E \\ \not{D}_E & 0 \end{pmatrix},$$

where $\not{D}_E \colon \mathrm{C}^\infty(V, \$_V \otimes E) \to \mathrm{C}^\infty(V, \$_V \otimes E)$ is the spinor Dirac operator on (V, g) twisted with the bundle E. As in the previous example, the curvature term from (2-5) is of the form

$$\mathcal{R} = \tfrac{1}{4}\,\mathrm{scal}_g + \mathcal{R}^E, \tag{2-20}$$

where

$$\mathcal{R}^E = \sum_{i<j} \mathrm{c}(e^i)\,\mathrm{c}(e^j)(\mathrm{id}_{\$_V \oplus \$_V} \otimes \mathrm{R}^{\nabla^E}_{e_i, e_j}).$$

3 Index theory for relative Dirac bundles

In this section, we introduce the Callias operators associated to relative Dirac bundles, and review the necessary analysis to develop an index theory for them.

We again start by briefly fixing the notation we are going to use. Let $E \to M$ be a Hermitian vector bundle over a smooth manifold M with *compact* boundary ∂M. The L^2–inner product of two sections $u, v \in \mathrm{C}^\infty_\mathrm{c}(M, E)$ is defined by

$$(u, v) := \int_M \langle u, v\rangle\, \mathrm{vol}_M.$$

The corresponding L^2–norm will be denoted by $\|u\| := (u, u)^{1/2}$. The space of *square-integrable sections*, denoted by $\mathrm{L}^2(M, E)$, can be identified with the completion of $\mathrm{C}^\infty_\mathrm{c}(M, E)$ with respect to this norm. We denote the space of *locally square-integrable sections* by $\mathrm{L}^2_\mathrm{loc}(M, E)$. Similarly, the latter comes endowed with the family of seminorms $\|-\|_K$ ranging over compact subsets $K \subseteq M$, where $\|u\|_K := \big(\int_K \langle u, u\rangle\, \mathrm{vol}_M\big)^{1/2}$.

We will also use Sobolev spaces to a limited extent; see for instance [40, Chapter 4] for a detailed introduction. The Sobolev space $\mathrm{H}^1_\mathrm{loc}(M, E)$ consists of all sections $u \in \mathrm{L}^2_\mathrm{loc}(M, E)$ such that ∇u, a priori defined as a distributional section over the interior of M, is also represented by a locally square integrable section. The space $\mathrm{H}^1_{\mathrm{loc}}(M, E)$ is topologized using the family of seminorms $\|-\|_{\mathrm{H}^1_K}$ defined by $\|u\|^2_{\mathrm{H}^1_K} := \|u\|^2_K + \|\nabla u\|^2_K$, where K ranges over all compact subsets of M. Since the boundary is assumed to be compact, the restriction $\mathrm{C}^\infty_{\mathrm{c}}(M, E) \to \mathrm{C}^\infty(\partial M, E|_{\partial M})$, $u \mapsto u|_{\partial M}$, extends to a continuous linear operator $\tau\colon \mathrm{H}^1_{\mathrm{loc}}(M, E) \to \mathrm{H}^{1/2}(\partial M, E|_{\partial M})$ by the trace theorem; see eg [40, Chapter 4, Proposition 4.5]. Here $\mathrm{H}^{1/2}(\partial M, E|_{\partial M})$ denotes a fractional Sobolev space for instance in the sense of [40, Chapter 4, Section 3]. An equivalent (and in our setup more relevant) description in the presence of a suitable Dirac operator on the boundary is given in [4, Section 4.1].

Next we turn to the main player, the Callias operator associated to a relative Dirac bundle and a suitable potential function. To this end, fix a complete Riemannian manifold M with compact boundary ∂M and $S \to M$ a relative Dirac bundle with support K. This means that $K \subset M^\circ$ is compact and the involution σ is defined on $M \setminus K$; see Definition 2.2.

Definition 3.1 A Lipschitz function $\psi\colon M \to \mathbb{R}$ is called an *admissible potential* if $\psi = 0$ on K and there exists a compact set $K \subseteq L \subseteq M$ such that ψ is equal to a nonzero constant on each component of $M \setminus L$.

Let ψ be an admissible potential. Then $\psi\sigma$ extends by zero to a continuous bundle map on all of M. The *Callias operator* associated to these data is the differential operator

$$\mathcal{B}_\psi := \mathcal{D} + \psi\sigma. \tag{3-1}$$

Since ψ is Lipschitz, the commutator $[\mathcal{D}, \psi]$ extends to a bounded operator on $\mathrm{L}^2(M, S)$; compare for example [3, Lemma 3.1]. In fact, by Rademacher's theorem [35], the differential $\mathrm{d}\psi$ exists almost everywhere and is an L^∞–section of the cotangent bundle. In this view, the commutator is given by the formula $[\mathcal{D}, \psi] = \mathrm{c}(\mathrm{d}\psi)$ as an element of $\mathrm{L}^\infty(M, \mathrm{End}(S))$. Using that the involution σ anticommutes with the Dirac operator, a direct computation then yields

$$\mathcal{B}_\psi^2 = \mathcal{D}^2 + \mathrm{c}(\mathrm{d}\psi)\sigma + \psi^2. \tag{3-2}$$

To be very precise, this expression implicitly uses the requirement in Definition 2.2 that σ extends to a smooth section on a neighborhood of $\overline{M \setminus K}$. However, as $\mathrm{d}\psi = 0$ on the interior of K, we only need the values of the continuous extension of σ to $\overline{M \setminus K}$ to specify $\mathrm{c}(\mathrm{d}\psi)\sigma$, and this is only relevant if the topological boundary of K has positive Lebesgue measure.

Observe that the distance function x from any fixed subset of M is 1–Lipschitz, so that $|\mathrm{d}x| \le 1$ almost everywhere. In our applications, we will consider potentials of the form $\psi = Y \circ x$, where Y is a smooth function on $\mathbb{R}$ and x is a distance function from a geometrically relevant compact region of M.

For a choice of signs $s\colon \partial M \to \{\pm 1\}$, we let $\mathcal{B}_{\psi,s}$ denote the operator $\mathcal{B}_\psi$ on the domain

$$\mathrm{C}^\infty_{\sigma,s}(M, S) := \{u \in \mathrm{C}^\infty_{\mathrm{c}}(M, S) \mid \chi(u|_{\partial M}) = u|_{\partial M}\},$$

where $\chi := s\,\mathrm{c}(\nu^\flat)\sigma$ as in (2-1).

By definition, $S = S^+ \oplus S^-$ is $\mathbb{Z}/2$–graded and the differential operator $\mathcal{B}_\psi$ can be written as

$$\mathcal{B}_\psi = \begin{pmatrix} 0 & \mathcal{B}_\psi^- \\ \mathcal{B}_\psi^+ & 0 \end{pmatrix},$$

where $\mathcal{B}_\psi^\pm$ are differential operators $\mathrm{C}^\infty(M, S^\pm) \to \mathrm{L}^2(M, S^\mp)$. Since χ is even with respect to the grading, $\mathcal{B}_{\psi,s}$ also decomposes similarly into operators

$$\mathcal{B}_{\psi,s}^\pm \colon \{u \in \mathrm{C}_\mathrm{c}^\infty(M, S^\pm) \mid \chi(u|_{\partial M}) = u|_{\partial M}\} \to \mathrm{L}^2(M, S^\mp).$$

Note that — even ignoring regularity — $\mathcal{B}_{\psi,s}$ does *not* preserve its domain.

In the following, we will discuss the necessary analysis and index theory for these operators. We will mostly rely on the general framework of elliptic boundary value problems for Dirac-type operators due to Bär and Ballmann [4; 3]. However, as we will need to allow potentials which are a priori only Lipschitz for some of our applications, we will take a slight detour and first apply the results of Bär and Ballmann [4] only to the Dirac operator $\mathcal{D}$ and Callias operators $\mathcal{B}_\psi$ with smooth potentials, before deducing the desired statements for the general case.

We first observe that the canonical boundary Dirac operator $\mathcal{A}$ from (2-9) is an adapted operator for $\mathcal{D}$ in the sense of [4, Section 3.2]. Since χ anticommutes with $\mathcal{A}$ (see (2-11)), χ is a *boundary chirality* in the sense of [4, Example 4.20]. Thus $B_\chi := \mathrm{H}^{1/2}(\partial M, S^\chi_{\partial M})$, where $S^\chi_{\partial M}$ is the $+1$–eigenbundle of χ, is an elliptic boundary condition in the sense of Bär and Ballmann [4]. Moreover, it is a consequence of Lemma 2.4 that the adjoint boundary condition of B_χ (see [4, Theorem 4.6] and [3, Section 7.2]) is B_χ itself (the crucial ingredient here is that χ anticommutes with $\mathrm{c}(\nu^\flat)$). In other words, B_χ is a selfadjoint boundary condition, and thus $\mathcal{D}$ is essentially selfadjoint on the domain $\mathrm{C}^\infty_{\sigma,s}(M, S)$ by [4, Theorem 4.11]. Moreover, [3, Lemma 7.3] implies that the domain of the closure is given by the Sobolev space

$$\mathrm{H}^1_{\sigma,s}(M, S) := \{u \in \mathrm{H}^1_{\mathrm{loc}}(M, S) \cap \mathrm{L}^2(M, S) \mid \mathcal{D}u \in \mathrm{L}^2(M, S),\ \sigma\tau(u) = s\tau(u)\}.$$

We will endow $\mathrm{H}^1_{\sigma,s}(M, S)$ with the norm defined by

$$\|u\|^2_{\mathrm{H}^1_{\sigma,s}(M,S)} := \|u\|^2 + \|\mathcal{D}u\|^2. \tag{3-3}$$

It also follows from ellipticity (also of the boundary condition) that the inclusion $\mathrm{H}^1_{\sigma,s}(M, S) \hookrightarrow \mathrm{H}^1_{\mathrm{loc}}(M, S)$ is continuous. In particular, the trace operator is continuous on $\mathrm{H}^1_{\sigma,s}(M, S)$. The next two lemmas allow us to also describe the domain of the closure of $\mathcal{B}_{\psi,s}$, temporarily denoted by $\overline{\mathcal{B}}_{\psi,s}$, and show that compact perturbations of the potential do not alter any relevant properties.

Lemma 3.2 *Let ψ be an admissible potential. Then $\overline{\mathcal{B}}_{\psi,s}$ is selfadjoint and* $\mathrm{dom}(\overline{\mathcal{B}}_{\psi,s}) = \mathrm{H}^1_{\sigma,s}(M, S)$. *Moreover, the graph norm induced by $\overline{\mathcal{B}}_{\psi,s}$ is equivalent to the norm given by (3-3).*

Proof This is a direct consequence of the fact that the difference $\mathcal{B}_\psi - \mathcal{D} = \psi\sigma$ is in $\mathrm{L}^\infty(M, \mathrm{End}(S))$ and (fiberwise) selfadjoint, and thus extends to a selfadjoint bounded operator on $\mathrm{L}^2(M, S)$. □

From now on, we will drop the notational distinction between $\mathcal{B}_{\psi,s}$ and its closure $\overline{\mathcal{B}}_{\psi,s}$, and simply view $\mathcal{B}_{\psi,s}$ as being defined on $\mathrm{H}^1_{\sigma,s}(M,S)$, unless specified otherwise.

Lemma 3.3 *Let ψ_1 and ψ_2 be admissible potentials coinciding outside a compact set L. Then $\mathcal{B}_{\psi_2}-\mathcal{B}_{\psi_1}$ defines a compact operator $\mathrm{H}^1_{\sigma,s}(M,S)\to\mathrm{L}^2(M,S)$. In particular, viewed as maps $\mathrm{H}^1_{\sigma,s}(M,S)\to\mathrm{L}^2(M,S)$, the operator $\mathcal{B}_{\psi_1,s}$ is Fredholm if and only if $\mathcal{B}_{\psi_2,s}$ is.*

Proof Observe that, since ψ_1 and ψ_2 are Lipschitz and coincide in $M\setminus L$, both $\eta:=\psi_2-\psi_1$ and its differential $\mathrm{d}\eta$ are essentially bounded and supported in the compact set L. Together with the estimate

$$\|\mathcal{D}(\mathcal{B}_{\psi_2}-\mathcal{B}_{\psi_1})u\|=\|\mathcal{D}\eta\sigma u\|\le\|\eta\sigma\mathcal{D}u\|+\|\mathrm{c}(\mathrm{d}\eta)u\|\le\|\eta\|_\infty\|\mathcal{D}u\|+\|\mathrm{d}\eta\|_\infty\|u\|,$$

this implies that $\mathcal{B}_{\psi_2}-\mathcal{B}_{\psi_1}=\eta\sigma$ defines a bounded operator $\mathrm{H}^1_{\sigma,s}(M,S)\to\mathrm{H}^1_{\sigma,s}(M,S)$. Moreover, since η is supported in the compact subset L and using continuity of the inclusion $\mathrm{H}^1_{\sigma,s}(M,S)\hookrightarrow\mathrm{H}^1_{\mathrm{loc}}(M,S)$, this actually means that $\mathcal{B}_{\psi_2}-\mathcal{B}_{\psi_1}$ is a bounded operator $\mathrm{H}^1_{\sigma,s}(M,S)\to\mathrm{H}^1(L,S)$. Finally, since the inclusion $\mathrm{H}^1(L,S)\hookrightarrow\mathrm{L}^2(M,S)$ is compact by the Rellich lemma, $\mathcal{B}_{\psi_2}-\mathcal{B}_{\psi_1}$ yields a compact operator $\mathrm{H}^1_{\sigma,s}(M,S)\to\mathrm{L}^2(M,S)$. The second claim follows from classical properties of Fredholm operators. □

Theorem 3.4 *Let M be a complete Riemannian manifold with compact boundary ∂M and $S\to M$ be a relative Dirac bundle. Let $\psi\colon M\to\mathbb{R}$ be an admissible potential and $s\colon\partial M\to\{\pm1\}$ a choice of signs. Then the operator $\mathcal{B}_{\psi,s}$ is selfadjoint and Fredholm.*

Proof Since for any given admissible potential we can always find a smooth admissible potential which agrees with the original one outside a compact subset, Lemmas 3.2 and 3.3 imply that we can assume without loss of generality that ψ is smooth. As $\mathcal{B}_\psi$ has the same principal symbol as $\mathcal{D}$, the canonical boundary Dirac operator $\mathcal{A}$ from (2-9) is also an adapted operator for $\mathcal{B}_\psi$ in the sense of [4, Section 3.2]. Thus the discussion in the paragraphs preceding Lemma 3.2 applies verbatim with $\mathcal{B}_\psi$ replaced by $\mathcal{D}$. Furthermore, since

$$\mathcal{B}_\psi^2=\mathcal{D}^2+\mathrm{c}(\mathrm{d}\psi)\sigma+\psi^2$$

and $\psi^2-|\mathrm{d}\psi|$ is uniformly positive outside a compact subset because ψ is admissible, the operator $\mathcal{B}_\psi$ is coercive at infinity in the sense of [4, Definition 5.1]. Thus $\mathcal{B}_{\psi,s}$ is a Fredholm operator by [4, Theorem 5.3]. □

In particular, we obtain an index

$$\mathrm{ind}(\mathcal{B}_{\psi,s}):=\mathrm{ind}(\mathcal{B}^+_{\psi,s}):=\dim\ker(\mathcal{B}^+_{\psi,s})-\dim\ker(\mathcal{B}^-_{\psi,s})\in\mathbb{Z}.$$

As another immediate consequence of Lemma 3.3, we obtain the following statement.

Lemma 3.5 *Let M be a complete Riemannian manifold with compact boundary ∂M and $S\to M$ be a relative Dirac bundle. Let $\psi_1,\psi_2\colon M\to\mathbb{R}$ be two admissible potentials and $s\colon\partial M\to\{\pm1\}$ a choice of signs. Suppose that ψ_1 and ψ_2 agree outside a compact subset of M. Then*

$$\mathrm{ind}(\mathcal{B}_{\psi_1,s})=\mathrm{ind}(\mathcal{B}_{\psi_2,s}).$$

The next theorem provides the main tool for computing $\operatorname{ind}(\mathcal{B}_{\psi,s})$.

Theorem 3.6 *Let M be a complete Riemannian manifold with $\partial M = \varnothing$ and $S \to M$ be a relative Dirac bundle of support $K \subseteq M$. Let $\psi\colon M \to \mathbb{R}$ be an admissible potential. Let $N \subset M \setminus K$ be a compact hypersurface with trivial normal bundle. Then we let M' be the manifold obtained from cutting M open along N so that $\partial M' = N_0 \sqcup N_1$, where N_0 and N_1 are two disjoint copies of N. Pulling back all data via the quotient map $M' \to M$ induces a relative Dirac bundle $S' \to M'$ and an admissible potential $\psi'\colon M' \to \mathbb{R}$, respectively. Let $s\colon \partial M' \to \{\pm 1\}$ be any choice of signs such that $s|_{N_0} = s|_{N_1}\colon N \to \{\pm 1\}$. Then*

$$\operatorname{ind}(\mathcal{B}'_{\psi',s}) = \operatorname{ind}(\mathcal{B}_{\psi}), \tag{3-4}$$

where $\mathcal{B}_{\psi} = \mathcal{D} + \psi\sigma$ denotes the Callias operator on M, and $\mathcal{B}'_{\psi'}$ the corresponding operator on M'.

Proof Using Lemmas 3.3 and 3.5, it is again enough to prove the claim in the case when ψ is smooth. We will prove this case as a consequence of the general splitting theorem due to Bär and Ballmann; see [4, Theorem 6.5; 3, Theorem 8.17].

First we note that, by the proof of Theorem 3.4 and the remarks preceding it, the boundary condition $B^+_\chi \subseteq \mathrm{H}^{1/2}(\partial M', S'^+)$ defined by the chirality $\chi = s\,\mathrm{c}(\nu^\flat)\sigma$ restricted to S^+ is elliptic and its adjoint is the corresponding boundary condition $B^-_\chi \subseteq \mathrm{H}^{1/2}(\partial M', S'^-)$. Thus we can also apply the theory of Bär and Ballmann [4] separately to $\mathcal{B}^{\pm}_{\psi}$ and $\mathcal{B}'^{\pm}_{\psi',s}$. In this light, the theorem is a direct consequence of the general splitting theorem; see [4, Theorem 6.5; 3, Theorem 8.17]. To see this, observe that $B^+_\chi = B^+_0 \oplus B^+_1$ with respect to the decomposition $\mathrm{H}^{1/2}(\partial M', S'^+) = \mathrm{H}^{1/2}(N, S^+) \oplus \mathrm{H}^{1/2}(N, S^+)$ coming from $\partial M = N_0 \sqcup N_1$, where

$$B^+_i = \{u \in \mathrm{H}^{1/2}(N, S^+) \mid s\,\mathrm{c}(\nu^\flat|_{N_i})\sigma u = u\} \quad \text{for } i = 0, 1.$$

By construction, the interior normal field of $\partial M'$ along N_0 is equal to the exterior normal field along N_1 and hence $\nu|_{N_0} = -\nu|_{N_1}$. Thus $(B^+_0)^\perp = B^+_1$ viewed as an L^2–orthogonal complement in $\mathrm{H}^{1/2}(N, S^+)$. Consequently, the hypotheses of [4, Theorem 6.5] are satisfied and we obtain $\operatorname{ind}(\mathcal{B}^+_{\psi',s}) = \operatorname{ind}(\mathcal{B}^+_{\psi})$. This concludes the proof because, by definition, $\operatorname{ind}(\mathcal{B}'_{\psi',s}) = \operatorname{ind}(\mathcal{B}'^+_{\psi',s})$ and $\operatorname{ind}(\mathcal{B}_{\psi}) = \operatorname{ind}(\mathcal{B}^+_{\psi})$. ◻

Lemma 3.7 *Let M be a complete Riemannian manifold with compact boundary ∂M and $S \to M$ be a relative Dirac bundle. Let $\psi\colon M \to \mathbb{R}$ be an admissible potential and $s\colon \partial M \to \{\pm 1\}$ a choice of signs such that*

(i) *there exists $C > 0$ such that $\psi^2 - |\mathrm{d}\psi| \geq C$ on all of M, and*

(ii) *$s\psi \geq 0$ along ∂M.*

Then $\mathcal{B}_{\psi,s}$ is invertible. In particular, $\operatorname{ind}(\mathcal{B}_{\psi,s}) = 0$.

Proof By Green's formula, for any $u \in \mathrm{C}^\infty_{\sigma,s}(M,S)$, we have

$$\begin{aligned}\int_M |\mathcal{B}_\psi u|^2 \operatorname{vol}_M &= \int_M \bigl(|\mathcal{D}u|^2 + \langle u, \mathrm{c}(\mathrm{d}\psi)\sigma u + \psi^2 u\rangle\bigr) \operatorname{vol}_M + \int_{\partial M} \langle u, \psi \underbrace{\mathrm{c}(\nu^\flat)\sigma u}_{=su}\rangle \operatorname{vol}_{\partial M} \\ &\geq \int_M (\psi^2 - |\mathrm{d}\psi|)|u|^2 \operatorname{vol}_M + \int_{\partial M} \underbrace{s\psi}_{\geq 0} |u|^2 \operatorname{vol}_{\partial M} \geq C \int_M |u|^2 \operatorname{vol}_M.\end{aligned}$$

By continuity, the final estimate holds for all $u \in \mathrm{H}^1_{\sigma,s}(M,S)$. Therefore, $\mathcal{B}_{\psi,s}$ is invertible; compare [4, Corollary 5.9]. □

Lemma 3.8 *Let M be a compact Riemannian manifold with boundary ∂M and $S \to M$ be a relative Dirac bundle with empty support* (*that is*, *the involution σ is defined on all of M*). *Let $s\colon \partial M \to \{\pm 1\}$ be a choice of signs. Then any Lipschitz function $\psi\colon M \to \mathbb{R}$ is an admissible potential and* $\operatorname{ind}(\mathcal{B}_{\psi,s})$ *does not depend on the choice of the potential ψ. Furthermore*, *if the sign $s \in \{\pm 1\}$ is constant on all of ∂M*, *then* $\operatorname{ind}(\mathcal{B}_{\psi,s}) = 0$ *for any potential ψ.*

Proof Since M is compact and σ is defined on of all M, the condition of being an *admissible* potential is vacuous. Moreover, by Lemma 3.5, any two potentials yield the same index. Finally, if the choice of signs s is constant, then Lemma 3.7 implies vanishing of the index for the constant potential $\psi = s$. Since the index does not depend on the potential, it must vanish for any choice of potential. □

Let us now specialize to the case of Example 2.5. Let M be a compact even-dimensional Riemannian manifold with boundary. Let S be the Dirac bundle associated to a GL pair (E,F) over M. For an admissible potential ψ, consider the Callias operator $\mathcal{B}_\psi$. With the choice of sign $s = 1$, let us consider the index of $\mathcal{B}_{\psi,1}$. In order to compute $\operatorname{ind}(\mathcal{B}_{\psi,1})$, we make use of the following construction. Form the double $\mathtt{d}M := M \cup_{\partial M} M^-$ of M, where M^- denotes the manifold M with opposite orientation. Observe that $\mathtt{d}M$ is a closed manifold equipped with a spin structure induced by the spin structure of M. Using condition (2-15), let $V(E,F) \to \mathtt{d}M$ be a bundle on $\mathtt{d}M$ which outside a small collar neighborhood coincides with E over M and with F over M^- defined using the bundle isomorphism implicit in a GL pair. The *relative index* of (E,F) is the index of the spin Dirac operator on $\mathtt{d}M$ twisted with the bundle $V(E,F)$, that is,

$$\operatorname{indrel}(M;E,F) := \operatorname{ind}(\not{D}_{\mathtt{d}M,V(E,F)}) \in \mathbb{Z}.$$

The computation of $\operatorname{ind}(\mathcal{B}_{\psi,1})$ is given by the next proposition.

Corollary 3.9 *Consider the setup of Example 2.5. Then for any choice of potential ψ, we have*

$$\operatorname{ind}(\mathcal{B}_{\psi,1}) = \operatorname{indrel}(M;E,F),$$

where the latter expression is described in the paragraph preceding this corollary.

Proof Let $E' = V(E,F)$. Moreover, extend the bundle F to a bundle with metric connection F' on $\mathtt{d}M$ such that $F'|_{M^-} = F$. Consider the $\mathbb{Z}/2$–graded Dirac bundle $W' := \$_{\mathtt{d}M} \mathbin{\widehat{\otimes}} (E' \oplus (F')^{\mathrm{op}})$ with associated Dirac operator $\mathcal{D}'$. Observe that

$$\operatorname{ind}\mathcal{D}' = \operatorname{indrel}(M; E, F) \tag{3-5}$$

because the index of the Dirac operator on $\$_{\mathrm{d}M} \otimes F'$ vanishes. Observe also that $\mathcal{D}' = \mathcal{B}_0^{E',F'}$. Cut $\mathrm{d}M$ open along ∂M as in Theorem 3.6. Pulling back all data, we obtain the operators $\mathcal{B}_{0,1}^{E,F}$ on M and $\mathcal{B}_{0,1}^{F,F}$ on M^-. By Lemma 3.8, $\operatorname{ind}\mathcal{B}_{0,1}^{F,F} = 0$. By Lemma 3.5, $\operatorname{ind}(\mathcal{B}_{\psi,1}^{E,F}) = \operatorname{ind}(\mathcal{B}_{0,1}^{E,F})$. Therefore, the thesis follows using identity (3-5) and Theorem 3.6. □

We now deal with Example 2.6 in a similar fashion.

Corollary 3.10 *Consider the setup of Example 2.6 and choose the signs so that $s|_{\partial_\pm V} = \pm 1$. Then for any choice of potential ψ, we have*

$$\operatorname{ind}(\mathcal{B}_{\psi,s}) = \operatorname{ind}(\not{D}_{\partial_- V, E|_{\partial_- V}}) = \int_{\partial_- V} \widehat{\mathbf{A}}(\partial_- V) \wedge \operatorname{ch}(E)|_{\partial_- V},$$

where $\not{D}_{\partial_- V, E|_{\partial_- V}}$ denotes the corresponding twisted spinor Dirac operator on $\partial_- V$, $\widehat{\mathbf{A}}(\partial_- V)$ is the $\widehat{\mathrm{A}}$–form of $\partial_- V$ and $\operatorname{ch}(E)$ the Chern character form associated to E.

Proof First of all, the index does not depend on ψ by Lemma 3.8 since V is compact. We can thus choose a function ψ suitable for our purposes. Furthermore, let

$$V_\infty = V_- \cup_{\partial_- V} V \cup_{\partial_+ V} V_+,$$

where

$$V_- := \partial_- V \times (-\infty, -1] \quad \text{and} \quad V_+ := \partial_+ V \times [1, \infty),$$

be the infinite band obtained from attaching infinite cylinders along the boundary parts $\partial_\pm V$. We extend the Riemannian metric on V to a complete metric on V_∞. Then the same construction as in Example 2.6 yields a relative Dirac bundle on V_∞ extending the data on V. Now choose a smooth function $\psi_\infty \colon V_\infty \to [-1, 1]$ such that $\psi_\infty(V_\pm) = \pm 1$. We will denote $\psi := \psi_\infty|_V$. Applying Theorem 3.6 to the splitting of V_∞ along $\partial V = \partial_- V \sqcup \partial_+ V$ implies that

$$\operatorname{ind}(\mathcal{B}_{V_\infty,\psi_\infty}) = \operatorname{ind}(\mathcal{B}_{V,\psi,s}) + \operatorname{ind}(\mathcal{B}_{V_-,-1,-1}) + \operatorname{ind}(\mathcal{B}_{V_+,+1,+1}).$$

However, Lemma 3.7 implies that $\operatorname{ind}(\mathcal{B}_{V_\pm,\pm1,\pm1}) = 0$, and so $\operatorname{ind}(\mathcal{B}_{V_\infty,\psi_\infty}) = \operatorname{ind}(\mathcal{B}_{V,\psi,s})$. Finally, it follows from [2, Corollary 1.9] that $\operatorname{ind}(\mathcal{B}_{V_\infty,\psi_\infty}) = \operatorname{ind}(\not{D}_{\partial_- V, E|_{\partial_- V}})$; for a more general context compare also the partitioned manifold index theorem from [43, Appendix A]. The last equality follows from the Atiyah–Singer index theorem [28, Theorem 13.10]. □

4 Spectral estimates

Our goal here is to establish spectral estimates for the Callias operator $\mathcal{B}_\psi$ from equation (3-1) associated to a relative Dirac bundle $S \to M$ and an admissible potential $\psi \colon M \to \mathbb{R}$. The parts only concerning the Dirac operator are similar to estimates of imaginary eigenvalues in the context of the "MIT bag boundary conditions" due to Raulot [36].

We start with a lemma computing the L^2–norm of the Dirac operator applied to a section in terms of the Bochner–Lichnerowicz–Weitzenböck curvature term and a boundary term. In the following, we make extensive use of the notation and formulas introduced in Section 2.

Lemma 4.1 *Let $S \to M$ be a relative Dirac bundle and $s\colon \partial M \to \{\pm 1\}$ a choice of signs. Then for every $u \in \mathrm{C}^\infty_\mathrm{c}(M, S)$, the following identity holds*:

$$\int_M |\mathcal{D}u|^2 \operatorname{vol}_M = \frac{n}{n-1}\int_M (|\mathcal{P}u|^2 + \langle u, \mathcal{R}u\rangle) \operatorname{vol}_M + \int_{\partial M} \Big\langle u, \Big(\frac{n}{2}\,\mathrm{H}_g - \frac{n}{n-1}\mathcal{A}\Big)u\Big\rangle \operatorname{vol}_{\partial M}.$$

Proof Using the Green and Weitzenböck formulas, we obtain

$$\begin{aligned}\int_M |\mathcal{D}u|^2 \operatorname{vol}_M &= \int_M \langle u, \mathcal{D}^2 u\rangle \operatorname{vol}_M + \int_{\partial M} \langle u, \nu^\flat \cdot \mathcal{D}u\rangle \operatorname{vol}_{\partial M} \\ &= \int_M |\nabla u|^2 \operatorname{vol}_M + \int_M \langle u, \mathcal{R}u\rangle \operatorname{vol}_M + \int_{\partial M} \langle u, (\underbrace{\mathrm{c}(\nu^\flat)\mathcal{D} + \nabla_\nu}_{=\frac{1}{2}(n-1)\,\mathrm{H}_g - \mathcal{A}})u\rangle \operatorname{vol}_{\partial M}.\end{aligned}$$

The identity now follows from equations (2-13) and (2-6). □

We now combine this with another application of Green's formula to get a corresponding computation for the Callias operator.

Proposition 4.2 *Let $S \to M$ be a relative Dirac bundle and $\psi\colon M \to \mathbb{R}$ be an admissible potential. Then for every $u \in \mathrm{C}^\infty_\mathrm{c}(M, S)$, the following identity holds.*

$$\begin{aligned}\text{(4-1)}\quad \int_M |\mathcal{B}_\psi u|^2 = \frac{n}{n-1}\int_M (|\mathcal{P}u|^2 + \langle u, \mathcal{R}u\rangle) \operatorname{vol}_M + \int_M \langle u, (\psi^2 + \mathrm{c}(\mathrm{d}\psi)\sigma)u\rangle \operatorname{vol}_M \\ + \int_{\partial M} \Big\langle u, \Big(\frac{n}{2}\,\mathrm{H}_g - \frac{n}{n-1}\mathcal{A} + \psi\,\mathrm{c}(\nu^\flat)\sigma\Big)u\Big\rangle \operatorname{vol}_{\partial M}.\end{aligned}$$

Proof We have

$$|\mathcal{B}_\psi u|^2 = |\mathcal{D}u|^2 + \langle \mathcal{D}u, \psi\sigma u\rangle + \langle \psi\sigma u, \mathcal{D}u\rangle + \psi^2|u|^2.$$

Using Green's formula (2-3) on the second term implies

$$\begin{aligned}\int_M |\mathcal{B}_\psi u|^2 \operatorname{vol}_M = \int_M |\mathcal{D}u|^2 \operatorname{vol}_M + \int_M \big(\langle u, \underbrace{(\mathcal{D}\psi\sigma + \psi\sigma\mathcal{D})}_{=\mathrm{c}(\mathrm{d}\psi)\sigma}\, u\rangle + \psi^2|u|^2\big) \operatorname{vol}_M \\ + \int_{\partial M} \langle u, \nu^\flat \cdot \psi\sigma u\rangle \operatorname{vol}_{\partial M}.\end{aligned}$$

Combining this with Lemma 4.1 yields the desired identity (4-1). □

Finally, this leads to the main result of this section.

Theorem 4.3 *Let $S \to M$ be a relative Dirac bundle over a compact manifold M, let $\psi\colon M \to \mathbb{R}$ be an admissible potential and $s\colon \partial M \to \{\pm 1\}$ be a choice of signs. Then for every $u \in \mathrm{H}^1_{\sigma,s}(M, S)$, the* *following estimate holds*:

$$\begin{aligned}(4\text{-}2)\quad \int_M |\mathcal{B}_\psi u|^2 \operatorname{vol}_M &\geq \frac{n}{n-1}\int_M \langle u, \mathcal{R}u\rangle \operatorname{vol}_M + \int_M \langle u, (\psi^2 + \mathrm{c}(\mathrm{d}\psi)\sigma)u\rangle \operatorname{vol}_M \\ &\qquad + \int_{\partial M} \bigl(\tfrac{1}{2}n \mathrm{H}_g + s\psi\bigr)|\tau(u)|^2 \operatorname{vol}_{\partial M} \\ &\geq \frac{n}{n-1}\int_M \langle u, \mathcal{R}u\rangle \operatorname{vol}_M + \int_M (\psi^2 - |\mathrm{d}\psi|)|u|^2 \operatorname{vol}_M \\ &\qquad + \int_{\partial M} \bigl(\tfrac{1}{2}n \mathrm{H}_g + s\psi\bigr)|\tau(u)|^2 \operatorname{vol}_{\partial M}.\end{aligned}$$

Moreover, *equality throughout both estimates in* (4-2) *holds if and only if*

$$(4\text{-}3)\qquad \begin{aligned}\mathcal{P}_\xi u = \nabla_\xi u + \frac{1}{n}\mathrm{c}(\xi^\flat)\mathcal{D}u = 0 \quad &\textit{for all } \xi \in \mathrm{C}^\infty(M, \mathrm{T}M),\\ \mathrm{c}(\mathrm{d}\psi)\sigma u = -|\mathrm{d}\psi| u.&\end{aligned}$$

Proof We first assume that $u \in \mathrm{C}^\infty_{\sigma,s}(M, S)$. Then

$$\mathrm{c}(\nu^\flat)\sigma u|_{\partial M} = s\chi u|_{\partial M} = s u|_{\partial M} \quad \text{and} \quad \langle u|_{\partial M}, \mathcal{A}u|_{\partial M}\rangle = 0;$$

see equations (2-1) and (2-12). Thus (4-1) simplifies to the equality

$$\begin{aligned}(4\text{-}4)\quad \int_M |\mathcal{B}_\psi u|^2 &= \frac{n}{n-1}\int_M (|\mathcal{P}u|^2 + \langle u, \mathcal{R}u\rangle) \operatorname{vol}_M + \int_M \langle u, (\psi^2 + \mathrm{c}(\mathrm{d}\psi)\sigma)u\rangle \operatorname{vol}_M \\ &\qquad + \int_{\partial M} \bigl\langle \tau(u), \bigl(\tfrac{1}{2}n \mathrm{H}_g + s\psi\bigr)\tau(u)\bigr\rangle \operatorname{vol}_{\partial M}.\end{aligned}$$

Now we observe that both sides of the identity (4-4) are continuous in u with respect to the topology of $\mathrm{H}^1_{\sigma,s}(M, S)$. Thus (4-4) still holds for all $u \in \mathrm{H}^1_{\sigma,s}(M, S)$.

The first estimate of (4-2) now follows directly because $|\mathcal{P}u|^2 \geq 0$. The second estimate follows from $\mathrm{c}(\mathrm{d}\psi)\sigma \geq -|\mathrm{d}\psi|$. Moreover, equality in the first estimate is equivalent to $\int_M |\mathcal{P}u|^2 = 0$ and thus $\mathcal{P}u = 0$. Equality for the second estimate is equivalent to

$$\int_M \langle u, (\mathrm{c}(\mathrm{d}\psi)\sigma + |\mathrm{d}\psi|)u\rangle \operatorname{vol}_M = 0.$$

Since the selfadjoint bundle endomorphism $\mathrm{c}(\mathrm{d}\psi)\sigma + |\mathrm{d}\psi|$ is fiberwise nonnegative, this is equivalent to $(\mathrm{c}(\mathrm{d}\psi)\sigma + |\mathrm{d}\psi|)u = 0$, as claimed. □

Remark 4.4 Suppose that u lies in the kernel of $\mathcal{B}_\psi$ and at the same time satisfies (4-3). Then $\mathcal{D}u = -\psi\sigma u$ and so

$$(4\text{-}5)\qquad \nabla_\xi u = \frac{\psi}{n}\mathrm{c}(\xi^\flat)\sigma u \quad \text{for all } \xi \in \mathrm{T}M.$$

On each point where $\mathrm{d}\psi \neq 0$, this implies

$$\nabla_\xi u = \frac{\psi}{n}\mathrm{c}(\xi^\flat)\,\mathrm{c}\left(\frac{\mathrm{d}\psi}{|\mathrm{d}\psi|}\right)u,$$

and in particular,

$$\nabla_{\nabla\psi/|\nabla\psi|} u = -\frac{\psi}{n}u.$$

In the final remark of this section, we prepare another technical observation in the context of the extremality case of Theorem 4.3, which we will use on multiple occasions. Here we assume that the potential has a special form relevant for our applications.

Remark 4.5 Suppose that there is a 1–Lipschitz function $x\colon M \to I$ for some compact interval $I \subseteq \mathbb{R}$ and let $\varphi\colon I \to (0,\infty)$ be a smooth strictly logarithmically concave function such that

$$\psi := f(x) := -\tfrac{1}{2}n\frac{\varphi'(x)}{\varphi(x)}$$

is an admissible potential for the relative Dirac bundle $S \to M$ we consider — the geometric meaning of this relationship will become apparent later; compare Section 8 below. Now suppose that we are in the same situation as in Remark 4.4, that is, we have an element $u \in \ker(\mathcal{B}_{\psi,s}) \subseteq \mathrm{H}^1_{\sigma,s}(M,S)$ that realizes equality in Theorem 4.3. Then the section $w := \varphi(x)^{-1/2}u$ still lies in $\mathrm{H}^1_{\sigma,s}(M,S)$ because $\varphi(x)^{-1/2}$ is a Lipschitz function. Using the elementary computation

$$\mathrm{d}\big(\varphi(x)^{-1/2}\big)(\xi) = \varphi(x)^{-1/2}\frac{\psi}{n}\,\mathrm{d}x(\xi),$$

we deduce from (4-5) that the (weak) covariant derivative of w satisfies, almost everywhere,

$$\nabla_\xi w = \frac{\psi}{n}\big(\mathrm{d}x(\xi) + \mathrm{c}(\xi^\flat)\sigma\big)w \tag{4-6}$$

for every smooth vector field ξ on M. In particular, $\nabla w = 0$ on K. Moreover, on the set

$$\{\mathrm{c}(\mathrm{d}x)u = \sigma u\} := \{p \in M \setminus K \mid \mathrm{c}(\mathrm{d}_p x)u_p = \sigma u_p\}$$

we deduce that

$$\nabla_\xi w = \frac{\psi}{n}(\mathrm{d}x(\xi) + \mathrm{c}(\xi^\flat)\sigma)w = \frac{\psi}{n}(\mathrm{d}x(\xi) + \mathrm{c}(\xi^\flat)\,\mathrm{c}(\mathrm{d}x))w = \frac{\psi}{n}\,\mathrm{c}(\xi^\flat \wedge \mathrm{d}x)w,$$

where in the last step we use the Clifford relation $\mathrm{c}(\mathrm{d}x)\,\mathrm{c}(\xi^\flat) + \mathrm{c}(\xi^\flat)\,\mathrm{c}(\mathrm{d}x) = -2\,\mathrm{d}x(\xi)$ and the notation $\mathrm{c}(\xi^\flat \wedge \mathrm{d}x) := \frac{1}{2}(\mathrm{c}(\xi^\flat)\,\mathrm{c}(\mathrm{d}x) - \mathrm{c}(\mathrm{d}x)\,\mathrm{c}(\xi^\flat))$.

In summary,

$$\nabla_\xi w = \frac{\psi}{n}\,\mathrm{c}(\xi^\flat \wedge \mathrm{d}x)w \quad \text{almost everywhere on } \{\mathrm{c}(\mathrm{d}x)u = \sigma u\}. \tag{4-7}$$

Finally, since $\mathrm{c}(\xi^\flat \wedge \mathrm{d}x)$ is an antiselfadjoint bundle endomorphism, (4-7) further implies that $\mathrm{d}(|w|^2)(\xi) = 2\langle w, \nabla_\xi w\rangle = 0$ almost everywhere on $\{\mathrm{c}(\mathrm{d}x)u = \sigma u\}$. Together with (4-6) this implies that

$$\mathrm{d}|w|^2 = 0 \quad \text{almost everywhere on } K \cup \{\mathrm{c}(\mathrm{d}x)u = \sigma u\}. \tag{4-8}$$

A priori, this only holds in the weak sense, but it still implies that $|w|^2$ is a constant function if M is connected and we have $K \cup \{\mathrm{c}(\mathrm{d}x)u = \sigma u\} = M$.

We also note that since $\mathrm{d}\psi = f'(x)\,\mathrm{d}x$ and $f'(x) > 0$, the second part of (4-3) implies that $\{\mathrm{c}(\mathrm{d}x)u = \sigma u\}$ contains the set of points in $M \setminus K$ where $|\mathrm{d}x| = 1$. But this is not a priori satisfied everywhere, just under the hypotheses of this remark, and so we will verify the condition $\mathrm{c}(\mathrm{d}x)u = \sigma u$ directly in the specific applications when needed.

5 A long neck principle with mean curvature

In this section, we establish our long neck principle for Riemannian spin manifolds with boundary, relating the length of the neck to the scalar curvature in the interior and the mean curvature of the boundary. Based on the technical preparations from the previous sections, we now directly enter into the proof of Theorem 1.4, starting with the following lemma.

Lemma 5.1 *Suppose that M is a compact spin manifold with boundary and $\Phi\colon M \to \mathrm{S}^n$ is a smooth map that is locally constant near ∂M and of nonzero degree, where $n = \dim M \geq 2$ is even. Set $l = \mathrm{dist}_g(\mathrm{supp}(\mathrm{d}\Phi), \partial M) > 0$. Then there exists a GL pair (E, F) in the sense of Example 2.5 such that*

- (i) $\mathcal{R}^{E\oplus F}_p \geq -a(p)\cdot \frac{1}{4}n(n-1)$ *at each point $p \in M$, where $a(p)$ is the area-contraction constant of Φ at p,*
- (ii) *(E, F) has support $K := \{p \in M \mid \mathrm{dist}_g(p, \partial M) \geq l\} \supseteq \mathrm{supp}(\mathrm{d}\Phi)$,*
- (iii) $\mathrm{indrel}(M; E, F) \neq 0$.

Proof We fix a basepoint $* \in \mathrm{S}^n$. Since Φ is locally constant on $M \setminus K$ and ∂M has only finitely many components, there exist finitely many distinct points $q_1, \dots, q_k \in \mathrm{S}^n$ such that $\Phi(M \setminus K) = \{q_1, \dots, q_k\}$. Let $\Omega_i = \Phi^{-1}(q_i) \cap M \setminus K$. Then each Ω_i is an open subset and $M \setminus K = \bigsqcup_{i=1}^k \Omega_i$. By continuity, $\Phi(\overline{\Omega}_i) = \{q_i\}$ and hence the closures $\overline{\Omega}_i$, $i = 1, \dots, k$, form a family of pairwise disjoint closed subsets of M. Thus there exist open neighborhoods $V_i \supset \overline{\Omega}_i$ that are still pairwise disjoint together with smooth functions $\nu_i\colon M \to [0,1]$ such that $\nu_i = 1$ on $\overline{\Omega}_i$ and $\nu_i = 0$ on $M \setminus V_i$. Next we choose geodesic curves $\gamma_i\colon [0,1] \to \mathrm{S}^n$ such that $\gamma_i(0) = *$ is the basepoint and $\gamma_i(1) = q_i$. From this we obtain a smooth map (see Figure 2)

$$\Psi\colon M \to \mathrm{S}^n, \qquad \Psi(p) = \begin{cases} \gamma_i(\nu_i(p)) & \text{if } p \in V_i, \\ * & \text{if } p \in M \setminus \bigsqcup_{i=1}^k V_i. \end{cases}$$

It follows that $\Psi = q_i = \Phi$ on each $\overline{\Omega}_i$ and, since Ψ by definition locally factors through smooth curves, the induced map $\Psi_*\colon \bigwedge^2 \mathrm{T}M \to \bigwedge^2 \mathrm{T}\mathrm{S}^n$ vanishes.

Llarull's argument [31] shows the existence of a Hermitian vector bundle $E_0 \to \mathrm{S}^n$ such that for any smooth map $\Theta\colon X \to \mathrm{S}^n$ with X any spin manifold, we have $\mathcal{R}^{\Theta^* E_0} \geq -a_\Theta \cdot \frac{1}{4}n(n-1)$, where a_Θ is the area-contraction function of Θ. Moreover, if X is closed and Θ has nonzero degree, then $\mathrm{ind}(\not{D}_{X,\Theta^* E_0}) \neq 0$. We now define Hermitian bundles $E = \Phi^* E_0$ and $F = \Psi^* E_0$ on M. Then F is a flat bundle because Ψ induces the zero map on 2–vectors. Thus Llarull's estimate for $E = \Phi^* E_0$ shows that (i) holds.

To see that (ii) holds, we choose pairwise disjoint open balls $U_1, \dots, U_k \subset \mathrm{S}^n$ such that $q_i \in U_i$ together with a unitary trivialization $\mathfrak{t}_0\colon E_0|_U \xrightarrow{\cong} U \times \mathbb{C}^r$, where $U = \bigsqcup_{i=1}^k U_i$. On the open set $\mathcal{N} := \Phi^{-1}(U) \cap \Psi^{-1}(U)$, this induces a unitary bundle isomorphism

$$\mathfrak{t}\colon E|_{\mathcal{N}} \xrightarrow{\Phi^*\mathfrak{t}_0} \mathcal{N} \times \mathbb{C}^r \xrightarrow{(\Psi^*\mathfrak{t}_0)^{-1}} F|_{\mathcal{N}}.$$

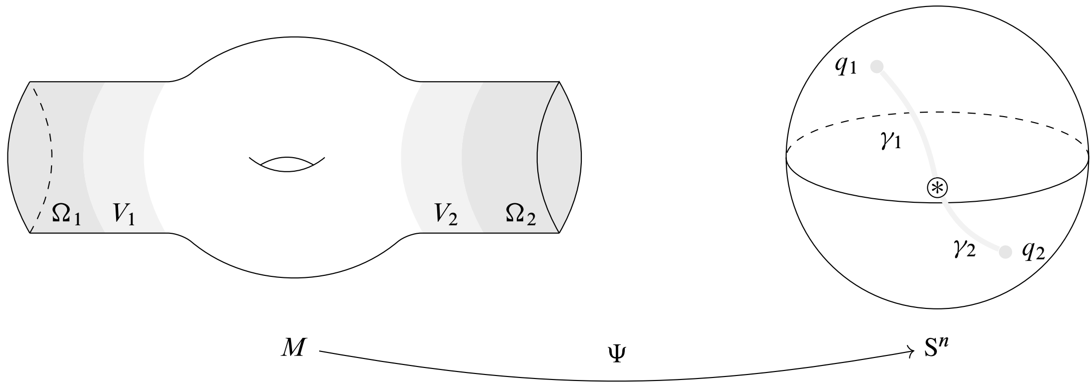


Figure 2: Construction of the map Ψ.

Note that $M \setminus K \subseteq \Phi^{-1}(\{q_1, \ldots, q_k\}) \cap \Psi^{-1}(\{q_1, \ldots, q_k\}) \subseteq \mathcal{N}$ by construction and thus $\mathcal{N}$ is an open neighborhood of $\overline{M \setminus K}$. The bundle isomorphism is not parallel on all of $\mathcal{N}$, but as $\Phi = \Psi$ is locally constant on $M \setminus K$, it follows that $\mathfrak{t}|_{M \setminus K}$ is parallel. Thus the condition (2-15) from Example 2.5 is satisfied, showing that (E, F) is a GL pair with support K.

Finally, let $\Theta\colon \mathrm{d}M = M \cup_{\partial M} M^- \to S^n$ be the smooth map defined by $\Theta|_M = \Phi$ and $\Theta|_{M^-} = \Psi$. Then $\deg(\Theta) = \deg(\Phi) \neq 0$ because Ψ has zero degree. By definition, we have $\Theta^* E_0 = V(E, F)$ and thus

$$\operatorname{indrel}(M; E, F) = \operatorname{ind}(\not{D}_{\mathrm{d}M, V(E,F)}) = \operatorname{ind}(\not{D}_{\mathrm{d}M, \Theta^* E_0}) \neq 0$$

by Llarull's argument. This shows that (iii) holds and concludes the proof of the lemma. □

Proof of Theorem 1.4 Suppose, by contradiction, that $\operatorname{dist}_g(\operatorname{supp}(\mathrm{d}\Phi), \partial M) \geq l$ and $\deg(\Phi) \neq 0$. We then pick a GL pair (E, F) satisfying the conditions (i)–(iii) from Lemma 5.1. Moreover, restating the hypotheses of the theorem, we have the curvature bounds

(iv) $\operatorname{scal}_g \geq n(n-1)$ on M,

(v) $\mathrm{H}_g \geq -\tan\bigl(\tfrac{1}{2} n l\bigr)$ on ∂M.

We then consider the relative Dirac bundle $S \to M$ constructed out of the GL pair (E, F) as in Example 2.5. Recall from (2-18) that the relevant curvature term here takes the form

$$\mathcal{R} = \tfrac{1}{4} \operatorname{scal}_g + \mathcal{R}^{E \oplus F}.$$

To construct a suitable admissible potential, we first define the 1–Lipschitz function

$$x\colon M \to [0, l], \quad x(p) := \min(\operatorname{dist}_g(K, p), l).$$

Since $\operatorname{dist}_g(K, \partial M) \geq l$, we have that $x|_{\partial M} = l$. Let $\varphi, f\colon [0, \pi/n) \to \mathbb{R}$ given by $\varphi(t) = \cos\bigl(\tfrac{1}{2} n t\bigr)^{2/n}$ and $f(t) = -\tfrac{1}{2} n \varphi'(t)/\varphi(t) = \tfrac{1}{2} n \tan\bigl(\tfrac{1}{2} n t\bigr)$; compare Remark 1.13. Then we set $\psi := f(x) := f \circ x$.

Since x is 1–Lipschitz, it follows that

$$\psi^2 - |\mathrm{d}\psi| = f(x)^2 - f'(x)|\mathrm{d}x| \geq f(x)^2 - f'(x) = -\tfrac{1}{4}n^2 \tag{5-1}$$

almost everywhere on M. Moreover, by (v),

$$\psi|_{\partial M} = \tfrac{1}{2}n\tan\bigl(\tfrac{1}{2}nl\bigr) \geq -\tfrac{1}{2}n\,\mathrm{H}_g. \tag{5-2}$$

Then ψ is an admissible potential for the relative Dirac bundle S. Corollary 3.9 together with (iii) implies that the corresponding Callias operator subject to the sign $s=1$ satisfies

$$\operatorname{ind}(\mathcal{B}_{\psi,1}) = \operatorname{indrel}(M;E,F) \neq 0.$$

In particular, there exists an element $0 \neq u \in \ker(\mathcal{B}_{\psi,1})$. To analyze u further, we let

$$U := \{p \in M \mid \mathrm{d}_p\Phi \neq 0\} \quad \text{and} \quad U' := \bigl\{p \in U \mid a(p) < \tfrac{1}{2}\bigr\}.$$

Then, by definition, $U' \subseteq U$ are open subsets in the interior of M and $U \subseteq K$. Since Φ has nonzero degree, the set U must be nonempty. Furthermore, as the map Φ is locally constant near the boundary and M is connected, the intermediate value theorem implies that U' is also nonempty. Then using our main spectral estimates (4-2) from Theorem 4.3, we obtain

$$0 \geq \frac{n}{n-1}\int_M \tfrac{1}{4}\operatorname{scal}_g|u|^2 + \langle u, \mathcal{R}^{E\oplus F}u\rangle\,\mathrm{vol}_M + \int_M \langle u, (\psi^2 + \mathrm{c}(\mathrm{d}\psi)\sigma)u\rangle\,\mathrm{vol}_M + \int_{\partial M}\underbrace{\bigl(\tfrac{1}{2}n\,\mathrm{H}_g + \psi\bigr)}_{\geq 0 \text{ by (5-2)}}|\tau(u)|^2\,\mathrm{vol}_{\partial M}$$

and continuing the estimate, using that $\psi = 0$ on U and $\mathcal{R}^{E\oplus F} = 0$ on $M \setminus U$, leads to

$$\begin{aligned}
&\geq \frac{n}{n-1}\int_U \underbrace{\tfrac{1}{4}\operatorname{scal}_g|u|^2 + \langle u, \mathcal{R}^{E\oplus F}u\rangle}_{\geq 0 \text{ by (iv) and (i)}}\mathrm{vol}_M \\
&\qquad + \int_{M\setminus U}\underbrace{\frac{n}{n-1}\cdot\tfrac{1}{4}\operatorname{scal}_g|u|^2 + (f^2(x) - f'(x))|u|^2}_{\geq 0 \text{ by (iv) and (5-1)}} + f'(x)\langle u, (1 - \mathrm{c}(\mathrm{d}x)\sigma)u\rangle\,\mathrm{vol}_M \\
&\geq \frac{n}{n-1}\int_{U'} \tfrac{1}{4}\operatorname{scal}_g|u|^2 + \langle u, \mathcal{R}^{E\oplus F}u\rangle\,\mathrm{vol}_M + \int_{M\setminus U} f'(x)\langle u, (1-\mathrm{c}(\mathrm{d}x)\sigma)u\rangle\,\mathrm{vol}_M \\
&\geq \int_{U'}\bigl(\tfrac{1}{4}n^2 - \tfrac{1}{8}n^2\bigr)|u|^2\,\mathrm{vol}_M + \int_{M\setminus U} f'(x)\langle u, (1-\mathrm{c}(\mathrm{d}x)\sigma)u\rangle\,\mathrm{vol}_M \geq 0,
\end{aligned}$$

where in the last step we used (iv) and (i) together with the fact that the area-contraction constant is at most $\frac{1}{2}$ on U' by definition. We conclude that we are in the equality situation of Theorem 4.3. Furthermore, since the last two integrands are separately nonnegative, we also deduce $u = 0$ on U' and $f'(x)\langle u, (1-\mathrm{c}(\mathrm{d}x)\sigma)u\rangle = 0$ on $M \setminus U$. Since $f'(x) > 0$ and $|\mathrm{d}x| \leq 1$, the latter implies that $\mathrm{c}(\mathrm{d}x)u = \sigma u$ almost everywhere on $M \setminus U$. Hence it follows from (4-8) in Remark 4.5 above that the modified section $w = \varphi^{-1/2}u$ has a constant norm. But since u vanishes on the nonempty open subset U' and $\varphi > 0$, this implies that u vanishes almost everywhere, a contradiction. ∎

The precise analysis of the equality situation in the proof above is necessary to rule out the case $\mathrm{dist}_g(\mathrm{d}\Phi, \partial M) = l$. If we only wanted to establish the nonstrict estimate $\mathrm{dist}_g(\mathrm{d}\Phi, \partial M) \leq l$ for $\deg(\Phi) \neq 0$, then this could be proved in a simpler way along the same lines as in [10, Proof of Theorem A] by directly showing that the operator $\mathcal{B}_{\psi,1}$ must be invertible if $\mathrm{dist}_g(\mathrm{d}\Phi, \partial M) > l$.

We now show that Theorem 1.4 is in fact sharp. The proof of the following proposition is an almost verbatim adaption of a construction due to Gromov and Lawson [23, Proposition 6.7]; see also [28, Chapter IV, Proposition 6.10].

Proposition 5.2 *For every $n \geq 2$, $\varepsilon > 0$ and $0 < l < \pi/n$, there exists a compact connected n–dimensional Riemannian spin manifold (M, g) and a smooth area-nonincreasing map $\Phi \colon M \to \mathrm{S}^n$ of nonzero degree such that*

(i) $\mathrm{scal}_g = n(n-1)$,

(ii) $\mathrm{H}_g = -\tan\big(\tfrac{1}{2}nl\big)$,

(iii) $\mathrm{dist}_g(\mathrm{supp}(\mathrm{d}\Phi), \partial M) \geq l - \varepsilon$.

In particular, Theorem 1.4 is sharp.

Proof We again work with the example from Remark 1.13; that is, consider $V := \mathrm{T}^{n-1} \times [-l, l]$ endowed with the metric $g_V = \varphi^2 g_{\mathrm{T}^{n-1}} + \mathrm{d}x \otimes \mathrm{d}x$, where $g_{\mathrm{T}^{n-1}}$ is the flat torus metric and $\varphi(t) = \cos\big(\tfrac{1}{2}nt\big)^{2/n}$. Then $\mathrm{scal}_{g_V} = n(n-1)$ and $\mathrm{H}_{g_V} = -\tan\big(\tfrac{1}{2}nl\big)$. Let S^1 be the circle of radius 1 and $* \in \mathrm{S}^1$ a basepoint. We choose a smooth map $\gamma \colon [-l, l] \to \mathrm{S}^1$ of degree one such that γ takes $[-l, l] \setminus (-\varepsilon, \varepsilon)$ to the basepoint $* \in \mathrm{S}^1$. For any $\delta > 0$ we can find a finite covering $\widetilde{\mathrm{T}}^{n-1} \to \mathrm{T}^{n-1}$ together with a δ–Lipschitz map $h \colon \widetilde{\mathrm{T}}^{n-1} \to \mathrm{S}^{n-1}$ of nonzero degree. Then we set $M := \widetilde{\mathrm{T}}^{n-1} \times [-l, l]$ endowed with the lifted metric $g := \varphi^2 g_{\widetilde{\mathrm{T}}^{n-1}} + \mathrm{d}x \otimes \mathrm{d}x$. This still satisfies $\mathrm{scal}_g = n(n-1)$ and $\mathrm{H}_g = -\tan\big(\tfrac{1}{2}nl\big)$. Let $\Theta \colon \mathrm{S}^{n-1} \times \mathrm{S}^1 \to \mathrm{S}^n$ be a smooth map of degree 1 which factors through the smash product $\mathrm{S}^{n-1} \wedge \mathrm{S}^1$. If δ is sufficiently small, then it follows that the composition

$$\Phi \colon M = \widetilde{\mathrm{T}}^{n-1} \times [-l, l] \xrightarrow{h \times \gamma} \mathrm{S}^{n-1} \times \mathrm{S}^1 \xrightarrow{\Theta} \mathrm{S}^n$$

is $(\delta \cdot C)$–area-contracting for some constant $C > 0$ which only depends on ε and the Lipschitz constants of γ and Θ; this can be seen using [23, Proposition 6.3]. By having chosen δ sufficiently small, we can thus arrange that f is area-nonincreasing. Moreover, it follows that $\deg(\Phi) = \deg(h) \neq 0$. Finally, the support of $\mathrm{d}\Phi$ is contained in $\widetilde{\mathrm{T}}^{n-1} \times [-\varepsilon, \varepsilon]$ by construction; hence $\mathrm{dist}_g(\mathrm{supp}(\mathrm{d}\Phi), \partial M) \geq l - \varepsilon$, as claimed. □

Note that unlike the extremal examples for the band and collar width estimates discussed in Remark 1.13, the bounds from Theorem 1.4 are only approximately realized. Indeed, it is not possible to do better because — unlike in our other results — Theorem 1.4 actually rules out the equality situation.

6 Relative K– and $\widehat{\mathrm{A}}$–area

In this section, we aim to prove Theorem 1.7 from the introduction. We begin with a version of K–area for manifolds with boundary. The classical case has been introduced by Gromov in [17, Section 4]. A slightly less general variant of the following was recently studied by Bär and Hanke [5].

Definition 6.1 Let (M, g) be a compact orientable Riemannian manifold with boundary. An *admissible pair* of bundles over M is a pair of Hermitian bundles $E, F \to M$ endowed with metric connections such that there exists a unitary parallel bundle isomorphism $E|_{\partial M} \xrightarrow{\cong} F|_{\partial M}$ along ∂M. We say that an admissible pair (E, F) has a nontrivial Chern number, if there exists a polynomial $p(\mathrm{c}_0, \mathrm{c}_1, \dots)$ in the Chern forms such that

$$\int_M p(\mathrm{c}_0(E), \mathrm{c}_1(E), \dots) - p(\mathrm{c}_0(F), \mathrm{c}_1(F), \dots) \neq 0. \tag{6-1}$$

The *relative* K*–area* of $(M, \partial M)$ is the supremum of the numbers

$$\|\mathrm{R}^{E \oplus F}\|_\infty^{-1},$$

where (E, F) ranges over all admissible pairs of bundles which have a nontrivial Chern number.

Remark 6.2 Since the bundles with all structures are isomorphic along the boundary, the Chern–Weil forms associated to E and F agree if pulled back to the boundary, and so (6-1) is a well-defined cohomological expression.

Remark 6.3 As in the classical case, the property of having *infinite* relative K–area does not depend on the Riemannian metric.

We now relate the notion of infinite relative K–area for a manifold with boundary M with the notion of infinite K–area for its double $\mathrm{d}M$. For a pair of Hermitian bundles $E, F \to M$ which are suitably identified in a neighborhood of ∂M, as in Section 3 we denote by $V(E, F)$ the Hermitian bundle on $\mathrm{d}M$ obtained by gluing E and F over a neighborhood of ∂M.

Proposition 6.4 *Let M be a compact manifold with boundary. The following conditions are equivalent*:

(a) *The double $\mathrm{d}M$ has infinite* K*–area in the classical sense.*

(b) *$(M, \partial M)$ has infinite relative* K*–area.*

Proof We fix a smooth Riemannian metric on $\mathrm{d}M$ and endow M with the restricted metric for the purposes of this argument. Then the statement would become immediate, if — in the definition of K–area — we only considered bundles which restricted to a fixed tubular neighborhood $\partial M \times [-1, 1] \cong U \subset \mathrm{d}M$ are of *product structure*, that is, they are of the form $\mathrm{pr}_1^* E_\partial$ for some bundle $E_\partial \to \partial M$ together with the pullback connection. Indeed, a bundle $E \to \mathrm{d}M$ of this form is just a pair of two bundles on M which are identified and of product structure on the corresponding collar neighborhood $U \cap M$ of ∂M.

Conversely, suppose $E, F \to M$ have product structure and are identified on the collar neighborhood $U \cap M$. Then the Chern numbers of $V(E, F)$ agree with the corresponding relative Chern numbers (6-1). The general case can be reduced to this observation by using a smooth map $\psi\colon \mathrm{d}M \to M$ such that ψ is smoothly homotopic to the identity map on $\mathrm{d}M$ through maps which are the identity on ∂M and preserve each half of $\mathrm{d}M$, and such that $\psi|_U$ agrees with the projection onto ∂M. In fact, ψ can be constructed on a slightly larger tubular neighborhood by manipulating the radial coordinate and setting it to be the identity on the rest of $\mathrm{d}M$. Then, for any bundle $E \to \mathrm{d}M$, the pullback $\psi^* E$ is of the special form described in the previous paragraph. Since ψ is homotopic to the identity, $\psi^* E$ and E have the same Chern numbers (and the same applies to the relative Chern number (6-1) when passing from a pair of bundles (E, F) to $(\psi|_M^* E, \psi|_M^* F)$). Moreover, ψ is a smooth map on a compact manifold, it admits some fixed Lipschitz constant $L > 0$, and hence passing from E to $\psi^* E$ only changes the norm of the curvature by at most a factor of L^2. Hence, for the purposes of detecting *infinite* K–area, it makes no difference to restrict to the class of (pairs of) bundles described in the previous paragraph, and the proposition follows. □

The notion of K–area is quite appealing because it is a purely bundle-theoretic property of the manifold and does not rely on spin structures or index theory. However, to apply it to positive scalar curvature geometry via spin geometry, the nonvanishing property (6-1) has to be translated (by potentially changing the bundle) into a property which can be used in the Atiyah–Singer index theorem for the Dirac operator. This is possible by the classical algebraic argument using Adams operations given in [17, Section $5\frac{3}{8}$]; see also Bär and Hanke [5, Lemma 7] for a situation closer to the present context. In the following, we introduce an explicit notion to capture the resulting property.

Definition 6.5 Let (M, g) be a compact Riemannian manifold with boundary. The *relative* $\widehat{\mathrm{A}}$*–area* of $(M, \partial M)$ is the supremum of the numbers

$$\|\mathrm{R}^{E \oplus F}\|_\infty^{-1},$$

where (E, F) ranges over all admissible pairs of bundles such that

$$\int_M \widehat{\mathbf{A}}(M) \wedge \mathrm{ch}(E) - \int_M \widehat{\mathbf{A}}(M) \wedge \mathrm{ch}(F) \neq 0. \tag{6-2}$$

Here $\widehat{\mathbf{A}}(M)$ denotes the $\widehat{\mathrm{A}}$–form of M and ch denotes the Chern character form.

Remark 6.6 The notion of $\widehat{\mathrm{A}}$–area for closed manifolds was discussed in Definition 1.6. An analogous argument as in Proposition 6.4 shows that $(M, \partial M)$ has infinite relative $\widehat{\mathrm{A}}$–area if and only if the double $\mathrm{d}M$ has infinite $\widehat{\mathrm{A}}$–area.

In the next proposition, we clarify the relationship between the notions of infinite relative K–area and infinite relative $\widehat{\mathrm{A}}$–area.

Proposition 6.7 *Let M be a manifold with boundary of infinite relative* K*–area and N a closed manifold such that $\widehat{\mathrm{A}}(N) \neq 0$. Then $M \times N$ has infinite relative $\widehat{\mathrm{A}}$–area. In particular, infinite relative* K*–area implies infinite relative $\widehat{\mathrm{A}}$–area.*

Proof Let $\varepsilon > 0$. We start with an admissible pair of bundles (E, F) on M such that $\|\mathrm{R}^{E\oplus F}\|_\infty < \varepsilon$ and such that (6-1) is satisfied. Arguing as in the proof of Proposition 6.4, we may assume without loss of generality that on a collar neighborhood U of the boundary the bundles E, F are of product structure and that there exists a parallel unitary bundle isomorphism $\Phi\colon E|_U \to F|_U$. Consider the bundle $V(E, F)$ on $\mathrm{d}M$. It follows that

$$\int_{\mathrm{d}M} p(\mathrm{c}_0(V(E,F)), \mathrm{c}_1(V(E,F)), \dots) \neq 0.$$

Now using the fact that $\widehat{\mathrm{A}}(\mathrm{d}M) = 0$ (as the double is nullbordant), the argument given in $\left[17, \text{Section } 5\tfrac{3}{8}\right]$ shows that (after altering the initial choice of ε, passing to a bundle associated to $V(E, F)$, and restricting the new bundle on $\mathrm{d}M$ to obtain a new pair of bundles on M), we can achieve that

$$\int_{\mathrm{d}M} \widehat{\mathbf{A}}(\mathrm{d}M) \wedge \mathrm{ch}(V(E,F)) \neq 0.$$

But this just means that (6-2) is satisfied for (E, F) on M and already shows that M has infinite $\widehat{\mathrm{A}}$–area. Finally, pulling back the bundles from M to $M \times N$ via the projection $\mathrm{pr}_M\colon M \times N \to M$ yields the desired result for $M \times N$ because

$$\begin{aligned}\int_{\mathrm{d}M\times N} \widehat{\mathbf{A}}(\mathrm{d}M\times N)\wedge\mathrm{ch}(V(\mathrm{pr}_M^* E, \mathrm{pr}_M^* F)) &= \int_{\mathrm{d}M\times N} \mathrm{pr}_{\mathrm{d}M}^* \widehat{\mathbf{A}}(\mathrm{d}M)\wedge\mathrm{pr}_N^* \widehat{\mathbf{A}}(N)\wedge\mathrm{pr}_{\mathrm{d}M}^* \mathrm{ch}(V(E,F)) \\ &= \left(\int_{\mathrm{d}M} \widehat{\mathbf{A}}(\mathrm{d}M)\wedge\mathrm{ch}(V(E,F))\right)\cdot\widehat{\mathrm{A}}(N) \neq 0. \qquad \square\end{aligned}$$

We can also relate this property to enlargeability. For an n–dimensional Riemannian manifold N with boundary, we say that $(N, \partial N)$ is *compactly area-enlargeable* if for any $\varepsilon > 0$, there exists a *finite* covering $\bar{N} \to N$ and an ε–area-contracting map $\bar{M} \to \mathrm{S}^n$ of nonzero degree which is locally constant outside a compact subset of the interior.

Proposition 6.8 *If N is an even-dimensional compact manifold with $(N, \partial N)$ compactly area-enlargeable and there exists a smooth map $\Phi\colon (M, \partial M) \to (N, \partial N)$ of nonzero $\widehat{\mathrm{A}}$–degree, then $(M, \partial M)$ has infinite relative $\widehat{\mathrm{A}}$–area.*

Proof By passing to doubles and Proposition 6.4 and Remark 6.6, it suffices to consider the case that M and N are closed manifolds. Since N is compactly area-enlargeable, for any $\varepsilon > 0$, there exists a Hermitian bundle $E_0 \to N$ such that $\|\mathrm{R}^{E_0}\|_\infty < \varepsilon/L^2$, where L is a Lipschitz constant for Φ, and such that $\mathrm{ch}(E_0)$ as a cohomology class is concentrated and nontrivial in the degrees 0 and $\dim N$; see for instance [25, page 313]. Let $E = \Phi^* E_0$ and F to be the trivial bundle of the same rank. Then it follows that $\|\mathrm{R}^{E\oplus F}\|_\infty < \varepsilon$ and

$$\int_M \widehat{\mathbf{A}}(M) \wedge (\mathrm{ch}(E) - \mathrm{ch}(F)) = \widehat{\mathrm{A}}\text{–}\deg(\Phi) \int_N \mathrm{ch}(E_0) \neq 0. \qquad \square$$

In the following lemma, we spell out a technical consequence of infinite $\widehat{\mathrm{A}}$–area that will be used in our applications.

Lemma 6.9 *Let (M,g) be an even-dimensional compact spin manifold of infinite relative $\widehat{\mathrm{A}}$–area and $U \cong \partial M \times (-1,0] \subseteq M$ be an open collar neighborhood. Then for any sufficiently small $\varepsilon_1, \varepsilon_2 > 0$, there exists a pair of Hermitian vector bundles (E,F) on M such that:*

(i) *There exists a parallel unitary bundle isomorphism $\mathfrak{t}\colon E|_{U_{\varepsilon_1}} \to F|_{U_{\varepsilon_1}}$, where U_{ε_1} corresponds to $\partial M \times (-1+\varepsilon_1, 0]$ in the tubular neighborhood.*

(ii) $\|\mathrm{R}^{E\oplus F}\|_\infty < \varepsilon_2$.

(iii) $\operatorname{ind}(\mathcal{B}_{\psi,1}) \neq 0$, *where $\mathcal{B}_{\psi,1}$ is the operator considered in Corollary 3.9 associated to the relative Dirac bundle constructed from (E,F) as in Example 2.5.*

Proof Let (E,F) be an admissible pair of bundles given by infinite $\widehat{\mathrm{A}}$–area satisfying (6-2) and (ii). Again arguing as in the proof of Proposition 6.4, we may assume that E and F are of product structure on the collar neighborhood and that also (i) is satisfied. We again consider the bundle $V(E,F)$ on $\mathrm{d}M$. Using Corollary 3.9 and the Atiyah–Singer index theorem, we deduce

$$\operatorname{ind}(\mathcal{B}_{\psi,1}) = \operatorname{indrel}(M;E,F) = \operatorname{ind}(\not{D}_{\mathrm{d}M,V(E,F)}) = \int_{\mathrm{d}M} \widehat{\mathbf{A}}(\mathrm{d}M) \wedge \operatorname{ch}(V(E,F)) \neq 0,$$

where the final nonvanishing statement is due to (6-2). Thus (iii) is also satisfied, completing the proof. □

After all these technical preparations, we are now ready to give a proof of Theorem 1.7 from the introduction.

Proof of Theorem 1.7 For all sufficiently small $d > 0$, denote by $\mathcal{N}_d$ the open geodesic collar neighborhood of ∂M of width d. Suppose, by contradiction, that $\mathcal{N}_{l'}$ exists for some $l < l' < \pi/(\sqrt{\kappa}n)$ such that $\operatorname{scal}_g \geq \kappa n(n-1)$ in $\mathcal{N}_{l'}$. Fix $\Lambda \in (l,l')$. Then $K_\Lambda := M \setminus \mathcal{N}_\Lambda$ is a compact manifold with boundary such that $\operatorname{scal}_g \geq \kappa n(n-1)$ in $M \setminus K_\Lambda$. For $r \in (0, \sqrt{\kappa}n/2)$, consider the function $Y_r(t) = r\tan(rt)$, with t varying in $[0, \pi/(\sqrt{\kappa}n))$. Observe that $Y_r(0) = 0$ and

$$\tfrac{1}{4}\kappa n^2 + Y_r^2 - |Y_r'| = \tfrac{1}{4}\kappa n^2 - r^2 > 0.$$

By choosing r close enough to $\sqrt{\kappa}n/2$, we can also ensure that

$$Y_r(\Lambda) > \tfrac{1}{2}\sqrt{\kappa}n\tan\bigl(\tfrac{1}{2}\sqrt{\kappa}nl\bigr). \tag{6-3}$$

Let $\kappa_0 := \inf_{p\in K_\Lambda} \operatorname{scal}_g(p) > 0$. By Lemma 6.9, Corollary 3.9 and Example 2.5, there exists a GL pair (E,F) and associated relative Dirac bundle $S \to M$ such that

(i) (E,F) and thus S have support K_Λ,

(ii) $4\|\mathcal{R}^{E\oplus F}\|_\infty < \kappa_0$,

(iii) $\dfrac{n}{n-1}\|\mathcal{R}^{E\oplus F}\|_\infty < \tfrac{1}{4}\kappa n^2 - r^2$,

(iv) $\operatorname{ind}(\mathcal{B}_{\psi,1}) = \operatorname{indrel}(M;E,F) \neq 0$ for any admissible potential ψ.

We now use the function Y_r to construct an admissible potential. Let $x\colon M \to [0,\Lambda]$ be the distance function from K_Λ. Consider the Lipschitz function $\psi := Y_r \circ x$ (which is actually smooth in the complement of ∂K_Λ). Then $\psi|_{K_\Lambda} = 0$ and $\psi|_{\partial M} = Y_r(\Lambda)$. By (i) and (6-3), ψ is an admissible potential satisfying

$$\psi|_{\partial M} > \tfrac{1}{2}\sqrt{\kappa}n\tan\bigl(\tfrac{1}{2}\sqrt{\kappa}nl\bigr). \tag{6-4}$$

Let $\mathcal{B}_{\psi,1}$ be the associated Callias operator subject to the boundary condition coming from the sign $s=1$ and $u \in \operatorname{dom}(\mathcal{B}_{\psi,1}) = \mathrm{H}^1_{\sigma,1}(M,S)$. From (6-4), we deduce

$$\int_{\partial M} \bigl(\tfrac{1}{2}n\,\mathrm{H}_g + s\psi\bigr)|u|^2 \operatorname{vol}_{\partial M} \geq 0.$$

Therefore, the estimates (4-2) in Theorem 4.3 imply

$$\int_M |\mathcal{B}_\psi u|^2 \operatorname{vol}_M \geq \int_M \Theta_{\psi,n}|u|^2 \operatorname{vol}_M,$$

where $\Theta_{\psi,n}$ is the L^∞–function defined by

$$\Theta_{\psi,n} := \frac{n}{n-1}\left(\tfrac{1}{4}\operatorname{scal}_g - |\mathcal{R}^{E\oplus F}|\right) + \psi^2 - |\mathrm{d}\psi|.$$

By (ii) and since ψ is constant on K_Λ, in the interior K_Λ° we have

$$\Theta_{\psi,n} = \frac{n}{n-1}\bigl(\tfrac{1}{4}\operatorname{scal}_g - |\mathcal{R}^{E\oplus F}|\bigr) \geq \frac{n}{n-1}\bigl(\tfrac{1}{4}\kappa_0 - \|\mathcal{R}^{E\oplus F}\|_\infty\bigr) > 0.$$

By (iii), $\operatorname{scal}_g \geq \kappa n(n-1)$ on $\mathcal{N}_{l'}$ and since x is 1–Lipschitz, in $M \setminus K_\Lambda^\circ$ we have

$$\Theta_{\psi,n} \geq \tfrac{1}{4}\kappa n^2 + \psi^2 - |\mathrm{d}\psi| - \frac{n}{n-1}\|\mathcal{R}^{E\oplus F}\|_\infty \geq \tfrac{1}{4}\kappa n^2 - r^2 - \frac{n}{n-1}\|\mathcal{R}^{E\oplus F}\|_\infty > 0.$$

Therefore, there exists a constant $c>0$ such that $\|\mathcal{B}_\psi u\| \geq c\|u\|$ for all $u \in \mathrm{H}^1_{\sigma,1}(M,S)$. It follows that $\operatorname{ind}(\mathcal{B}_{\psi,1}) = 0$, contradicting (iv). □

7 Estimates of bands

In this section, we prove our statements related to Conjecture 1.3, that is, estimates of Riemannian bands under lower bounds on scalar and mean curvature. We start with reviewing Gromov's notion of a Riemannian band [18] and other relevant concepts to formulate our results more conveniently.

Definition 7.1 (i) A *band* is a compact manifold V together with a decomposition $\partial V = \partial_- V \sqcup \partial_+ V$, where $\partial_\pm V$ are unions of components.

(ii) A map $\Phi\colon V \to V'$ between bands is called a *band map* if $\Phi(\partial_\pm V) \subseteq \partial_\pm V'$.

(iii) The *width* $\operatorname{width}(V,g)$ of a Riemannian band (V,g) is the distance between $\partial_- V$ to $\partial_+ V$ with respect to g.

(iv) A *width function* for a Riemannian band (V,g) is a 1–Lipschitz function $x\colon V \to [t_-,t_+]$ for some real numbers $t_- < t_+$ such that $\partial_\pm V \subseteq x^{-1}(t_\pm)$.

In other words, a width function is a band map $V \to [t_-, t_+]$ which is 1–Lipschitz. A width function $x\colon V \to [t_-, t_+]$ always satisfies $t_+ - t_- \le \operatorname{width}(V, g)$. We also have the following converse:

Lemma 7.2 *Let (V, g) be a band. Then there exists a width function $x\colon V \to [t_-, t_+]$ which is smooth near the boundary of V and is such that $t_+ - t_- = \operatorname{width}(V, g)$. Moreover, for every $t_- < t_+$ satisfying $t_+ - t_- < \operatorname{width}(V, g)$, there exists a smooth width function $x\colon V \to [t_-, t_+]$.*

Proof Let $w = \operatorname{width}(V, g)$. Then we obtain the desired Lipschitz width function by setting

$$x\colon V \to [0, w], \quad x(p) := \begin{cases} d(p, \partial_- V) & \text{if } d(p, \partial_- V) \le \frac{1}{2} w, \\ w - d(p, \partial_+ V) & \text{if } d(p, \partial_+ V) \le \frac{1}{2} w, \\ \frac{1}{2} w & \text{otherwise.} \end{cases}$$

Note that this x is already smooth in a neighborhood of ∂V. Moreover, if $t_+ - t_- < w$, we can find $\varepsilon > 0$ such that $(1+\varepsilon)^{-1} w = t_+ - t_-$. We can then approximate x by a smooth function $\tilde{x}\colon V \to [0, w]$ which agrees with x near ∂V and is $(1+\varepsilon)$–Lipschitz. Then $(1+\varepsilon)^{-1}\tilde{x}\colon V \to [0, t_+ - t_-]$ is a smooth width function. Translating to the interval $[t_-, t_+]$ yields the desired result. □

The following notion is an adaption of the ideas from Section 6 to the situation of bands.

Definition 7.3 A band V is said to have *infinite vertical* $\hat{\mathrm{A}}$*–area* if for every $\varepsilon > 0$, there exists a Hermitian vector bundle $E \to V$ such that $\|\mathrm{R}^E\|_\infty < \varepsilon$ and such that we have

$$\int_{\partial_- V} \hat{\mathbf{A}}(\partial_- V) \wedge \operatorname{ch}(E|_{\partial_- V}) \ne 0. \tag{7-1}$$

Example 7.4 A simple example of an infinite vertical $\hat{\mathrm{A}}$–area is an $\hat{\mathrm{A}}$–*band*; that is, a band such that $\hat{\mathrm{A}}(\partial_- V) \ne 0$. Another is a band of the form $V = M \times [-1, 1]$, where M has infinite $\hat{\mathrm{A}}$–area.

Example 7.5 If N is a closed manifold that is compactly area-enlargeable and there exists a band map $V \to N \times [0, 1]$ of nonzero $\hat{\mathrm{A}}$–degree, then V has infinite vertical $\hat{\mathrm{A}}$–area; compare Proposition 6.8. In particular, this includes the classes of *overtorical bands* introduced by Gromov in [18], and the generalization of $\hat{\mathrm{A}}$–*overtorical bands* studied in [42].

Theorem 7.6 *Let (V, g) be a spin band of infinite vertical $\hat{\mathrm{A}}$–area. Suppose that $\operatorname{scal}_g \ge n(n-1)$ and let $-\pi/n < t_- < t_+ < \pi/n$ such that the mean curvature of ∂V satisfies*

$$\mathrm{H}_g|_{\partial_\pm V} \ge \mp\tan\left(\tfrac{1}{2} n t_\pm\right). \tag{7-2}$$

Then $\operatorname{width}(V, g) \le t_+ - t_-$.

Proof Assume by contradiction that $\operatorname{width}(V, g) > t_+ - t_-$. Let $t_0 := \frac{1}{2}(t_+ + t_-)$ be the midpoint between t_- and t_+. Then we can find $d > 0$ such that $\operatorname{width}(V, g) > 2d > t_+ - t_-$ and $-\pi/n < t_0 - d < t_0 + d < \pi/n$. Lemma 7.2 implies that there exists a smooth width function $x\colon V \to [-d, d]$.

Now for $0 < \lambda \leq 1$, we set

$$f_\lambda : [-d, d] \to \mathbb{R}, \quad f_\lambda(s) = \tfrac{1}{2}\lambda n \tan\bigl(\tfrac{1}{2}n(t_0 + \lambda s)\bigr).$$

We now fix a $\lambda_0 < 1$ such that

$$\mp\tfrac{1}{2}n \tan\bigl(\tfrac{1}{2}nt_\pm\bigr) \geq \mp f_{\lambda_0}(\pm d). \tag{7-3}$$

This is possible because the tangent function is increasing and $t_0 - d < t_- < t_+ < t_0 + d$ by our choice of d. We now choose a Hermitian bundle $E \to V$ satisfying (7-1) and such that the corresponding Weitzenböck curvature endomorphism satisfies $\mathcal{R}^E \geq -\delta$ for some $\delta < \frac{1}{4}n^2(1-\lambda_0^2)$. Now we consider the Callias operator $\mathcal{B}_{\psi,s}$ associated to the relative Dirac bundle from Example 2.6 (using the twisting bundle E) and with potential $\psi := f_{\lambda_0} \circ x$ and subject to the boundary conditions coming from the choice of signs $s(\partial_\pm V) = \pm 1$. Now (7-1) and Corollary 3.10 imply that $\operatorname{ind}(\mathcal{B}_{\psi,s}) \neq 0$. On the other hand, we have

$$\psi^2 - |\mathrm{d}\psi| = f_{\lambda_0}(x)^2 - f'_{\lambda_0}(x)|\mathrm{d}x| \geq f_{\lambda_0}(x)^2 - f'_{\lambda_0}(x) = -\lambda_0^2 \cdot \tfrac{1}{4}n^2.$$

Thus (4-2) together with $\mathrm{scal}_g \geq n(n-1)$ implies that each u in the domain of $\mathcal{B}_{\psi,s}$ satisfies

$$\int_V |\mathcal{B}_{\psi,s}u|^2 \,\mathrm{vol}_V \geq \bigl(\tfrac{1}{4}n^2(1-\lambda_0^2) - \delta\bigr)\int_V |u|^2 \,\mathrm{vol}_M + \int_{\partial_- V}\bigl(\tfrac{1}{2}n\,\mathrm{H}_g - f_{\lambda_0}(-d)\bigr)|u|^2 \,\mathrm{vol}_{\partial_- V} + \int_{\partial_- V}\bigl(\tfrac{1}{2}n\,\mathrm{H}_g + f_{\lambda_0}(d)\bigr)|u|^2 \,\mathrm{vol}_{\partial_+ V}.$$

Since the terms $\frac{1}{2}n\,\mathrm{H}_g|_{\partial_\pm V} \pm f_{\lambda_0}(\pm d)$ at the boundary are nonnegative by (7-2) and (7-3), this implies that

$$\int_V |\mathcal{B}_{\psi,s}u|^2 \,\mathrm{vol}_V \geq C\int_V |u|^2 \,\mathrm{vol}_V,$$

where $C = \frac{1}{4}n^2(1-\lambda_0^2) - \delta > 0$ by the choice of $\delta > 0$. This shows that the operator $\mathcal{B}_{\psi,s}$ is invertible, a contradiction to $\operatorname{ind}(\mathcal{B}_{\psi,s}) \neq 0$. □

Corollary 7.7 *Let (V, g) be a spin band of infinite vertical $\widehat{\mathrm{A}}$–area. Suppose that $\mathrm{scal}_g \geq n(n-1)$ and that the mean curvature of ∂V satisfies*

$$\mathrm{H}_g \geq -\tan\bigl(\tfrac{1}{4}nl\bigr) \quad \textit{for some } 0 < l < \frac{2\pi}{n}.$$

Then $\mathrm{width}(V, g) \leq l$.

Proof This is a consequence of Theorem 7.6 by setting $t_\pm := \pm\frac{1}{2}l$. □

Corollary 7.8 *Let (V, g) be a spin band of infinite vertical $\widehat{\mathrm{A}}$–area. Suppose that $\mathrm{scal}_g \geq n(n-1)$ and that the mean curvature of ∂V satisfies*

$$\mathrm{H}_g|_{\partial_- V} \geq 0 \quad \textit{and} \quad \mathrm{H}_g|_{\partial_+ V} \geq -\tan\bigl(\tfrac{1}{2}nl\bigr) \quad \textit{for some } 0 < l < \frac{\pi}{n}.$$

Then $\mathrm{width}(V, g) \leq l$.

Proof This is also a consequence of Theorem 7.6 by setting $t_- := 0$ and $t_+ := l$. □

Corollary 7.9 *Let (V, g) be a spin band of infinite vertical $\widehat{\mathrm{A}}$–area. Suppose that $\operatorname{scal}_g \geq n(n-1)$. Then we always have $\operatorname{width}(V, g) < 2\pi/n$. Moreover, if $\partial_- V$ is mean-convex, then $\operatorname{width}(V, g) < \pi/n$.*

Proof The first statement follows from Corollary 7.7 and the second is a consequence of Corollary 7.8, in both cases because $-\tan(\vartheta) \to -\infty$ as $\vartheta \nearrow \frac{\pi}{2}$, but $\inf_{p \in \partial V} \mathrm{H}_g > -\infty$ by compactness. □

8 The general warped product rigidity theorem

In this section, we establish a general rigidity theorem which allows us to compare metrics on bands with certain warped products.

Setup 8.1 We consider the following setup. Let $n \in \mathbb{N}$, let $I \subseteq \mathbb{R}$ be an interval, V a band, and $x \colon V \to I$ a continuous function. Furthermore, let $\varphi \colon I \to (0, \infty)$ and $\kappa, \nu \colon V \to \mathbb{R}$ be smooth functions. We define the auxiliary functions $h, f \colon I \to \mathbb{R}$ by

$$h(t) := \frac{\varphi'(t)}{\varphi(t)} \quad \text{and} \quad f(t) := -\tfrac{1}{2} n h(t),$$

and suppose that the following two conditions are satisfied:

- The function φ is strictly logarithmically concave, that is, $h'(t) < 0$ for all $t \in I$.
- For all $p \in V$, the following inequality holds:

$$\tfrac{1}{4} n^2 \kappa(p) + \frac{n\nu(p)}{n-1} + f(x(p))^2 - f'(x(p)) \geq 0. \tag{8-1}$$

Remark 8.2 The geometric motivation for (8-1) is the following. Suppose that we have a warped product metric $g = \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x$ on an n–dimensional band of the form $V = M \times [t_-, t_+]$. Then, if we choose $\kappa \colon V \to \mathbb{R}$ and $\nu \colon V \to \mathbb{R}$ such that $n(n-1)\kappa = \operatorname{scal}_g$ and $-\nu = \frac{1}{4} \operatorname{scal}_{\varphi^2 g_M} = (1/(4\varphi^2)) \operatorname{scal}_{g_M}$, we precisely obtain the relation

$$\tfrac{1}{4} n^2 \kappa(y, t) + \frac{n\nu(y, t)}{n-1} + f(t)^2 - f'(t) = 0 \quad \text{for all } (y, t) \in M \times [t_-, t_+],$$

where f is defined as in Setup 8.1 above. In this sense, (8-1) abstractly models the scalar curvature equation for warped products.

Theorem 8.3 *Let (V, g) be an n–dimensional Riemannian spin band and $E \to V$ be a Hermitian vector bundle endowed with a metric connection. We suppose that we have chosen smooth functions $x \colon V \to I$, $\varphi, h, f \colon I \to \mathbb{R}$ and $\kappa, \nu \colon V \to \mathbb{R}$ satisfying Setup 8.1. Furthermore, we assume that the following conditions hold:*

(i) *$\operatorname{scal}_g \geq n(n-1)\kappa$ and $\mathcal{R}^E \geq \nu$.*

(ii) *$x \colon V \to [t_-, t_+]$ is a width function for some $t_-, t_+ \in I$.*

(iii) *$\mathrm{H}_g|_{\partial_\pm V} \geq \pm h(t_\pm)$, where H_g denotes the mean curvature of ∂V.*

(iv) $\ker(\mathcal{B}_{\psi,s}) \neq 0$, *where* $\mathcal{B}_{\psi,s}$ *is the Callias operator associated to the relative Dirac bundle from* Example 2.6 *with potential* $\psi := f \circ x$ *and subject to the boundary conditions coming from the choice of signs* $s(\partial_\pm V) = \pm 1$.

Then $\mathrm{scal}_g = n(n-1)\kappa$, *and* (V, g) *is isometric to the warped product*

$$(M \times [t_-, t_+], \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x),$$

where $M = x^{-1}(t_0)$ *for an arbitrary fixed* $t_0 \in [t_-, t_+]$ *and* $g_M := \varphi(t_0)^{-2} g|_M$ *on* M. *Furthermore, for any* $u \in \ker(\mathcal{B}_{\psi,s})$ *and* $t \in [t_-, t_+]$, *the restriction* $u|_{M\times\{t\}}$ *is parallel with respect to the connection*

$$\widetilde{\nabla}_\xi := \nabla_\xi + \tfrac{1}{2}\,\mathrm{c}(\nabla_\xi \mathrm{d}x)\,\mathrm{c}(\mathrm{d}x), \quad \textit{with } \xi \in \mathrm{T}(M \times \{t\}).$$

The proof of this theorem is based on the following lemma, where we include the slightly more general situation of x being a not necessarily smooth Lipschitz function. This will be of importance later for the rigidity of bands.

Lemma 8.4 *Let* (V, g) *be an* n*–dimensional Riemannian spin band and let* $E \to V$ *be a Hermitian vector bundle endowed with a metric connection. Let* $x\colon V \to I$ *be a* 1*–Lipschitz function, and* $\varphi, h, f\colon I \to \mathbb{R}$ *and* $\kappa, \nu\colon V \to \mathbb{R}$ *be smooth functions satisfying* Setup 8.1. *Suppose furthermore that the conditions* (i)–(iv) *from the statement of* Theorem 8.3 *are satisfied. Then, for any* $u \in \ker(\mathcal{B}_{\psi,s})$, *the section* $w := \varphi(x)^{-1/2} u$ *lies in* $\mathrm{H}^1_{\sigma,s}(V, S)$ *and satisfies almost everywhere*

$$\mathrm{c}(\mathrm{d}x)w = \sigma w, \tag{8-2}$$

$$\nabla_\xi w = \frac{f(x)}{n}\,\mathrm{c}(\xi^\flat \wedge \mathrm{d}x) w \tag{8-3}$$

for every vector field ξ *on* V, *where* $\mathrm{c}(\xi^\flat \wedge \mathrm{d}x) = \frac{1}{2}(\mathrm{c}(\xi^\flat)\,\mathrm{c}(\mathrm{d}x) - \mathrm{c}(\mathrm{d}x)\,\mathrm{c}(\xi^\flat))$. *In particular,* $|w|^2$ *is a constant function and thus* $|u|^2$ *is a constant multiple of the function* $\varphi(x)$. *Moreover,* $|\mathrm{d}x| = 1$ *almost everywhere and* $\mathrm{scal}_g = n(n-1)\kappa$.

Proof We first observe that $|\mathrm{d}x| \le 1$ almost everywhere because x is 1–Lipschitz. Let $u \in \ker(\mathcal{B}_{\psi,s})$. By (iv), we can assume that $u \neq 0$. Then, using (4-2) from Theorem 4.3, we obtain

$$\begin{aligned}
0 &= \int_V |\mathcal{B}_\psi u|^2 \,\mathrm{vol}_V \\
&\ge \frac{n}{4(n-1)} \int_V (\mathrm{scal}_g - n(n-1)\kappa)|u|^2 \,\mathrm{vol}_V + \frac{n}{n-1}\int_V (\langle u, \mathcal{R}^E u\rangle - \nu|u|^2)\,\mathrm{vol}_V \\
&\quad + \int_V \Big(\tfrac{1}{4}n^2\kappa + \frac{n\nu}{n-1} + f(x)^2 - f'(x)\Big)|u|^2\,\mathrm{vol}_V + \int_V f'(x)\langle u, (1 + \mathrm{c}(\mathrm{d}x)\sigma)u\rangle\,\mathrm{vol}_V \\
&\quad + \int_{\partial_- V} \big(\tfrac{1}{2}n\,\mathrm{H}_g - f(t_-)\big)|u|^2\,\mathrm{vol}_{\partial_- V} + \int_{\partial_+ V}\big(\tfrac{1}{2}n\,\mathrm{H}_g + f(t_+)\big)|u|^2\,\mathrm{vol}_{\partial_+ V}.
\end{aligned} \tag{8-4}$$

Note that in the latter six integrals each integrand is nonnegative — in the first two cases due to (i), in the third this follows from (8-1), in the fourth from $|\mathrm{d}x| \le 1$, and for the two boundary integrals this is a consequence of item (iii). Thus all integrands appearing in (8-4) vanish (almost everywhere) and we are in the equality situation of Theorem 4.3. Since $f' > 0$ and $1 + \mathrm{c}(\mathrm{d}x)\sigma \geq 0$ almost everywhere, we furthermore deduce from vanishing of the fourth integrand that $\mathrm{c}(\mathrm{d}x)\sigma u = -u$ almost everywhere and thus

$$\mathrm{c}(\mathrm{d}x)u = \sigma u. \tag{8-5}$$

This already proves (8-2) for $w := \varphi(x)^{-1/2}u$. The identity (8-3) and the fact that $|u|^2 = c\varphi$ for $c = |w|^2$ constant is now a consequence of Remark 4.5 because (8-5) holds almost everywhere. Moreover, since $|u|^2 = c\varphi > 0$, vanishing of the first integrand in (8-4) implies that $\mathrm{scal}_g = n(n-1)\kappa$. It also follows from (8-2) that $|\mathrm{d}x|^2|w|^2 = |\,\mathrm{c}(\mathrm{d}x)w|^2 = |\sigma w|^2 = |w|^2$, and thus $|\mathrm{d}x| = 1$ almost everywhere. □

Proof of Theorem 8.3 We will refer to Lemma 8.4 and its proof in the following argument. First of all, we obtain that $|\mathrm{d}x| = 1$. Since x is assumed to be smooth, this already implies that V is diffeomorphic to $M \times [t_-, t_+]$ with g corresponding to $g_x + \mathrm{d}x \otimes \mathrm{d}x$ for some family of Riemannian metrics $(g_x)_{x \in [t_-, t_+]}$ on M. To prove the theorem, it remains to compute the Hessian of x. To this end, we let $0 \neq u \in \ker(\mathcal{B}_{\psi,\sigma})$ and let $w := \varphi(x)^{-1/2}u$ as in Lemma 8.4. Since $\psi = f(x)$ is smooth, boundary elliptic regularity (see for example [4, Theorem 4.4]) implies that u and thus w are smooth sections. Using equations (8-2) and (8-3) and the fact that σ is parallel, we compute for any smooth vector field ξ on V,

$$\begin{aligned}
\mathrm{c}(\nabla_\xi \mathrm{d}x)w = \underbrace{\nabla_\xi(\mathrm{c}(\mathrm{d}x)w)}_{=\sigma\nabla_\xi w} - \mathrm{c}(\mathrm{d}x)\nabla_\xi w \underset{(8\text{-}3)}{=} &\frac{f(x)}{n}\sigma\, \mathrm{c}(\xi^\flat \wedge \mathrm{d}x)w - \frac{f(x)}{n}\mathrm{c}(\mathrm{d}x)\,\mathrm{c}(\xi^\flat \wedge \mathrm{d}x)w \\
= &\frac{f(x)}{n}(\mathrm{c}(\xi^\flat \wedge \mathrm{d}x)\sigma w + \mathrm{c}(\xi^\flat \wedge \mathrm{d}x)\,\mathrm{c}(\mathrm{d}x)w) \\
= &\frac{2f(x)}{n}\mathrm{c}(\xi^\flat \wedge \mathrm{d}x)\,\mathrm{c}(\mathrm{d}x)w \\
= &-\frac{2f(x)}{n}\mathrm{c}(\xi^\flat - \mathrm{d}x(\xi)\,\mathrm{d}x)w = h(x)\,\mathrm{c}(\xi^\flat - \mathrm{d}x(\xi)\,\mathrm{d}x)w.
\end{aligned}$$

Since w vanishes nowhere, this implies $\nabla_\xi \mathrm{d}x = h(x)(\xi^\flat - \mathrm{d}x(\xi)\mathrm{d}x)$ for each vector field ξ. Hence the Hessian of x is given by

$$\nabla^2 x = h(x)(g - \mathrm{d}x \otimes \mathrm{d}x). \tag{8-6}$$

Using standard formulas for Riemannian distance functions (in the sense of [34, Section 3.2.2]), the identity (8-6) implies for the Lie derivative of g_x that

$$\mathrm{L}_{\partial_x} g_x = 2h(x)g_x;$$

see for instance [34, Proposition 3.2.11(1)]. Let $\tilde{g}_x = \varphi(x)^2 g_M$, where $g_M := \varphi(t_0)^{-2} g|_M$. Then $\tilde{g}_x$ satisfies the same differential equation because

$$\mathrm{L}_{\partial_x} \tilde{g}_x = 2\varphi'(x)\varphi(x)g_M = 2\frac{\varphi'(x)}{\varphi(x)}\tilde{g}_x = 2h(x)\tilde{g}_x.$$

Since $\tilde{g}_{t_0} = g_{t_0} = g_M$ this implies that $g_x = \tilde{g}_x = \varphi(x)^2 g_M$, as desired.

To see the final statement, observe that by Remark 4.4 we have

$$\nabla_\xi u = \frac{f(x)}{n}\,\mathrm{c}(\xi^\flat)\sigma u = \frac{f(x)}{n}\,\mathrm{c}(\xi^\flat)\,\mathrm{c}(\mathrm{d}x)u.$$

Since $\nabla_\xi \mathrm{d}x = h(x)\xi = -(2/n)f(x)\xi$ for any vertical tangent vector ξ, this implies that $u|_{M\times\{t\}}$ is parallel with respect to $\tilde{\tilde{\nabla}}$. □

9 Rigidity of bands

In this section, we use the general results from the previous section to deduce a rigidity result for $\hat{\mathrm{A}}$–bands subject to scalar and mean curvature bounds.

Theorem 9.1 *Let (V,g) be an n–dimensional band which is a spin manifold and satisfies $\hat{\mathrm{A}}(\partial_- V)\neq 0$. Suppose furthermore that there exist $-\pi/n < t_- < t_+ < \pi/n$ such that*

(i) $\mathrm{scal}_g \geq n(n-1)$,

(ii) $\mathrm{width}(V,g)\geq t_+ - t_-$,

(iii) $\mathrm{H}_g|_{\partial_\pm V} \geq \mp\tan\big(\tfrac12 n t_\pm\big)$, *where H_g denotes the mean curvature of ∂V.*

Then (V,g) is isometric to a warped product $(M\times[t_-,t_+], \varphi^2 g_M + \mathrm{d}x\otimes\mathrm{d}x)$, where $\varphi(t)=\cos\big(\tfrac12 nt\big)^{2/n}$ and g_M is some Riemannian metric on M which carries a nontrivial parallel spinor. In particular, g_M is Ricci-flat.

Proof We first prepare a particular case of Setup 8.1 we wish to apply. To this end, let

$$\varphi\colon \Big(-\frac{\pi}{n},\frac{\pi}{n}\Big)\to(0,\infty),\qquad \varphi(t):=\cos\big(\tfrac12 nt\big)^{2/n}.$$

Then

$$h(t):=\frac{\varphi'(t)}{\varphi(t)} = -\tan\big(\tfrac12 nt\big),\qquad h'(t) = -\tfrac12 n\frac{1}{\cos\big(\tfrac12 nt\big)^2} < 0,$$

and

$$\tfrac14 n^2 + f(t)^2 - f'(t) = 0, \tag{9-1}$$

where $f=-\tfrac12 nh$. Thus (8-1) is satisfied with $\kappa\equiv 1$ and $\nu\equiv 0$ and we are in Setup 8.1. Next we use Lemma 7.2 to choose a width function $x\colon V\to[t_-,t_+]$ which is smooth near the boundary of V. We let $\psi = f\circ x$ and form the Callias-type operator $\mathcal{B}_\psi=\mathcal{D}+\psi\sigma$.

We choose signs $s\colon \partial M\to\{\pm1\}$ as in Corollary 3.10. Then, since $\hat{\mathrm{A}}(\partial_- V)\neq 0$, we deduce from Corollary 3.10 that $\mathrm{ind}(\mathcal{B}_{\psi,s})\neq 0$. In particular, $\ker(\mathcal{B}_{\psi,s})\neq 0$.

At this point, if x was smooth everywhere, the result would follow readily from Theorem 8.3 because (iii) says that $\mathrm{H}_g|_{\partial_\pm V}\geq \pm h(t_\pm)$. However, since we do not know this a priori, we need to supply an argument ensuring that x is indeed smooth everywhere. To this end, we fix $0\neq u\in\ker(\mathcal{B}_{\psi,s})$ and apply Lemma 8.4. From equations (3-2), (8-2) and (9-1), we thus deduce

$$0=\mathcal{B}_\psi^2 u = \mathcal{D}^2 u + f'(x)\,\mathrm{c}(\mathrm{d}x)\sigma u + f(x)^2 u = \mathcal{D}^2 u - f'(x)u + f(x)^2 u = \mathcal{D}^2 u - \tfrac14 n^2 u.$$

Then interior elliptic regularity for the operator $\mathcal{D}^2 - \frac{1}{4}n^2$ implies that u is smooth in the interior of V. Since, by Lemma 8.4, u is nowhere vanishing and we have the equality $\mathrm{c}(\mathrm{d}x)u = \sigma u$, this implies that the covector field $\mathrm{d}x$ must also be smooth in the interior of V. Since we already know that x is smooth near the boundary, this just means that the function x is smooth everywhere and so we can indeed apply Theorem 8.3.

We conclude that (V, g) is isometric to a warped product $(M \times [t_-, t_+], \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x)$ and $\mathrm{scal}_g \equiv n(n-1)$. Moreover, the final statement of Theorem 8.3 implies that any $0 \neq u \in \ker(\mathcal{B}_{\psi,s})$ restricts to a nowhere-vanishing parallel spinor on each fiber; compare (2-8). It is a well-known fact that the existence of a parallel spinor forces the Ricci curvature to vanish; see for instance [8, Corollary 2.8]. □

Corollary 9.2 *Let (V, g) be an n–dimensional band which is a spin manifold and satisfies $\hat{\mathrm{A}}(\partial_- V) \neq 0$. Suppose that $\mathrm{scal}_g \geq n(n-1)$. Let $0 < d < \pi/n$ and assume furthermore that one of the following conditions holds: either*

- $\mathrm{width}(V, g) \geq 2d$ *and we have* $\mathrm{H}_g|_{\partial V} \geq -\tan\left(\frac{1}{2}nd\right)$, *or*
- $\mathrm{width}(V, g) \geq d$ *and we have* $\mathrm{H}_g|_{\partial_- V} \geq 0$ *and* $\mathrm{H}_g|_{\partial_+ V} \geq -\tan\left(\frac{1}{2}nd\right)$.

Then (V, g) is isometric to a warped product $(M \times I, \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x)$, where either $I = [-d, d]$ or $I = [0, d]$, and we have $\varphi(t) = \cos\left(\frac{1}{2}nt\right)^{2/n}$, while g_M is some Riemannian metric on M which carries a nontrivial parallel spinor. In particular, g_M is Ricci-flat.

Proof This follows immediately from Theorem 9.1 by setting $t_\pm = \pm d$ in the first case, and $t_- = 0$ and $t_+ = d$ in the second. □

10 Scalar-mean extremality and rigidity of warped products

In this section, we prove our general extremality and rigidity results for logarithmically concave warped products.

As a preparation for the proof of the main theorem, we discuss a particularly relevant example of a twisting bundle $E \to V$ to be used in Theorem 8.3, namely the fiberwise spinor bundle on a warped product.

Remark 10.1 (twisting with the fiberwise spinor bundle) Consider a warped product band

$$(V := M \times [t_-, t_+], g := \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x),$$

and let $\$ \to V$ be the spinor bundle with respect to the metric g. Then we let $E_0 := \$$ be the same bundle endowed with the same bundle metric but with the "fiberwise spinor connection"

$$\nabla^{E_0}_X := \nabla^{\$}_X + \tfrac{1}{2}\mathrm{c}(\nabla_X \mathrm{d}x)\,\mathrm{c}(\mathrm{d}x) \quad \text{for } X \in \mathrm{T}V.$$

With this connection, each restriction $E_0|_{M\times\{t\}}$ is precisely the spinor bundle of $M\times\{t\}$ (if n is odd) or two copies of it (if n is even); compare also (2-8). Moreover, a direct calculation using the warped product structure shows that the curvature tensor of ∇^{E_0} satisfies $\mathrm{R}^{E_0}_{\partial_x,X}=0$ for any tangent vector $X\in \mathrm{T}V$. In other words, the connection ∇^{E_0} is chosen in such a way that only the vertical directions contribute to its curvature. Consequently, if we form the twisted spinor bundle $\$\otimes E_0$, the corresponding curvature endomorphism from the Bochner–Lichnerowicz–Weitzenböck formula (see equations (2-5) and (2-20)) satisfies

$$\mathcal{R}^{E_0}|_{M\times\{t\}}=\sum_{i,j=1}^{n-1}\mathrm{c}(e^i)\,\mathrm{c}(e^j)\otimes\mathrm{R}^{\nabla^{E_0}}_{e_i,e_j}=\sum_{i,j=1}^{n-1}\mathrm{c}(e^i)\,\mathrm{c}(\mathrm{d}x)\,\mathrm{c}(e^j)\,\mathrm{c}(\mathrm{d}x)\otimes\mathrm{R}^{\nabla^{E_0}}_{e_i,e_j}=\mathcal{R}^{E_0|_{M\times\{t\}}},$$

where $e_1,\dots,e_{n-1}$ is a local orthonormal frame of $\mathrm{T}M$ and $\mathcal{R}^{E_0|_{M\times\{t\}}}$ denotes the Weitzenböck curvature endomorphism on the fiber $M\times\{t\}$ of the twisting bundle $E_0|_{M\times\{t\}}$ (or two copies thereof). Finally, since the metric on $M\times\{t\}$ is simply the constant multiple $\varphi(t)^2g_M$ of the metric g_M, we can identify each restriction $E_0|_{M\times\{t\}}$ with the spinor bundle on M (or two copies of it) and with respect to this identification, we obtain

$$\mathcal{R}^{E_0}|_{M\times\{t\}}=\varphi(t)^{-2}\mathcal{R}^{M,E_0}, \tag{10-1}$$

where $\mathcal{R}^{M,E_0}$ denotes the curvature endomorphism on (M,g_M) associated to using the spinor bundle on M itself as a twisting bundle (or two copies of each).

We now state and prove the main result of this section.

Theorem 10.2 *Let n be odd and let (N,g_N) be an $(n-1)$–dimensional Riemannian spin manifold of nonvanishing Euler-characteristic whose Riemannian curvature operator is nonnegative. Moreover, let $\varphi\colon[t_-,t_+]\to(0,\infty)$ be a strictly logarithmically concave function and consider the warped product metric $g_0=\varphi^2g_N+\mathrm{d}y\otimes\mathrm{d}y$ on $V_0:=N\times[t_-,t_+]$. Let (V,g) be an n–dimensional Riemannian spin band and $\Phi\colon(V,g)\to(V_0,g_0)$ a smooth band map such that*

- (i) *Φ is 1–Lipschitz and of nonzero degree,*
- (ii) $\mathrm{scal}_g\geq\mathrm{scal}_{g_0}\circ\Phi$,
- (iii) $\mathrm{H}_g|_{\partial_\pm V}\geq\mathrm{H}_{g_0}|_{\partial_\pm V_0}=\pm h(t_\pm)$, *where $h=\varphi'/\varphi$.*

Then $\mathrm{scal}_g=\mathrm{scal}_{g_0}\circ\Phi$, and (V,g) is isometric to a warped product

$$(M\times[t_-,t_+],\varphi^2g_M+\mathrm{d}x\otimes\mathrm{d}x),$$

where $x=y\circ\Phi$, $M=x^{-1}(t_0)$ for any $t_0\in[t_-,t_+]$, and $g_M:=\varphi(t_0)^{-2}g|_M$ on M. Moreover, we have $\mathrm{scal}_{g_M}=\mathrm{scal}_{g_N}\circ\Phi|_M$ in this case.

If, furthermore, the metric on N satisfies $\mathrm{Ric}_{g_N}>0$, then Φ is an isometry under the above hypotheses.

Proof The main idea of the proof is to apply the argument of Goette and Semmelmann [16] fiberwise in combination with Theorem 8.3. Since Φ is a 1–Lipschitz band map, the function $x = y \circ \Phi \colon V \to [t_-, t_+]$ is a width function. We need to verify that we are in an instance of Setup 8.1. To this end, we set

$$\kappa := \frac{\mathrm{scal}_{g_0} \circ \Phi}{n(n-1)} \quad \text{and} \quad \nu := -\frac{\mathrm{scal}_{g_N} \circ \mathrm{pr}_N \circ \Phi}{4\varphi(x)^2}, \tag{10-2}$$

where $\mathrm{pr}_N \colon V_0 \to N$ is the projection onto the first factor. It is now a consequence of the discussion in Remark 8.2 that

$$\tfrac{1}{4}n^2\kappa + \frac{n\nu}{n-1} + f(x)^2 - f'(x) = 0, \tag{10-3}$$

where $f := -\frac{1}{2}nh$. In particular, this choice of functions satisfies Setup 8.1.

Now we consider the fiberwise spinor bundle $E_0 \to V_0$ constructed as in Remark 10.1 and let $E := \Phi^* E_0$ be the pullback bundle on V. Now the main estimate of Goette and Semmelmann [16, Section 1.1] together with the description of the Weitzenböck curvature endomorphism of E_0 from (10-1) shows that we precisely have the estimate

$$\mathcal{R}^E \geq \nu.$$

For each $t \in [t_-, t_+]$, the twisted Dirac operator $\not{D}_{N \times \{t\}, E_0|_{N \times \{t\}}}$ is the Euler characteristic operator of N and thus has nontrivial index because $n-1$ is even. Since the degree of Φ is nonzero, it follows that the index of $\not{D}_{\partial_- V, E|_{\partial_- V}}$ is also nonzero. Hence, Corollary 3.10 shows that the Callias operator $\mathcal{B}_{\psi,s}$ considered in Theorem 8.3 has nontrivial index and hence nontrivial kernel. Thus Theorem 8.3 applies and we obtain $\mathrm{scal}_g = \kappa n(n-1)$ and (V, g) is isometric to a warped product

$$(M \times [t_-, t_+], \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x),$$

where $M = x^{-1}(t_0)$ for some arbitrary but fixed $t_0 \in [t_-, t_+]$ and some Riemannian metric g_M on M. Using the warped product structure, we obtain

$$\kappa n(n-1) = \mathrm{scal}_g = \frac{\mathrm{scal}_{g_M}}{\varphi(x)^2} - 2(n-1)h'(x) - n(n-1)h(x)^2.$$

Together with (10-3), this completely determines the scalar curvature of g_M, and we obtain $\mathrm{scal}_{g_M} = \mathrm{scal}_{g_N} \circ \Phi|_M$. This proves the first part of the theorem.

To prove the second part, we first observe that due to $x = y \circ \Phi$, under the isometry $V \cong M \times [t_-, t_+]$, the map Φ is of the form $(p, x) \mapsto (\Phi_x(p), x)$, where $\Phi_t := \Phi|_{M \times \{t\}} \colon M \to N$. Since $\|\mathrm{T}\Phi\| \leq 1$ and $\mathrm{T}\Phi(0, \partial_x) = (\partial \Phi_x / \partial x, \partial_x)$, it follows that $\partial \Phi_x / \partial x = 0$; that is, $\Phi_t = \Phi_{t_0}$ for all t. Moreover, $\Phi_{t_0} \colon M \to N$ is 1–Lipschitz with respect to the metric $\varphi(t_0)^2 g_M$ and $\varphi(t_0)^2 g_N$. Thus the same holds with respect to the metrics g_M and g_N. If we assume that $\mathrm{Ric}_{g_N} > 0$, then the rigidity argument of Goette and Semmelmann [16, Section 1.2] (and note that our Φ_{t_0} is length-nonincreasing), implies that Φ_{t_0} is an isometry. Together with the warped product structure, all of this implies that Φ itself is an isometry. □

Restricting to the special case where Φ is the identify map immediately yields the following corollary. The notions of *scalar-mean extremality* and *scalar-mean rigidity* are defined in the introduction in Section 1.3.

Corollary 10.3 *Let n be odd and (M, g_M) be an $(n-1)$–dimensional Riemannian spin manifold of nonvanishing Euler characteristic, whose Riemannian curvature operator is nonnegative. Let $\varphi\colon [t_-, t_+] \to (0, \infty)$ be a smooth strictly logarithmically concave function and consider the warped product metric $g_V = \varphi^2 g_M + \mathrm{d}x \otimes \mathrm{d}x$ on $V := M \times [t_-, t_+]$. Then any metric g on V which satisfies*

(i) $g \geq g_V$,

(ii) $\mathrm{scal}_g \geq \mathrm{scal}_{g_V}$,

(iii) $\mathrm{H}_g \geq \mathrm{H}_{g_V}$

is itself a warped product $g = \varphi^2 \tilde{g}_M + \mathrm{d}x \otimes \mathrm{d}x$ for some metric $\tilde{g}_M$ on M which satisfies $\mathrm{scal}_{\tilde{g}_M} = \mathrm{scal}_{g_M}$. In particular, g_V is scalar-mean extremal.

If, in addition, the metric g_M satisfies $\mathrm{Ric}_{g_M} > 0$, then g_V is scalar-mean rigid.

In particular, the main theorem and corollary of this section are fully applicable to strictly log-concave warped products over even-dimensional spheres. This corresponds to a fiberwise application of Llarull's result [31]. In the following, we single out one important special class of examples, namely annuli in simply connected space forms.

Indeed, let $\kappa \in \mathbb{R}$ be fixed and (M_κ, g_κ) be the n–dimensional simply connected space form of constant sectional curvature κ. To apply the theorem, we recall the description of (M_κ, g_κ) as a warped product over the sphere. Choose a basepoint $p_0 \in M_\kappa$. Let sn_κ be the unique solution to the initial value problem $\varphi'' + \kappa\varphi = 0$ with $\varphi(0) = 0$ and $\varphi'(0) = 1$. Similarly, cs_κ denotes the unique solution to the same differential equation but with initial values $\varphi(0) = 1$ and $\varphi'(0) = 0$. If $\kappa > 0$, we let $p_\infty \in M_\kappa$ be the point opposite to p_0 and set $M'_\kappa := M_\kappa \setminus \{p_0, p_\infty\}$. If $\kappa \leq 0$, we let $M'_\kappa = M_\kappa \setminus \{p_0\}$. On $M'_\kappa \cong \mathrm{S}^{n-1} \times (0, t_\infty)$ the metric g_κ appears as the warped product

$$g_\kappa = \mathrm{sn}_\kappa^2\, g_{\mathrm{S}^{n-1}} + \mathrm{d}x \otimes \mathrm{d}x,$$

where t_∞ is chosen so that $I := (0, t_\infty)$ is a maximal interval on which sn_κ remains positive. This means $t_\infty = +\infty$ for $\kappa \leq 0$ and $t_\infty = \pi/\sqrt{\kappa}$ for $\kappa > 0$. Moreover, $\log(\mathrm{sn}_\kappa)'' = -1/\mathrm{sn}_\kappa^2 < 0$; that is, sn_κ is strictly logarithmically concave. This means that Theorem 10.2 is applicable to the metric g_κ. Given $0 < t_- < t_+ < t_\infty$, we consider the annulus

$$\mathrm{A}_{t_-,t_+} := \{p \in M_\kappa \mid t_- \leq d_{g_\kappa}(p, p_0) \leq t_+\} \subset M_\kappa.$$

We will view A_{t_-,t_+} as a band with $\partial_\pm \mathrm{A}_{t_-,t_+} = \mathrm{S}_{t_\pm} := \{p \in M_\kappa \mid d_{g_\kappa}(p, p_0) = t_\pm\}$. Furthermore, we set $\mathrm{ct}_\kappa = \mathrm{cs}_\kappa/\mathrm{sn}_\kappa$. Then the mean curvature of $\partial_\pm \mathrm{A}_{t_-,t_+}$ is equal to $\pm\mathrm{ct}_\kappa(t_\pm)$. We thus deduce the following consequences of Theorem 10.2.

Corollary 10.4 *Let $n \geq 3$ be odd and (M_κ, g_κ) be the n–dimensional simply connected space form of constant sectional curvature $\kappa \in \mathbb{R}$. Let $0 < t_- < t_+ < t_\infty$ and consider an annulus A_{t_-,t_+} as above. Let (V, g) be an n–dimensional spin band and let $\Phi\colon V \to \mathrm{A}_{t_-,t_+}$ be a smooth band map such that*

(i) *Φ is 1–Lipschitz and of nonzero degree,*

(ii) $\mathrm{scal}_g \geq \mathrm{scal}_{g_\kappa} = \kappa n(n-1)$,

(iii) $\mathrm{H}_g|_{\partial_\pm V} \geq \mathrm{H}_{g_\kappa}|_{\partial_\pm \mathrm{A}_{t_-,t_+}} = \pm \mathrm{ct}_\kappa(t_\pm)$.

Then Φ is an isometry.

Corollary 10.5 *Let $n \geq 3$ be odd and (M_κ, g_κ) the n–dimensional simply connected space form of constant sectional curvature $\kappa \in \mathbb{R}$. Let $0 < t_- < t_+ < t_\infty$ and consider the annulus*

$$\mathrm{A}_{t_-,t_+} := \{p \in M_\kappa \mid t_- \leq d_{g_\kappa}(p, p_0) \leq t_+\}$$

around some basepoint $p_0 \in M_\kappa$. Then any Riemannian metric g on A_{t_-,t_+} which satisfies

(i) $g \geq g_\kappa$,

(ii) $\mathrm{scal}_g \geq \mathrm{scal}_{g_\kappa} = \kappa n(n-1)$, *and*

(iii) $\mathrm{H}_g|_{\mathrm{S}_{t_\pm}} \geq \mathrm{H}_{g_\kappa}|_{\mathrm{S}_{t_\pm}} = \pm \mathrm{ct}_\kappa(t_\pm)$

is equal to g_κ. That is, g_κ is scalar-mean rigid on A_{t_-,t_+}.

References

[1] **L Andersson**, **M Dahl**, *Scalar curvature rigidity for asymptotically locally hyperbolic manifolds*, Ann. Global Anal. Geom. 16 (1998) 1–27 MR Zbl

[2] **N Anghel**, *On the index of Callias-type operators*, Geom. Funct. Anal. 3 (1993) 431–438 MR Zbl

[3] **C Bär**, **W Ballmann**, *Boundary value problems for elliptic differential operators of first order*, from "Surveys in differential geometry, XVII" (H-D Cao, S-T Yau, editors), Surv. Differ. Geom. 17, International, Boston, MA (2012) 1–78 MR Zbl

[4] **C Bär**, **W Ballmann**, *Guide to elliptic boundary value problems for Dirac-type operators*, from "Arbeitstagung Bonn 2013" (W Ballmann, C Blohmann, G Faltings, P Teichner, D Zagier, editors), Progr. Math. 319, Springer (2016) 43–80 MR Zbl

[5] **C Bär**, **B Hanke**, *Boundary conditions for scalar curvature*, from "Perspectives in scalar curvature, II" (M L Gromov, J Lawson, H Blaine, editors), World Sci., Hackensack, NJ (2023) 325–377 MR Zbl

[6] **B Booß-Bavnbek**, **K P Wojciechowski**, *Elliptic boundary problems for Dirac operators*, Birkhäuser, Boston, MA (1993) MR Zbl

[7] **B Botvinnik**, **J Ebert**, **O Randal-Williams**, *Infinite loop spaces and positive scalar curvature*, Invent. Math. 209 (2017) 749–835 MR Zbl

[8] **J-P Bourguignon**, **O Hijazi**, **J-L Milhorat**, **A Moroianu**, **S Moroianu**, *A spinorial approach to Riemannian and conformal geometry*, Eur. Math. Soc., Zürich (2015) MR Zbl

[9] **C Callias**, *Axial anomalies and index theorems on open spaces*, Comm. Math. Phys. 62 (1978) 213–234 MR Zbl

[10] **S Cecchini**, *A long neck principle for Riemannian spin manifolds with positive scalar curvature*, Geom. Funct. Anal. 30 (2020) 1183–1223 MR Zbl

[11] **S Cecchini**, **R Zeidler**, *Scalar curvature and generalized Callias operators*, from "Perspectives in scalar curvature, I" (M L Gromov, J Lawson, H Blaine, editors), World Sci., Hackensack, NJ (2023) 515–542 MR Zbl

[12] **O Chodosh**, **C Li**, *Generalized soap bubbles and the topology of manifolds with positive scalar curvature*, preprint (2020) arXiv 2008.11888

[13] **P T Chruściel**, **M Herzlich**, *The mass of asymptotically hyperbolic Riemannian manifolds*, Pacific J. Math. 212 (2003) 231–264 MR Zbl

[14] **D S Freed**, *Two index theorems in odd dimensions*, Comm. Anal. Geom. 6 (1998) 317–329 MR Zbl

[15] **T Friedrich**, *Der erste Eigenwert des Dirac-Operators einer kompakten, Riemannschen Mannigfaltigkeit nichtnegativer Skalarkrümmung*, Math. Nachr. 97 (1980) 117–146 MR Zbl

[16] **S Goette**, **U Semmelmann**, *Scalar curvature estimates for compact symmetric spaces*, Differential Geom. Appl. 16 (2002) 65–78 MR Zbl

[17] **M Gromov**, *Positive curvature, macroscopic dimension, spectral gaps and higher signatures*, from "Functional analysis on the eve of the 21st century, II" (S Gindikin, J Lepowsky, R L Wilson, editors), Progr. Math. 132, Birkhäuser, Boston, MA (1996) 1–213 MR Zbl

[18] **M Gromov**, *Metric inequalities with scalar curvature*, Geom. Funct. Anal. 28 (2018) 645–726 MR Zbl

[19] **M Gromov**, *Four lectures on scalar curvature*, preprint (2020) arXiv 1908.10612v4

[20] **M Gromov**, *No metrics with positive scalar curvatures on aspherical 5–manifolds*, preprint (2020) arXiv 2009.05332

[21] **M Gromov**, **H B Lawson, Jr**, *The classification of simply connected manifolds of positive scalar curvature*, Ann. of Math. 111 (1980) 423–434 MR Zbl

[22] **M Gromov**, **H B Lawson, Jr**, *Spin and scalar curvature in the presence of a fundamental group, I*, Ann. of Math. 111 (1980) 209–230 MR Zbl

[23] **M Gromov**, **H B Lawson, Jr**, *Positive scalar curvature and the Dirac operator on complete Riemannian manifolds*, Publ. Math. Inst. Hautes Études Sci. 58 (1983) 83–196 MR Zbl

[24] **H Guo**, **Z Xie**, **G Yu**, *Quantitative K–theory, positive scalar curvature, and bandwidth*, from "Perspectives in scalar curvature, II" (M L Gromov, J Lawson, H Blaine, editors), World Sci., Hackensack, NJ (2023) 763–798 MR Zbl

[25] **B Hanke**, **T Schick**, *Enlargeability and index theory*, J. Differential Geom. 74 (2006) 293–320 MR Zbl

[26] **N Higson**, **J Roe**, *Analytic K–homology*, Oxford Univ. Press (2000) MR Zbl

[27] **O Hijazi**, **S Montiel**, **A Roldán**, *Eigenvalue boundary problems for the Dirac operator*, Comm. Math. Phys. 231 (2002) 375–390 MR Zbl

[28] **H B Lawson, Jr**, **M-L Michelsohn**, *Spin geometry*, Princeton Mathematical Series 38, Princeton Univ. Press (1989) MR Zbl

[29] **M Lesourd**, **R Unger**, **S-T Yau**, *The positive mass theorem with arbitrary ends*, preprint (2021) arXiv 2103.02744

[30] **A Lichnerowicz**, *Spineurs harmoniques*, C. R. Acad. Sci. Paris 257 (1963) 7–9 MR Zbl

[31] **M Llarull**, *Sharp estimates and the Dirac operator*, Math. Ann. 310 (1998) 55–71 MR Zbl

[32] **J Lott**, *Index theory for scalar curvature on manifolds with boundary*, Proc. Amer. Math. Soc. 149 (2021) 4451–4459 MR Zbl

[33] **M Min-Oo**, *Scalar curvature rigidity of asymptotically hyperbolic spin manifolds*, Math. Ann. 285 (1989) 527–539 MR Zbl

[34] **P Petersen**, *Riemannian geometry*, 3rd edition, Graduate Texts in Math. 171, Springer (2016) MR Zbl

[35] **H Rademacher**, *Über partielle und totale differenzierbarkeit von Funktionen mehrerer Variabeln und über die Transformation der Doppelintegrale*, Math. Ann. 79 (1919) 340–359 MR Zbl

[36] **S Raulot**, *Optimal eigenvalues estimate for the Dirac operator on domains with boundary*, Lett. Math. Phys. 73 (2005) 135–145 MR Zbl

[37] **A Sakovich**, **C Sormani**, *Almost rigidity of the positive mass theorem for asymptotically hyperbolic manifolds with spherical symmetry*, Gen. Relativity Gravitation 49 (2017) art. id. 125 MR Zbl

[38] **R Schoen**, **S T Yau**, *On the structure of manifolds with positive scalar curvature*, Manuscripta Math. 28 (1979) 159–183 MR Zbl

[39] **S Stolz**, *Simply connected manifolds of positive scalar curvature*, Ann. of Math. 136 (1992) 511–540 MR Zbl

[40] **M E Taylor**, *Partial differential equations, I: Basic theory*, 2nd edition, Applied Mathematical Sciences 115, Springer (2011) MR Zbl

[41] **E Witten**, *A new proof of the positive energy theorem*, Comm. Math. Phys. 80 (1981) 381–402 MR Zbl

[42] **R Zeidler**, *Width, largeness and index theory*, Symmetry Integrability Geom. Methods Appl. 16 (2020) art. id. 127 MR Zbl

[43] **R Zeidler**, *Band width estimates via the Dirac operator*, J. Differential Geom. 122 (2022) 155–183 MR Zbl

[44] **W Zhang**, *Nonnegative scalar curvature and area decreasing maps*, Symmetry Integrability Geom. Methods Appl. 16 (2020) art. id. 033 MR Zbl

[45] **J Zhu**, *Rigidity results for complete manifolds with nonnegative scalar curvature*, preprint (2020) arXiv 2008.07028

[46] **J Zhu**, *Width estimate and doubly warped product*, Trans. Amer. Math. Soc. 374 (2021) 1497–1511 MR Zbl

Department of Mathematics, Texas A&M University
College Station, TX, United States

Mathematical Institute, University of Münster
Münster, Germany

cecchini@tamu.edu, math@rzeidler.eu

https://simonececchini.org, https://www.rzeidler.eu

Proposed: John Lott Received: 28 July 2021
Seconded: Bruce Kleiner, Tobias H Colding Revised: 14 March 2022

Geometry & Topology **28**:3 (2024) 1213–1285
DOI: 10.2140/gt.2024.28.1213
Published: 10 May 2024

Symplectic capacities, unperturbed curves and convex toric domains

DUSA MCDUFF
KYLER SIEGEL

We use explicit pseudoholomorphic curve techniques (without virtual perturbations) to define a sequence of symplectic capacities analogous to those defined recently by the second author using symplectic field theory. We then compute these capacities for all four-dimensional convex toric domains. This gives various new obstructions to stabilized symplectic embedding problems, which are sometimes sharp.

53D35, 57K43

1 Introduction

1.1 Overview

Symplectic capacities have long played an important role in symplectic geometry, providing a systematic tool for studying nonsqueezing phenomena. Let us mention here just two prominent sequences of symplectic capacities: the Ekeland–Hofer capacities [12; 13] and the embedded contact homology (ECH) capacities of Hutchings [20]. The former are defined in any dimension and they provide obstructions which can be viewed as refinements of Gromov's celebrated nonsqueezing theorem [14]. The latter are defined only in dimension four, but they often give very strong obstructions, eg they give sharp obstructions for symplectic embeddings between four-dimensional ellipsoids.

Higher-dimensional symplectic embeddings remain rather poorly understood, but there has been considerable recent interest in so-called "stabilized symplectic embedding problems", in which one studies symplectic embeddings of the form $X \times \mathbb{C}^N \overset{s}{\hookrightarrow} X' \times \mathbb{C}^N$ for four-dimensional Liouville domains X and X', and for $N \in \mathbb{Z}_{\geq 1}$; see for instance Hind and Kerman [17], Cristofaro-Gardiner and Hind [9], Cristofaro-Gardiner, Hind and McDuff [10], McDuff [29], Siegel [37; 38] and Irvine [26]. In order to systematize and generalize these results, the second author introduced in [37] a sequence of symplectic capacities $\mathfrak{g}_1, \mathfrak{g}_2, \mathfrak{g}_3, \ldots$ which are "stable" in the sense that $\mathfrak{g}_k(X \times \mathbb{C}^N) = \mathfrak{g}_k(X)$ for any Liouville domain X and $k, N \in \mathbb{Z}_{\geq 1}$. These capacities are defined using symplectic field theory (SFT), more specifically the (chain level) filtered $\mathscr{L}_\infty$ structure on linearized contact homology, and their definition also involves curves satisfying local tangency constraints. As a proof of concept, [37] shows that these capacities perform quite well in toy problems, for instance they recover the sharp obstructions from [29] and they often outperform the Ekeland–Hofer capacities. In fact, the capacities $\mathfrak{g}_1, \mathfrak{g}_2, \mathfrak{g}_3, \ldots$ are a specialization of a more general family of capacities $\{\mathfrak{g}_\mathfrak{b}\}$ which are expected to give sharp obstructions to the stabilized ellipsoid embedding problem.

However, two broad questions naturally become apparent:

(1) What is the role of symplectic field theory? Namely, it is known that SFT typically requires virtually perturbing moduli spaces of pseudoholomorphic curves, and yet ultimately all of the data of $\mathfrak{g}_k(X)$ should be carried by honest pseudoholomorphic curves in $\widehat{X}$ and $\mathbb{R} \times \partial X$, so does one really need the full SFT package?[1]

(2) How does one actually compute $\mathfrak{g}_1, \mathfrak{g}_2, \mathfrak{g}_3, \ldots$ for Liouville domains of interest? Note that even computing $\mathfrak{g}_k$ for a four-dimensional ellipsoid is a nontrivial problem.

Note that these questions are coupled, since a concrete answer to (1) could open up new direct avenues for computations as in (2).

The primary purpose of this paper is to address both of these questions. In short:

(1) We give an ersatz definition of $\mathfrak{g}_k$, denoted by $\widetilde{\mathfrak{g}}_k$, which is simple and explicit and does not require any virtual perturbations.

(2) We compute (or at least reduce to elementary combinatorics) $\widetilde{\mathfrak{g}}_k$ for all four-dimensional convex toric domains. This gives a large family of examples, which includes ellipsoids and polydisks as special cases.

Combining these, one can directly extract many new symplectic embedding obstructions. As an illustration, the recent work of Cristofaro-Gardiner, Hind and Siegel [11] applies our computations for ellipsoids and polydisks in order to obstruct various stabilized symplectic embeddings between these. Remarkably, these obstructions are often sharp, at least when certain aspect ratios are integral; see Example 1.3.3 and Remark 1.3.5.

1.2 Statement of main results

We now describe our results in more detail. In Section 3, we define the capacity $\widetilde{\mathfrak{g}}_k(M)$ for all symplectic manifolds M and $k \in \mathbb{Z}_{\geq 1}$. Roughly, if X is a Liouville domain with nondegenerate contact boundary, then $\widetilde{\mathfrak{g}}_k(X)$ is the maximum over all suitable almost complex structures J of the minimum energy of any asymptotically cylindrical rational J–holomorphic curve in $\widehat{X}$ which satisfies a local tangency constraint $\lll \mathcal{T}^{(k)} p \ggg$. The latter means that the curve has contact order k (or equivalently tangency order $k-1$) to a chosen local divisor D defined near a point $p \in X$. Note that we do not require the curves entering into the definition of $\widetilde{\mathfrak{g}}_k(X)$ to be regular or even index zero. This definition of $\widetilde{\mathfrak{g}}_k(X)$ is extended to $\widetilde{\mathfrak{g}}_k(M)$ for M an arbitrary symplectic manifold by taking a supremum over all Liouville domains which symplectically embed into M.[2]

[1] As outlined in [37, Section 1], we also expect an alternative definition of $\mathfrak{g}_k$ using (S^1–equivariant) Floer theory instead of symplectic field theory. Since this involves Hamiltonian perturbations and many associated choices, it is also quite difficult to compute directly from the definition.

[2] After a first draft of this paper was completed, the authors learned from G Mikhalkin about independent work defining a similar capacity directly for all symplectic manifolds using an even broader class of almost complex structures and pseudoholomorphic curves. It seems likely that these two definitions are equivalent, but they may have slightly different realms of utility.

Remark 1.2.1 In the special case of the first capacity $\tilde{\mathfrak{g}}_1$, our definition essentially coincides with Gromov's original definition [15, Section 4.1] of "symplectic width" via a maxi-min procedure.

The following summarizes some of the key properties of $\tilde{\mathfrak{g}}_k$:

Theorem 1.2.2 *For each $k \in \mathbb{Z}_{\geq 1}$, the capacity $\tilde{\mathfrak{g}}_k$ is independent of the choice of local divisor and is a symplectomorphism invariant. It satisfies the following properties*:

- ***Scaling*** *It scales like area, ie $\tilde{\mathfrak{g}}_k(M, \mu\omega) = \mu\tilde{\mathfrak{g}}_k(M,\omega)$ for any symplectic manifold (M,ω) and $\mu \in \mathbb{R}_{>0}$.*
- ***Nondecreasing*** *We have $\tilde{\mathfrak{g}}_1(M) \leq \tilde{\mathfrak{g}}_2(M) \leq \tilde{\mathfrak{g}}_3(M) \leq \cdots$ for any symplectic manifold M.*
- ***Subadditivity*** *We have $\tilde{\mathfrak{g}}_{i+j}(M) \leq \tilde{\mathfrak{g}}_i(M) + \tilde{\mathfrak{g}}_j(M)$ for any $i, j \in \mathbb{Z}_{\geq 1}$.*
- ***Symplectic embedding monotonicity*** *It is monotone under equidimensional symplectic embeddings, ie $M \overset{s}{\hookrightarrow} M'$ implies $\tilde{\mathfrak{g}}_k(M) \leq \tilde{\mathfrak{g}}_k(M')$ for any symplectic manifolds M and M'.*
- ***Closed curve upper bound*** *If (M,ω) is a closed semipositive symplectic manifold satisfying $N_{M,A}\langle\!\langle \mathcal{T}^{(k)}p \rangle\!\rangle \neq 0$ for some $A \in H_2(M)$, then we have $\tilde{\mathfrak{g}}_k(M) \leq [\omega]\cdot A$.*
- ***Stabilization*** *For any Liouville domain X we have $\tilde{\mathfrak{g}}_k(X \times B^2(c)) = \tilde{\mathfrak{g}}_k(X)$ for any $c \geq \tilde{\mathfrak{g}}_k(X)$, provided that the hypotheses of Proposition 3.7.1 are satisfied.* (This holds, for instance, for X any four-dimensional convex toric domain.)

In the penultimate point, $N_{M,A}\langle\!\langle \mathcal{T}^{(k)}p \rangle\!\rangle$ denotes the Gromov–Witten type invariant which counts closed rational pseudoholomorphic curves in M in homology class A satisfying the local tangency constraint $\langle\!\langle \mathcal{T}^{(k)}p \rangle\!\rangle$, as defined in our paper [30]. Also, $B^2(c)$ denotes the closed two-ball of area c (ie radius $\sqrt{c/\pi}$), equipped with its standard symplectic form. For more detailed explanations and proofs, see Sections 2 and 3.

Remark 1.2.3 (stabilization hypotheses) The hypotheses of Proposition 3.7.1 roughly amount to the assumption that $\tilde{\mathfrak{g}}_k(X)$ is represented by a moduli space of curves which is sufficiently robust that it cannot degenerate in generic one-parameter families. When this holds, we can iteratively stabilize to obtain $\tilde{\mathfrak{g}}_k(X \times B^2(c) \times \cdots \times B^2(c)) = \tilde{\mathfrak{g}}_k(X)$ for $c \geq \tilde{\mathfrak{g}}_k(X)$, and in particular we have $\tilde{\mathfrak{g}}_k(X \times \mathbb{C}^N) = \tilde{\mathfrak{g}}_k(X)$ for $N \in \mathbb{Z}_{\geq 1}$. Compared with $\mathfrak{g}_k$, the extra hypotheses in the stabilization property is one place where we "pay the price" for such a simple definition of $\tilde{\mathfrak{g}}_k$, although we do not know whether the extra hypotheses are truly essential.

Remark 1.2.4 (relationship with $\mathfrak{g}_k$) As we explain in Section 3.4, we must have $\tilde{\mathfrak{g}}_k(X) = \mathfrak{g}_k(X)$ whenever X is a Liouville domain satisfying the hypotheses of Proposition 3.7.1. In particular, this is the case for all four-dimensional convex toric domains, and we are not aware of any examples with $\tilde{\mathfrak{g}}_k(X) \neq \mathfrak{g}_k(X)$.

Remark 1.2.5 (relationship with Gutt–Hutchings capacities) In Section 3.1, we define (following [37]) a refined family of capacities $\widetilde{\mathfrak{g}}_k^{\le l}$ for $k, l \in \mathbb{Z}_{\ge 1}$, using the same prescription as for $\widetilde{\mathfrak{g}}_k$ except that we now only allow curves having at most l positive ends. Note that the case $l = \infty$ recovers $\widetilde{\mathfrak{g}}_k = \widetilde{\mathfrak{g}}_k^{\le\infty}$. The capacities $\{\widetilde{\mathfrak{g}}_k^{\le l}\}$ satisfy most of the properties in Theorem 1.2.2, except that the closed curve upper bound no longer holds, and monotonicity for $\widetilde{\mathfrak{g}}_k^{\le l}$ only holds for *generalized Liouville embeddings*, ie smooth embeddings $\iota\colon (X,\lambda) \hookrightarrow (X',\lambda')$ of equidimensional Liouville domains such that the closed 1–form $(\iota^*(\lambda') - \lambda)|_{\partial X}$ is exact; cf Gutt and Hutchings [16, Section 1.4]. In Section 5.6 we show that, at least for four-dimensional convex toric domains, the $l = 1$ specialization $\widetilde{\mathfrak{g}}_k^{\le 1}$ coincides with the k^{th} Gutt–Hutchings capacity c_k^{GH} from [16]. The latter is in turn known to agree with the k^{th} Ekeland–Hofer capacity c_k^{EH} in all examples where both are computed, eg ellipsoids and polydisks.

Remark 1.2.6 (nondecreasing property) Curiously, for the analogous SFT capacities the nondecreasing property $\mathfrak{g}_1 \le \mathfrak{g}_2 \le \mathfrak{g}_3 \le \cdots$ is not at all obvious from the definition.

Remark 1.2.7 (generalizations) The approach taken in this paper to define $\{\widetilde{\mathfrak{g}}_k\}$ naturally generalizes to define various other families of capacities, eg by replacing the local tangency constraint $\lessdot\mathcal{T}^{(k)}p\gtrdot$ with k generic point constraints, and/or by allowing curves of higher genus. In this spirit, the very recent preprint of Hutchings [23] adapts our approach to define (without relying on Seiberg–Witten theory) a sequence of four-dimensional capacities, which agree in many cases with the ECH capacities.

With the capacities $\widetilde{\mathfrak{g}}_1, \widetilde{\mathfrak{g}}_2, \widetilde{\mathfrak{g}}_3, \dots$ at hand, we turn to computations. Given a compact convex domain $\Omega \subset \mathbb{R}^n$, put $X_\Omega := \mu^{-1}(\Omega)$, where $\mu\colon \mathbb{C}^n \to \mathbb{R}^n_{\ge 0}$ is given by

$$\mu(z_1,\dots,z_n) = (\pi|z_1|^2,\dots,\pi|z_n|^2).$$

Define $\|{-}\|^*_\Omega\colon \mathbb{R}^n \to \mathbb{R}$ by $\|\vec{v}\|^*_\Omega := \max_{\vec{w}\in\Omega}\langle\vec{v},\vec{w}\rangle$, where $\langle -,- \rangle$ denotes the standard dot product. Note that if $\partial\Omega$ is smooth, then the maximizer $\vec{w}$ lies in $\partial\Omega$ and is such that the hyperplane through $\vec{w}$ normal to $\vec{v}$ is tangent to $\partial\Omega$. If Ω contains the origin in its interior, then $\|{-}\|^*_\Omega$ is a (nonsymmetric) norm, dual to the norm having Ω as its unit ball. Otherwise, $\|{-}\|^*_\Omega$ is not generally nondegenerate or even nonnegative, although it is still convenient to treat it like a norm. Recall that X_Ω is a "convex toric domain" if the symmetrization of Ω about the axes is itself convex; see Section 4.1 for more details.

Theorem 1.2.8 *Let X_Ω be a four-dimensional convex toric domain. For $k \in \mathbb{Z}_{\ge 1}$, we have*

$$\widetilde{\mathfrak{g}}_k(X_\Omega) = \min \sum_{s=1}^{q} \|(i_s, j_s)\|^*_\Omega, \tag{1-2-1}$$

where the minimization is over all $(i_1,j_1),\dots,(i_q,j_q) \in \mathbb{Z}^2_{\ge 0} \setminus \{(0,0)\}$ such that

- $\sum_{s=1}^{q}(i_s + j_s) + q - 1 = k$, *and*
- *if $q \ge 2$, then $(i_1,\dots,i_q) \ne (0,\dots,0)$ and $(j_1,\dots,j_q) \ne (0,\dots,0)$.*

Using results from Section 4, we have the following appealing reformulation, which we prove at the end of Section 4.3. If $P \subset \mathbb{R}^2$ is a convex lattice polygon, ie a convex polygon such that each vertex lies at an integer lattice point, let $\ell_\Omega(\partial P)$ denote the length of its boundary as measured by $\|-\|_\Omega^*$, and let $|\partial P \cap \mathbb{Z}^2|$ denote the number of lattice points along the boundary. Here we allow the degenerate case where P is a line segment, in which case by definition $\partial P = P$. Note that $\ell_\Omega(\partial P)$ is unaffected if we translate Ω so that it contains the origin in its interior, after which $\|-\|_\Omega^*$ becomes nondegenerate.

Corollary 1.2.9 *For X_Ω a four-dimensional convex toric domain and $k \in \mathbb{Z}_{\geq 1}$, we have:*

(1-2-2) $\widetilde{\mathfrak{g}}_k(X_\Omega) = \min\{\ell_\Omega(\partial P) \mid P \subset \mathbb{R}^2 \text{ is a convex lattice polygon such that } |\partial P \cap \mathbb{Z}^2| = k+1\}.$

Remark 1.2.10 (i) The k^{th} ECH capacity $c_k^{\text{ECH}}(X_\Omega)$ is given by the exact same formula except that we replace $|\partial P \cap \mathbb{Z}^2|$ with $|P \cap \mathbb{Z}^2|$, ie the number of lattice points in both the interior and boundary of P; see Hutchings [20]. Under the correspondence between lattice polygons and generators, $|\partial P \cap \mathbb{Z}^2|$ corresponds to the (half) Fredholm index, whereas $|P \cap \mathbb{Z}^2|$ corresponds to the (half) ECH index. It is interesting to ask whether Corollary 1.2.9 holds for more general domains $\Omega \subset \mathbb{R}^2$. One can also ask about extensions to higher dimensions, with lattice polygons in $\mathbb{R}^2$ replaced by lattice paths in $\mathbb{R}^n$.

(ii) Corollary 1.2.9 involves arbitrary lattice points, whereas Theorem 1.2.8 involves only nonnegative ones. Conceptually this mirrors the fact that X_Ω has the same values for $\widetilde{\mathfrak{g}}_k$ as its associated "free toric domain" $\mathbb{T}^2 \times \Omega$, thanks to the "Traynor trick"; see eg Landry, McMillan and Tsukerman [28].

(iii) Closely related formulas appear in the recent work of Chaidez and Wormleighton [5]. In particular, [5, Corollary 1] computes $\mathfrak{g}_k(X_\Omega)$ under the additional assumption that the lengths of Ω along the x– and y–axes agree, which holds for instance if X is the round ball $B^4(c)$ or the cube $B^2(c) \times B^2(c)$. Whereas our upper bounds come from curves constructed via the ECH cobordism map and iterated obstruction bundle gluing (see Section 5), the upper bounds in [5] come from cocharacter curves in (possibly singular) closed toric surfaces.

(iv) The second author's work [38] offers another combinatorial computation of $\mathfrak{g}_k(X_\Omega)$ for any four-dimensional convex toric domain X_Ω, and in fact it also computes the full family of capacities $\{\mathfrak{g}_{\mathfrak{b}}(X_\Omega)\}$. However, since that framework involves a nontrivial recursive algorithm, it is not clear how to use it to extract the above formulas.

1.3 Examples and applications

In Section 4.3 we significantly simplify the combinatorial optimization problem involved in Theorem 1.2.8 by showing that there are only a few possibilities for the minimizers. Indeed, Corollary 4.3.9 implies the following simplification of Theorem 1.2.8:

Corollary 1.3.1 *Let X_Ω be a four-dimensional convex toric domain as in Theorem 1.2.8, and assume that Ω has sides of length a and b along the x– and y–axes, respectively, with $a \geq b$. Then there is a minimizer $(i_1, j_1), \dots, (i_q, j_q)$ taking one of the following forms:*

(i) $(0,1)^{\times i} \times (1,1)^{\times j}$ *for* $i \geq 0$ *and* $j \geq 1$.

(ii) $(0,1)^{\times i} \times (1,s)$ *for* $i \geq 0$ *and* $s \geq 2$.

(iii) $(0,1)^{\times i} \times (1,0)$ *for* $i \geq 1$.

(iv) $(0,s)$ *for* $s \geq 1$.

This formulation is particularly useful for extracting closed-form expressions for $\tilde{\mathfrak{g}}_k$ in various families of examples, as in the following results.

Let

$$E(a_1,a_2) := \left\{(z_1,z_2) \in \mathbb{C}^2 \;\middle|\; \frac{1}{a_1}\pi|z_1|^2 + \frac{1}{a_2}\pi|z_2|^2 \leq 1\right\}$$

denote the ellipsoid with area factors a_1, a_2. Up to scaling and symplectomorphism, we can assume that $a_2 = 1$ and $a_1 \geq 1$.

Theorem 1.3.2 (i) *For* $1 \leq a \leq \frac{3}{2}$, *we have*

$$\tilde{\mathfrak{g}}_k(E(a,1)) = \begin{cases} 1 + ia & \text{for } k = 1 + 3i \text{ with } i \geq 0, \\ a + ia & \text{for } k = 2 + 3i \text{ with } i \geq 0, \\ 2 + ia & \text{for } k = 3 + 3i \text{ with } i \geq 0. \end{cases} \tag{1-3-1}$$

(ii) *For* $a > \frac{3}{2}$, *we have*

$$\tilde{\mathfrak{g}}_k(E(a,1)) = \begin{cases} k & \text{for } 1 \leq k \leq \lfloor a \rfloor, \\ a + i & \text{for } k = \lceil a \rceil + 2i \text{ with } i \geq 0, \\ \lceil a \rceil + i & \text{for } k = \lceil a \rceil + 2i + 1 \text{ with } i \geq 0. \end{cases} \tag{1-3-2}$$

Example 1.3.3 We illustrate Theorem 1.3.2 with a simple embedding example which is a special case of [11, Theorem 1.1]. The first few $\tilde{\mathfrak{g}}_k$ capacities are:

k	1	2	3	4	5	6	7	8	9	10	11	12
$\tilde{\mathfrak{g}}_k(E(1,7))$	1	2	3	4	5	6	7	7	8	8	9	9
$\tilde{\mathfrak{g}}_k(E(1,2))$	1	2	2	3	3	4	4	5	5	6	6	7

This gives a lower bound for stabilized embeddings $E(1,7) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \mu \cdot E(1,2) \times \mathbb{C}^N$ (with $N \geq 1$) of $\mu \geq \frac{7}{4}$. By [11, Corollary 3.4] this is optimal, ie there exists a stabilized symplectic embedding realizing this lower bound. In particular, this outperforms the Gutt–Hutchings (or Ekeland–Hofer) capacities, the first few of which are:

k	1	2	3	4	5	6	7	8	9	10	11	12
$c_k^{\mathrm{GH}}(E(1,7))$	1	2	3	4	5	6	7	7	8	9	10	11
$c_k^{\mathrm{GH}}(E(1,2))$	1	2	2	3	4	4	5	6	6	7	8	8

In fact the best bound obtained by the full infinite sequence is $\mu \geq \frac{3}{2}$. By contrast, the ECH capacities give a stronger lower bound, which evidently cannot stabilize. Indeed, we have

k	1	2	3	4	5	6	7	8	9	10	11	12
$c_k^{\mathrm{ECH}}(E(1,7))$	1	2	3	4	5	6	7	7	8	8	9	9
$c_k^{\mathrm{ECH}}(E(1,2))$	1	2	2	3	3	4	4	4	5	5	5	6

giving the lower bound $\mu \geq \frac{9}{5} > \frac{7}{4}$ for the unstabilized problem $E(1,7) \overset{s}{\hookrightarrow} \mu \cdot E(1,2)$. Note that the volume bound is $\mu \geq \sqrt{7/2} \approx 1.87 > \frac{9}{5}$, and this is necessarily recovered by the full sequence of ECH capacities since these are known to give sharp obstructions for four-dimensional ellipsoid embeddings (and also their asymptotics recover the volume).

Now let $P(a_1,a_2) := B^2(a_1) \times B^2(a_2)$ denote the polydisk with area factors a_1 and a_2. Again, without loss of generality we can assume $a_2 = 1$ and $a_1 \geq 1$.

Theorem 1.3.4 *For $k \in \mathbb{Z}_{\geq 1}$ and $a \geq 1$ we have*

$$\widetilde{\mathfrak{g}}_k(P(a,1)) = \min\big(k, a + \lceil \tfrac{1}{2}(k-1) \rceil\big). \tag{1-3-3}$$

Remark 1.3.5 (sharp obstructions) Example 1.3.3 generalizes as follows. By complementing Theorem 1.3.2 with explicit embedding constructions, [11, Theorem 1.1] shows that the capacities $\{\widetilde{\mathfrak{g}}_k\}$ are sharp for embeddings of the form $E(a,1) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \mu \cdot E(b,1) \times \mathbb{C}^N$ with $a \geq b+1 \geq 3$ integers of the opposite parity and $\mu \in \mathbb{R}_{>0}$, $N \in \mathbb{Z}_{\geq 1}$, and such an embedding exists if and only if $\mu \geq 2a/(a+b-1)$. Similarly, [11, Theorem 1.3] shows that the capacities $\{\widetilde{\mathfrak{g}}_k\}$ are sharp for embeddings of the form

$$E(a,1) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \mu \cdot P(b,1) \times \mathbb{C}^N$$

with $b \in \mathbb{R}_{\geq 1}$ (not necessarily an integer), $a \geq 2b-1$ any odd integer and $\mu \in \mathbb{R}_{>0}$, $N \in \mathbb{Z}_{\geq 1}$, and such an embedding exists if and only if $\mu \geq 2a/(a+2b-1)$.

For embeddings of the form $E(a,1) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \mu \cdot B^4(b) \times \mathbb{C}^N$ with $N \in \mathbb{Z}_{\geq 1}$, it was observed in [37] that the capacities $\{\mathfrak{g}_k\}$ (and hence also $\{\widetilde{\mathfrak{g}}_k\}$ by the results of this paper) give sharp obstructions when $a \in 3\mathbb{Z}_{\geq 1} - 1$. On the other hand, for all other $a \in \mathbb{R}_{>1}$ we do not expect optimal obstructions from the capacities $\{\mathfrak{g}_k\}$, but rather from the full family $\{\mathfrak{g}_{\mathfrak{b}}\}$; see the discussion at the end of [37, Section 6.3]. It is natural to ask whether a "naive" analogue $\{\widetilde{\mathfrak{g}}_{\mathfrak{b}}\}$ could be defined and computed in the spirit of this paper.

Remark 1.3.6 The formulas (1-3-3) also appeared for $\mathfrak{g}_k$ in [37] in the case of odd k, based on a slightly different computational framework.

Next, we consider a more complicated family of examples. Given $c \geq 1$ and $(a,b) \in \mathbb{R}^2_{>0}$, denote by $Q(a,b,c) \subset \mathbb{R}^2_{\geq 0}$ the quadrilateral with vertices $(0,0)$, $(0,1)$, $(c,0)$, (a,b). We note that $X_{Q(a,b,c)} \subset \mathbb{C}^2$ is a convex toric domain if and only if we have $a \leq c$, $b \leq 1$ and $a + bc \geq c$. The next result gives the formula for $\widetilde{\mathfrak{g}}_k$ when $\max(a+b,c) \leq 2$. The case $\max(a+b,c) > 2$ is similar to case (ii) below; see Remark 6.0.1.

Theorem 1.3.7 *Let $X := X_{Q(a,b,c)}$ be a convex toric domain for some $c \geq 1$ and $(a,b) \in \mathbb{R}^2_{>0}$, and put $M := \max(a+b, c)$.*

(i) *For $M \leq \frac{3}{2}$, we have*

$$\begin{gathered} \widetilde{\mathfrak{g}}_1(X) = 1, \qquad \widetilde{\mathfrak{g}}_2(X) = M, \qquad \widetilde{\mathfrak{g}}_3(X) = \min(\max(2, a+2b), 1+c), \\ \widetilde{\mathfrak{g}}_4 = 1+M, \qquad \widetilde{\mathfrak{g}}_5 = 2M, \qquad \widetilde{\mathfrak{g}}_6 = 2+M, \\ \widetilde{\mathfrak{g}}_{k+3}(X) = \widetilde{\mathfrak{g}}_k(X) + M \quad \text{for } k \geq 4. \end{gathered} \tag{1-3-4}$$

(ii) *For $\frac{3}{2} \leq M \leq 2$, then $\widetilde{\mathfrak{g}}_k(X)$ is as above for $k \leq 4$, and*

$$\widetilde{\mathfrak{g}}_5(X) = \min(\max(3, 1+a+2b), 2M, 2+c), \qquad \widetilde{\mathfrak{g}}_{k+2}(X) = 1 + \widetilde{\mathfrak{g}}_k(X) \quad \text{for } k \geq 4. \tag{1-3-5}$$

For our last family of examples, take $p \in \mathbb{R}_{\geq 1} \cup \{\infty\}$ and consider the L^p norm $\|-\|_p$ defined by $\|(x,y)\|_p := (x^p + y^p)^{1/p}$. Put

$$\Omega_p := \{(x,y) \in \mathbb{R}^2_{\geq 0} \mid \|(x,y)\|_p \leq 1\}.$$

Note that Ω_1 is the right triangle with vertices $(0,0)$, $(1,0)$, $(0,1)$, and Ω_∞ is the square with vertices $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$, ie the corresponding family of convex toric domains $\{X_{\Omega_p}\}$ interpolates between the round ball and the cube. Also, note that for $(x,y) \in \mathbb{R}^2_{\geq 0}$, we have

$$\|(x,y)\|^*_{\Omega_p} = \|(x,y)\|_q,$$

where $q \in \mathbb{R}_{\geq 1} \cup \{\infty\}$ is such that $1/p + 1/q = 1$.

Theorem 1.3.8 (i) *For $p \leq \ln(2)/\ln\left(\frac{4}{3}\right)$ we have*

$$\widetilde{\mathfrak{g}}_k(X_{\Omega_p}) = \begin{cases} 1 + i\sqrt[q]{2} & \text{for } k = 1+3i \text{ with } i \geq 0, \\ (i+1)\sqrt[q]{2} & \text{for } k = 2+3i \text{ with } i \geq 0, \\ 2 + i\sqrt[q]{2} & \text{for } k = 3+3i \text{ with } i \geq 0. \end{cases} \tag{1-3-6}$$

(ii) *For $p > \ln(2)/\ln\left(\frac{4}{3}\right)$ we have*

$$\widetilde{\mathfrak{g}}_k(X_{\Omega_p}) = \begin{cases} 1 + i & \text{for } k = 1+2i \text{ with } i \geq 0, \\ \sqrt[q]{2} + i & \text{for } k = 2+2i \text{ with } i \geq 0. \end{cases} \tag{1-3-7}$$

Remark 1.3.9 Incidentally, $\widetilde{\mathfrak{g}}_k(X_{\Omega_p}) = \widetilde{\mathfrak{g}}_k(E(1, \sqrt[q]{2}))$. Moreover, one can show using Corollary 1.3.1 that the capacities $\widetilde{\mathfrak{g}}_k(X_\Omega)$ of any four-dimensional convex toric domain normalized as in Corollary 1.3.1 are eventually either 2–periodic or 3–periodic in k, depending on which of $3\|(0,1)\|^*_\Omega$ or $2\|(1,1)\|^*_\Omega$ is smaller. Intuitively, domains which are "rounder" have 3–periodic capacities while domains which are "skinnier" have 2–periodic ones.

Example 1.3.10 For concreteness let us flesh out a simple implication of Theorem 1.3.8 for the symplectic embedding problem $E(1,5)\times\mathbb{C}^N \overset{s}{\hookrightarrow} \mu\cdot X_{\Omega_2}\times\mathbb{C}^N$ with $N\in\mathbb{Z}_{\geq 0}$. Using Theorem 1.6 of Gutt and Hutchings [16] (see also Kerman and Liang [27]), it is easy to check that we have

$$c_k^{\mathrm{GH}}(X_{\Omega_2}) = \begin{cases} \sqrt{\tfrac{1}{2}k^2} & \text{for } k \text{ even,} \\ \sqrt{\tfrac{1}{2}(k^2+1)} & \text{for } k \text{ odd,} \end{cases}$$

that is,

k	1	2	3	4	5	6	7	8	9	10	11	12
$c_k^{\mathrm{GH}}(E(1,5))$	1	2	3	4	5	5	6	7	8	9	10	10
$c_k^{\mathrm{GH}}(X_{\Omega_2})$	1	$\sqrt{2}$	$\sqrt{5}$	$2\sqrt{2}$	$\sqrt{13}$	$3\sqrt{2}$	5	$4\sqrt{2}$	$\sqrt{41}$	$5\sqrt{2}$	$\sqrt{61}$	$6\sqrt{2}$

and the capacities $\{c_k^{\mathrm{GH}}\}$ give the lower bound $\mu\geq 2/\sqrt{2}\approx 1.414$. Meanwhile, we have

k	1	2	3	4	5	6	7	8	9	10	11
$\tilde{\mathfrak{g}}_k(E(1,5))$	1	2	3	4	5	5	6	6	7	7	8
$\tilde{\mathfrak{g}}_k(X_{\Omega_2})$	1	$\sqrt{2}$	2	$1+\sqrt{2}$	$2\sqrt{2}$	$2+\sqrt{2}$	$1+2\sqrt{2}$	$3\sqrt{2}$	$2+2\sqrt{2}$	$1+3\sqrt{2}$	$4\sqrt{2}$

and the $\{\tilde{\mathfrak{g}}_k\}$ capacities give the lower bound $\mu\geq 5/(2\sqrt{2})\approx 1.768$.

We end this introduction with a brief outline of the proof of Theorem 1.2.8, deferring the reader to the body of the paper for the details. Firstly, as in Siegel [38; 36], we "fully round" our convex toric domain. This is a small perturbation and so leaves $\tilde{\mathfrak{g}}_k$ essentially unaffected, while it standardizes the Reeb dynamics on the boundary. Next, we obtain a lower bound on $\tilde{\mathfrak{g}}_k$ by mostly action and index considerations, with the second condition in Theorem 1.2.8 coming from the relative adjunction formula and writhe bounds. To obtain a corresponding upper bound, we first study the combinatorial optimization problem in Theorem 1.2.8 more carefully and arrive at the simplifications described in Section 4.3. We then inductively construct a curve for each minimizer. The base cases are cylinders or pairs of pants which we produce using the ECH cobordism map, while the inductive step is based on an iterated application of obstruction bundle gluing based on the work of Hutchings and Taubes.

Acknowledgements Siegel is partially supported by NSF grant DMS-2105578 and a visiting membership at the Institute for Advanced Study.

2 Preliminaries on pseudoholomorphic curves

The main purpose of this section is to briefly recall some requisite background on pseudoholomorphic curves and to establish notation, conventions and terminology for the rest of the paper. In Section 2.1 we discuss moduli spaces of punctured pseudoholomorphic curves in symplectic cobordisms. In Section 2.2

we recall the notion of local tangency constraints and the equivalence with skinny ellipsoidal constraints as in [30]. In Section 2.3 we introduce the notion of formal curves, which provides a convenient language and bookkeeping tool in SFT compactness arguments. Lastly, in Section 2.4 we discuss the extent to which our moduli spaces persist in one-parameter families, and we introduce the notion of "formal perturbation invariance", which will be particularly relevant for us.

2.1 Asymptotically cylindrical curves and their moduli

Our exposition in this subsection will be somewhat brief; we refer the reader to [40; 2] for more details.

2.1.1 Symplectic and contact manifolds Recall that a *Liouville cobordism* (X, λ) is a compact manifold-with-boundary X, equipped with a one-form λ whose exterior derivative $\omega := d\lambda$ is symplectic, and whose restriction to ∂X is a contact form. We have a natural decomposition $\partial X = \partial^+ X \sqcup \partial^- X$, where $\lambda|_{\partial^+ X}$ is a positive contact form and $\lambda|_{\partial^- X}$ is a negative contact form. When no confusion should arise, we will typically suppress λ from the notation and denote such a Liouville cobordism simply by X; a similar convention will apply to most other mathematical objects. We view $\partial^+ X$ and $\partial^- X$ as strict (ie equipped with a preferred contact form) contact manifolds.

Quite often we will have $\partial^- X = \varnothing$, in which case X is a *Liouville domain*. We say that a Liouville domain X has *nondegenerate contact boundary* if the contact form $\alpha := \lambda|_{\partial X}$ has nondegenerate Reeb orbits. The *action* of a Reeb orbit γ in ∂X is its period, ie the integral $\mathcal{A}_{\partial X}(\gamma) := \int_\gamma \alpha$, assuming γ is parametrized so that its velocity always agrees with the Reeb vector field R_α on ∂X.

More generally, a *compact symplectic cobordism* is a compact manifold-with-boundary X equipped with a symplectic form ω and a primitive one-form λ defined on $\mathcal{O}p(\partial X)$ whose restriction to ∂X is a contact form. As before we have a natural decomposition $\partial X = \partial^+ X \sqcup \partial^- X$. We will refer to the case $\partial^- X = \varnothing$ as a *symplectic filling* and the case $\partial^+ X = \varnothing$ as a *symplectic cap*. Note that the case with $\partial X = \varnothing$ is simply a closed symplectic manifold.

Convention If X and X' are Liouville domains and $\iota\colon X \overset{s}{\hookrightarrow} X'$ is a symplectic embedding, we will by slight abuse of notation write $X' \setminus X$ to denote the compact symplectic cobordism $X' \setminus \operatorname{Int} \iota(X)$, after attaching a small collar $[0, \delta) \times \partial X'$ to X' if necessary (ie if $\iota(X) \cap \partial X' \neq \varnothing$).

2.1.2 Admissible almost complex structures Let Y be a strict contact manifold with contact form α. Recall that the *symplectization* of Y is the symplectic manifold $\mathbb{R}_r \times Y$ with symplectic form given by $d(e^r \alpha)$. We denote by $\mathcal{J}(Y)$ the space of *admissible almost complex structures* on the symplectization $\mathbb{R} \times Y$. That is, $J_Y \in \mathcal{J}(Y)$ is a compatible almost complex structure on $\mathbb{R} \times Y$ which is r–translation invariant, maps ∂_r to the Reeb vector field R_α, and restricts to a compatible almost complex structure on each contact hyperplane.

Given a compact symplectic cobordism X with $Y^\pm := \partial^\pm X$, its *symplectic completion* $\widehat{X}$ is given attaching a positive half-symplectization $\mathbb{R}_{\geq 0} \times Y^+$ to its positive boundary and a negative half-symplectization $\mathbb{R}_{\leq 0} \times Y^-$ to its negative boundary. There is a natural symplectic form on $\widehat{X}$ which extends that of X and looks like the restriction of the symplectic form on a symplectization on the cylindrical ends. We denote by $\mathcal{J}(X)$ the space of *admissible almost complex structures* on the symplectic completion of X. That is, $J_X \in \mathcal{J}(X)$ is a compatible almost complex structure on $\widehat{X}$ which is symplectization-admissible on the cylindrical ends, ie we have $J_X|_{\mathbb{R}_{\geq 0}\times Y^+} = J_{Y^+}|_{\mathbb{R}_{\geq 0}\times Y^+}$ for some $J_{Y^+} \in \mathcal{J}(Y^+)$, and $J_X|_{\mathbb{R}_{\leq 0}\times Y^-} = J_{Y^-}|_{\mathbb{R}_{\leq 0}\times Y^-}$ for some $J_{Y^-} \in \mathcal{J}(Y^-)$. In particular, J_X is translation-invariant on each cylindrical end. Given fixed $J_{Y^+} \in \mathcal{J}(Y^+)$ and $J_{Y^-} \in \mathcal{J}(Y^-)$ as above, we denote by

$$\mathcal{J}^{J_{Y^+}}_{J_{Y^-}}(X) \subset \mathcal{J}(X)$$

the subspace consisting of almost complex structures J which satisfy $J|_{\mathbb{R}_{\geq 0}\times Y^+} = J_{Y^+}|_{\mathbb{R}_{\geq 0}\times Y^+}$ and $J|_{\mathbb{R}_{\leq 0}\times Y^-} = J_{Y^-}|_{\mathbb{R}_{\leq 0}\times Y^-}$. By slight abuse of notation, for $J \in \mathcal{J}^{J_{Y^+}}_{J_{Y^-}}$ we also use the notation $J|_{Y^\pm} := J_{Y^\pm}$ to denote the "restriction" of J to $Y^\pm$.

2.1.3 Moduli spaces of pseudoholomorphic curves Let X be a compact symplectic cobordism, and consider $J \in \mathcal{J}(X)$. A *J–holomorphic curve* C in $\widehat{X}$ consists of a Riemann surface Σ, with almost complex structure j, and a map $u\colon \Sigma \to \widehat{X}$ satisfying $du \circ j = J \circ du$. We will often refer to C as a "pseudoholomorphic curve" (or simply "curve") if J is implicit or unspecified. Such a curve C is *asymptotically cylindrical* if Σ is a closed Riemann surface minus a finite set of puncture points, such that u is positively or negatively asymptotic to a Reeb orbit in the ideal boundary at each puncture; see eg [40, Section 6.4] for a more precise formulation. All pseudoholomorphic curves considered in this paper will be asymptotically cylindrical in either the symplectic completion of a compact symplectic cobordisms (closed symplectic manifolds being a special case), or in the symplectization of a contact manifold. Strictly speaking the latter is a special case of the former, but it is helpful to distinguish between these two cases since in the latter case we work with almost complex structures having an additional translation symmetry.

Convention All pseudoholomorphic curves in this paper are asymptotically cylindrical, and for brevity we often refer to curves in $\widehat{X}$ as simply "curves in X", with the process of symplectically completing tacitly understood.

Consider tuples of nondegenerate Reeb orbits $\Gamma^+ = (\gamma_1^+, \dots, \gamma_a^+)$ in $\partial^+ X$ and $\Gamma^- = (\gamma_1^-, \dots, \gamma_b^-)$ in $\partial^- X$. Given $J \in \mathcal{J}(X)$, we denote by $\mathcal{M}^J_X(\Gamma^+;\Gamma^-)$ the moduli space of asymptotically cylindrical rational J–holomorphic curves in $\widehat{X}$ with positive asymptotics Γ^+ and negative asymptotics Γ^-, equipped with the Gromov topology. Here the conformal structure on the domain varies over the moduli space of genus-zero Riemann surfaces with a (resp. b) ordered positive (resp. negative) punctures. If $\partial^- X = \varnothing$, we write $\mathcal{M}^J_X(\Gamma^+)$ as a shorthand for $\mathcal{M}^J_X(\Gamma^+;\varnothing)$, and similarly in the case $\partial^+ X = \varnothing$ we write $\mathcal{M}^J_X(\Gamma^-)$ in place of $\mathcal{M}^J_X(\varnothing;\Gamma^-)$. We will sometimes suppress J from the notation and write simply $\mathcal{M}(\Gamma^+;\Gamma^-)$ if the almost complex structure is implicit or unspecified.

Convention *By default all curves in this paper have genus zero unless otherwise stated.*

Similarly, given a strict contact manifold Y, $J \in \mathcal{J}(Y)$, and Reeb orbits $\Gamma^+ = (\gamma_1^+, \dots, \gamma_a^+)$ and $\Gamma^- = (\gamma_1^-, \dots, \gamma_b^-)$ in Y, we denote by $\mathcal{M}^J_Y(\Gamma^+;\Gamma^-)$ the moduli space of asymptotically cylindrical curves in $\mathbb{R} \times Y$ with positive asymptotics Γ^+ and negative asymptotics Γ^-. There is a natural $\mathbb{R}$–action on $\mathcal{M}_Y^J(\Gamma^+;\Gamma^-)$ induced by translations in the first factor of $\mathbb{R} \times Y$, and this is free away from the *trivial cylinders*, ie cylinders of the form $\mathbb{R} \times \gamma$ with γ a Reeb orbit in Y. We denote the quotient by $\mathcal{M}_Y^J(\Gamma^+;\Gamma^-)/\mathbb{R}$.

We will consider moduli spaces associated to one-parameter families of almost complex structures. For instance, given a one-parameter family $\{J_t\}_{t\in[0,1]}$ in $\mathcal{J}(X)$, we denote by $\mathcal{M}_X^{\{J_t\}}(\Gamma^+;\Gamma^-)$ the corresponding *parametrized moduli space* consisting of pairs (t, C) with $t \in [0,1]$ and $C \in \mathcal{M}_X^{J_t}(\Gamma^+;\Gamma^-)$.

We will assume throughout that suitable choices have been made so that every regular moduli space of curves is oriented. In particular, any curve C which is regular and isolated in $\mathcal{M}_X(\Gamma^+;\Gamma^-)$ or $\mathcal{M}_Y(\Gamma^+;\Gamma^-)/\mathbb{R}$ has an associated sign $\varepsilon(C) \in \{-1, 1\}$. We briefly recall the procedure for orienting moduli spaces in Section 5.2.

2.1.4 SFT compactifications The above moduli spaces admit SFT compactifications as in [2], which we denote by replacing $\mathcal{M}$ with $\overline{\mathcal{M}}$. For example, let X be a compact symplectic cobordism with $J_\pm \in \mathcal{J}(\partial^\pm X)$ and $J_X \in \mathcal{J}_{J_-}^{J_+}(X)$. Elements of $\overline{\mathcal{M}}_X^{J_X}(\Gamma^+;\Gamma^-)$ are *stable pseudoholomorphic buildings* in $\widehat{X}$, which consist of

- some number (possibly zero) of J_+–holomorphic levels in the symplectization $\mathbb{R} \times \partial^+ X$,
- a "main" J_X–holomorphic level in $\widehat{X}$, and
- some number (possibly zero) of J_-–holomorphic levels in the symplectization $\mathbb{R} \times \partial^- X$,

such that for each pair of adjacent levels the positive asymptotic Reeb orbits of the lower level are paired with the negative asymptotic Reeb orbits of the upper level. The symplectization levels are always defined modulo target translations. Note that each level consists of one or more connected components, each of which is a nodal punctured Riemann surface. The stability condition states that each component of the domain on which the map is constant must have negative Euler characteristic after removing all special points; also there are no symplectization levels consisting entirely of trivial cylinders. See [2] for details.

We will use the following language in this paper. (Note that the slightly different notion of *matched component* employed in [30] serves a similar purpose.)

Definition 2.1.1 We say that a (rational) curve in a given level is *connected* if its domain is connected but possibly nodal, *smooth* if its domain is without nodes, and *irreducible* if it is both connected and smooth. By *curve component* we mean a (rational) curve which is irreducible.

Note that each level of a pseudoholomorphic building can be decomposed into its constituent (irreducible) components.

We will also frequently make use of *neck stretching*. If X^+ and X^- are compact symplectic cobordisms with a common contact boundary $\partial^- X^+ = \partial^+ X^- = Y$, we denote the glued compact symplectic cobordism by $X^- \circledcirc X^+$. Given almost complex structures $J_Y \in \mathcal{J}(Y)$, $J_{X^+} \in \mathcal{J}_{J_Y}(X^+)$, $J_{X^-} \in \mathcal{J}^{J_Y}(X^-)$, we can consider the corresponding neck-stretching family of almost complex structures $J_t \in \mathcal{J}(X)$, where $t \in [0,1)$. The limit $t \to 1$ corresponds to the broken cobordism which we denote by $X^- \oplus X^+$. The compactification $\bar{\mathcal{M}}_X^{\{J_t\}}(\Gamma^+;\Gamma^-)$ consists of pairs (t, C) for $t \in [0,1)$ and $C \in \bar{\mathcal{M}}_X^{J_t}(\Gamma^+;\Gamma^-)$, as well as limiting configurations for $t = 1$, which are pseudoholomorphic buildings with

- some number (possibly zero) of $J_{\partial^+ X^+}$–holomorphic levels in the symplectization $\mathbb{R} \times \partial^+ X^+$,
- a J_{X^+}–holomorphic level in $\hat{X}^+$,
- some number (possibly zero) of J_Y–holomorphic levels in the symplectization $\mathbb{R} \times Y$,
- a J_{X^-}–holomorphic level in $\hat{X}^-$, and
- some number (possibly zero) of $J_{\partial^- X^-}$–holomorphic levels in the symplectization $\mathbb{R} \times \partial^- X^-$,

subject to suitable matching and stability conditions. Here we have put $J_{\partial^+ X^+} := J_{X^+}|_{\partial^+ X^+}$ and $J_{\partial^- X^-} := J_{X^-}|_{\partial^- X^-}$.

2.1.5 Homology classes and energy Given a compact symplectic cobordism X and Reeb orbits $\Gamma^+ = (\gamma_1^+, \dots, \gamma_a^+)$ in $\partial^+ X$ and $\Gamma^- = (\gamma_1^-, \dots, \gamma_b^-)$ in $\partial^- X$, we let $H_2(X, \Gamma^+ \cup \Gamma^-)$ denote the group of potential homology classes of curves in $\mathcal{M}_X(\Gamma^+;\Gamma^-)$. Namely, $H_2(X, \Gamma^+ \cup \Gamma^-)$ is the abelian group freely generated by 2–chains Σ in X with $\partial\Sigma = \sum_{i=1}^a \gamma_i^+ - \sum_{j=1}^b \gamma_j^-$, modulo boundaries of 3–chains in X; see also [40, Section 6.4] for a slightly more homological perspective. Given $A \in H_2(X, \Gamma^+ \cup \Gamma^-)$, we denote by $\mathcal{M}_{X,A}(\Gamma^+;\Gamma^-) \subset \mathcal{M}_X(\Gamma^+;\Gamma^-)$ the subspace of curves lying in homology class A.

Similarly, given a strict contact manifold Y and Reeb orbits $\Gamma^+ = (\gamma_1^+, \dots, \gamma_a^+)$ and $\Gamma^- = (\gamma_1^-, \dots, \gamma_b^-)$ in Y, let $H_2(Y, \Gamma^+ \cup \Gamma^-)$ denote the homology group of 2–chains Σ in Y with $\partial\Sigma = \sum_{i=1}^a \gamma_i^+ - \sum_{j=1}^b \gamma_j^-$, modulo boundaries of 3–chains in Y. Given $A \in H_2(Y, \Gamma^+ \cup \Gamma^-)$, we denote by $\mathcal{M}_{Y,A}(\Gamma^+;\Gamma^-) \subset \mathcal{M}_Y(\Gamma^+;\Gamma^-)$ the subspace of curves in $\mathbb{R} \times Y$ lying in homology class A.

There are also natural subspaces $\bar{\mathcal{M}}_{X,A}(\Gamma^+;\Gamma^-) \subset \bar{\mathcal{M}}_X(\Gamma^+;\Gamma^-)$ and $\bar{\mathcal{M}}_{Y,A}(\Gamma^+;\Gamma^-) \subset \mathcal{M}_Y(\Gamma^+;\Gamma^-)$ and so on. These are defined by required the total homology class of a building, which is defined in a natural way by concatenating the levels, to be A.

If (Y, α) is strict contact manifold, define the *energy*[3] of a curve $C \in \mathcal{M}_{Y,A}(\Gamma^+;\Gamma^-)$ to be $E_Y(C) := \int_C d\alpha$. By Stokes' theorem, we have

$$E_Y(C) = \sum_{i=1}^a \mathcal{A}_Y(\gamma_i^+) - \sum_{j=1}^b \mathcal{A}_Y(\gamma_j^-).$$

Note that this depends only the homology class $A \in H_2(Y, \Gamma^+ \cup \Gamma^-)$, so we can also put $E_Y(C) = E_Y(A) := \int_A d\alpha$. Similarly, if X is a compact symplectic cobordism with symplectic form ω and locally defined Liouville one-form λ, the energy $E_X(C)$ of a curve $C \in \mathcal{M}_{Y,A}^J(\Gamma^+;\Gamma^-)$ is defined to

[3]This is called the *ω–energy in* [2], their full energy having this as one of its two summands.

be the integral over C of the piecewise smooth two-form which agrees with ω on X and with $d\lambda$ on the cylindrical ends $\widehat{X} \setminus X$. If X is further a Liouville cobordism (ie λ is globally defined), then Stokes' theorem gives

$$E_X(C) = \sum_{i=1}^{a} \mathcal{A}_{\partial^+ X}(\gamma_i^+) - \sum_{j=1}^{b} \mathcal{A}_{\partial^- X}(\gamma_j^-).$$

This again depends only on $A \in H_2(X, \Gamma^+ \cup \Gamma^-)$, and we have $E_X(C) = E_X(A) := \int_A \omega$.

2.2 Local tangency and skinny ellipsoidal constraints

Let X be a compact symplectic cobordism. Recall that the *local tangency constraint* $\langle\mathcal{T}^{(m)}p\rangle$ with $m \in \mathbb{Z}_{\geq 1}$ is imposed by choosing a point $p \in \operatorname{Int} X$ and a smooth symplectic divisor $D \subset \mathcal{O}p(p)$ and considering curves with an additional marked point required to pass through p with contact order (at least) m to D; see eg [7; 8; 30]. We will also denote this constraint by $\langle\mathcal{T}^{m-1}p\rangle$, with $m-1$ representing the tangency order (in particular $\langle p\rangle$ corresponds simply to a marked point passing through p).

Let $\mathcal{J}(X; D) \subset \mathcal{J}(X)$ denote the space of admissible almost complex structures on $\widehat{X}$ which are integrable near p and preserve the germ of D near p. Given tuples of Reeb orbits Γ^+ and Γ^- in $\partial^+ X$ and $\partial^- X$ respectively and $J \in \mathcal{J}(X; D)$, we define the moduli space $\mathcal{M}_X^J(\Gamma^+; \Gamma^-)\langle\mathcal{T}^{(m)}p\rangle$ as before, but now the local tangency constraint $\langle\mathcal{T}^{(m)}p\rangle$ is imposed on each curve.

Some care is needed when compactifying $\mathcal{M}_X^J(\Gamma^+; \Gamma^-)\langle\mathcal{T}^{(m)}p\rangle$, due to the possibility of a ghost (ie constant) component inheriting the marked point. Indeed, strictly speaking a constant component is tangent to D to infinite order, and hence ghost configurations always appear with much higher than expected dimension. To get around this, first note, as in the proof of [30, Proposition 2.2.2], that there is a natural inclusion

$$\mathcal{M}_X^J(\Gamma^+; \Gamma^-)\langle\mathcal{T}^{(m)}p\rangle \subset \overline{\mathcal{M}}_X^J(\Gamma^+; \Gamma^-)\langle p\rangle,$$

where the codomain is the usual SFT compactification of $\mathcal{M}_X^J(\Gamma^+; \Gamma^-)\langle p\rangle$ by stable pseudoholomorphic buildings. Let $\overline{\mathcal{M}}_X^J(\Gamma^+; \Gamma^-)\langle T^{(m)}p\rangle$ denote the closure of $\mathcal{M}_X^J(\Gamma^+; \Gamma^-)\langle\mathcal{T}^{(m)}p\rangle$ in this compact ambient space. To understand what this amounts to, consider a pseudoholomorphic building C in $\overline{\mathcal{M}}_X^J(\Gamma^+; \Gamma^-)\langle T^{(m)}p\rangle$ such that the marked point z_0 mapping to p lies on a ghost component C_0. Let $N_1, \dots, N_a$ denote those nodes connecting a nonconstant component of C to C_0, or more generally connecting a nonconstant component of C to some ghost component which is nodally connected through ghost components to C_0. Let $z_1, \dots, z_a$ denote the corresponding special points in the domain of C which are "near z_0", ie participate in the nodes $N_1, \dots, N_a$ and lie on nonconstant components of C. Let $C_1, \dots, C_a$ denote the respective nonconstant components of C on which $z_1, \dots, z_a$ lie. According to [7, Lemma 7.2], in this situation the marked points $z_1, \dots, z_a$ satisfy local tangency constraints $\langle\mathcal{T}^{(m_1)}p\rangle, \dots, \langle\mathcal{T}^{(m_a)}p\rangle$, respectively, such that we have

$$m_1 + \dots + m_a \geq m.$$

In this way, elements of $\overline{\mathcal{M}}_X^J(\Gamma^+; \Gamma^-)\langle\mathcal{T}^{(m)}p\rangle$ "remember" the constraint $\langle\mathcal{T}^{(m)}p\rangle$.

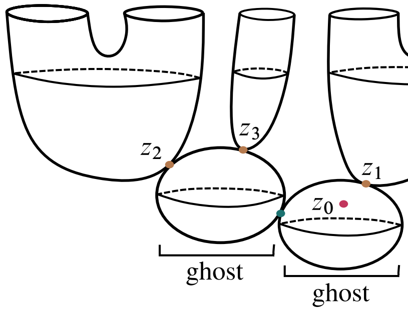


Figure 1: A configuration which could potentially arise in $\overline{\mathcal{M}}_X^J(\Gamma^+;\Gamma^-)\llangle\mathcal{T}^{(m)}p\rrangle$. Here the marked point z_0 mapping to p lies on a ghost component, and z_1, z_2 and z_3 are the special points near z_0 lying on nonconstant components. These satisfy respective constraints $\llangle\mathcal{T}^{(m_1)}p\rrangle$, $\llangle\mathcal{T}^{(m_2)}p\rrangle$ and $\llangle\mathcal{T}^{(m_3)}p\rrangle$ such that $m_1+m_2+m_3\geq m$. Such a configuration is also included in $\overline{\overline{\mathcal{M}}}_X^J(\Gamma^+;\Gamma^-)\llangle\mathcal{T}^{(m)}p\rrangle$ even if it does not arise as a limit of curves in $\mathcal{M}_X^J(\Gamma^+;\Gamma^-)\llangle\mathcal{T}^{(m)}p\rrangle$.

We will also need to consider a potentially larger compactification of $\mathcal{M}_X^J(\Gamma^+;\Gamma^-)\llangle T^{(m)}p\rrangle$ which allows all ghost configurations as described above, even if they do not arise as a limit of smooth curves:

Definition 2.2.1 Let $\overline{\overline{\mathcal{M}}}_X^J(\Gamma^+;\Gamma^-)\llangle\mathcal{T}^{(m)}p\rrangle$ denote the subset of $\overline{\mathcal{M}}_X^J(\Gamma^+;\Gamma^-)\llangle p\rrangle$ given by the union of $\mathcal{M}_X^J\llangle\mathcal{T}^{(m)}p\rrangle$ with the set of all buildings C such that the marked point z_0 mapping to p lies on a ghost component and the special points $z_1,\dots,z_a$ near z_0 (as above) satisfy respective constraints $\llangle\mathcal{T}^{(m_1)}p\rrangle,\dots,\llangle\mathcal{T}^{(m_a)}p\rrangle$ such that $m_1+\dots+m_a\geq m$. See Figure 1.

Remark 2.2.2 It is worth emphasizing that the extra buildings C involving ghost components which appear in Definition 2.2.1 have virtual codimension at least two (cf the proof of [30, Proposition 2.2.2]), and hence are not expected to appear whenever sufficient transversality holds. This is essentially why such configurations do not contribute to the local tangency constraint counts $N_{M,A}\llangle\mathcal{T}^{(m)}p\rrangle$ defined in [30] for semipositive closed symplectic manifolds M.

For $m\in\mathbb{Z}_{\geq 1}$, let $\llangle(m)\rrangle_E$ denote the *skinny ellipsoidal constraint* of order m, defined as follows. Let E_{sk} denote a *skinny ellipsoid*, ie a symplectic ellipsoid whose first area factor is sufficiently small compared to the others. After possibly shrinking (ie replacing E_{sk} by ϵE_{sk} for $0<\epsilon\ll 1$) we can assume that E_{sk} symplectically embeds into X in an essentially unique way, and we typically denote this embedding by an inclusion $E_{\mathrm{sk}}\subset X$. Let η_m denote the m–fold cover of the simple Reeb orbit of least action in ∂E_{sk}. For curves in X, the constraint $\llangle(m)\rrangle$ is imposed by replacing $\widehat{X}$ with $\widehat{X\setminus E_{\mathrm{sk}}}$, and considering curves with one additional negative puncture which is asymptotic to η_m. We define the moduli space $\mathcal{M}_X^J(\Gamma^+;\Gamma^-)\llangle(m)\rrangle_E$ by analogy with $\mathcal{M}_X^J(\Gamma^+;\Gamma^-)\llangle\mathcal{T}^{(m)}p\rrangle$, replacing the local tangency constraint $\llangle\mathcal{T}^{(m)}p\rrangle$ with the skinny ellipsoidal constraint $\llangle(m)\rrangle_E$. Note that both of these moduli spaces have the same index, namely

$$\text{(2-2-1)}\qquad \mathrm{ind}=(n-3)(2-a-b)+2c_1^\tau(A)+\sum_{i=1}^{a}\mathrm{CZ}_\tau(\gamma_i^+)-\sum_{j=1}^{b}\mathrm{CZ}_\tau(\gamma_j^-)-2n-2m+4,$$

where $2n = \dim_{\mathbb{R}}(X)$. Here τ is a choice of trivialization (up to homotopy) of the symplectic vector bundle over each Reeb orbit, $c_1^\tau(A)$ is the corresponding relative first Chern class evaluated on A, and CZ_τ denotes the Conley–Zehnder index measured with respect to τ. Recall that the index does not depend on the choice of τ, even though the individual terms do.

If M is a closed symplectic four-manifold with homology class $A \in H_2(M)$, [30, Section 4.1] establishes an equivalence of signed counts

$$\#\mathcal{M}_{M,A}\lessdot\mathcal{T}^{(m)}p\gtrdot = \#\mathcal{M}_{M,A}\lessdot(m)\gtrdot_E.$$

The basic idea is to place the tangency constraint in E_{sk} and stretch the neck along ∂E_{sk}, and then to argue that only degenerations of the expected type can arise. Although [30] only proves this in dimension four in order to invoke an argument which sidesteps any technicalities about gluing curves with tangency constraints, this is expected to hold for closed manifolds of all dimensions. In the context of a symplectic cobordism X, it is not quite reasonable to expect in general an equality of signed curve counts

$$\#\mathcal{M}^J_{X,A}(\Gamma^+;\Gamma^-)\lessdot(m)\gtrdot_E = \#\mathcal{M}^J_{X,A}(\Gamma^+;\Gamma^-)\lessdot\mathcal{T}^{(m)}p\gtrdot.$$

Indeed, these counts might not be particularly robust, eg they could depend on J and the embedding $E_{\mathrm{sk}} \overset{s}{\hookrightarrow} X$. However, the argument in [30, Theorem 4.1.1] does extend to this setting to prove:

Proposition 2.2.3 *If* $\dim X = 4$, *we have*

$$\#\mathcal{M}^J_{X,A}(\Gamma^+;\Gamma^-)\lessdot(m)\gtrdot_E = \#\mathcal{M}^J_{X,A}(\Gamma^+;\Gamma^-)\lessdot\mathcal{T}^{(m)}p\gtrdot$$

provided that the following conditions hold:

(i) *The moduli space* $\#\mathcal{M}^J_{X,A}(\Gamma^+;\Gamma^-)\lessdot\mathcal{T}^{(m)}p\gtrdot$ *is formally perturbation invariant.* (See Section 2.4 below.)

(ii) *Each Reeb orbit in* $\Gamma^+ \cup \Gamma^-$ *is nondegenerate and either elliptic or negative hyperbolic.*

Indeed, the first condition guarantees that curve counts remain constant over a generic one-parameter family of almost complex structures (cf Proposition 2.4.2), and the second condition ensures that the relevant curves count positively (cf Section 5.2 and Remark 5.2.3(ii)).

2.3 Formal curves

In this subsection we introduce the notion of a "formal curve", which is a convenient device for storing combinatorial curve data, but without requiring that this data be represented by any actual solution to the pseudoholomorphic curve equation. We also define "formal buildings", which are analogous to pseudoholomorphic buildings but with each pseudoholomorphic curve component replaced by a formal curve component. This will allow us to discuss "formal perturbation invariance" of moduli spaces in the next subsection.

2.3.1 Formal curve components To begin, we define:

Definition 2.3.1 A *formal curve component* C in a compact symplectic cobordism (X,ω) is a triple (Γ^+,Γ^-,A), where

- $\Gamma^+=(\gamma_1^+,\dots,\gamma_a^+)$ is a tuple of Reeb orbits in ∂^+X,
- $\Gamma^-=(\gamma_1^-,\dots,\gamma_b^-)$ is a tuple of Reeb orbits in ∂^-X,
- $A\in H_2(X,\Gamma^+\cup\Gamma^-)$ is a homology class, and
- we require the energy $E_X(C):=E_X(A)=\int_A\omega$ to be nonnegative.

Similarly, a formal curve component C in a strict contact manifold (Y,α) is a triple (Γ^+,Γ^-,A), where Γ^+ and Γ^- are tuples of Reeb orbits in Y and $A\in H_2(Y,\Gamma^+\cup\Gamma^-)$ is a homology class, and we require the energy $E_Y(C):=E(A)=\int_A d\alpha$ to be nonnegative.

We view C as representing a hypothetical genus-zero[4] irreducible asymptotically cylindrical curve in $\widehat{X}$ or $\mathbb{R}\times Y$. Note that a formal curve component also has a well-defined index $\operatorname{ind}(C)$, defined by the same formula (2-2-1). We will say that a formal curve component in Y is a "trivial cylinder" (or just "trivial") if $a=b=1$ and $\Gamma^+=\Gamma^-=(\gamma)$ for some Reeb orbit γ in Y. A formal curve component C is "closed" if $\Gamma^+=\Gamma^-=\varnothing$, and it is moreover "constant" if $E_X(C)=0$.

It will also be convenient to speak about formal curve components in X carrying a constraint $\langle\!\langle\mathcal{T}^{(m)}p\rangle\!\rangle$ for some $m\in\mathbb{Z}_{\ge1}$. Here the constraint $\langle\!\langle\mathcal{T}^{(m)}p\rangle\!\rangle$ is an extra piece of formal data which has the effect of decreasing the index by $2n-4+2m$ (here $2n=\dim_{\mathbb{R}}(X)$).

Given a formal curve $C=(\Gamma^+,\Gamma^-,A)$ in X and $J_X\in\mathcal{J}(X)$, we introduce the shorthand notation $\mathcal{M}_X^{J_X}(C):=\mathcal{M}_{X,A}^{J_X}(\Gamma^+;\Gamma^-)$ for the corresponding space of J_X–holomorphic curves representing C. As before, we will often omit the almost complex structure from the notation. Similarly, if $C=(\Gamma^+,\Gamma^-,A)$ is a formal curve in Y and $J_Y\in\mathcal{J}(Y)$, we put $\mathcal{M}_Y^{J_Y}(C):=\mathcal{M}_{Y,A}^{J}(\Gamma^+;\Gamma^-)$. This shorthand also applies when C carries a local tangency constraint, which is then implicit in, say, the notation $\mathcal{M}_X(C)$.

2.3.2 Formal nodal curves and buildings We now extend the above definition in order to model elements of the SFT compactification. Firstly, a *connected formal nodal curve* C in X or Y is roughly the same as a pseudoholomorphic nodal curve, but with each pseudoholomorphic curve component replaced by a formal curve component. More precisely:

Definition 2.3.2 A *connected formal nodal curve* C in X (resp. Y) consists of

- a tree T, and
- for each vertex v of T, a formal curve component C_v in X (resp. Y).

More generally, we drop the "connected" condition by allowing T to be a forest (ie disjoint union of trees).

[4] One could of course extend the definition to allow for higher-genus curves, but we will not need this.

We view the edges as representing nodes. We say that C is *stable* if, for each nonconstant component C_v, the number of punctures plus the number of edges connected to v is at least three.

Definition 2.3.3 A *formal building* in X consists of

- formal nodal curves $C_1, \dots, C_a$ in $\partial^+ X$ for some $a \in \mathbb{Z}_{\geq 0}$,
- a formal nodal curve C_0 in X, and
- formal nodal curves $C_{-1}, \dots, C_{-b}$ in $\partial^- X$ for some $b \in \mathbb{Z}_{\geq 0}$,

such that the tuple of positive Reeb orbits for C_i coincides with the tuple of negative Reeb orbits for C_{i+1} for $i = -b, \dots, a-1$. We also assume that the graph given naturally by concatenating the forest of each level is acyclic.

Similarly, a formal building in Y consists of formal nodal curves $C_1, \dots, C_a$ in Y for some $a \in \mathbb{Z}_{\geq 1}$ such that the tuple of positive Reeb orbits for C_i coincides with the tuple of negative Reeb orbits for C_{i+1} for $i = 1, \dots, a-1$, and such that the underlying graph is acyclic.

We view a formal building as modeling a rational pseudoholomorphic building in X or $\mathbb{R} \times Y$, with each constituent formal nodal curve representing a level. Note that the acyclicity condition ensures total genus zero and could be relaxed, but for our purposes we will keep it. Such a building has a *total homology class* in $H_2(X; \Gamma^+ \cup \Gamma^-)$ or $H_2(Y; \Gamma^+ \cup \Gamma^-)$, where Γ^+ (resp. Γ^-) is the tuple of positive Reeb orbits of the top (resp. bottom) level. We will say that a formal building is *stable* if each constituent formal nodal curve is stable, and no level is a union of trivial cylinders. We denote the set of stable formal buildings in X whose top (resp. bottom) level has positive (resp. negative) Reeb orbits Γ^+ (resp. Γ^-) by $\overline{\mathcal{F}}_{X,A}(\Gamma^+; \Gamma^-)$. The set $\overline{\mathcal{F}}_{Y,A}(\Gamma^+; \Gamma^-)$ of stable formal buildings in Y is defined similarly.

We denote the formal analogue of $\overline{\overline{\mathcal{M}}}_{X,A}(\Gamma^+; \Gamma^-)\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$ by $\overline{\overline{\mathcal{F}}}_{X,A}(\Gamma^+; \Gamma^-)\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$. This consists of two types of stable formal buildings, modeling curves where the marked point z_0 mapping to p lies on a nonconstant component or constant component, respectively. In the first case, we have all stable formal buildings such that one of the components in X is formally endowed with a constraint $\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$. In the second case, we have all stable formal buildings such that some constant component C_0 in X is formally endowed with a constraint $\langle\!\langle p \rangle\!\rangle$, and the nearby nonconstant components $C_1, \dots, C_a$ (ie those nonconstant components which are nodally connected through constant components to C_0) are formally endowed with constraints $\langle\!\langle \mathcal{T}^{(m_1)} p \rangle\!\rangle, \dots, \langle\!\langle \mathcal{T}^{(m_a)} p \rangle\!\rangle$, respectively, such that $m_1 + \cdots + m_a \geq m$ (cf Section 2.2). Note that the extra constraint $\langle\!\langle p \rangle\!\rangle$ is taken into account as a marked point when formulating stability, whereas the constraints $\langle\!\langle \mathcal{T}^{(m_1)} p \rangle\!\rangle, \dots, \langle\!\langle \mathcal{T}^{(m_a)} p \rangle\!\rangle$ do not affect stability since they lie on nonconstant components.

2.3.3 Formal covers Next, we define the formal analogue of multiple covers of pseudoholomorphic curves. Let X be a symplectic filling, and let $\Gamma = (\gamma_1, \dots, \gamma_a)$ and $\overline{\Gamma} = (\overline{\gamma}_1, \dots, \overline{\gamma}_{\overline{a}})$ be tuples of Reeb orbits in $Y := \partial X$. Let $C = (\Gamma, \varnothing, A)$ and $\overline{C} = (\overline{\Gamma}, \varnothing, \overline{A})$ be formal curve components in X satisfying constraints $\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$ and $\langle\!\langle \mathcal{T}^{(\overline{m})} p \rangle\!\rangle$, respectively. We say that C is a κ–fold *formal cover* of $\overline{C}$ if there exist

- a sphere Σ with marked points $(z_0, \dots, z_a)$,
- a sphere $\overline{\Sigma}$ with marked points $(\overline{z}_0, \dots, \overline{z}_{\overline{a}})$, and
- a κ–fold branched cover $\pi \colon \Sigma \to \overline{\Sigma}$

such that

- $\pi^{-1}(\{\overline{z}_1, \dots, \overline{z}_{\overline{a}}\}) = \{z_1, \dots, z_a\}$,
- $\pi(z_0) = \overline{z}_0$,
- for each $i = 1, \dots, a$, γ_i is the κ_i–fold cover of $\overline{\gamma}_j$, where j is such that $\pi(z_i) = \overline{z}_j$ and κ_i is the ramification order of π at z_i, and
- we have $\kappa \overline{m} \geq m$, where κ is the ramification order of π at z_0.

A formal curve component is *simple* if it cannot be written as a nontrivial (ie with $\kappa \geq 2$) formal cover of any other formal curve component.

2.4 Formal perturbation invariance

The following is our main criterion for establishing upper bounds and proving stabilization for the capacities defined in Section 3.

Definition 2.4.1 Let X be a Liouville domain with nondegenerate contact boundary Y, and let C be an index zero simple formal curve component in X with positive asymptotics $\Gamma = (\gamma_1, \dots, \gamma_a)$, homology class $A \in H_2(X, \Gamma)$, and carrying a constraint $\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$ for some $m \in \mathbb{Z}_{\geq 1}$. We say that C is *formally perturbation invariant* if there exists a generic $J_Y \in \mathcal{J}(Y)$ such that the following holds. Suppose that $C' \in \overline{\overline{\mathcal{F}}}_{X,A}(\Gamma)\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$ is any stable formal building satisfying:

(A1) Each nonconstant component of C' in X is a formal cover of some formal curve component $\overline{C}'$ with $\mathrm{ind}(\overline{C}') \geq -1$.

(A2) Each nonconstant component of C' in Y is a formal cover of some formal curve component $\overline{C}'$ which is either trivial or else satisfies $\mathrm{ind}(\overline{C}') \geq 1$.

Then either

(B1) C' consists of a single component, ie $C' = C$, or else

(B2) C' is a two-level building, with bottom level in X consisting of a single component C_X which is simple with index -1, and with top level in Y represented by a union of some trivial cylinders with a simple index 1 component C_Y in $\mathbb{R} \times Y$; moreover we require that $\mathcal{M}_Y^{J_Y}(C_Y)$ is regular and satisfies $\#\mathcal{M}_Y^{J_Y}(C_Y)/\mathbb{R} = 0$.

More generally, if C is any formal curve component in X, we say that it is formally perturbation invariant if it is a formal cover of an index zero simple formal curve component $\overline{C}$ which is formally perturbation invariant as above.

We will also say that the associated moduli space $\mathcal{M}_X(C)$ is formally perturbation invariant if the formal curve component C is. Roughly, this means that for "purely formal reasons" the moduli space $\mathcal{M}_X(C)$ cannot degenerate in a generic one-parameter family. More precisely, the condition is "formal in X but not in Y", ie it takes into account pseudoholomorphic curves in $\mathbb{R} \times Y$ (via the last condition about $\mathcal{M}_Y^{J_Y}(C_Y)$) but only formal curves in X.[5] We will also say that C is "formally perturbation invariant with respect to J_Y" when we wish to emphasize the role of J_Y in Definition 2.4.1.

The following is a consequence of structure transversality and gluing techniques for simple curves:

Proposition 2.4.2 *Let X be a Liouville domain with nondegenerate contact boundary Y, and let C be a simple index-zero formal curve component X which carries a local tangency constraint $\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$. Assume that C is formally perturbation invariant with respect to some generic $J_Y \in \mathcal{J}(Y)$. Then the associated moduli space $\mathcal{M}_X^{J_X}(C)$ is regular and finite for generic $J_X \in \mathcal{J}^{J_Y}(X; D)$, and moreover the signed count $\#\mathcal{M}_X^{J_X}(C)$ is independent of J_X provided that $\mathcal{M}_X^{J_X}(C)$ is regular.*

Proof If $J_Y \in \mathcal{J}(Y)$ and $J_X \in \mathcal{J}^{J_Y}(X; D)$ are generic, it follows by standard transversality techniques (cf [40, Section 8]) that

- every simple J_Y–holomorphic curve component in $\mathbb{R} \times Y$ is either trivial or else has index at least one, and
- every simple J_X–holomorphic curve component in $\widehat{X}$ has nonnegative index.

In particular, since C is simple, $\mathcal{M}_X^{J_X}(C)$ is regular and hence a zero-dimensional smooth oriented manifold. It also follows by formal perturbation invariance of C and the SFT compactness theorem (plus the discussion in Section 2.2) that we must have $\overline{\overline{\mathcal{M}}}{}_X^{J_X}(C) = \mathcal{M}_X^{J_X}(C)$, whence $\mathcal{M}_X^{J_X}(C)$ is finite. Indeed, any element C' of $\overline{\overline{\mathcal{M}}}{}_X^{J_X}(C)$ defines a stable formal building in $\overline{\overline{\mathcal{F}}}_{X,A}(\Gamma)\langle\!\langle \mathcal{T}^{(m)} p \rangle\!\rangle$ satisfying (A1) and (A2), and since (B2) is impossible when J_X is regular we must have $C' \in \mathcal{M}_X^{J_X}(C)$.

Now assume that $J_0, J_1 \in \mathcal{J}^{J_Y}(X; D)$ are chosen such that $\mathcal{M}_X^{J_i}(C)$ is regular for $i = 0, 1$, and let $\{J_t\}_{t\in[0,1]}$ be a generic one-parameter family in $\mathcal{J}^{J_Y}(X; D)$ interpolating between them. Standard transversality techniques imply that $\mathcal{M}_X^{\{J_t\}}(C)$ is regular and hence a smooth oriented one-dimensional manifold. By formal perturbation invariance and the SFT compactness theorem, the compactification $\overline{\overline{\mathcal{M}}}{}_X^{\{J_t\}}(C)$ (defined similarly to Definition 2.2.1) is a smooth cobordism between $\mathcal{M}_X^{J_0}(C)$ and $\mathcal{M}_X^{J_1}(C)$, with possibly some additional boundary configurations as in (B2). Our goal is to prove $\#\mathcal{M}_X^{J_0}(C) = \#\mathcal{M}_X^{J_1}(C)$. Note that this would be immediate if there were none of these additional boundary configurations.

[5] In our application, C_Y will occur as a low-energy cylinder between an elliptic orbit $e_{i,j}$ and the corresponding hyperbolic orbit $h_{i,j}$ in $\partial \widetilde{X}_\Omega$; cf Lemma 5.1.3 below.

Each of these additional boundary configurations occurs at some time $t_b \in (0,1)$ and consists of a two-level building, with

- a top-level J_Y–holomorphic curve in $\mathbb{R}\times Y$ having a single nontrivial component C_Y, which satisfies $\mathrm{ind}(C_Y)=1$ and is such that $\mathcal{M}_Y^{J_Y}(C_Y)$ is regular with $\#\mathcal{M}_Y^{J_Y}(C_Y)/\mathbb{R}=0$, and
- bottom level having a single component C_X, which has index -1 and is simple.

By standard transversality techniques we can assume that C_X is regular in the parametrized sense.

We now invoke SFT gluing, using for instance the general formulation given in [32, Theorem 2.54]; see also [34, Section 2.5.3] for the simpler Morse homology analogue of our setting. For ease of discussion let us make the following simplifying assumptions:

- All of the additional boundary configurations occur at the same time $t_b \in (0,1)$.
- All of these configurations involve the same -1 component C_X.
- $\mathcal{M}_Y^{J_Y}(C_Y)/\mathbb{R}$ consists of just two elements $C_{Y,1}, C_{Y,2}$ that have opposite signs.

For $i=1,2$, gluing realizes the configuration $(C_{Y,i}, C_X)$ as an end of the moduli space $\mathcal{M}_X^{\{J_t\}}(C)$, with gluing applying for $|t-t_b|$ sufficiently small and either $t<t_b$ or $t>t_b$ (but not both). That is, an end of the moduli space $\mathcal{M}_X^{\{J_t\}}(C)$ with $\pm(t-t_b)>0$ is compactified by the point $(C_{Y,i}, C_X)$ at $t=t_b$, and it does not extend to $\pm(t-t_b)<0$.

We assume orientation choices have been made as in Section 5.2. Together with the canonical orientation on $[0,1]$ this induces an orientation on the one-dimensional manifold $\mathcal{M}_X^{\{J_t\}}(C)$, and hence also its compactification $\overline{\overline{\mathcal{M}}}{}_X^{\{J_t\}}(C)$, such that $\mathcal{M}_X^{J_0}(C)$ appears as a negative boundary component (ie its sign as a boundary point is the opposite of its sign coming from the orientation on $\mathcal{M}_X^{J_0}(C)$), and similarly $\mathcal{M}_X^{J_1}(C)$ appears as a positive boundary component. The curves $C_{Y,i}, C_X$ also inherit signs $\varepsilon(C_{Y,i}), \varepsilon(C_X) \in \{-1,1\}$, and by gluing compatibility the sign of each configuration $(C_{Y,i}, C_X)$ as a boundary point of $\overline{\overline{\mathcal{M}}}{}_X^{\{J_t\}}(C)$ matches the product sign $\varepsilon(C_{Y,i})\varepsilon(C_X)$. Concretely, the sign associated with the boundary orientation of a boundary point on an oriented one-manifold is positive or negative according to whether the orientation points in the outgoing or incoming direction, respectively. Since $C_{Y,1}$ and $C_{Y,2}$ have opposite signs, we have also $\varepsilon(C_{Y,1})\varepsilon(C_X) \neq \varepsilon(C_{Y,2})\varepsilon(C_X)$, and hence as boundary points the configurations $(C_{Y,1}, C_X)$ and $(C_{Y,2}, C_X)$ have opposite orientations. We then have four possibilities:

(i) One gluing applies for $t<t_b$ with the corresponding boundary point outgoing, while the other gluing applies for $t>t_b$ with the corresponding boundary point incoming.

(ii) One gluing applies for $t<t_b$ with the corresponding boundary point incoming, while the other gluing applies for $t>t_b$ with the corresponding boundary point outgoing.

(iii) Both gluings apply for $t<t_b$, with one corresponding boundary point incoming and the other outgoing.

(iv) Both gluings apply for $t>t_b$, with one corresponding boundary point incoming and the other outgoing.

In case (i), by following the cobordism we get a sign-preserving identification of $\mathcal{M}_X^{J_1}(C)$ with $\mathcal{M}_X^{J_0}(C)$; case (ii) is similar. In case (iii), we get a sign-preserving identification of $\mathcal{M}_X^{J_0}(C)$ with $\mathcal{M}_X^{J_1}(C)$, plus two extra points of opposite signs; case (iv) is similar. In any case, $\#\mathcal{M}_X^{J_0}(C) = \#\mathcal{M}_X^{J_1}(C)$. □

Remark 2.4.3 One could imagine defining a weaker condition than Definition 2.4.1, which is neither formal in X nor in Y. However, this would not suffice for our proof of stabilization (see Section 3.7), since a priori there could be certain bad degenerations which are ruled out in dimension four for reasons which do not carry over to higher dimensions.

One could also imagine defining a stronger condition, which is formal in both X and Y. However, this would be insufficient for our study of convex toric domains, since "low-energy cylinders" joining an elliptic to a corresponding hyperbolic orbit always occur in the perturbed full rounding $\mathbb{R} \times \partial \widetilde{X}_\Omega$; cf Lemma 5.1.3.

3 The capacity $\widetilde{\mathfrak{g}}_k$

In this section we define the main object of study in this paper and establish some of its fundamental properties, in particular proving Theorem 1.2.2. In Section 3.1 we give the precise definition of $\widetilde{\mathfrak{g}}_k$ and point out its invariance properties. We then briefly compare $\widetilde{\mathfrak{g}}_k$ with its SFT analogue in Section 3.4. Sections 3.2 and 3.5 cover the symplectic embedding monotonicity and closed-curve upper bound properties, while the proof of the stabilization property occupies Sections 3.6 and 3.7.

3.1 Definition and basic properties

Given a Liouville domain (X, λ) and a positive constant $c \in \mathbb{R}_{>0}$, we use the shorthand $c \cdot X$ to denote the Liouville domain $(X, c\lambda)$.

Definition 3.1.1 Let X be a Liouville domain with nondegenerate contact boundary, and let D be a smooth local symplectic divisor passing through $p \in \operatorname{Int} X$. We put

$$\widetilde{\mathfrak{g}}_k(X) := \sup_{J \in \mathcal{J}(X;D)} \inf_{\Gamma} \mathcal{A}_{\partial X}(\Gamma),$$

where the infimum is over all tuples $\Gamma = (\gamma_1, \dots, \gamma_b)$ of Reeb orbits such that

$$\overline{\overline{\mathcal{M}}}{}_X^J(\Gamma)\lessdot\mathcal{T}^{(k)}p\gtrdot \neq \varnothing.$$

Here we put $\mathcal{A}_{\partial X}(\Gamma) := \sum_{i=1}^{a} \mathcal{A}_{\partial X}(\gamma_i)$, which is equivalently the energy of any curve with positive ends Γ. Recall that $\overline{\overline{\mathcal{M}}}{}_X^J(\Gamma)\lessdot\mathcal{T}^{(k)}p\gtrdot$ and $\mathcal{J}(X; D)$ are defined in Section 2.2. We emphasize that the moduli spaces $\mathcal{M}_X^J(\Gamma_i)\lessdot\mathcal{T}^{(k_i)}p\gtrdot$ are not required to be regular or to have index zero.

Remark 3.1.2 In Definition 3.1.1, we could alternatively put

$$\widetilde{\mathfrak{g}}_k(X) := \sup_{J \in \mathcal{J}(X;D)} \inf_{\Gamma_1,\dots,\Gamma_a} (\mathcal{A}_{\partial X}(\Gamma_1) + \cdots + \mathcal{A}_{\partial X}(\Gamma_a)),$$

where the infimum is over all tuples $\Gamma_1 = (\gamma_1^1, \dots, \gamma_{b_1}^1), \dots, \Gamma_a = (\gamma_1^a, \dots, \gamma_{b_a}^a)$ of Reeb orbits in ∂X for which the moduli spaces $\mathcal{M}_X^J(\Gamma_1)\langle\!\langle \mathcal{T}^{(k_1)} p\rangle\!\rangle, \dots, \mathcal{M}_X^J(\Gamma_a)\langle\!\langle \mathcal{T}^{(k_a)} p\rangle\!\rangle$ are nonempty and $k_1, \dots, k_a \in \mathbb{Z}_{\geq 0}$ satisfy $k_1 + \cdots + k_a \geq k$. This definition is equivalent and conceptually (if not notationally) cleaner. Indeed, consider some $C \in \overline{\overline{\mathcal{M}}}_X^J(\Gamma)\langle\!\langle \mathcal{T}^{(k)} p\rangle\!\rangle$. If the marked point z_0 mapping to p lies on a nonconstant component C_0, then we simply note that C_0 lies in $\mathcal{M}_X^J(\Gamma')\langle\!\langle \mathcal{T}^{(k)} p\rangle\!\rangle$ for some tuple of Reeb orbits Γ' satisfying $\mathcal{A}_{\partial X}(\Gamma') \leq \mathcal{A}_{\partial X}(\Gamma)$. On the other hand, if z_0 lies on a ghost component C_0, then as in Definition 2.2.1 we can consider the nearby nonconstant components $C_i \in \mathcal{M}_X^J(\Gamma_i)\langle\!\langle \mathcal{T}^{(k_i)} p\rangle\!\rangle$ for $i = 1, \dots, a$, and we necessarily have $\sum_{i=1}^a \mathcal{A}_{\partial X}(\Gamma_i) \leq \mathcal{A}_{\partial X}(\Gamma)$ and $\sum_{i=1}^a k_i \geq k$.

Conversely, any tuple of curves as above can viewed as an element of the compactified moduli space considered in Definition 3.1.1.

The quantity $\tilde{\mathfrak{g}}_k(X)$ is manifestly independent of any choice of almost complex structure, and the scaling property $\tilde{\mathfrak{g}}_k(X, \mu\omega) = \mu\tilde{\mathfrak{g}}_k(X, \omega)$ is immediate from the corresponding property for symplectic action. The nondecreasing property $\tilde{\mathfrak{g}}_1 \leq \tilde{\mathfrak{g}}_2 \leq \tilde{\mathfrak{g}}_3 \leq \cdots$ also follows directly, since by definition any curve satisfying the constraint $\langle\!\langle \mathcal{T}^{(k)} p\rangle\!\rangle$ for $k \in \mathbb{Z}_{\geq 2}$ also satisfies the constraint $\langle\!\langle \mathcal{T}^{(k-1)} p\rangle\!\rangle$. Note that the subadditivity property in Theorem 1.2.2 is also immediate from Definition 3.1.1.

A priori $\tilde{\mathfrak{g}}_k$ does depend on the choice of local divisor D, but we have:

Lemma 3.1.3 *Let X be a Liouville domain with nondegenerate contact boundary. Then $\tilde{\mathfrak{g}}_k(X)$ is independent of the choice of point $p \in \operatorname{Int} X$ and the local divisor D.*

Proof If p and D are fixed, then there is a contractible family of choices for J_D. Further, given two local symplectic divisors D, D' near $p, p' \in \operatorname{Int} X$ respectively, using Moser's trick we can find a symplectomorphism $\Phi\colon X \to X$ which is the identity near ∂X and which maps the germ of D near p to the germ of D' near p'. This induces a bijection $\mathcal{J}(X; D) \xrightarrow{\approx} \mathcal{J}(X; D')$ sending J to $\Phi_* J := (d\Phi) \circ J \circ (d\Phi)^{-1}$, and we get a corresponding bijection

$$\overline{\overline{\mathcal{M}}}_X^J\langle\!\langle \mathcal{T}^{(k)} p\rangle\!\rangle(\Gamma) \xrightarrow{\approx} \overline{\overline{\mathcal{M}}}_X^{\Phi_* J}\langle\!\langle \mathcal{T}^{(k)} p\rangle\!\rangle(\Gamma)$$

sending C to $\Phi \circ C$. □

In the next subsection we prove that $\tilde{\mathfrak{g}}_k(X) \leq \tilde{\mathfrak{g}}_k(X')$ whenever X, X' are Liouville domains of the same dimension with nondegenerate contact boundaries for which there is a symplectic embedding $X \overset{s}{\hookrightarrow} X'$. Taking this on faith for the moment, we extend the definition of $\tilde{\mathfrak{g}}_k$ to all symplectic manifolds:

Definition 3.1.4 If M is any symplectic manifold, we put

$$\tilde{\mathfrak{g}}_k(M) := \sup_X \tilde{\mathfrak{g}}_k(X),$$

where the supremum is over all Liouville domains X with nondegenerate contact boundary for which there exists a symplectic embedding $X \overset{s}{\hookrightarrow} M$.

Evidently the above definition is consistent with Definition 3.1.1 when X is a Liouville domain with nondegenerate contact boundary (assuming Proposition 3.2.1 below). It is also immediate that $\tilde{\mathfrak{g}}_k(M)$ is a symplectomorphism invariant; in particular, in the case of a Liouville domain (X,λ), $\tilde{\mathfrak{g}}_k(X)$ depends on the symplectic form $d\lambda$ but not on its primitive λ.

Remark 3.1.5 (local tangency versus skinny ellipsoidal constraints) In light of Section 2.2, to first approximation we can trade (at least in dimension four) the local tangency constraint $\lll \mathcal{T}^{(m)}p \ggg$ in Definition 3.1.1 with a skinny ellipsoidal constraint $\lll (m) \ggg_E$. However, the resulting invariant is not immediately equivalent without additional assumptions, and in fact our proof of monotonicity in Section 3.2 does not a priori apply to skinny ellipsoidal constraints due to the possibility of extra negative ends which bound pseudoholomorphic planes in lower levels. Nevertheless, it will be fruitful to utilize skinny ellipsoidal constraints in Section 5 when computing $\tilde{\mathfrak{g}}_k$ for convex toric domains, and in that setting the relevant moduli spaces are sufficiently nice that Proposition 2.2.3 applies.

3.2 Monotonicity under symplectic embeddings

Proposition 3.2.1 *Let X and X' be Liouville domains of the same dimension with nondegenerate contact boundaries, and suppose there is a symplectic embedding $X \overset{s}{\hookrightarrow} \operatorname{Int} X'$. Then for $k \in \mathbb{Z}_{\geq 1}$ we have $\tilde{\mathfrak{g}}_k(X) \leq \tilde{\mathfrak{g}}_k(X')$.*

Proof Let $\iota\colon X \overset{s}{\hookrightarrow} \operatorname{Int} X'$ be a symplectic embedding, let D be a local symplectic divisor near $p \in \operatorname{Int} X$, and put $p' := \iota(p)$ and $D' := \iota(D)$. Given $J \in \mathcal{J}(X;D)$, let $J' \in \mathcal{J}(X',D')$ be an admissible almost complex structure on $\hat{X}'$ which restricts to $\iota_* J$ on $\iota(X)$. Let $\{J'_t\}_{t\in[0,1)}$ be a family of almost complex structures in $\mathcal{J}(X';D')$ which realizes neck stretching along $\partial\iota(X)$, with $J'_0 = J'$. By definition of $\tilde{\mathfrak{g}}_k(X')$, for each $t \in [0,1)$ there is some collection of Reeb orbits $\Gamma^t = (\gamma_1^t,\dots,\gamma_k^t)$ in $\partial X'$ satisfying $\mathcal{A}_{\partial X'}(\Gamma^t) \leq \tilde{\mathfrak{g}}_k(X')$ and $\overline{\overline{\mathcal{M}}}_{X'}^{J'_t}(\Gamma^t)\lll\mathcal{T}^{(k)}p'\ggg \neq \varnothing$. Since $\partial X'$ has nondegenerate Reeb orbits, there are only finitely many Reeb orbits of action less than any given value, and hence we can find an increasing sequence $t_1,t_2,t_3,\ldots \in [0,1)$ with $\lim_{t\to\infty} t_i = 1$ such that $\Gamma^{t_i} = \Gamma^{t_1}$ is independent of i. By the SFT compactness theorem there is some element in the compactified moduli space $\overline{\overline{\mathcal{M}}}_{X'}^{\{J'_t\}}(\Gamma^{t_1})\lll\mathcal{T}^{(k)}p'\ggg$ corresponding to $t=1$. This is a pseudoholomorphic building in the broken cobordism $X \circledcirc (X' \setminus X)$, and in particular by looking at the components mapping to X we get an element in $\overline{\overline{\mathcal{M}}}_{X}^{J}(\Gamma^{t_1})\lll\mathcal{T}^{(k)}p\ggg$ with energy at most $\tilde{\mathfrak{g}}_k(X')$. Since J was arbitrary, we then have $\tilde{\mathfrak{g}}_k(X) \leq \tilde{\mathfrak{g}}_k(X')$. □

Remark 3.2.2 Fix any $J_{\partial X'} \in \mathcal{J}(\partial X')$, and put

$$\tilde{\mathfrak{g}}_k^{J_{\partial X'}}(X') := \sup_{J_{X'} \in \mathcal{J}^{J_{\partial X'}}(X')} \inf_{\Gamma} \mathcal{A}_{\partial X'}(\Gamma),$$

the infimum taken over all tuples $\Gamma = (\gamma_1,\dots,\gamma_a)$ of Reeb orbits in $\partial X'$ for which $\overline{\overline{\mathcal{M}}}_{X'}^{J_{X'}}(\Gamma)\lll\mathcal{T}^{(k)}p\ggg \neq \varnothing$. In other words, $\tilde{\mathfrak{g}}_k^{J_{\partial X'}}(X)$ is defined just like $\tilde{\mathfrak{g}}_k(X')$ except that we take the supremum over almost complex structures having fixed form on the cylindrical end. Then the above proof actually shows that we have $\tilde{\mathfrak{g}}_k(X) \leq \tilde{\mathfrak{g}}_k^{J_{\partial X'}}(X')$.

As a consequence of the above remark, by considering symplectic embeddings of X into a slight enlargement of itself we have:

Corollary 3.2.3 *For any Liouville domain X with nondegenerate contact boundary and any $J_{\partial X} \in \mathcal{J}(\partial X)$, we have $\widetilde{\mathfrak{g}}_k^{J_{\partial X}}(X) = \widetilde{\mathfrak{g}}_k(X)$.*

The symplectic embedding monotonicity property of Theorem 1.2.2 is now an immediate consequence of Proposition 3.2.1 and Definition 3.1.4:

Corollary 3.2.4 *If M and M' are symplectic manifolds of the same dimension with a symplectic embedding $M \overset{s}{\hookrightarrow} M'$, then we have $\widetilde{\mathfrak{g}}_k(M) \le \widetilde{\mathfrak{g}}_k(M')$ for any $k \in \mathbb{Z}_{\ge 1}$.*

Remark 3.2.5 (i) By a standard observation, it also follows that $\widetilde{\mathfrak{g}}_k$ is continuous with respect to C^0 deformations of X within $\widehat{X}$.

(ii) One could also in principle directly extend Definition 3.1.1 to include all (not necessarily exact) symplectic fillings with nondegenerate contact boundary. However, a priori our proof of Proposition 3.2.1 does not extend, since in principle there could be infinitely many homology classes with bounded energy.

3.3 Word-length filtration

As in [37], we can also define a refinement $\widetilde{\mathfrak{g}}_k^{\le l}$ of $\widetilde{\mathfrak{g}}_k$ for any $k, l \in \mathbb{Z}_{\ge 1}$ by restricting the allowed number of positive ends. This gives a more general framework, which includes, at least for four-dimensional convex toric domains, both $\{\widetilde{\mathfrak{g}}_k\}$ and $\{c_k^{\mathrm{GH}}\}$ as special cases; see Section 5.6 for more details.

Definition 3.3.1 Let X be a Liouville domain with nondegenerate contact boundary, and let D be a smooth local symplectic divisor passing through $p \in \operatorname{Int} X$. We put

$$\widetilde{\mathfrak{g}}_k^{\le l}(X) := \sup_{J \in \mathcal{J}(X;D)} \inf_{\Gamma} \mathcal{A}_{\partial X}(\Gamma),$$

the infimum taken over all tuples $\Gamma = (\gamma_1, \dots, \gamma_a)$ of Reeb orbits in ∂X for which $\overline{\overline{\mathcal{M}}}{}_X^J(\Gamma)\langle\!\langle \mathcal{T}^{(k)} p \rangle\!\rangle \ne \varnothing$, and such that $a \le l$.

With only minor modifications, our proof of Theorem 1.2.2 also gives the following:

Theorem 3.3.2 *For each $k, l \in \mathbb{Z}_{\ge 1}$, $\widetilde{\mathfrak{g}}_k^{\le l}$ is independent of the choice of local divisor and is a symplectomorphism invariant. It satisfies the following properties:*

- ***Scaling*** *It scales like area, ie $\widetilde{\mathfrak{g}}_k^{\le l}(M, \mu\omega) = \mu \widetilde{\mathfrak{g}}_k^{\le l}(M, \omega)$ for any symplectic manifold (M, ω) and $\mu \in \mathbb{R}_{>0}$.*
- ***Nondecreasing*** *We have $\widetilde{\mathfrak{g}}_1^{\le l}(M) \le \widetilde{\mathfrak{g}}_2^{\le l}(M) \le \widetilde{\mathfrak{g}}_3^{\le l}(M) \le \cdots$ for any symplectic manifold M.*

- ***Generalized Liouville embedding monotonicity*** *Given a pair of equidimensional Liouville domains X and X', and a generalized Liouville embedding of X into X'* (see Remark 1.2.5), *we have $\widetilde{\mathfrak{g}}_k^{\leq l}(X) \leq \widetilde{\mathfrak{g}}_k^{\leq l}(X')$.*
- ***Stabilization*** *For any Liouville domain X we have $\widetilde{\mathfrak{g}}_k(X \times B^2(c)) = \widetilde{\mathfrak{g}}_k(X)$ for any $c \geq \widetilde{\mathfrak{g}}_k(X)$, provided that the hypotheses of Proposition 3.7.1, substituting $\widetilde{\mathfrak{g}}_k$ with $\widetilde{\mathfrak{g}}_k^{\leq l}$, are satisfied.*

Compared with Theorem 1.2.2, for a general symplectic embedding $X \overset{s}{\hookrightarrow} X'$ there may be curves in $X' \setminus X$ having no positive ends, and a curve with l positive ends in X' may produce a curve in X with a greater number of positive ends after neck stretching, since the top of the limiting building might contain a component with no positive ends. Generalized Liouville embeddings carry an additional an exactness condition which precisely rules out curves in $X' \setminus X$ without positive ends via Stokes' theorem.

Note that if $X^{2n\geq 4}$ is a star-shaped domain then a symplectic embedding $X \overset{s}{\hookrightarrow} X'$ is automatically a generalized Liouville embedding, but this does not necessarily extend to cases with $H^1(\partial X; \mathbb{R})$ nontrivial. Moreover, if ∂X has no contractible Reeb orbits then we have $\widetilde{\mathfrak{g}}_k^{\leq 1}(X) = c_k^{\mathrm{GH}}(X) = \infty$, and hence these capacities contain no quantitative information; $\widetilde{\mathfrak{g}}_k^{\leq l}(X)$ is more often finite for l sufficiently large.

3.4 Comparison with SFT counterpart

At first glance the definitions of $\widetilde{\mathfrak{g}}_k$ and $\mathfrak{g}_k$ look rather different, despite involving the same types of curves. Recall that $\mathfrak{g}_k(X)$ is defined in [37] using the $\mathscr{L}_\infty$ algebra structure on the linearized contact homology chain complex $\mathrm{CH}_{\mathrm{lin}}(X)$ of a Liouville domain X, along with the induced $\mathscr{L}_\infty$ homomorphism $\varepsilon_{\mathrm{lin}}\langle\!\langle\mathcal{T}^{(k)}p\rangle\!\rangle : \mathrm{CH}_{\mathrm{lin}}(X) \to \mathbb{K}$ defined by counting rational curves with a local tangency constraint $\langle\!\langle\mathcal{T}^{(k)}p\rangle\!\rangle$. In brief, $\mathfrak{g}_k(X)$ is the minimal action of an element of the bar complex $\mathcal{B}\mathrm{CH}_{\mathrm{lin}}(X)$ which is closed under the bar differential and whose image under the chain map $\mathcal{B}\mathrm{CH}_{\mathrm{lin}}(X) \to \mathbb{K}$ induced by $\varepsilon_{\mathrm{lin}}\langle\!\langle\mathcal{T}^{(k)}p\rangle\!\rangle$ is nonzero. Here $\mathcal{B}\mathrm{CH}_{\mathrm{lin}}(X)$ as a vector space is the (appropriately graded) symmetric tensor algebra on the vector space $\mathrm{CH}_{\mathrm{lin}}(X)$ spanned by good Reeb orbits in ∂X, and the bar differential is built out of the $\mathscr{L}_\infty$ structure maps ℓ^1, ℓ^2 and ℓ^3 which count pseudoholomorphic buildings in $\mathbb{R} \times \partial X$, anchored in X, with one negative and several positive ends. In particular, this definition of $\mathfrak{g}_k(X)$ typically requires virtual perturbations in order to set up the chain complex $\mathrm{CH}_{\mathrm{lin}}(X)$ along with its $\mathscr{L}_\infty$ structure, and its basic invariance and structural properties follow naturally from SFT functoriality.

The precise virtual perturbation framework is not important for our present discussion, but we mention two important axioms:

(a) a structure coefficient can only be nonzero if the corresponding SFT compactified moduli space is nonempty, and

(b) if the naive pseudoholomorphic curve count for a given structure coefficient is already regular and there are other representatives in its corresponding SFT compactified moduli space, then this count remains valid after turning on virtual perturbations.

It is then easy to deduce that $\tilde{\mathfrak{g}}_k(X) \le \mathfrak{g}_k(X)$ for any Liouville domain X. Indeed, for any J, by (a) and the definition of $\mathfrak{g}_k(X)$ there must be a pseudoholomorphic building $C \in \overline{\overline{\mathcal{M}}}{}^J_X(\Gamma)\lll\mathcal{T}^{(k)}p\ggg$ having total energy at most $\mathfrak{g}_k(X)$. Since J is arbitrary, we therefore have $\tilde{\mathfrak{g}}_k(X) \le \mathfrak{g}_k(X)$.

In principle we could have $\tilde{\mathfrak{g}}_k(X) < \mathfrak{g}_k(X)$, if all curves in $\widehat{X}$ with energy $\tilde{\mathfrak{g}}_k(X)$ are undetected by $\mathfrak{g}_k(X)$. However, this cannot occur if $\tilde{\mathfrak{g}}_k(X)$ is carried by a suitably nice moduli space, eg as in Proposition 3.7.1. In particular, it follows from the results of this paper that $\tilde{\mathfrak{g}}_k(X) = \mathfrak{g}_k(X)$ whenever X is a four-dimensional convex toric domain; we are not currently aware of any Liouville domain X for which $\tilde{\mathfrak{g}}_k(X) \neq \mathfrak{g}_k(X)$.

3.5 Upper bounds from closed curves

Here we prove the closed curve upper bound part of Theorem 1.2.2. Recall from the introduction that $N_{M,A}\lll\mathcal{T}^{(k)}p\ggg \neq 0$ counted the number of curves in class A that are tangent to the local divisor D at p to order k.

Proposition 3.5.1 *If (M,ω) is a closed semipositive symplectic manifold satisfying $N_{M,A}\lll\mathcal{T}^{(k)}p\ggg \neq 0$ for some $A \in H_2(M)$, then we have $\tilde{\mathfrak{g}}_k(M) \le [\omega]\cdot A$.*

Proof This is quite similar to the proof of Proposition 3.2.1. It suffices to show that for any Liouville domain X with nondegenerate contact boundary which admits a symplectic embedding $\iota\colon X \overset{s}{\hookrightarrow} M$, we have $\tilde{\mathfrak{g}}_k(X) \le [\omega]\cdot A$. Given $J \in \mathcal{J}(X;D)$, we extend $\iota_* J$ to a compatible almost complex structure J' on M. Let $\{J_t\}_{t\in[0,1)}$ be a family of compatible almost complex structures on M realizing neck stretching along $\partial\iota(X)$, with $J_0 = J'$. Note that $\mathcal{M}^{J_t}_{M,A}\lll\mathcal{T}^{(k)}p\ggg$ is nonempty for all $t \in [0,1)$, since otherwise this moduli space would be empty and in particular regular, contradicting the invariance of $N_{M,A}\lll\mathcal{T}^{(k)}p\ggg$; see [30, Section 2.2]. Then, as in the proof of Proposition 3.2.1, the SFT compactness theorem implies that there must be a limiting building corresponding to $t = 1$, and in particular in the bottom level we can find $C \in \overline{\overline{\mathcal{M}}}{}^J_X(\Gamma)\lll\mathcal{T}^{(k)}p\ggg$ for some tuple of Reeb orbits satisfying $\mathcal{A}_{\partial X}(\Gamma) \le [\omega]\cdot A$. □

3.6 Stabilization lower bounds

Proposition 3.6.1 *For any Liouville domain X, we have $\tilde{\mathfrak{g}}_k(X \times B^2(c)) \ge \tilde{\mathfrak{g}}_k(X)$ for all $k \ge 1$ provided that $c \ge \tilde{\mathfrak{g}}_k(X)$.*

As a preliminary step, the next lemma allows us to identify the Reeb orbits after stabilizing (and suitably smoothing the corners) with those before stabilizing, plus additional orbits of large action. We denote by $\lambda_{\mathrm{std}} = \frac{1}{2}(x\,dy - y\,dx)$ the standard Liouville form on $B^2(c)$. Given a Liouville form λ, recall that the Liouville vector field V_λ is characterized by $d\lambda(V_\lambda, -) = \lambda$.

Suppose that (Y,α) is a strict contact manifold and $Z \subset Y$ is a submanifold of codimension 2 such that $\alpha|_Z$ is a contact form on Z and the Reeb vector field R_α is tangent to Z. Let $\xi_Y := \ker\alpha$ and $\xi_Z := \ker\alpha|_Z$ denote the contact hyperplane distributions of Y and Z, respectively. Since ξ_Z

is a subbundle of ξ_Y, we can consider its orthogonal complement $\xi_Z^{\perp}$ with respect to the symplectic form $d\alpha|_{\xi_Y}$. Let γ be a nondegenerate Reeb orbit of Y which lies in Z, and let τ be a trivialization of the symplectic vector bundle $\gamma^*\xi_Y$ which splits as $\tau = \tau_Z + \tau_Z^{\perp}$ with respect to the direct-sum decomposition $\xi_Y = \xi_Z \oplus \xi_Z^{\perp}$. Since the latter decomposition is also preserved by the linearized Reeb flow of Y along γ, the trivialization $\tau_Z^{\perp}$ in the normal direction identifies the linearized Reeb flow along $\gamma^*\xi_Z^{\perp}$ with a loop of 2×2 symplectic matrices which starts at the identity and ends at a matrix without 1 as an eigenvalue. Such a loop has a well-defined Conley–Zehnder index, called the *normal Conley–Zehnder index* of γ, denoted by $\mathrm{CZ}^{\perp}_{\tau_Z^{\perp}}(\gamma)$.

In the following we show that Reeb orbits of ∂X can be viewed as Reeb orbits in a suitable smoothing of $\partial(X \times B^2(c))$, and we apply the above discussion with Y given by the smoothing of $\partial(X \times B^2(c))$ and Z given by ∂X. In this situation, there is a canonical trivialization of $\xi_Z^{\perp}$, coming from its identification with the normal bundle of $Z \subset Y$, which in turn is naturally identified with the restriction to Z of $\{0\} \times TB^2(c) \subset TX \times TB^2(c)$. By default we will always measure normal Conley–Zehnder indices by working with a split trivialization $\tau = \tau_Z + \tau_Z^{\perp}$ of $\gamma^*\xi_Y$, where $\tau_Z^{\perp}$ comes from this canonical trivialization of $\xi_Z^{\perp}$.

Lemma 3.6.2 *Let (X, λ) be a Liouville domain. For any $c, \epsilon \in \mathbb{R}_{>0}$, there is a subdomain with smooth boundary $\widetilde{X} \subset X \times B^2(c)$ such that*

- *the Liouville vector field $V_\lambda + V_{\lambda_{\mathrm{std}}}$ is outwardly transverse along $\partial\widetilde{X}$,*
- *$X \times \{0\} \subset \widetilde{X}$ and the Reeb vector field of $\partial\widetilde{X}$ is tangent to $\partial X \times \{0\}$, and*
- *any Reeb orbit of the contact form $(\lambda + \lambda_{\mathrm{std}})|_{\partial\widetilde{X}}$ with action less than $c - \epsilon$ is entirely contained in $\partial X \times \{0\}$ and has normal Conley–Zehnder index equal to 1.*

Proof For notational convenience put $X_1 := X$ and $X_2 := B^2(c)$. For $i = 1, 2$ we denote the associated Liouville forms by λ_i, the associated contact forms by $\alpha_i := \lambda_i|_{\partial X_i}$, and the associated Liouville vector fields by V_{λ_i}. Note that every closed Reeb orbit of ∂X_2 has action at least c.

Recall that we can use the Liouville flow to identify a collar neighborhood U_i of ∂X_i with $(-\eta, 0] \times \partial X_i$ for some small $\eta > 0$, and under this identification we have $\lambda_i = e^{r_i}\alpha_i$, where r_i denotes the coordinate on the first factor. Given a smooth function $H_i \colon (-\eta, 0] \times \partial X_i \to \mathbb{R}$ of the form $H(r_i, y_i) = h(e^{r_i})$ for some $h_i \colon (e^{-\eta}, 1] \to \mathbb{R}$, the Hamiltonian vector field takes the form $\mathscr{X}_{H_i} = h_i'(e^{r_i})R_{\alpha_i}$, where R_{α_i} is the Reeb vector field of α_i. Note that for such a Hamiltonian we have $V_{\lambda_i}(H_i) = \lambda_i(\mathscr{X}_{H_i}) = e^{r_i}h_i'(e^{r_i})$.

By considering functions which depend only on the Liouville flow coordinate r_i near the boundary and are otherwise sufficiently small, we can find smooth functions $H_i \colon X_i \to [0, 1]$ for $i = 1, 2$ such that

(a) $\partial X_i = H_i^{-1}(1)$ is a regular level set,

(b) $H_i^{-1}(0) = \{p_i\}$ is a nondegenerate minimum, where we assume $p_2 = 0 \in B^2(c)$,

(c) on $U_i \approx (-\eta, 0] \times \partial X_i$ we have $H_i(r_i, y_i) = h_i(e^{r_i})$ for some $h_i \colon (e^{-\eta}, 1] \to [0, 1]$ with $h_i' > 0$,

(d) on $X_i \setminus U_i$ we have $|V_{\lambda_i}(H_i)| < \frac{1}{2}\epsilon$, and

(e) we have $H_i^{-1}([\delta,1]) \subset U_i$ for some small $\delta > 0$, and on $H_i^{-1}([\delta,1])$ we have $V_{\lambda_i}(H_i) > c + \epsilon$.

We can further arrange that

(f) $V_{\lambda_2}(H_2) > 0$ on $B^2(c) \setminus \{0\}$,

(g) for every T–periodic γ_2 orbit of $\mathscr{X}_{H_2}$ with $T \le 1$, we have $\int_{\gamma_2} \lambda_2 > c - \frac{1}{2}\epsilon$, and

(h) using standard symplectic coordinates x, y, on a small neighborhood of $0 \in B^2(c)$ we have
$$H_2(x,y) = \tfrac{1}{2}\rho(x^2+y^2),$$
with $\rho < \pi$.

Put $\widetilde{X} := \{(x_1,x_2) \in X_1 \times X_2 \mid H_1(x_1) + H_2(x_2) \le 1\}$. It follows from the above properties that $\widetilde{X}$ has smooth boundary, and we have
$$(V_{\lambda_1} + V_{\lambda_2})(H_1 + H_2) > 0 \quad \text{along } \partial\widetilde{X}.$$

Indeed, consider $(x_1,x_2) \in \partial\widetilde{X}$. Suppose first that $x_1 \in U_1$. Then we have $(V_{\lambda_1})_{x_1}(H_1) = e^{r_1}h_1'(e^{r_1}) > 0$ by (c) and $(V_{\lambda_2})_{x_2}(H_2) \ge 0$ by (f). On the other hand, if $x_1 \in X_1 \setminus U_1$, then we must have $H_1(x_1) \in [0,\delta]$ by (e) and $|(V_{\lambda_1})_{x_1}(H_1)| < \frac{1}{2}\epsilon$ by (d). In this case we have $H_2(x_2) = 1 - H_1(x_1) \in [1-\delta,1]$, whence $(V_{\lambda_2})_{x_2}(H_2) > c + \epsilon$ and therefore $(V_{\lambda_1})_{x_1}(H_1) + (V_{\lambda_2})_{x_2}(H_2) > 0$.

It follows from the above discussion that $\lambda + \lambda_{\mathrm{std}}$ is a Liouville form on $\widetilde{X}$, and in particular it restricts to a positive contact form on $\partial\widetilde{X}$. Observe that the corresponding Reeb vector field is at each point in $\partial\widetilde{X}$ proportional to the Hamiltonian vector field of $H_1 + H_2$. In particular, this is tangent to $\partial(X \times \{0\})$, since along $\partial(X \times \{0\})$ we have $\mathscr{X}_{H_2} \equiv 0$.

We now prove the assertion about actions of Reeb orbits. Suppose that γ is a T–periodic Reeb orbit of $\partial\widetilde{X}$ for some $T \in \mathbb{R}_{>0}$. Let γ_i denote its projection to X_i for $i = 1,2$. Note that we have $\gamma_i \subset H_i^{-1}(C_i)$ for some $C_i \in [0,1]$ with $C_1 + C_2 = 1$. If γ_2 is constant, then γ lies in $X_1 \times \{0\}$. Otherwise, if $C_1 \in [0,\delta]$, then $C_2 \in [1-\delta,1]$, and we have $\left|\int_{\lambda_1} \gamma_1\right| < \frac{1}{2}T\epsilon$ by (d) and $\int_{\lambda_2} \gamma_2 > \max\bigl(c - \frac{1}{2}\epsilon, T(c+\epsilon)\bigr)$ by (g), and therefore we have
$$\int_\gamma \lambda = \int_{\gamma_1} \lambda_1 + \int_{\gamma_2} \lambda_2 > \max\bigl(c - \tfrac{1}{2}\epsilon, T(c+\epsilon)\bigr) - \tfrac{1}{2}T\epsilon > c - \epsilon.$$

Lastly, if $C_1 \in [\delta,1]$ and γ_2 is not constant, then if $T \ge 1$ we have
$$\int_\gamma \lambda \ge \int_{\gamma_1} \lambda_1 > T(c+\epsilon) > c - \epsilon,$$
whereas if $T < 1$ then we have
$$\int_\gamma \lambda \ge \int_{\gamma_2} \lambda_2 > c - \tfrac{1}{2}\epsilon > c - \epsilon.$$

As for the assertion about normal Conley–Zehnder indices, suppose that γ is a Reeb orbit in $\partial(X \times \{0\})$ with action $T \le c$. Observe that Reeb vector field on $\partial \widetilde{X}$ is given by

$$\frac{1}{\lambda_1(\mathscr{X}_{H_1}) + \lambda_2(\mathscr{X}_{H_2})}(\mathscr{X}_{H_1} + \mathscr{X}_{H_2}),$$

and along $\partial(X \times \{0\})$ we have $\lambda_1(\mathscr{X}_{H_1}) > c + \epsilon$ and $\lambda_2(\mathscr{X}_{H_2}) = 0$. We can therefore identify the linearized Reeb flow along γ in the normal direction with the time-T linearized Hamiltonian flow of $(1/\lambda_1(\mathscr{X}_{H_1}))\mathscr{X}_{H_2}$ at 0. By design, this is rotation by the angle $T\rho/\lambda_1(\mathscr{X}_{H_1})$. In particular, the Conley–Zehnder contribution for each factor is 1 provided that we have $T\rho/\lambda_1(\mathscr{X}_{H_1}) < \pi$, for which $\rho < \pi$ suffices. $\square$

In the sequel, we will denote any Liouville domain $\widetilde{X}$ satisfying the properties of Lemma 3.6.2 for some $\epsilon > 0$ sufficiently small by $X \mathbin{\widetilde{\times}} B^2(c)$.

Lemma 3.6.3 *Let X be a Liouville domain, and let $X \mathbin{\widetilde{\times}} B^2(c)$ be a smoothing of $X \times B^2(c)$ as in Lemma 3.6.2.*

(i) *Let $J \in \mathcal{J}(X \mathbin{\widetilde{\times}} B^2(c))$ be an admissible almost complex structure on the symplectic completion of $X \mathbin{\widetilde{\times}} B^2(c)$ for which $\widehat{X} \times \{0\}$ is J–holomorphic. Let C be an asymptotically cylindrical J–holomorphic curve in $\widehat{X}$, all of whose asymptotic Reeb orbits are nondegenerate and lie in $\partial X \times \{0\}$ with normal Conley–Zehnder index 1. Then C is either disjoint from the slice $\widehat{X} \times \{0\}$ or entirely contained in it.*

(ii) *Let $J \in \mathcal{J}(\partial(X \mathbin{\widetilde{\times}} B^2(c)))$ be an admissible almost complex structure on the symplectization of $\partial(X \mathbin{\widetilde{\times}} B^2(c))$ for which $\mathbb{R} \times \partial X \times \{0\}$ is J–holomorphic. Let C be an asymptotically cylindrical J–holomorphic curve in $\mathbb{R} \times \partial(X \mathbin{\widetilde{\times}} B^2(c))$, all of whose asymptotic Reeb orbits are nondegenerate and lie in $\partial X \times \{0\}$ with normal Conley–Zehnder index 1. Then C is either disjoint from the slice $\mathbb{R} \times \partial X \times \{0\}$ or entirely contained in it. Moreover, only the latter is possible of C has at least one negative puncture.*

To prove Lemma 3.6.3, we invoke the higher-dimensional extension of [35]; compare with the exposition in [31, Section 2]. Namely, let C be an asymptotically cylindrical curve in the symplectic completion of $X \mathbin{\widetilde{\times}} B^2(c)$ or the symplectization of $\partial(X \mathbin{\widetilde{\times}} B^2(c))$, and let Q denote the divisor $\widehat{X} \times \{0\}$ or $\mathbb{R} \times \partial(X \times \{0\})$, respectively. Assume that each puncture of C is asymptotic to a nondegenerate Reeb orbit in $\partial X \times \{0\}$, and that C is not entirely contained in Q. For each puncture z of C, we can consider the corresponding asymptotic winding number wind_z around Q as we approach the puncture, as measured by the canonical trivialization discussed in the leadup to Lemma 3.6.2.

We will need the following facts:

(a) The curve C intersects Q in only finitely many points, each of which has a positive local intersection number.

(b) If z is a positive puncture and γ_z is the corresponding asymptotic Reeb orbit, then we have $\mathrm{wind}_z \le \lfloor \frac{1}{2}\mathrm{CZ}^{\perp}(\gamma_z) \rfloor$.

(c) If z is a negative puncture and γ_z is the corresponding asymptotic Reeb orbit, then we have $\mathrm{wind}_z \ge \lceil \frac{1}{2}\mathrm{CZ}^{\perp}(\gamma_z) \rceil$.

(d) We have

$$\mathrm{push}(C)\cdot Q = C\cdot Q - \sum_{\substack{z \text{ positive}\\ \text{puncture}}} \mathrm{wind}_z + \sum_{\substack{z \text{ negative}\\ \text{puncture}}} \mathrm{wind}_z, \tag{3-6-1}$$

where $\mathrm{push}(C)$ is a pushoff of C whose direction near each puncture is a nonzero constant with respect to the canonical trivialization of the normal bundle.

Here $C\cdot Q$ and $\mathrm{push}(C)\cdot Q$ denote homological intersection numbers, ie the sum of local homological intersection numbers over all (necessarily finitely many) intersection points. In particular, we have $\mathrm{push}(C)\cdot Q = 0$ since there is an obvious displacement of C from Q which takes the specified form near each of the punctures.

The last fact (d) is elementary topology. The proof of (a) follows from an asymptotic description of C in the normal direction near each puncture, which is written in terms of an eigenfunction of the corresponding normal asymptotic operator. Properties (b) and (c) follow from a characterization of normal Conley–Zehnder indices in terms of the corresponding normal asymptotic operators, together with bounds on the winding numbers of their eigenfunctions.

Proof of Lemma 3.6.3 To prove (i), suppose that C is not contained in $Q := \widehat{X}\times\{0\}$. Since each puncture of C is positively asymptotic to a Reeb orbit in $\partial X\times\{0\}$ with normal Conley–Zehnder index 1, using (3-6-1) and (b) we have

$$0 = \mathrm{push}(C)\cdot Q = C\cdot Q - \sum_{\substack{z \text{ positive}\\ \text{puncture}}} \mathrm{wind}_z \ge C\cdot Q - \sum_{\substack{z \text{ positive}\\ \text{puncture}}} \lfloor \tfrac{1}{2} \rfloor = C\cdot Q,$$

and hence $C\cdot Q \le 0$. Since each local intersection between C and Q counts positively, this is only possible if C is disjoint from Q.

The proof of (ii) is similar. Assume that C is not contained in $Q := \mathbb{R}\times\partial X\times\{0\}$. Using (3-6-1) we have

$$0 \ge C\cdot Q - \sum_{\substack{z \text{ positive}\\ \text{puncture}}} \lfloor \tfrac{1}{2} \rfloor + \sum_{\substack{z \text{ negative}\\ \text{puncture}}} \lceil \tfrac{1}{2} \rceil = C\cdot Q + \sum_{\substack{z \text{ negative}\\ \text{puncture}}} 1.$$

This is only possible if C has no negative punctures and C is disjoint from Q. $\square$

Proof of Proposition 3.6.1 We can assume $c > \widetilde{\mathfrak{g}}_k(X)$ and that ∂X is nondegenerate, since then the result follows by continuity; cf Remark 3.2.5(i). Let $X \mathbin{\widetilde{\times}} B^2(c)$ be a smoothing of $X\times B^2(c)$ as in Lemma 3.6.2, with $\epsilon > 0$ chosen sufficiently small so that $c-\epsilon > \widetilde{\mathfrak{g}}_k(X)$. Let D be a local divisor near $p\in \operatorname{Int} X$, and let us take the local divisor $\widetilde{D}$ in $X \mathbin{\widetilde{\times}} B^2(c)$ near $\widetilde{p} := (p,0)$ to be of the form $D\times B^2(\delta) \subset X \mathbin{\widetilde{\times}} B^2(c)$ for some small $\delta > 0$.

Let $J_X \in \mathcal{J}(X; D)$ be such that for every tuple of Reeb orbits Γ satisfying $\overline{\overline{\mathcal{M}}}_X^{J_X}(\Gamma)\langle\mathcal{T}^{(k)}p\rangle \neq \varnothing$, we have $\mathcal{A}_{\partial X}(\Gamma) \geq \widetilde{\mathfrak{g}}_k(X)$. Pick $\widetilde{J} \in \mathcal{J}(X \mathbin{\widehat{\times}} B^2(c); \widetilde{D})$ such that $\widehat{X} \times \{0\}$ is $\widetilde{J}$–holomorphic with $\widetilde{J}|_{\widehat{X}\times\{0\}} = J_X$. It suffices to show that for any tuple of Reeb orbits Γ' for which $\overline{\overline{\mathcal{M}}}_{X \mathbin{\widehat{\times}} B^2(c)}^{\widetilde{J}}(\Gamma')\langle\mathcal{T}^{(k)}p\rangle \neq \varnothing$, we have $\mathcal{A}_{\partial(X \mathbin{\widehat{\times}} B^2(c))}(\Gamma') \geq \widetilde{\mathfrak{g}}_k(X)$, since then we have

$$\widetilde{\mathfrak{g}}_k(X \times B^2(c)) \geq \widetilde{\mathfrak{g}}_k(X \mathbin{\widehat{\times}} B^2(c)) \geq \widetilde{\mathfrak{g}}_k(X).$$

Consider $C \in \overline{\overline{\mathcal{M}}}_{X \mathbin{\widehat{\times}} B^2(c)}^{\widetilde{J}}(\Gamma')\langle\mathcal{T}^{(k)}p\rangle$. For some $a \in \mathbb{Z}_{\geq 1}$, let $C_i \in \mathcal{M}_{X \mathbin{\widehat{\times}} B^2(c)}^{\widetilde{J}}(\Gamma_i)\langle\mathcal{T}^{(k_i)}p\rangle$ for $i = 1, \dots, a$ be nonconstant components of C with $\sum_{i=1}^a k_i \geq k$ and $\sum_{i=1}^a E(C_i) \leq E(C)$ as in Remark 3.1.2. We need to establish the bound $\sum_{i=1}^a E(C_i) \geq \widetilde{\mathfrak{g}}_k(X)$. If any positive end of some C_i is not asymptotic to the slice $\widehat{X} \times \{0\}$, then the corresponding Reeb orbit must have action at least $c - \epsilon$, and hence $E(C_i) \geq c - \epsilon > \widetilde{\mathfrak{g}}_k(X)$. Otherwise, by Lemma 3.6.3, each C_i must be entirely contained in $\widehat{X} \times \{0\}$ — note that it cannot be disjoint from the slice due to the local tangency constraint at $p \in \widehat{X} \times \{0\}$. By our choice of $\widetilde{D}$, each C_i then corresponds to a J_X–holomorphic curve in $\widehat{X}$ satisfying the constraint $\langle\mathcal{T}^{(k_i)}p\rangle$ with local divisor D, from which the desired bound readily follows. □

3.7 Stabilization upper bounds

In order to prove the stabilization property in Theorem 1.2.2, we need to complement Proposition 3.6.1 by proving an upper bound. Our proof will require some additional assumptions, which amount to saying that the capacity $\widetilde{\mathfrak{g}}_k(X)$ is represented by elements in a well-behaved moduli space of curves. Indeed, without such an assumption, after stabilizing and perturbing the almost complex structure it is conceivable that all curves with energy equal to $\widetilde{\mathfrak{g}}_k(X)$ disappear, resulting in $\widetilde{\mathfrak{g}}_k(X \times B^2(c)) > \widetilde{\mathfrak{g}}_k(X)$.

Proposition 3.7.1 *Let X be a Liouville domain, put $Y := \partial X$, and let C be a simple index-zero formal curve component in X with constraint $\langle\mathcal{T}^{(k)}p\rangle$ for some $k \in \mathbb{Z}_{\geq 1}$, such that $E_X(C) = \widetilde{\mathfrak{g}}_k(X)$. Assume further that the following conditions hold*:

(a) *C is formally perturbation invariant with respect to some generic $J_Y \in \mathcal{J}(Y)$ (cf Section 2.4).*

(b) *The moduli space $\mathcal{M}_X^{J_X}(C)$ is regular and finite with nonzero signed count $\#\mathcal{M}_X^{J_X}(C)$ for some $J_X \in \mathcal{J}^{J_Y}(X; D)$.*

Then we have $\widetilde{\mathfrak{g}}_k(X \times B^2(c)) \leq \widetilde{\mathfrak{g}}_k(X)$ for any $c \in \mathbb{R}_{>0}$. The same conclusion also holds if we instead assume that the hypotheses hold with k replaced by some divisor l of k such that $\widetilde{\mathfrak{g}}_k(X) = (k/l)\widetilde{\mathfrak{g}}_l(X)$.

The last part of Proposition 3.7.1 follows easily from the existence of multiple covers, or as a special case of subadditivity.

Proof By monotonicity of $\widetilde{\mathfrak{g}}_k$ under symplectic embeddings, it suffices to prove establish $\widetilde{\mathfrak{g}}_k(\widetilde{X}) \leq \widetilde{\mathfrak{g}}_k(X)$ for $\widetilde{X} := X \mathbin{\widehat{\times}} B^2(c)$ with c arbitrarily large. In particular, we can assume that any Reeb orbit in $\widetilde{Y} := \partial\widetilde{X}$ which is not contained in $Y \times \{0\}$ has action greater than $\widetilde{\mathfrak{g}}_k(X)$.

Let $J_{\widetilde{Y}} \in \mathcal{J}(\widetilde{Y})$ be an almost complex structure which agrees with J_Y on $\mathbb{R} \times Y \times \{0\}$. By Corollary 3.2.3 we have $\widetilde{\mathfrak{g}}_k(\widetilde{X}) = \widetilde{\mathfrak{g}}_k^{J_{\widetilde{Y}}}(\widetilde{X})$, so it suffices to prove $\widetilde{\mathfrak{g}}_k^{J_{\widetilde{Y}}}(\widetilde{X}) \leq \widetilde{\mathfrak{g}}_k(X)$.

Let $J_{\widetilde{X}} \in \mathcal{J}^{J_{\widetilde{Y}}}(\widetilde{X}; \widetilde{D})$ be an admissible almost complex structure which agrees with J_X on $\widehat{X} \times \{0\}$. Here we put $\widetilde{D} := D \times B^2(\delta)$, with $\delta > 0$ small as in the proof of Proposition 3.6.1. Since Reeb orbits of Y can also be viewed as Reeb orbits of $\widetilde{Y}$, C naturally corresponds to a formal curve component $\widetilde{C}$ in $\widetilde{X}$. Note that $\widetilde{C}$ is again simple and has index zero, the latter being a consequence of the index formula and the fact that the Reeb orbits of C have normal Conley–Zehnder index 1 by Lemma 3.6.2.

Moreover, we claim that $\widetilde{C}$ is formally perturbation invariant with respect to $J_{\widetilde{Y}}$. Indeed, let Γ (resp. $\widetilde{\Gamma}$) denote the positive asymptotic orbits of C (resp. $\widetilde{C}$), let A (resp. $\widetilde{A}$) denote its homology class, and let $\widetilde{C}' \in \overline{\overline{\mathcal{F}}}_{\widetilde{X},\widetilde{A}}(\widetilde{\Gamma})\lessdot\mathcal{T}^{(k)}p\gtrdot$ be a hypothetical stable formal building satisfying conditions (A1) and (A2) of Definition 2.4.1. By action considerations we can assume that each asymptotic Reeb orbit involved in $\widetilde{C}'$ lies in $Y \times \{0\}$, and hence $\widetilde{C}'$ naturally corresponds to a stable formal building $C' \in \overline{\overline{\mathcal{F}}}_{X,A}(\Gamma)\lessdot\mathcal{T}^{(k)}p\gtrdot$. In particular, by formal perturbation invariance of C, we have either $C' = C$ (whence $\widetilde{C}' = \widetilde{C}$) or else C' is a two-level building as in Definition 2.4.1(B2), with top level consisting of a union of a simple index-one component C_Y and possibly some trivial cylinders, and moreover $\mathcal{M}_Y^{J_Y}(C_Y)$ is regular and satisfies $\#\mathcal{M}_Y^{J_Y}(C_Y)/\mathbb{R} = 0$. Let $\widetilde{C}_Y$ denote the analogue of C_Y in $\widetilde{Y}$. By Lemma 3.6.3(ii), every curve in $\mathcal{M}_{\widetilde{Y}}^{J_{\widetilde{Y}}}(\widetilde{C}_Y)$ must be contained in the slice $\mathbb{R} \times Y \times \{0\}$ because it has a negative end. In particular, we have a natural identification

$$\mathcal{M}_{\widetilde{Y}}^{J_{\widetilde{Y}}}(\widetilde{C}_Y) \approx \mathcal{M}_Y^{J_Y}(C_Y),$$

and since each curve in $\mathcal{M}_{\widetilde{Y}}^{J_{\widetilde{Y}}}(\widetilde{C}_Y)$ is also regular by Proposition A.4 we have $\#\mathcal{M}_{\widetilde{Y}}^{J_{\widetilde{Y}}}(\widetilde{C}_Y)/\mathbb{R} = 0$. This establishes the above claim that $\widetilde{C}$ is formally perturbation invariant with respect to $J_{\widetilde{Y}}$.

Invoking now Lemma 3.6.3(i), we have a natural identification

$$\mathcal{M}_{\widetilde{X}}^{J_{\widetilde{X}}}(\widetilde{C}) \approx \mathcal{M}_X^{J_X}(C),$$

and the former is also regular by Proposition A.1. In particular, we have $\#\mathcal{M}_{\widetilde{X}}^{J_{\widetilde{X}}}(\widetilde{C}) \neq 0$, so we conclude by Proposition 2.4.2 that $\mathcal{M}_{\widetilde{X}}^{\widetilde{J}}(\widetilde{C}) \neq \varnothing$ for all $\widetilde{J} \in \mathcal{J}^{J_{\widetilde{Y}}}(\widetilde{X}; \widetilde{D})$. In particular, it follows that we have

$$\tilde{\mathfrak{g}}_k^{J_{\widetilde{Y}}}(\widetilde{X}) \le E_{\widetilde{X}}(\widetilde{C}) = E_X(C) = \tilde{\mathfrak{g}}_k(X),$$

as needed. □

4 Fully rounding, permissibility and minimality

In this section we develop our main tools for getting lower bounds on the capacities of convex toric domains. In Section 4.1 we explain the fully rounding procedure, which standardizes the Reeb dynamics. In Section 4.2 we discuss the extent to which curves are obstructed by the relative adjunction formula and writhe bounds. Lastly, in Section 4.3 we analyze those words of Reeb orbits having minimal action for a given index. The proof that these minimal action words can all be represented by curves is deferred to Section 5.

4.1 The fully rounding procedure

We consider a four-dimensional[6] *convex toric domain*, ie a subdomain of $\mathbb{C}^2$ of the form $X_\Omega := \mu^{-1}(\Omega)$, where

- $\mu\colon \mathbb{C}^2 \to \mathbb{R}^2_{\geq 0}$ is the standard moment map defined by $\mu(z_1, z_2) = (\pi|z_1|^2, \pi|z_2|^2)$, and
- $\Omega \subset \mathbb{R}^2_{\geq 0}$ is a subdomain such that

$$\widehat{\Omega} := \{(x_1, x_2) \in \mathbb{R}^2 \mid (|x_1|, |x_2|) \in \Omega\} \subset \mathbb{R}^2$$

  is compact and convex.

We equip X_Ω with the restriction of the standard Liouville form $\lambda_{\mathrm{std}} = \frac{1}{2}(x\,dy - y\,dx)$ on $\mathbb{C}^2$. For example, if $\Omega \subset \mathbb{R}^2$ is a rational triangle with vertices $(0,0), (a,0), (0,b)$, then X_Ω is the ellipsoid $E(a,b) \subset \mathbb{C}^2$.

The "fully rounding procedure" replaces X_Ω with a C^0–small perturbation whose Reeb orbits are indexed in a straightforward way which is essentially insensitive to the shape of Ω. We proceed in two steps:

(1) Replace X_Ω with another convex toric domain $X_\Omega^{\mathrm{FR}} := X_{\Omega^{\mathrm{FR}}}$, where $\Omega^{\mathrm{FR}} \subset \mathbb{R}^2_{\geq 0}$ is a C^0–small perturbation of Ω with smooth boundary as in [38, Figure 5.1]; see also [16, Section 2.2].

(2) Let $\widetilde{X}_\Omega$ denote the result after a further C^0–small smooth perturbation of X_Ω^{FR} which replaces each Morse–Bott circle of Reeb orbits of action less than some large constant K with two nondegenerate Reeb orbits, one elliptic and one positive hyperbolic; see also [1] or [22, Section 5.3].

In more detail, we assume Ω^{FR} is bounded by the axes and a smooth function $h\colon [0,a] \to [0,b]$ for some $a, b \in \mathbb{R}_{>0}$ such that

- h is strictly decreasing and strictly concave down,
- $h(0) = b$ and $h(a) = 0$, and
- $-\nu < h'(0) < 0$ and $h'(a) < -1/\nu$ for some $\nu > 0$ sufficiently small, and $h'(0), h'(a) \in \mathbb{R} \setminus \mathbb{Q}$.

The Reeb orbits after fully rounding are as follows. For each $(i,j) \in \mathbb{Z}^2_{\geq 1}$ with $\nu < j/i < 1/\nu$, there is an S^1–family of Reeb orbits lying in the two-torus $\mu^{-1}(p_{i,j}) \subset \partial X_\Omega^{\mathrm{FR}}$, where $p_{i,j} \in \partial\Omega^{\mathrm{FR}}$ is such that the outward normal to $\partial\Omega^{\mathrm{FR}}$ at $p_{i,j}$ is parallel to (i,j). The Reeb orbits in this family are $\gcd(i,j)$–fold covers of their underlying simple orbits. In $\partial\widetilde{X}_\Omega$, these S^1–families having action less than K get replaced by a corresponding pair of nondegenerate elliptic and hyperbolic orbits, which we denote by $e_{i,j}$ and $h_{i,j}$, respectively. There are also nondegenerate elliptic Reeb orbits of $\partial X_\Omega^{\mathrm{FR}}$ which lie in $\mu^{-1}(a,0)$ and $\mu^{-1}(0,b)$. We denote these by $e_{i,0}$ and $e_{0,j}$, respectively, for $j \in \mathbb{Z}_{\geq 1}$, and we use the same notation for their natural analogues in $\partial\widetilde{X}_\Omega$. We refer to the Reeb orbits of $\partial\widetilde{X}_\Omega$ of the form $e_{i,j}$ or $h_{i,j}$ as above as *acceptable*. Note that each acceptable orbit has action less than K.

[6] We note that the discussion in this subsection generalizes very naturally to higher dimensions, but for concreteness we restrict our exposition to dimension four.

For the acceptable Reeb orbits in $\widetilde{X}_\Omega$ described above, we have

$$\text{(4-1-1)} \qquad \mathrm{CZ}(e_{i,j}) = 2i + 2j + 1 \quad \text{and} \quad \mathrm{CZ}(h_{i,j}) = 2i + 2j,$$

where over each Reeb orbit γ we use by default the trivialization of the contact distribution that extends over a disc in ∂X with boundary γ.[7] There are also three slightly different associated action filtrations. We denote by $\|-\|^*_\Omega$ the dual of the norm on $\mathbb{R}^2$ whose unit ball is Ω. Viewing $e_{i,j}$ and $h_{i,j}$ as formal symbols, we put:

- $\mathcal{A}_\Omega(e_{i,j}) = \mathcal{A}_\Omega(h_{i,j}) = \|(i,j)\|^*_\Omega = \max_{\vec{v}\in\Omega}\langle \vec{v}, (i,j)\rangle$, the *idealized action*.
- $\mathcal{A}^{\mathrm{FR}}_\Omega(e_{i,j}) = \mathcal{A}^{\mathrm{FR}}_\Omega(h_{i,j}) = \|(i,j)\|^*_{\Omega^{\mathrm{FR}}} = \max_{\vec{v}\in\Omega^{\mathrm{FR}}}\langle \vec{v}, (i,j)\rangle$, the *fully rounded action*.
- $\widetilde{\mathcal{A}}_\Omega(e_{i,j})$ and $\widetilde{\mathcal{A}}_\Omega(h_{i,j})$ denote the actions of the corresponding Reeb orbits in the domain $\widetilde{X}_\Omega$, the *perturbed action*.

We will sometimes refer to any of these as simply "the action" if which one we are referring to is clear from the context or irrelevant, and we will often omit Ω from the notation if it is implicit. Note that $\widetilde{\mathcal{A}}_\Omega$ is a small perturbation of $\mathcal{A}^{\mathrm{FR}}_\Omega$, although its precise values are sensitive to the choices involved in constructing $\widetilde{X}_\Omega$.

Let $w = \gamma_1 \times \cdots \times \gamma_k$ be an (unordered) tuple of acceptable Reeb orbits in $\partial\widetilde{X}_\Omega$. We will refer to such a w as a *word*, and we often view it as simply a collection of formal symbols of the form $e_{i,j}$ or $h_{i,j}$. As a convenient shorthand we define the *index* of w to be the sum

$$\text{(4-1-2)} \qquad \mathrm{ind}(w) := \sum_{i=1}^{k} \mathrm{CZ}(\gamma_i) + k - 2.$$

More generally, for any trivialization τ, the Fredholm index of a curve C with top ends on the orbits $\gamma_1,\dots,\gamma_k$ and negative ends on $\gamma'_1,\dots,\gamma'_{k'}$ is given by

$$\text{(4-1-3)} \qquad \mathrm{ind}(u) = -\chi(C) + 2c_\tau(C) + \sum_{i=1}^{k} \mathrm{CZ}_\tau(\gamma_i) - \sum_{j=1}^{k'} \mathrm{CZ}_\tau(\gamma'_j).$$

Note that the relative first Chern class term in (4-1-3) vanishes if we use the trivialization τ_{ex}, so the formula in (4-1-2) is the contribution of the top end of a curve to its Fredholm index. In particular, $\mathrm{ind}(\gamma_1 \times \cdots \times \gamma_k) = 2m$ is an even integer, a (rational) curve in $\widetilde{X}_\Omega$ with top ends $\gamma_1,\dots,\gamma_k$ and satisfying the constraint $\lessdot\mathcal{T}^{(m)}p\gtrdot$ has Fredholm index zero. As we will see in Section 5, the strong permissibility condition introduced below ensures that every connected curve with strongly permissible top end is somewhere injective.

We note also that if w is "elliptic", meaning that all of the constituent Reeb orbits are elliptic, then its half-index is given by

$$\tfrac{1}{2}\,\mathrm{ind}(e_{i_1,j_1} \times \cdots \times e_{i_k,j_k}) = \sum_{s=1}^{q} (i_s + j_s) + k - 1.$$

[7] This is the trivialization called τ_{ex} in [30, Section 3.2].

We extend the definition of idealized action to words by putting

$$\mathcal{A}(\gamma_1,\dots,\gamma_k) := \sum_{i=1}^{k} \mathcal{A}(\gamma_i),$$

and similarly for the fully rounded action $\mathcal{A}^{\mathrm{FR}}$ and perturbed action $\widetilde{\mathcal{A}}$. We will say that a word w is acceptable if each of its constituent orbits is.

Lemma 4.1.1 *We can arrange the fully rounding procedure such that the following further conditions are satisfied*:

(a) *For each pair of acceptable orbits $e_{i,j}$ and $h_{i,j}$, we have*

$$0 < \widetilde{\mathcal{A}}(e_{i,j}) - \widetilde{\mathcal{A}}(h_{i,j}) < \tfrac{1}{2}|\widetilde{\mathcal{A}}(e_{i',j'}) - \widetilde{\mathcal{A}}(e_{i'',j''})|$$

for any pair of acceptable orbits $e_{i',j'}$ and $e_{i'',j''}$ with $(i',j') \neq (i'',j'')$.

(b) *Given any two acceptable words w, w' such that $\mathcal{A}(w) < \mathcal{A}(w')$, we have also $\mathcal{A}^{\mathrm{FR}}(w) < \mathcal{A}^{\mathrm{FR}}(w')$.*

(c) *Given any two acceptable words such that $\mathcal{A}^{\mathrm{FR}}(w) < \mathcal{A}^{\mathrm{FR}}(w')$, we have also $\widetilde{\mathcal{A}}(w) < \widetilde{\mathcal{A}}(w')$.*

(d) *For any two distinct acceptable orbits γ and γ', we have $\widetilde{\mathcal{A}}(\gamma) \neq \widetilde{\mathcal{A}}(\gamma')$, and moreover the set of $\widetilde{\mathcal{A}}$ values of acceptable orbits which are simple* (ie have $\gcd(i,j) = 1$) *is linearly independent over $\mathbb{Q}$.*

In the sequel, we will take $K > 0$ (the upper bound of the energy of acceptable orbits) sufficiently large and $\nu > 0$ (which measures the size of the perturbation) sufficiently small that for action reasons the unacceptable Reeb orbits play essentially no role; thus without much harm we can pretend that the Reeb orbits of $\partial \widetilde{X}_\Omega$ are precisely $e_{i,j}$ for any $(i,j) \in \mathbb{Z}^2_{\geq 0}$ with i, j not both zero, and $h_{i,j}$ for any $(i,j) \in \mathbb{Z}^2_{\geq 1}$.

4.2 Strong and weak permissibility

In this subsection we prove Lemma 4.2.2, which states that the positive orbits of a somewhere injective curve in a fully rounded convex toric domain must be strongly permissible in the sense of the following definition:

Definition 4.2.1 Consider a word $w = \gamma_1 \times \cdots \times \gamma_q$, where for each $s = 1,\dots,q$ we have that $\gamma_i = e_{i_s,j_s}$ or $\gamma_i = h_{i_s,j_s}$ for some i_s, j_s. We say that w is *strongly permissible* if one of the following holds:

- $w = e_{1,0}$ or $w = e_{0,1}$, or else
- $i_1,\dots,i_q$ are not all zero, and similarly $j_1,\dots,j_q$ are not all zero.

We say w is *weakly permissible* if it is either strongly permissible or it is of the form $e_{k,0}$ or $e_{0,k}$ for some $k \in \mathbb{Z}_{\geq 2}$.

Lemma 4.2.2 *Let C be an asymptotically cylindrical J–holomorphic rational curve in $\widetilde{X}_\Omega$, where $\widetilde{X}_\Omega$ is a fully rounded four-dimensional convex toric domain and $J \in \mathcal{J}(\widetilde{X}_\Omega)$. If C is somewhere injective, then its word of positive orbits is strongly permissible.*

Before proving the lemma, we recall how to compute the terms in the relative adjunction formula in the case of a four-dimensional fully rounded convex toric domain. Following [21, Section 3.3], the relative adjunction formula for a somewhere injective curve asymptotically cylindrical curve in a four-dimensional symplectic cobordism reads

$$c_\tau(C) = \chi(C) + Q_\tau(C) + w_\tau(C) - 2\delta(C).$$

Here τ denotes a choice of trivialization over each Reeb orbit, $\chi(C)$ is the Euler characteristic of the curve C, and $\delta(C)$ is a count of singularities which is necessarily nonnegative. The computation of the remaining terms for $\widetilde{X}_\Omega$ with respect to a certain choice[8] of trivialization τ_{Hut}, is described in [22, Section 5.3], which we briefly summarize as follows. Let C be a curve in $\widetilde{X}_\Omega$, and let $\Gamma = (\gamma_1, \dots, \gamma_k)$ denote its positive asymptotic Reeb orbits.

- **Relative self-intersection** We have $Q_{\tau_{\mathrm{Hut}}}(C) = Q_{\tau_{\mathrm{Hut}}}(\Gamma) = 2\,\mathrm{Area}(R)$, where:
  - For each constituent orbit (including repeats) of Γ of the form $e_{i,j}$ or $h_{i,j}$, we consider the corresponding "edge vector" $(j, -i)$.
  - We reorder the collection of edge vectors and place them end-to-end so that they form a concave down path $\Lambda \subset \mathbb{R}^2_{\geq 0}$ from $(0, y(\Lambda))$ to $(x(\Lambda), 0)$ for some $x(\Lambda), y(\Lambda) \in \mathbb{Z}_{\geq 0}$.
  - R is the lattice polygon bounded by Λ and the axes.

  For example, we have $Q_{\tau_{\mathrm{Hut}}}(h_{i,j}) = Q_{\tau_{\mathrm{Hut}}}(e_{i,j}) = ij$.
- **Relative first Chern class** We have $c_{\tau_{\mathrm{Hut}}}(C) = c_{\tau_{\mathrm{Hut}}}(\Gamma) = \sum_{i=1}^k c_{\tau_{\mathrm{Hut}}}(\gamma_i)$, where
$$c_{\tau_{\mathrm{Hut}}}(h_{i,j}) = c_{\tau_{\mathrm{Hut}}}(e_{i,j}) = i + j.$$
- **Asymptotic writhe** The term $w_{\tau_{\mathrm{Hut}}}(C)$ measures the total asymptotic writhe of C around its asymptotic Reeb orbits. Although this is difficult to compute directly, we have the writhe bound (3.2.9) in [30, Section 3.2]; see [22, Section 5.1] for more details. This is formulated in terms of the *monodromy angle* θ of each simple Reeb orbit. In particular, since we can take this to be 0 for the hyperbolic orbits $h_{i,j}$ and positive but very small for the elliptic orbits $e_{i,j}$, the writhe bound implies that the top writhe of any curve with positive ends on a word in $e_{i,j}$, $h_{i,j}$ is always ≤ 0.

Proof of Lemma 4.2.2 Without loss of generality, consider a somewhere injective curve in $\widetilde{X}_\Omega$ with positive ends $(\gamma_1, \dots, \gamma_k)$, and suppose that for each $s = 1, \dots, k$ we have $\gamma_s = e_{i_s,0}$ for some $i_s \in \mathbb{Z}_{\geq 1}$. The writhe bound gives $w_{\tau_{\mathrm{Hut}}}(C) \leq 0$. Meanwhile, we have $c_{\tau_{\mathrm{Hut}}}(C) = \sum_{s=1}^k i_s$ and $Q_{\tau_{\mathrm{Hut}}}(C) = 0$, and hence

$$w_{\tau_{\mathrm{Hut}}}(C) = c_{\tau_{\mathrm{Hut}}}(C) - \chi(C) - Q_{\tau_{\mathrm{Hut}}}(C) + 2\delta(C) = \sum_{s=1}^k i_s - (2-k) + 2\delta \leq 0,$$

and consequently $\sum_{s=1}^k (i_s + 1) \leq 2$, which forces $k = i_1 = 1$. A similar calculation rules out the possibility that $i_s = 0$ for all s. □

[8] This is different from the trivialization used before, in which $c_\tau(e_{i,j}) = c_\tau(h_{i,j}) = 0$.

Using Lemma 4.2.2, we prove the following lower bound on $\widetilde{\mathfrak{g}}_k(\widetilde{X}_\Omega)$, which will be further refined in the next subsection.

Lemma 4.2.3 *For any four-dimensional convex toric domain X_Ω we have*

$$\widetilde{\mathfrak{g}}_k(X_\Omega) \geq \min_w \mathcal{A}(w),$$

where we minimize over all weakly permissible words w satisfying $\operatorname{ind}(w) \geq 2k$.

Proof By C^0–continuity it suffices to prove the analogous lower bound after fully rounding, namely $\widetilde{\mathfrak{g}}_k(\widetilde{X}_\Omega) \geq \min_w \widetilde{\mathcal{A}}(w)$, still minimizing over all weakly permissible words w satisfying $\operatorname{ind}(w) \geq 2k$. Pick a generic $J \in \mathcal{J}(\widetilde{X}_\Omega; D)$. By definition of $\widetilde{\mathfrak{g}}_k(\widetilde{X}_\Omega)$, we can find a curve C in $\widetilde{X}_\Omega$ satisfying the constraint $\lll \mathcal{T}^{(k)} p \ggg$ with $E(C) \leq \widetilde{\mathfrak{g}}_k(\widetilde{X}_\Omega)$ (a priori we should also consider the case $a \geq 2$ as in Remark 3.1.2, but it is easy to check that these do not affect the infimum). Let w denote the word of positive orbits corresponding to C. Note that the underlying simple curve $\overline{C}$ is somewhere injective and has nonnegative index by genericity of J, and therefore its word $\overline{w}$ of positive orbits is strongly permissible by Lemma 4.2.2. Then the word w is also strongly permissible unless we have $\overline{w} = e_{1,0}$ or $\overline{w} = e_{0,1}$. Moreover, we have $\operatorname{ind}(C) \geq \kappa \operatorname{ind}(\overline{C}) \geq 0$ by Lemma 5.1.2 below, where κ is the covering index of C over $\overline{C}$, and hence we have $\operatorname{ind}(w) \geq 2k$. If w is strongly permissible then it is also weakly permissible and we have $\widetilde{\mathfrak{g}}_k(\widetilde{X}_\Omega) \geq \widetilde{\mathcal{A}}(w) \geq \min_w \widetilde{\mathcal{A}}(w)$, with the minimum taken over all weakly permissible words w satisfying $\operatorname{ind}(w) \geq 2k$.

We can therefore assume $\overline{w} = e_{1,0}$ or $\overline{w} = e_{0,1}$, since otherwise the proof is already complete. Observe that since C satisfies the constraint $\lll \mathcal{T}^{(k)} p \ggg$, we must have $k \leq \kappa$. Then $\widetilde{\mathcal{A}}(w) = \kappa \widetilde{\mathcal{A}}(e_{1,0}) \geq \widetilde{\mathcal{A}}(e_{k,0})$ or $\widetilde{\mathcal{A}}(w) = \kappa \widetilde{\mathcal{A}}(e_{0,1}) \geq \widetilde{\mathcal{A}}(e_{0,k})$, respectively. Since $e_{k,0}$ and $e_{0,k}$ are weakly permissible with index $2k$, this again implies the desired result. □

Definition 4.2.4 We will denote by $w_{\min}$ the weakly permissible word with minimal $\widetilde{\mathcal{A}}_\Omega$ value subject to $\operatorname{ind}(w_{\min}) = 2k$.

Since distinct words have different actions by condition (d) in Lemma 4.1.1, $w_{\min}$ is unique for each k.

4.3 Minimal words

As before, let X_Ω be a four-dimensional convex toric domain with full rounding $\widetilde{X}_\Omega$. In light of Lemma 4.2.3, we seek to understand which weakly permissible words have minimal $\widetilde{\mathcal{A}}_\Omega$ value. We begin with some preliminary lemmas. In the following, put

$$a := \max\{x \mid (x, 0) \in \Omega^{\mathrm{FR}}\} \quad \text{and} \quad b := \max\{y \mid (0, y) \in \Omega^{\mathrm{FR}}\}$$

as in Section 4.1.

Lemma 4.3.1 *For any $(i,j) \in \mathbb{Z}^2_{\geq 1}$, we have $\max(ia, jb) < \|(i,j)\|^*_{\Omega^{\mathrm{FR}}} < ai + jb$.*

Proof Let $\vec{v} = (v_1, v_2) \in \partial\Omega^{\mathrm{FR}} \cap \mathbb{R}^2_{>0}$ be such that $\|(i,j)\|^*_{\Omega^{\mathrm{FR}}} = \langle \vec{v}, (i,j)\rangle$. Then the line in $\mathbb{R}^2$ passing through $\vec{v}$ and orthogonal to (i,j) is tangent to $\partial\Omega^{\mathrm{FR}}$, and is given by

$$\{(x,y) \in \mathbb{R}^2 \mid \langle (x,y), (i,j)\rangle = \langle \vec{v}, (i,j)\rangle = iv_1 + jv_2 < ai + bj\}.$$

This gives the upper bound. To derive the lower bound, notice that the y intercept is given by $(iv_1 + jv_2)/j$, and this is strictly greater than b since $h\colon [0,a] \to [0,b]$ is strictly concave down. That is, we have $\|(i,j)\|^*_{\Omega^{\mathrm{FR}}} = iv_1 + jv_2 > jb$. Similarly, the x intercept is given by $(iv_1 + jv_2)/i$, and by strict convexity this is strictly greater than a, ie we have $\|(i,j)\|^*_{\Omega^{\mathrm{FR}}} = iv_1 + jv_2 > ia$. □

Lemma 4.3.2 *Given distinct pairs $(i,j), (i',j') \in \mathbb{Z}^2_{\geq 0}$ with $i' \leq i$ and $j' \leq j$, we have*

$$\|(i',j')\|^*_{\Omega^{\mathrm{FR}}} < \|(i,j)\|^*_{\Omega^{\mathrm{FR}}}.$$

Proof Without loss of generality we can assume $(i',j') = (i-1,j)$, since the case $(i',j') = (i,j-1)$ is completely analogous and then the general case follows by induction. Let $\vec{v} = (v_1,v_2) \in \partial\Omega^{\mathrm{FR}}$ be such that $\|(i-1,j)\|^*_{\Omega^{\mathrm{FR}}} = \langle \vec{v}, (i-1,j)\rangle$. Then we have

$$\|(i-1,j)\|^*_{\Omega^{\mathrm{FR}}} = (i-1)v_1 + jv_2 \leq iv_1 + jv_2 \leq \|(i,j)\|^*_{\Omega^{\mathrm{FR}}},$$

and the inequality is strict unless $v_1 = 0$, which is only possible if $(i-1,j)$ lies on the y–axis, ie $i = 1$. In this case, by Lemma 4.3.1 we have

$$\|(i,j)\|^*_{\Omega^{\mathrm{FR}}} = \|(1,j)\|^*_{\Omega^{\mathrm{FR}}} > \max(a, jb) \geq jb = \|(0,j)\|^*_{\Omega^{\mathrm{FR}}} = \|(i-1,j)\|^*_{\Omega^{\mathrm{FR}}},$$

as desired. □

We next show that we can effectively ignore the hyperbolic orbits. Recall that a word $w = \gamma_1 \times \cdots \times \gamma_k$ is called "elliptic" if each constituent orbit γ_i is elliptic.

Lemma 4.3.3 *Given any word w which is not elliptic, we can find an elliptic word w' with*

$$\mathrm{ind}(w') \geq \mathrm{ind}(w) - 1 \quad \textit{and} \quad \widetilde{\mathcal{A}}(w') < \widetilde{\mathcal{A}}(w).$$

Moreover, if w is strongly (resp. weakly) permissible, then we can arrange that the same is true for w'.

Proof Firstly, if $\frac{1}{2}\,\mathrm{ind}(w)$ is not an integer, then we replace some hyperbolic orbit $h_{i,j}$ by $e_{i',j'}$ with $(i',j') = (i-1,j)$ or $(i',j') = (i,j-1)$. Note that this replacement decreases the index by 1. Moreover, we have $\|(i',j')\|^*_{\Omega^{\mathrm{FR}}} < \|(i,j)\|^*_{\Omega^{\mathrm{FR}}}$ by Lemma 4.3.2, and hence $\widetilde{\mathcal{A}}(e_{i',j'}) < \widetilde{\mathcal{A}}(e_{i,j})$ by Lemma 4.1.1(c). Then by Lemma 4.1.1(a) we also have

$$\widetilde{\mathcal{A}}(e_{i,j}) - \widetilde{\mathcal{A}}(h_{i,j}) < \widetilde{\mathcal{A}}(e_{i,j}) - \widetilde{\mathcal{A}}(e_{i',j'}),$$

and hence $\widetilde{\mathcal{A}}(e_{i',j'}) < \widetilde{\mathcal{A}}(h_{i,j})$, so this shows that the above replacement strictly decreases $\widetilde{\mathcal{A}}$.

Now suppose there are $2l$ hyperbolic orbits in w for some $l \in \mathbb{Z}_{\geq 0}$. For l of these replace $h_{i,j}$ with $e_{i,j}$, and for the other l replace $h_{i,j}$ with $e_{i-1,j}$ or $e_{i,j-1}$. Each pair of such replacements strictly deceases $\widetilde{\mathcal{A}}$ by the same lemma, and the total index is unchanged. For example, we have $\widetilde{\mathcal{A}}(h_{i,j} \times h_{i',j'}) > \widetilde{\mathcal{A}}(e_{i,j} \times e_{i'-1,j'})$ using

$$\widetilde{\mathcal{A}}(e_{i',j'}) - \widetilde{\mathcal{A}}(h_{i',j'}) < \tfrac{1}{2}(\widetilde{\mathcal{A}}(e_{i',j'}) - \widetilde{\mathcal{A}}(e_{i'-1,j'})) \quad \text{and} \quad \widetilde{\mathcal{A}}(e_{i,j}) - \widetilde{\mathcal{A}}(h_{i,j}) < \tfrac{1}{2}(\widetilde{\mathcal{A}}(e_{i',j'}) - \widetilde{\mathcal{A}}(e_{i'-1,j'})).$$

Lastly, it is straightforward to check that each of these replacements can be done so as to preserve strong or weak permissibility. □

Remark 4.3.4 For future reference, note that in Lemma 4.3.3 if $\operatorname{ind}(w) = 2k$ for some $k \in \mathbb{Z}_{\geq 1}$ then we must also have $\operatorname{ind}(w') = 2k$ since the index of any elliptic word is even. If particular, if $\operatorname{ind}(w) \geq 2k$ for some $\mathbb{Z}_{\geq 1}$, then we have $\operatorname{ind}(w') \geq 2k$ as well.

The following lemma will be our most useful tool for iteratively reducing the action of a word:

Lemma 4.3.5 *Assume $a > b$. Then we have $\widetilde{\mathcal{A}}(e_{0,1} \times e_{i,j}) < \widetilde{\mathcal{A}}(e_{i+1,j+1})$ for any $(i,j) \in \mathbb{Z}^2_{\geq 0} \setminus \{(0,0)\}$.*

Proof Let $\vec{v} = (v_1, v_2) \in \partial\Omega^{\mathrm{FR}}$ be such that $\|(i,j)\|^*_{\Omega^{\mathrm{FR}}} = \langle \vec{v}, (i,j)\rangle$. Suppose first that we have $v_1, v_2 \geq 1$. Note that (v_1, v_2) lies above or on the line joining $(a,0)$ and $(0,b)$, ie we have $av_2 + bv_1 \geq ab$. Since $a > b$, we have $v_1 + v_2 \geq b$, with equality only if $v_1 = 0$. We then have

$$\begin{aligned} \mathcal{A}^{\mathrm{FR}}(e_{0,1} \times e_{i,j}) &= \|(0,1)\|^*_{\Omega^{\mathrm{FR}}} + \|(i,j)\|^*_{\Omega^{\mathrm{FR}}} = b + iv_1 + jv_2 \\ &\leq v_1 + v_2 + iv_1 + jv_2 = \langle \vec{v}, (i+1, j+1)\rangle \\ &\leq \|(i+1, j+1)\|^*_{\Omega^{\mathrm{FR}}} = \mathcal{A}^{\mathrm{FR}}(e_{i+1,j+1}), \end{aligned}$$

where the first inequality is strict unless v lies on the y–axis, in which case we must have $i = 0$. If $i = 0$, by Lemma 4.3.1 we have

$$\mathcal{A}^{\mathrm{FR}}(e_{1,j+1}) > \max(a, (j+1)b) \geq (j+1)b = \mathcal{A}^{\mathrm{FR}}(e_{0,1} \times e_{0,j}).$$

Thus in any case we have $\mathcal{A}^{\mathrm{FR}}(e_{0,1} \times e_{i,j}) < \mathcal{A}^{\mathrm{FR}}(e_{i+1,j+1})$, and by Lemma 4.1.1(c) we also have $\widetilde{\mathcal{A}}(e_{0,1} \times e_{i,j}) < \widetilde{\mathcal{A}}(e_{i+1,j+1})$. □

Remark 4.3.6 The assumption $a > b$ is not very restrictive, since if $a < b$ we can simply replace Ω^{FR} by its reflection about the diagonal.

Using the above tools, we first consider ways to reduce action without any regard to permissibility:

Lemma 4.3.7 *Assume $a > b$. Given any elliptic word w, there is another elliptic word w' with $\operatorname{ind}(w') = \operatorname{ind}(w)$ and $\widetilde{\mathcal{A}}(w') \leq \widetilde{\mathcal{A}}(w)$, where w takes one of the following forms*:

(a) $e_{0,1}^{\times i}$ *for $i \geq 1$.*

(b) $e_{0,1}^{\times i} \times e_{1,1}^{\times j}$ *for $i \geq 0$ and $j \geq 1$.*

(c) $e_{0,1}^{\times i} \times e_{0,2}$ *for $i \geq 0$.*

Moreover, we have $\widetilde{\mathcal{A}}(w') < \widetilde{\mathcal{A}}(w)$ unless w and w' differ by a reordering.

Proof We first iteratively apply Lemma 4.3.5 as many times as possible, replacing $e_{i+1,j+1}$ with $e_{0,1} \times e_{i,j}$ if $(i,j) \in \mathbb{Z}^2_{\geq 0} \setminus \{(0,0\}$. Note that the resulting word contains only orbits of the forms $e_{1,1}, e_{k,0}, e_{0,k}$ for $k \geq 1$, and each replacement strictly decreases $\widetilde{\mathscr{A}}$.

Next, we replace each $e_{k,0}$ for $k \geq 1$ with $e_{0,k}$. Similarly, we replace each $e_{0,2k-1}$ such that $2k-1 \geq 3$ with $e_{0,1}^{\times k}$, and we replace each $e_{0,2k}$ such that $2k \geq 4$ with $e_{0,1}^{\times(k-1)} \times e_{0,2}$. We also replace each $e_{0,2} \times e_{0,2}$ with $e_{0,1} \times e_{0,1} \times e_{0,1}$. Each of these replacements strictly decreases $\widetilde{\mathscr{A}}$.

The resulting word is of the form $e_{0,1}^{\times i} \times e_{1,1}^{\times j} \times e_{0,2}^{\times k}$ for some $i, j \in \mathbb{Z}_{\geq 0}$ and $k \in \{0,1\}$. By Lemma 4.1.1(d) we have either $\widetilde{\mathscr{A}}(e_{1,1}) < \widetilde{\mathscr{A}}(e_{0,2})$ or else $\widetilde{\mathscr{A}}(e_{1,1}) > \widetilde{\mathscr{A}}(e_{0,2})$. In the former case, we also replace any remaining $e_{0,2}$ with $e_{1,1}$. In the latter case, we replace each $e_{1,1}$ with $e_{0,2}$, and then further replace each $e_{0,2} \times e_{0,2}$ with $e_{0,1} \times e_{0,1} \times e_{0,1}$ as above.

The resulting word w' satisfies $\mathrm{ind}(w') = \mathrm{ind}(w)$ and $\widetilde{\mathscr{A}}(w') \leq \widetilde{\mathscr{A}}(w)$ and takes one of the forms (a)–(c). Moreover, up to reordering these are the only cases when none of the above reductions are applicable, and otherwise we have $\widetilde{\mathscr{A}}(w') < \widetilde{\mathscr{A}}(w)$. □

Next, we investigate reductions in actions which preserve the strong permissibility condition. Perhaps surprisingly, there are only a few possibilities for minimal words, regardless of Ω:

Proposition 4.3.8 *Assume $a > b$. Given any strongly permissible elliptic word w with $\frac{1}{2}\,\mathrm{ind}(w) > 1$, there is another strongly permissible elliptic word w' with $\mathrm{ind}(w') = \mathrm{ind}(w)$ and $\widetilde{\mathscr{A}}(w') \leq \widetilde{\mathscr{A}}(w)$, where w' takes one of the following forms*:

1. $e_{0,1}^{\times i} \times e_{1,1}^{\times j}$ *for $i \geq 0$ and $j \geq 1$.*
2. $e_{0,1}^{\times i} \times e_{1,s}$ *for $i \geq 0$ and $s \geq 2$.*
3. $e_{0,1}^{\times i} \times e_{1,0}$ *for $i \geq 1$.*

Moreover, we have $\widetilde{\mathscr{A}}(w') < \widetilde{\mathscr{A}}(w)$ unless w and w' differ by a reordering.

Proof To start, we iteratively apply Lemma 4.3.5 as many times as possibly without spoiling strong permissibility, and let $w = e_{i_1,j_1} \times \cdots \times e_{i_q,j_q}$ denote the resulting word. Note that we must have $i_s \leq 1$ or $j_s \leq 1$ for each $s = 1, \dots, q$, since otherwise a further application of Lemma 4.3.5 would be possible. Furthermore, we can assume without loss of generality that we have $i_1 \neq 0$ (note that $e_{0,1}$ is ruled out using $\frac{1}{2}\,\mathrm{ind}(w) > 1$).

Next, by applying Lemma 4.3.7 to the subword $e_{i_2,j_2} \times \cdots \times e_{i_q,j_q}$, we obtain a word of the form $e_{i_1,j_1} \times w''$, where w'' is a word having one of the forms (a)–(c). This replacement leaves the index unchanged and strictly decreases $\widetilde{\mathscr{A}}$ (unless it is vacuous). Moreover, by our assumption $i_1 \neq 0$ and inspection of the forms (a)–(c), the word $e_{i_1,j_1} \times w''$ is strongly permissible. We also have $(i_1, j_1) \notin \mathbb{Z}_{\geq 2} \times \mathbb{Z}_{\geq 2}$, and hence (i_1, j_1) must be one of the following:

$$(1,0),\ (1,1),\ (1,s),\ (s,0),\ (s,1), \quad \text{where } s \geq 2. \tag{4-3-1}$$

Our goal is to apply further reductions to $e_{i_1,j_1} \times w''$ which decrease $\widetilde{\mathcal{A}}$ and leave the index unchanged, in order to arrive at one of the forms (1)–(3). Observe that for $s \geq 2$ we have

$$\|(s-1,1)\|^*_{\Omega^{\mathrm{FR}}} \leq \|(s-1,0)\|^*_{\Omega^{\mathrm{FR}}} + \|(0,1)\|^*_{\Omega^{\mathrm{FR}}} = (s-1)a + b < sa = \|(s,0)\|^*_{\Omega^{\mathrm{FR}}},$$

so $\widetilde{\mathcal{A}}(e_{s-1,1}) < \widetilde{\mathcal{A}}(e_{s,0})$ by Lemma 4.1.1(c). Also, by Lemma 4.3.5 we have $\widetilde{\mathcal{A}}(e_{0,1} \times e_{s-1,0}) < \widetilde{\mathcal{A}}(e_{s,1})$ for $s \geq 2$. This shows that we can ignore the last two items in (4-3-1).

We now consider each of the remaining possibilities for $e_{i_1,j_1} \times w''$ and explain the necessary reductions:

- $(i_1, j_1) = (1,0)$:
  (a) $e_{1,0} \times e_{0,1}^{\times i}$ for $i \geq 1$: already of form (3).
  (b) $e_{1,0} \times e_{0,1}^{\times i} \times e_{1,1}^{\times j}$ for $i \geq 0$ and $j \geq 1$: replace $e_{1,0}$ with $e_{0,1}$, becomes of form (1).
  (c) $e_{1,0} \times e_{0,1}^{\times i} \times e_{0,2}$ for $i \geq 0$: replace $e_{1,0} \times e_{0,2}$ with $e_{0,1} \times e_{1,1}$ by (i) below, becomes of form (1).
- $(i_1, j_1) = (1,1)$:
  (a) $e_{1,1} \times e_{0,1}^{\times i}$ for $i \geq 1$: already of form (1).
  (b) $e_{1,1} \times e_{0,1}^{\times i} \times e_{1,1}^{\times j}$ for $i \geq 0$ and $j \geq 1$: already of form (1).
  (c) $e_{1,1} \times e_{0,1}^{\times i} \times e_{0,2}$ for $i \geq 0$: replace $e_{1,1} \times e_{0,2}$ with $e_{1,0} \times e_{0,1} \times e_{0,1}$ by (ii) below, becomes of form (3).
- $(i_1, j_1) = (1,s)$ for $s \geq 2$:
  (a) $e_{1,s} \times e_{0,1}^{\times i}$ for $i \geq 1$: already of form (2).
  (b) $e_{1,s} \times e_{0,1}^{\times i} \times e_{1,1}^{\times j}$ for $i \geq 0$ and $j \geq 1$: replace $e_{1,s}$ with $e_{0,1}^{\times(k+1)}$ if $s = 2k$ is even, or with $e_{1,1} \times e_{0,1}^{\times k}$ if $s = 2k+1$ is odd by (iii), becomes of form (1).
  (c) $e_{1,s} \times e_{0,1}^{\times i} \times e_{0,2}$ for $i \geq 0$: replace $e_{0,2} \times e_{1,s}$ with $e_{0,1} \times e_{1,s+1}$ by (iv), becomes of form (2).

We justify the above replacements by applying Lemma 4.3.1 as follows:

(i) We have $\widetilde{\mathcal{A}}(e_{1,1} \times e_{0,1}) < \widetilde{\mathcal{A}}(e_{1,0} \times e_{0,2})$ since

$$\|(1,1)\|^*_{\Omega^{\mathrm{FR}}} + \|(0,1)\|^*_{\Omega^{\mathrm{FR}}} < \|(1,0)\|^*_{\Omega^{\mathrm{FR}}} + 2\|(0,1)\|^*_{\Omega^{\mathrm{FR}}} = a + 2b = \|(1,0)\|^*_{\Omega^{\mathrm{FR}}} + \|(0,2)\|^*_{\Omega^{\mathrm{FR}}}.$$

(ii) We have $\widetilde{\mathcal{A}}(e_{1,0} \times e_{0,1} \times e_{0,1}) < \widetilde{\mathcal{A}}(e_{1,1} \times e_{0,2})$ since

$$\|(1,0)\|^*_{\Omega^{\mathrm{FR}}} + 2\|(0,1)\|^*_{\Omega^{\mathrm{FR}}} = a + 2b < \|(1,1)\|^*_{\Omega^{\mathrm{FR}}} + 2b = \|(1,1)\|^*_{\Omega^{\mathrm{FR}}} + \|(0,2)\|^*_{\Omega^{\mathrm{FR}}}.$$

(iii) We have $\widetilde{\mathcal{A}}(e_{0,1}^{\times(k+1)}) < \widetilde{\mathcal{A}}(e_{1,2k})$ since

$$(k+1)\|(0,1)\|^*_{\Omega^{\mathrm{FR}}} = (k+1)b \leq 2kb < \|(1,2k)\|^*_{\Omega^{\mathrm{FR}}},$$

and $\widetilde{\mathcal{A}}(e_{1,1} \times e_{0,1}^{\times k}) < \widetilde{\mathcal{A}}(e_{1,2k+1})$ since

$$\|(1,1)\|^*_{\Omega^{\mathrm{FR}}} + k\|(0,1)\|^*_{\Omega^{\mathrm{FR}}} \leq (k+2)b \leq (2k+1)b < \|(1,2k+1)\|^*_{\Omega^{\mathrm{FR}}}.$$

(iv) We have $\widetilde{\mathcal{A}}(e_{0,1} \times e_{1,s+1}) < \widetilde{\mathcal{A}}(e_{0,2} \times e_{1,s})$ since

$$\|(0,1)\|^*_{\Omega^{\mathrm{FR}}} + \|(1,s+1)\|^*_{\Omega^{\mathrm{FR}}} < 2b + \|(1,s)\|^*_{\Omega^{\mathrm{FR}}} = \|(0,2)\|^*_{\Omega^{\mathrm{FR}}} + \|(1,s)\|^*_{\Omega^{\mathrm{FR}}}. \tag{4-3-2}$$

This completes the proof. □

Corollary 4.3.9 *Given any weakly permissible elliptic word w, there is another weakly permissible elliptic word w' with $\operatorname{ind}(w') = \operatorname{ind}(w)$ and $\widetilde{\mathcal{A}}(w') \le \widetilde{\mathcal{A}}(w)$, where w' takes one of the following forms*:

(1) $e_{0,1}^{\times i} \times e_{1,1}^{\times j}$ *for $i \ge 0$ and $j \ge 1$.*

(2) $e_{0,1}^{\times i} \times e_{1,s}$ *for $i \ge 0$ and $s \ge 2$.*

(3) $e_{0,1}^{\times i} \times e_{1,0}$ *for $i \ge 1$.*

(4) $e_{0,s}$ *for $s \ge 1$.*

Moreover, we have $\widetilde{\mathcal{A}}(w') < \widetilde{\mathcal{A}}(w)$ unless w and w' differ by a reordering.

We next refine Lemma 4.2.3 so that the minimization involves only words which are elliptic and satisfy $\operatorname{ind}(w) = 2k$ (rather than $\operatorname{ind}(w) \ge 2k$). This completes the proof of half of Theorem 1.2.8.

Corollary 4.3.10 *For any four-dimensional convex toric domain X_Ω we have*

$$\widetilde{\mathfrak{g}}_k(X_\Omega) \ge \min_w \mathcal{A}(w),$$

where we minimize over all weakly permissible elliptic words w satisfying $\operatorname{ind}(w) = 2k$.

Proof The restriction to elliptic words follows from Remark 4.3.4. Now it suffices to show that given any weakly permissible elliptic word w with $\frac{1}{2}\operatorname{ind}(w) > 1$, there is another weakly permissible elliptic word w' with $\frac{1}{2}\operatorname{ind}(w') = \frac{1}{2}\operatorname{ind}(w) - 1$ and $\widetilde{\mathcal{A}}(w') < \widetilde{\mathcal{A}}(w)$. After applying Corollary 4.3.9, we can assume that w has one of the forms (1)–(4), and we then make the following respective replacements:

(1) $e_{0,1}^{\times i} \times e_{1,1}^{\times (j-1)} \times e_{1,0}$.

(2) $e_{0,1}^{\times i} \times e_{1,s-1}$.

(3) $e_{0,1}^{\times (i-1)} \times e_{1,1}$.

(4) $e_{0,s-1}$. □

We end this section by proving Corollary 1.2.9, which claims that (in dimension four) $\widetilde{\mathfrak{g}}_k(X_\Omega)$ is the minimal length $\ell_\Omega(\partial P)$ of the boundary ∂P of a convex lattice polygon P such that ∂P contains exactly $k+1$ lattice points. For the moment we assume Theorem 1.2.8, the proof of which is completed in Section 5 below.

Proof of Corollary 1.2.9 We first prove that the right-hand side of (1-2-2) is less than or equal to the right-hand side of (1-2-1); in other words, for each minimal word w there is a lattice polygon P with $\ell_\Omega(\partial P)$ less than or equal to $\mathcal{A}(w)$. To this end, let $(i_1, j_1), \dots, (i_q, j_q) \in \mathbb{Z}^2_{\ge 0} \setminus \{(0,0)\}$ be a minimizer, which we can assume takes one of the forms (1)–(4) given in Corollary 4.3.9. Then we have

$$\sum_{s=1}^{q}(i_s + j_s) + q - 1 = k \quad \text{and} \quad \tilde{\mathfrak{g}}_k(X_\Omega) = \sum_{s=1}^{q} \|(i_s, j_s)\|_\Omega^*,$$

and we seek a convex lattice polygon P with $\ell_\Omega(\partial P) \le \sum_{s=1}^{q} \|(i_s, j_s)\|_\Omega^*$ such that $|\partial P \cap \mathbb{Z}^2| = k+1$. In case (4), we take P to be the degenerate polygon given by the convex hull of $(0,0), (0,k)$, which contains $k+1$ lattice points and satisfies $\ell_\Omega(\partial P) = \|(0,k)\|_\Omega^*$. In cases (1)–(3), let $p_1, \dots, p_{q+1} \in \mathbb{Z}^2_{\le 0}$ be the unique ordered list of lattice points such that

(1) the displacement vectors $p_2 - p_1, \dots, p_{q+1} - p_q$ equal $(i_1, j_1), \dots, (i_q, j_q)$ up to order,

(2) $p_1 = (0, -\beta)$ and $p_{q+1} = (\alpha, 0)$ for $\alpha = \sum_{s=1}^{q} i_s$ and $\beta = \sum_{s=1}^{q} j_s$, and

(3) the lower boundary G of the convex hull of $p_1, \dots, p_{q+1}$ is the graph of a convex piecewise linear function $[0, \alpha] \to [0, -\beta]$.

Let $P \subset \mathbb{R}^2_{\le 0}$ be the convex lattice polygon given by the convex hull of $(0,0), p_1, \dots, p_{q+1}$, ie P is the union of G with the line segments joining $(\alpha, 0)$ to $(0,0)$ and $(0,0)$ to $(0, -\beta)$. Using the definition of $\|-\|_\Omega^*$ and the fact that X_Ω is a convex toric domain, observe that for any $(v_x, v_y) \in \mathbb{R}^2$ we have

$$\|(v_x, v_y)\|_\Omega^* = \|(\max(v_x, 0), \max(v_y, 0)\|_\Omega^*. \tag{4-3-3}$$

In particular, we have $\|(v_x, v_y)\|_\Omega^* = 0$ if $(v_x, v_y) \in \mathbb{R}^2_{\le 0}$, and hence

$$\ell_\Omega(\partial P) = \ell_\Omega(G) = \sum_{s=1}^{q} \|(i_s, j_s)\|_\Omega^*.$$

Moreover, since $\gcd(i_s, j_s) = 1$ for $s = 1, \dots, q$, the number of lattice points along G is $q+1$, and hence

$$|\partial P \cap \mathbb{Z}^2| = q + 1 + \alpha + \beta - 1 = q + \sum_{s=1}^{q}(i_s + j_s) = k+1.$$

Now we prove that the reverse inequality. Let P be a convex lattice polygon which is a minimizer for the right-hand side of (1-2-2), that is, it minimizes $\ell_\Omega(\partial P)$. We will assume that P is nondegenerate, the degenerate case being a straightforward extension. Let A (resp. B) denote the minimal (resp. maximal) x coordinate of any point in P, and similarly let C (resp. D) denote the minimal (resp. maximal) y coordinate of any point in P. Let P' denote the convex lattice polygon given by the convex hull of P with the additional points $(A, D), (B, D), (A, C)$. Note that we have $P \subset P'$, and moreover $|P \cap \mathbb{Z}^2| \le |P' \cap \mathbb{Z}^2|$. Let $p_1, \dots, p_{q+1} \in \mathbb{Z}^2$ denote the lattice points encountered as we traverse $\partial P'$ in the counterclockwise direction from (A, C) to (B, D). For $s = 1, \dots, q$, let $(i_s, j_s) := p_{s+1} - p_s$ denote the corresponding displacement vectors. Then we have

$$k+1 = |\partial P \cap \mathbb{Z}^2| \le |\partial P' \cap \mathbb{Z}^2| = \sum_{s=1}^{q}(i_s + j_s) + q.$$

Moreover, using (4-3-3) we have $\ell_\Omega(\partial P) = \ell_\Omega(\partial P')$. Therefore the right-hand side of (1-2-1) is less than or equal to $\sum_{s=1}^{q} \|(i_s, j_s)\|_\Omega^* = \ell_\Omega(\partial P') = \ell_\Omega(\partial P)$. $\square$

5 Constructing curves in four-dimensional convex toric domains

In this section we complement Corollary 4.3.10 by proving a corresponding upper bound for $\tilde{\mathfrak{g}}_k(X_\Omega)$, thereby completing the proof of Theorem 1.2.8. In Section 5.1 we prove that the formal curve component C in $\widetilde{X}_\Omega$ with local tangency constraint $\lessdot\mathcal{T}^{(k)}p\gtrdot$ and positive asymptotics the minimal word $w_{\min}$ of index $2k$ is formally perturbation invariant with respect to some generic $J_{\partial\widetilde{X}_\Omega} \in \mathcal{J}(\partial\widetilde{X}_\Omega)$. After establishing this, we then show that the moduli space $\mathcal{M}^J_{\widetilde{X}_\Omega}(w_{\min})\lessdot\mathcal{T}^{(k)}p\gtrdot$ is in fact nonempty for some (and hence any) $J \in \mathcal{J}^{J_{\partial\widetilde{X}_\Omega}}(\widetilde{X}_\Omega)$, thereby achieving our desired upper bound.

More precisely, we show in Proposition 5.4.5 that (except for the case $w_{\min} = e_{0,k}$ with $k \geq 2$) there is $J \in \mathcal{J}^{J_{\partial\widetilde{X}_\Omega}}(\widetilde{X}_\Omega)$ such that $\mathcal{M}^J_{\widetilde{X}_\Omega}(w_{\min})$ is regular with nonzero signed count. By Proposition 2.4.2, this implies that $\mathcal{M}^J_{\widetilde{X}_\Omega}(w_{\min}) \neq \varnothing$ for any $J \in \mathcal{J}^{J_{\partial\widetilde{X}_\Omega}}(\widetilde{X}_\Omega)$ — recall that the empty moduli space is automatically regular. Since we will then have verified all the hypotheses of Proposition 3.7.1, this also proves the stabilization property for four-dimensional convex toric domains.

To prove that suitable curves exist we argue as follows. By Proposition 2.2.3 we have

$$\#\mathcal{M}_{\widetilde{X}_\Omega}(w_{\min})\lessdot\mathcal{T}^{(k)}p\gtrdot = \#\mathcal{M}_{\widetilde{X}_\Omega}(w_{\min})\lessdot(k)\gtrdot_E,$$

ie we can swap the local tangency constraint with a skinny ellipsoidal constraint. In Section 5.2, we show that every curve in the latter moduli space counts positively, so it suffices to show that it is nonempty. In Section 5.3 we give a biased summary of Hutchings–Taubes' obstruction bundle gluing, adapted to the case of cobordisms, and in Section 5.4 we explain how to apply obstruction bundle gluing in order to piece together the curves we need inductively from certain basic curves with very simple top ends. Finally, in Section 5.5 we use the cobordism map in ECH to establish the base cases for our induction.

5.1 Invariance of minimal word counts

Our main goal in this subsection is to prove Proposition 5.1.4, which establishes formal perturbation invariance for those moduli spaces corresponding to weakly permissible words $w_{\min}$ of minimal action; see Definition 4.2.4. At first glance it seems plausible that we can rule out degenerations using minimality, but some care is needed due to the possibility of multiply covered curves of negative index. Recall that we are considering degenerations that might occur for a generic path J_t, with $0 \leq t \leq 1$, in $\mathcal{J}^{J_{\partial\widetilde{X}_\Omega}}(\widetilde{X}_\Omega)$. Thus the (fixed) almost complex structure $J_{\partial\widetilde{X}_\Omega}$ on the symplectization levels can be assumed to be generic. If a curve with top $w_{\min}$ does degenerate, the resulting building has a main component C_0 in $\widetilde{X}_\Omega$ that satisfies the tangency constraint, as well as some other components that may be assembled into representatives of a union of connected formal buildings each of which has one negative end that attaches to C_0.

We first consider the properties of such a formal building. For definitions of the language used here, see Definitions 2.1.1, 2.3.2 and 2.3.3.

Lemma 5.1.1 *Let C be a connected formal building with main level in $\widetilde{X}_\Omega$ and some number of symplectization levels in $\mathbb{R} \times \partial \widetilde{X}_\Omega$, except that exactly one negative end of some curve component is not paired with any positive end of a curve component in a lower level. Assume that each component of C in a symplectization level is a (possibly trivial) formal cover of some formal curve component $\overline{C}$ which is either trivial or else satisfies $\operatorname{ind}(\overline{C}) \geq 1$. Then we have $\operatorname{ind}(C) \geq 0$, with equality if and only if every component of C is trivial.*

Note that Lemma 5.1.1 does not involve any local tangency constraints.

Proof Let $C_1, \dots, C_q$ denote the components of C which have at least one negative end, and let $b_1, \dots, b_q$ denote the corresponding numbers of negative ends. Observe that since C has genus zero, there must have at least $\sum_{i=1}^q (b_i - 1)$ components without any negative ends, and by (4-1-1) and (4-1-3) each of these has index at least 3. Therefore we have

$$\operatorname{ind}(C) \geq \sum_{i=1}^q (\operatorname{ind}(C_i) + 3(b_i - 1)).$$

We will show that for $i = 1, \dots, q$ we have $\operatorname{ind}(C_i) + 3(b_i - 1) \geq 0$, with equality if and only if C_i is a trivial cylinder, from which the result immediately follows.

Let D denote one of the components[9] $C_1, \dots, C_q$. Let a and b denote the respective numbers of positive and negative ends D, and let $e^+ \leq a$ and $e^- \leq b$ denote the numbers of positive and negative ends which are elliptic. We assume that D is a κ–fold cover of $\overline{D}$ for some $\kappa \in \mathbb{Z}_{\geq 1}$, where by assumption $\overline{D}$ is either trivial or satisfies $\operatorname{ind}(\overline{D}) \geq 1$. We denote by $\overline{a}, \overline{b}, \overline{e}^+, \overline{e}^-$ the analogues of the above for $\overline{D}$.

For each puncture or point in the domain of D, let us define its *excess branching* to be one less than its ramification order as a cover of $\overline{D}$.[10] Let $E^\pm$ be the total excess branching at all positive (resp. negative) elliptic ends of D, and similarly let $H^\pm$ be the total excess branching at all positive (resp. negative) hyperbolic ends of D. By elementary Riemann–Hurwitz considerations we have the following:

- $a = \kappa \overline{a} - E^+ - H^+$ and $b = \kappa \overline{b} - E^- - H^-$.
- $e^+ = \kappa \overline{e}^+ - E^+$ and $e^- = \kappa \overline{e}^- - E^-$.
- $0 \leq E^+, E^-, H^+, H^- \leq \kappa - 1$.
- $E^+ + E^- + H^+ + H^- = 2(\kappa - 1)$.

By (4-1-3) we then have

$$\operatorname{ind}(D) - \kappa \operatorname{ind}(\overline{D}) = (a + b - 2) - \kappa(\overline{a} + \overline{b} - 2) + (e^+ - e^-) - \kappa(\overline{e}^+ - \overline{e}^-) = 2\kappa - 2 - 2E^+ - H^+ - H^-.$$

[9]In this paper each component lies in a single level; it is not a "matched component" in the sense of [30].

[10]That is, if u is locally given by $z \mapsto z^k$, then the point in the domain corresponding to the origin has excess branching $k - 1$.

Consider first the case that $\overline{D}$ is trivial. Then we have $\mathrm{ind}(\overline{D}) = 0$ and $\overline{a} = \overline{b} = 1$. If the ends of $\overline{D}$ are elliptic, then we have $H^+ = H^- = 0$, and hence

$$\mathrm{ind}(D) = 2\kappa - 2 - 2E^+ \geq 0.$$

Similarly, if the ends of $\overline{D}$ are hyperbolic, then we have $E^+ = E^- = 0$, and hence

$$\mathrm{ind}(D) = 2\kappa - 2 - H^+ - H^- \geq 0.$$

In either case we have $\mathrm{ind}(D) + 3(b-1) = 0$ if and only if $a = b = 1$, in which case D is a trivial cylinder.

Now consider the case that $\overline{D}$ is nontrivial, and hence $\mathrm{ind}(\overline{D}) \geq 1$. We have

$$\begin{aligned}
\mathrm{ind}(D) + 3(b-1) &\geq 3\kappa - 2 - 2E^+ - H^+ - H^- + 3(b-1) \\
&= 3\kappa - 2 - E^+ - (E^+ - H^+ - H^-) + 3(b-1) \\
&\geq 3\kappa - 2 - (\kappa - 1) - 2(\kappa - 1) + 3(b-1) \\
&= 1 + 3(b-1) \geq 1.
\end{aligned}$$

□

Lemma 5.1.2 *Let C be curve in $\widetilde{X}_\Omega$ satisfying a constraint $\lll\mathcal{T}^{(m)}p\ggg$ for some $m \in \mathbb{Z}_{\geq 1}$, and assume that C is a κ–fold cover of its underlying simple curve $\overline{C}$ for some $\kappa \in \mathbb{Z}_{\geq 1}$. Let*

- *e (resp. $\overline{e}$) denote the number of elliptic positive ends of C (resp. $\overline{C}$),*
- *h (resp. $\overline{h}$) denote the number of hyperbolic positive ends of C (resp. $\overline{C}$),*
- *$q = e + h$ (resp. $\overline{q} = \overline{e} + \overline{h}$) denote the total number of positive ends of C (resp. $\overline{C}$).*

Then we have

$$\mathrm{ind}(C) - \kappa\,\mathrm{ind}(\overline{C}) \geq \max(q - 2 - \kappa\overline{q} + 2\kappa + e - \kappa\overline{e}, \kappa\overline{h} - h).$$

Note that in particular we have $h \leq \kappa\overline{h}$ and hence $\mathrm{ind}(C) \geq \kappa\,\mathrm{ind}(\overline{C})$.

Proof Let E (resp. H) denote the sum of the excess branching at all elliptic (resp. hyperbolic) punctures of C, and let B denote the excess branching of the point in the domain of C which satisfies the constraint $\lll\mathcal{T}^{(m)}p\ggg$. The curve $\overline{C}$ satisfies a constraint $\lll\mathcal{T}^{(\overline{m})}p\ggg$ for some $\overline{m} \in \mathbb{Z}_{\geq 1}$. With the help of the Riemann–Hurwitz formula we have

$$B \leq \kappa - 1, \quad e = \kappa\overline{e} - E \quad h = \kappa\overline{h} - H, \quad B + E + H \leq 2\kappa - 2, \quad m \leq (B+1)\overline{m}.$$

For $s = 1, \dots, q$, let γ_s be the s^{th} positive end of C, which we take to be either e_{i_s,j_s} or h_{i_s,j_s}. Similarly, for $s = 1, \dots, \overline{q}$, let $\overline{\gamma}$ be the s^{th} positive end of $\overline{C}$, which we take to be either $e_{\overline{i}_s,\overline{j}_s}$ or $h_{\overline{i}_s,\overline{j}_s}$.

By (4-1-3), we have

$$\mathrm{ind}(C) = q - 2 + 2\sum_{s=1}^{q}(i_s + j_s) + e - 2m \quad \text{and} \quad \mathrm{ind}(\overline{C}) = \overline{q} - 2 + 2\sum_{s=1}^{\overline{q}}(\overline{i}_s + \overline{j}_s) + \overline{e} - 2\overline{m},$$

and therefore

$$\begin{aligned}\operatorname{ind}(C)-\kappa\operatorname{ind}(\overline{C}) &= (q-2)-\kappa(\overline{q}-2)+e-\kappa\overline{e}-2m+2\overline{m}\kappa \\ &\geq q-2-\kappa\overline{q}+2\kappa+e-\kappa\overline{e}-2(B+1)\overline{m}+2\overline{m}\kappa \\ &= q-2-\kappa\overline{q}+2\kappa+e-\kappa\overline{e}+2\overline{m}(\kappa-B-1) \\ &\geq q-2-\kappa\overline{q}+2\kappa+e-\kappa\overline{e}+2\kappa-2B-2 \\ &\geq q-2-\kappa\overline{q}+2\kappa+e-\kappa\overline{e}.\end{aligned}$$

Note that we have $q-\kappa\overline{q}=-E-H\geq B-2\kappa+2$, so we also have

$$\begin{aligned}\operatorname{ind}(C)-\kappa\operatorname{ind}(\overline{C}) &\geq B-2\kappa+2-2+2\kappa+e-\kappa\overline{e}+2\kappa-2B-2 \\ &= -B+e-\kappa\overline{e}+2\kappa-2=-B-E+2\kappa-2 \\ &\geq H=\kappa\overline{h}-h.\end{aligned}$$

□

Recall that, for $(i,j)\in\mathbb{Z}^2_{\geq 0}\setminus\{(0,0)\}$, the pair of acceptable Reeb orbits $e_{i,j}, h_{i,j}$ come from perturbing an S^1–family of Reeb orbits in $\mu^{-1}(p_{i,j})\subset\partial X^{\mathrm{FR}}_\Omega$. The precise perturbation is controlled by a choice of Morse function $f\colon S^1\to\mathbb{R}$, which we can assume is perfect. We take $\widetilde{X}_\Omega$ to be an arbitrarily small perturbation of X^{FR}_Ω and, fixing $J_{\mathrm{MB}}\in\mathcal{J}(X^{\mathrm{FR}}_\Omega)$, we can correspondingly consider $J\in\mathcal{J}(\widetilde{X}_\Omega)$ which is a small perturbation of J_{MB}. Then by the standard correspondence between Morse gradient flowlines and Morse–Bott cascades, one expects J–pseudoholomorphic cylinders with positive asymptotic $e_{i,j}$ and negative asymptotic $h_{i,j}$ to correspond to gradient flow lines for f, of which there are precisely two, and they have canceling signs. Indeed, by the Morse–Bott techniques developed in [1; 4] — see also [40, Section 10.3] for a detailed discussion and also an alternative perspective — we have the following standard result:

Lemma 5.1.3 *There exists generic $J_{\partial\widetilde{X}_\Omega}\in\mathcal{J}(\partial\widetilde{X}_\Omega)$ such that for each acceptable pair $e_{i,j}$ and $h_{i,j}$, there are precisely two J–holomorphic cylinders in $\mathbb{R}\times\partial\widetilde{X}_\Omega$ with positive asymptotic $e_{i,j}$ and negative asymptotic $h_{i,j}$. Moreover, these are regular and count with opposite signs.*

The cylinders in Lemma 5.1.3 have energy $\widetilde{\mathcal{A}}(e_{i,j})-\widetilde{\mathcal{A}}(h_{i,j})$, which by Lemma 4.1.1(a) is very small; we will refer to them as *low-energy cylinders*.

Proposition 5.1.4 *Assume $a>b$. For $k\in\mathbb{Z}_{\geq 1}$, let $w_{\min}$ be the weakly permissible word of index $2k$ with minimal $\widetilde{\mathcal{A}}_\Omega$ value. Then the formal curve component in $\widetilde{X}_\Omega$ having positive asymptotics $w_{\min}$ is formally perturbation invariant with respect to any generic $J_{\partial\widetilde{X}_\Omega}\in\mathcal{J}(\partial\widetilde{X}_\Omega)$ as in Lemma 5.1.3.*

Proof By Lemma 4.3.3, $w_{\min}$ must be elliptic, and must take one of the forms (1)–(4) from Corollary 4.3.9. Let C denote the formal curve component in $\widetilde{X}_\Omega$ having positive asymptotics $w_{\min}$. After possibly replacing C with another formal curve component which it formally covers, we can assume that C is simple, ie we can ignore the case $w_{\min}=e_{0,k}$ with $k\geq 2$. Now consider a stable formal building $C'\in\overline{\mathcal{F}}_{X,A}(\Gamma)\lll\mathcal{T}^{(k)}p\ggg$ satisfying conditions (A1) and (A2) from Definition 2.4.1. We seek to show that C' satisfies either (B1) or (B2) with respect to $J_{\partial\widetilde{X}_\Omega}$.

Let C_0 denote the main component of C', ie the one in $\widetilde{X}_\Omega$ which carries the local tangency constraint. We can assume that C' involves at least one symplectization level, since otherwise we must have $C' = C_0$, whence (B1) holds. Let $q \in \mathbb{Z}_{\geq 1}$ denote the number of positive ends of C_0. Excluding C_0, we can view C' as some number q of connected buildings with one unpaired negative end, precisely as in Lemma 5.1.1. Denote these by $C_1, \dots, C_q$. We have $\operatorname{ind}(C_s) \geq 0$, with equality if and only if C_s consists entirely of trivial cylinders. In particular, if the unpaired negative end of C_s is hyperbolic, the fact that its top is elliptic implies that $\operatorname{ind}(C_s) \geq 1$. Thus if C_s has a hyperbolic end, we have $\operatorname{ind}(C_s) \geq 1$ so that $\sum_{s=1}^q \operatorname{ind}(C_s) \geq h$, where h denotes the number of hyperbolic ends of C_0.

Next suppose D is one of $C_1, \dots, C_q$ with $\operatorname{ind}(D) = 1$. Then we claim that D is a low-energy cylinder (that is, a cylinder connecting some $e_{i,j}$ and $h_{i,j}$), possibly along with extra trivial cylinders in other levels. Indeed, for parity reasons the unpaired negative end must be hyperbolic, say $h_{i,j}$ for some $(i,j) \in \mathbb{Z}^2_{\geq 1}$. Let $w'_{\min}$ denote the word obtained from $w_{\min}$ by replacing the set of the positive ends of D by $e_{i,j}$. Then $w'_{\min}$ is strongly permissible and satisfies $\operatorname{ind}(w'_{\min}) = \operatorname{ind}(w_{\min})$ and $\widetilde{\mathscr{A}}(w'_{\min}) \leq \widetilde{\mathscr{A}}(w_{\min})$, with equality only if $w_{\min} = e_{i,j}$. Then by minimality of $w_{\min}$ we must have $w_{\min} = e_{i,j}$, and the claim follows by energy considerations.

Assume now that C_0 is a κ–fold cover of a simple formal curve component $\overline{C}_0$ for some $\kappa \in \mathbb{Z}_{\geq 1}$. By assumption we have $\operatorname{ind}(\overline{C}_0) \geq -1$. Let e denote the number of elliptic ends of C_0 and define $\overline{q}, \overline{e}, \overline{h}$ analogously for $\overline{C}_0$. Suppose first that we have $\overline{h} = 0$ and hence $h = 0$. In this case, $\overline{C}_0$ has only elliptic ends and hence its index must be even, so we have a fortiori $\operatorname{ind}(\overline{C}_0) \geq 0$. Applying Lemma 5.1.2 yields

$$0 = \operatorname{ind}(C_0) + \sum_{s=1}^{q} \operatorname{ind}(C_s) \geq \kappa \operatorname{ind}(\overline{C}_0) + \kappa\overline{h} - h + \sum_{s=1}^{q} \operatorname{ind}(C_s) \geq \sum_{s=1}^{q} \operatorname{ind}(C_s).$$

This is only possible if C_s consists entirely of trivial cylinders for $s = 1, \dots, q$, but this contradicts the stability of C'.

Now suppose that $\overline{h} \geq 1$, and moreover that the covering $C_0 \to \overline{C}_0$ is not ramified at any positive punctures. In this case we have $q = \kappa\overline{q}$, $e = \kappa\overline{e}$, and $h = \kappa\overline{h}$, and hence

$$\operatorname{ind}(C_0) - \kappa \operatorname{ind}(\overline{C}_0) \geq q - 2 - \kappa\overline{q} + 2\kappa + e - \kappa\overline{e} = -2 + 2\kappa.$$

We then have

$$0 = \operatorname{ind}(C_0) + \sum_{s=1}^{q} \operatorname{ind}(C_s) \geq \kappa \operatorname{ind}(\overline{C}_0) - 2 + 2\kappa + h \geq \kappa - 2 + \kappa\overline{h} \geq 2\kappa - 2.$$

This is only possible if $\kappa = 1$, and hence $\operatorname{ind}(C_0) = -1$. Then we have $\operatorname{ind}(C_s) \leq 1$ for $s = 1, \dots, q$, with equality for at most one s. By the above discussion and stability considerations, we conclude that C' is a breaking of the form (B2).

Finally, suppose that $\overline{h} \geq 1$ and also one of the positive punctures of C_0 is ramified. Then the corresponding component C_s cannot be a low-energy cylinder, and so as explained above, it must then satisfy $\operatorname{ind}(C_s) \geq 2$.

Thus we have $\sum_{s=1}^{q} \operatorname{ind}(C_s) \geq h+1$, and hence

$$0 = \operatorname{ind}(C) + \sum_{s=1}^{q} \operatorname{ind}(C_s) \geq -\kappa + \kappa \bar{h} - h + \sum_{s=1}^{q} \operatorname{ind}(C_s) \geq \kappa(\bar{h}-1) + 1 \geq 1, \tag{5-1-1}$$

which is impossible. □

5.2 Automatic transversality and positive signs

Our main goal in this subsection is to prove Proposition 5.2.2, which roughly states that rigid curves in dimension four count with positive sign as long as none of the punctures are asymptotic to positive hyperbolic Reeb orbits (such as $h_{i,j}$). This will later allow us conclude that certain moduli spaces have nonzero signed counts simply by showing that they are nonempty. The content of this subsection is likely well-known to experts, but we include a precise statement and proof for the sake of completeness.

To begin, let us recall a version of the automatic transversality criterion from [39]. A pseudoholomorphic curve satisfying this criterion is regular even without any genericity assumption on the almost complex structure. It is natural to state the results in this subsection in arbitrary genus.

Theorem 5.2.1 *Let X be a four-dimensional compact symplectic cobordism, take $J \in \mathcal{J}(X)$, and let C be a nonconstant asymptotically cylindrical J–holomorphic curve component of genus $g(C)$ in $\widehat{X}$ such that all of the asymptotic Reeb orbits are nondegenerate. Let $h_+(C)$ denote the number of punctures (positive or negative) which are asymptotic to positive hyperbolic Reeb orbits, and let $Z(C)$ be the count (with multiplicities) of zeros of the derivative of a map representing C. If*

$$2g(C) - 2 + h_+(C) + 2Z(C) < \operatorname{ind}(C),$$

then C is regular.

We point out that the quantity $Z(C)$ is always nonnegative, and is zero if and only if C is immersed.

As above let X be a four-dimensional compact symplectic cobordism with $\partial^{\pm} X$ nondegenerate. Let us pick coherent orientations for all moduli spaces of immersed asymptotically cylindrical pseudoholomorphic curves in X following the framework of [25, Section 9] — this is quite to similar to the approach of [3], see also [40, Section 11]. This involves the following main ingredients. An *orientation triple* is a triple $(\Sigma, E, \{S_k\})$, where

- Σ is a Riemann surface with positive and negative cylindrical ends;
- E is a Hermitian complex line bundle over Σ, trivialized over each end;
- at the k^{th} end we have a smooth family of symmetric 2×2 matrices, $S_k \in C^\infty(S^1, \operatorname{End}_{\mathbb{R}}^{\text{sym}}(\mathbb{R}^2))$, such that the asymptotic operator

  $$\mathbb{A} : C^\infty(S^1, \mathbb{C}) \to C^\infty(S^1, \mathbb{C}), \quad \eta(t) \mapsto -J_0 \partial_t \eta(t) - S_k(t)\eta(t),$$

  is nondegenerate, ie does not have 0 as an eigenvalue.

Here J_0 denotes the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

For each orientation triple $(\Sigma, E, \{S_k\})$, we denote by $\mathfrak{D}(\Sigma, E, \{S_k\})$ the space of differential operators $D\colon C^\infty(E) \to C^\infty(T^{0,1}\Sigma \otimes E)$ which look locally like a zeroth order perturbation of the Cauchy–Riemann operator $\overline{\partial}$ on E for some choice of conformal structure on Σ, and where on the k^{th} end in cylindrical coordinates D has the form

$$\psi(s,t) \mapsto (\overline{\partial} + M_k(s,t))\psi(s,t)\,d\overline{z}, \quad \text{with} \lim_{|s|\to\infty} M_k(s,t) = S_k(t).$$

Each $D \in \mathfrak{D}(\Sigma, E, \{S_k\})$ extends to an operator $W^{1,2}(E) \to L^2(T^{0,1}\Sigma \otimes E)$, and this is Fredholm since the corresponding asymptotic operators are nondegenerate. Moreover, the space of such operators is an affine space and thus contractible, and hence the set of orientations of the determinant lines of any two elements of $\mathfrak{D}(\Sigma, E, \{S_k\})$ are naturally identified. We denote the set of these two possible orientations by $\mathfrak{O}(\Sigma, E, \{S_k\})$.

Now, to orient moduli spaces of curves we choose preferred orientations in $\mathfrak{O}(\Sigma, E, \{S_k\})$ ranging over all possible orientation triples $(\Sigma, E, \{S_k\})$, subject to axioms (OR1), (OR2), (OR3) and (OR4). These axioms roughly correspond to compatibility under gluing and disjoint unions and agreement with the natural complex orientation whenever D happens to be complex-linear. Henceforth we will implicitly assume that a choice of coherent orientations has been made. Given such a choice, any moduli space of regular, immersed, asymptotically cylindrical curves in X naturally inherits an orientation. Indeed, for a curve C in such a moduli space we have an associated orientation triple $(\Sigma, E, \{S_k\})$, where Σ is the domain of the curve, $E = N_C$ its normal bundle, and $\{S_k\}$ is given by the induced asymptotic operators at each puncture; see eg [40, Section 3]. Then the associated deformation operator D_C lies in $\mathfrak{D}(\Sigma, E, \{S_k\})$, and by regularity its determinant line is its kernel, which is also the tangent space to the corresponding moduli space.

In the special case of Fredholm index zero, surjectivity of D_C means that we have an identification $\det(D_C) = \mathbb{R}$, and the associated sign $\varepsilon(C) \in \{1, -1\}$ is determined by whether our chosen orientation of $\det(D_C)$ agrees or disagrees with the canonical orientation of $\mathbb{R}$.

Proposition 5.2.2 *Suppose that C is an immersed, somewhere injective, asymptotically cylindrical J–holomorphic rational curve in a four-dimensional symplectic cobordism X. Assume that we have $\operatorname{ind}(C) = 0$, and all of the asymptotic Reeb orbits of C are nondegenerate and are either elliptic or negative hyperbolic. Then we have $\varepsilon(C) = 1$.*

Proof Since C is immersed, it has a well-defined normal bundle $N_C \to C$ and associated deformation operator D_C, which we can view as a Fredholm operator $W^{1,2}(N_C) \to L^2(T^{0,1}\Sigma \otimes E)$ (here Σ denotes the domain of C). According to [40, Theorem 3.53], any two nondegenerate asymptotic operators with the same Conley–Zehnder index are homotopic through nondegenerate asymptotic operators. In particular, if γ is an elliptic or negative hyperbolic Reeb orbit, we can deform its asymptotic operator $\mathbb{A}_\gamma$ through nondegenerate asymptotic operators to be of the form given in [40, Example 3.60], ie $\mathbb{A} = -J_0\partial_t - \epsilon$ for some $\epsilon \in \mathbb{R} \setminus 2\pi\mathbb{Z}$. Note that in this case the associated symplectic parallel transport rotates the contact planes along γ by total angle ϵ in the chosen trivialization. It follows that we can deform D_C through Fredholm operators, after which the asymptotic operator at each end is complex-linear. The resulting Cauchy–Riemann type operator might not be complex-linear, but we can further deform it to its complex-linear part. We can take this latter deformation to be along an affine line and hence asymptotically constant on each end, meaning that it is a deformation through Fredholm operators. Combining these two deformations, the corresponding $\mathbb{Z}/2$ spectral flow gives the sign $\varepsilon(C)$.

At the same time, by automatic transversality, the Fredholm operators in this deformation are isomorphisms throughout, and hence the spectral flow is trivial. Indeed, this follows by invoking the criterion $2g(\Sigma)-2+h_+(C) < \operatorname{ind}(C)$, after noting that Theorem 5.2.1 holds also on the level of operators in $\mathfrak{D}(\Sigma, E, \{S_k\})$. Finally, observe that we have endowed the determinant line of the complex linear operator at the end of the deformation with its canonical complex orientation, which is necessarily positive. □

Remark 5.2.3 (i) The above discussion has a natural analogue in a symplectization $\mathbb{R}\times Y$, in which we consider the signed count of index-one curves modulo target translations. Note that positivity does not hold for the low-energy cylinders in $\mathbb{R}\times\partial\widetilde{X}_\Omega$ that connect $e_{i,j}$ to $h_{i,j}$, and indeed in that case the negative end is positive hyperbolic.

(ii) In Proposition 2.2.3 we assert that each curve in the moduli space $\#\mathcal{M}^J_{X,A}(\Gamma^+;\Gamma^-)\lll\mathcal{T}^{(m)}p\ggg$ also counts positively when the orbits in Γ^+ and Γ^- are elliptic or negative hyperbolic. To prove this, one must check that the tangency constraint is always compatible with the orientation. This is proved in [30, Lemma 2.3.5].

5.3 Obstruction bundle gluing

In this subsection we briefly review the Hutchings–Taubes theory [24; 25] of obstruction bundle gluing, after making the minor adaptations necessary to glue curves in cobordisms rather than symplectizations. As noted also in [29], since the gluing is essentially local to the neck region, which is the same in both cases, the underlying analysis of [24; 25] still applies in the cobordism setting.

Let X^+ and X^- be four-dimensional compact symplectic cobordisms with common strict contact boundary $Y := \partial^- X^+ = \partial^+ X^-$. We will assume that all Reeb orbits of Y under discussion are nondegenerate. By concatenating, we can form the compact symplectic cobordism $X := X^+ \circledcirc X^-$. Fix a generic admissible almost complex structure $J_Y \in \mathcal{J}(Y)$, and let $J^\pm$ be generic admissible almost complex structures on $\widehat{X}^\pm$ which restrict to J_Y on the corresponding ends, ie we have $J^+ \in \mathcal{J}_{J_Y}(X^+)$ and $J^- \in \mathcal{J}^{J_Y}(X^-)$. Let α^+, β^+, β^- and α^- be tuples of Reeb orbits in ∂^+X^+, Y, Y and ∂^-X^-, respectively.

Definition 5.3.1 (cf [24, Definition 1.9]) A *gluing pair* is a pair (u_+, u_-) consisting of immersed pseudoholomorphic curves $u^+ \in \mathcal{M}^{J^+}_{X^+}(\alpha^+;\beta^+)$ and $u_- \in \mathcal{M}^{J^-}_{X^-}(\beta^-;\alpha^-)$ such that:

(a) $\operatorname{ind}(u_+) = \operatorname{ind}(u_-) = 0$.

(b) u_+ and u_- are simple.[11]

(c) For each simple Reeb orbit γ in Y, the total covering multiplicity of Reeb orbits covering γ in the list β^+ is the same as the total for β^-.

(d) Each component of u_+ has exactly one negative end, and each component of u_- has exactly one positive end.

Here we consider a possibly disconnected curve to be simple if and only if each component is simple and no two components have the same image.

Remark 5.3.2 Condition (d) is a somewhat artificial simplifying assumption, which is used to ensure that we do not encounter higher-genus curves after gluing rational curves. Alternatively, the following discussion holds equally well if we drop this condition and simply allow $u_\pm$ and also the gluing result to have higher genus.

J^+ and J^- can also be concatenated to give $J \in \mathcal{J}(X)$ satisfying $J|_{X^\pm} = J^\pm|_{X^\pm}$. For each $R > 0$, let

$$X_R := X^+ \circledcirc ([-R, R] \times Y) \circledcirc X^-$$

denote the compact symplectic cobordism given by inserting a finite piece of the symplectization of Y in between X^+ and X^-. Let also $J_R \in \mathcal{J}(X_R)$ denote the concatenated almost complex structure which satisfies $J_R|_{X^\pm} = J^\pm|_{X^\pm}$ and $J_R|_{[-R,R]\times Y} = J_Y|_{[-R,R]\times Y}$. Note that the family $\{J_R\}_{R\in[0,\infty)}$ realizes neck-stretching along Y, with the limit $R \to \infty$ corresponding to (J^+, J^-)–holomorphic buildings in the broken cobordism $X^+ \circledcirc X^-$. We denote the corresponding parametrized moduli space by $\mathcal{M}_X^{\{J_R\}}(\alpha^+; \alpha^-)$ and its SFT compactification by $\overline{\mathcal{M}}_X^{\{J_R\}}(\alpha^+; \alpha^-)$.

Given a gluing pair (u_+, u_-), Hutchings and Taubes glue together u_+ and u_- after possibly inserting a union u_0 of index-zero branched covers of trivial cylinders in an intermediate symplectization level $\mathbb{R} \times Y$. This is more complicated than the typical gluing encountered in SFT, where the intermediate level u_0 would be barred from participating in the gluing since it is irregular. Indeed, note that u_0 lives in a moduli space $\mathcal{M}_Y^{J_Y}(\beta^+; \beta^-)$ of branched covers which has dimension $2b$, where b corresponds to the number of interior branch points. The main computation of [24] determines the signed number $\#G(u_+, u_-)$ of ends of $\mathcal{M}_X^{\{J_R\}}(\alpha^+; \alpha^-)$ which arise by gluing (u_+, u_-) in this way.

Analogously to [25, Section 5], one can perform pregluing to produce an approximately J_R–holomorphic curve in $\widehat{X}_R$ which interpolates via cutoff functions between u_+ on $\widehat{X}^+$, u_0 on $\mathbb{R} \times Y$, and u_- on the $\widehat{X}^-$. The index of the normal deformation operator of u_0 is $-2b$ and the kernel can be shown to be trivial, so the cokernels as u_0 varies form a well-defined rank $2b$ real vector bundle over (a large compact subspace of) $\mathcal{M}_Y^{J_Y}(\beta^+; \beta^-)$, called the "obstruction bundle". From the gluing analysis we get a section $\mathfrak{s}$ such that the gluing successfully goes through for $u_0 \in \mathcal{M}_Y^{J_Y}(\beta^+; \beta^-)$ precisely if $\mathfrak{s}(u_0) = 0$. The computation of $\#G(u_-, u_+)$ therefore amounts to counting zeros of $\mathfrak{s}$.

[11] In the symplectization setting, Hutchings and Taubes also allow some components of u_+ and u_- to be trivial cylinders, subject to a certain combinatorial condition.

More precisely, the number $\#G(u_-,u_+)$ is defined in several steps as follows. For each $R \geq 0$, fix a metric on $\widehat{X}_R$ which is a product metric on the cylindrical ends $(-\infty,0]\times\partial^- X$ and $[0,\infty)\times\partial^+ X$ and on the neck region $[-R,R]\times Y$. We assume this metric does not depend on R except for the varying length of the neck.

Definition 5.3.3 (cf [24, Definition 1.10]) For $\delta>0$, let $\mathscr{C}_\delta(u_+,u_-)$ denote the union over $R\in(1/\delta,\infty)$ of the set of surfaces in $\widehat{X}_R$ which are immersed apart from finitely many points and can be decomposed as $C_-\cup C_0\cup C_+$, where:

- There is a section ψ_+ of the normal bundle of u_+ restricted to
$$\left(\left[-\frac{1}{\delta},0\right]\times Y\right)\cup X^+\cup([0,\infty)\times\partial^+X^+)$$
such that $\|\psi_+\|<\delta$ and C_+ is the exponential map image of ψ_+ after identifying $[-1/\delta,0]\times Y$ with $[R-1/\delta,R]\times Y$.
- There is a section ψ_- of the normal bundle of u_- restricted to
$$((-\infty,0]\times\partial^-X^-)\cup X^+\cup\left(\left[0,\frac{1}{\delta}\right]\times Y\right)$$
such that $\|\psi_-\|<\delta$ and C_+ is the exponential map image of ψ_- after identifying $[0,1/\delta]\times Y$ with $[-R,-R+1/\delta]\times Y$.
- C_0 lies in the δ–tubular neighborhood of $[-R,R]\times(\beta^+\cup\beta^-)\subset[R,R]\times Y$, and we have $\partial C_0=\partial C_-\cup\partial C_+$, with the positive boundary of C_0 coinciding with the negative boundary of C_+ and the negative boundary of C_0 coinciding with the positive boundary of C_-.

Definition 5.3.4 Let $\mathscr{G}_\delta(u_+,u_-)$ denote the set of index-zero curves in $\mathcal{M}_X^{\{J_R\}}(\alpha^+;\alpha^-)\cap\mathscr{C}_\delta(u_+,u_-)$.

By the following lemma, $\mathscr{G}_\delta(u_+,u_-)$ represents curves in $\mathcal{M}_X^{\{J_R\}}(\alpha^+;\alpha^-)$ which are "δ–close" to breaking into an SFT building corresponding to the gluing pair (u_+,u_-):

Lemma 5.3.5 (cf [24, Lemma 1.11]) *Given a gluing pair (u_+,u_-), there exists $\delta_0>0$ such that for any $\delta\in(0,\delta_0)$ and any sequence of curves $u_1,u_2,u_3,\ldots\in\mathscr{G}_\delta(u_+,u_-)$, there is a subsequence which converges in the SFT sense to either a curve in $\mathcal{M}_{X_{R_\infty}}^{J_{R_\infty}}(\alpha^+;\alpha^-)$ for some $R_\infty\in[0,\infty)$, or else to an SFT building with top level u_+ in $\widehat{X}^+$, bottom level u_- in $\widehat{X}^-$, and some number (possibly zero) of intermediate symplectization levels in $\mathbb{R}\times Y$ each consisting entirely of unions of index-zero branched covers of trivial cylinders.*

Finally, we define the count of ends $\#G(u_+,u_-)$:

Definition 5.3.6 For a gluing pair (u_+,u_-) and δ_0 as above, choose $0<\delta'<\delta<\delta_0$ and an open subset $U\subset\mathscr{G}_\delta(u_+,u_-)$ containing $\mathscr{G}_{\delta'}(u_+,u_-)$ such that $\overline{U}$ has finitely many boundary points. We then define $\#G(u_-,u_+)$ to be minus the signed count of boundary points of $\overline{U}$.

By Lemma 5.3.5, the count $\#G(u_+,u_-)$ is independent of the choice of δ', δ and U.

The analogue of the main result of Hutchings and Taubes is as follows:

Theorem 5.3.7 (cf [24, Theorem 1.13]) *If $J^+ \in \mathcal{J}(X^+)$ and $J^- \in \mathcal{J}(X^-)$ are generic and (u_+, u_-) is a gluing pair, then we have*

$$\#G(u_+, u_-) = \varepsilon(u_+)\varepsilon(u_-) \prod_{\gamma} c_\gamma(u_+, u_-),$$

where the product is over all simple Reeb orbits whose covers appear in β^+ and β^-, and $c_\gamma(u_+, u_-)$ depends only on γ, the multiplicities of the negative ends of u_+ at covers of γ, and the multiplicities of the positive ends of u_- at covers of γ.

For simplicity, let us now assume that the orbits in β^+ and β^- are all covers of the same simple Reeb orbit γ which is elliptic. Denote the corresponding partitions by $(a_1, \dots, a_k)$ and $(b_1, \dots, b_l)$, where $\sum_{i=1}^k a_i = \sum_{j=1}^l b_j$. Following [24, Section 1], there is a purely combinatorial algorithm for computing $c_\gamma(u_+, u_-)$ in terms of the monodromy angle θ of γ and the partitions $(a_1, \dots, a_k)$ and $(b_1, \dots, b_l)$, but it is rather elaborate to state. For our purposes, it is enough to observe that, by [24, Remark 1.21], $c_\gamma(u_-, u_+)$ is a positive integer provided that there is a branched cover u_0 of the trivial cylinder $\mathbb{R} \times \gamma \subset \mathbb{R} \times Y$ which is connected with genus zero and index zero (this is the analogue of $\kappa_\theta = 1$ in [24]). Namely, this criterion holds exactly if

$$k + l - 2 + \sum_{i=1}^{k} \mathrm{CZ}_\tau(\gamma^{a_i}) - \sum_{j=1}^{l} \mathrm{CZ}(\gamma^{b_j}) = 0. \tag{5-3-1}$$

Here τ is any choice of trivialization along γ, and the left-hand side of (5-3-1) is simply the index of u_0, noting that the first Chern class term vanishes since we are using the same trivialization along γ at the positive and negative ends. Explicitly, if θ denotes the monodromy angle of γ with respect to τ, then we have $\mathrm{CZ}_\tau(\gamma^m) = \lfloor m\theta \rfloor + \lceil m\theta \rceil$, and hence (5-3-1) is equivalent to

$$l - 1 + \sum_{i=1}^{k} \lceil a_j\theta \rceil - \sum_{j=1}^{l} \lceil b_j\theta \rceil = 0. \tag{5-3-2}$$

Note that the left-hand side of (5-3-2) is indeed independent of the choice of trivialization, since θ modulo the integers is independent of τ and by assumption we have $\sum_{i=1}^k a_i = \sum_{j=1}^l b_j$.

We summarize the above discussion as follows:

Theorem 5.3.8 *Suppose that $X^\pm$ are four-dimensional compact symplectic cobordisms with common nondegenerate strict contact boundary $Y := \partial^- X^+ = \partial^+ X^-$. Let $J_Y \in \mathcal{J}(Y)$, $J^+ \in \mathcal{J}_{J_Y}(X^+)$ and $J^- \in \mathcal{J}^{J_Y}(X^-)$ be generic admissible almost complex structures. For $R \geq 0$, let $J_R \in \mathcal{J}(X_R)$ be the concatenated almost complex structure on the symplectic completion of $X_R := X^+ \circledcirc ([-R, R] \times Y) \circledcirc X^-$ which satisfies $J_R|_{X^\pm} = J^\pm|_{X^\pm}$ and $J_R|_{[-R,R]\times Y} = J_Y|_{[-R,R]\times Y}$. Let $u_\pm$ be simple immersed $J^\pm$–holomorphic curves in $\widehat{X}^\pm$ such that each component of u_+ has exactly one negative end and each component of u_- has exactly one positive end. Assume that the negative ends of u_+ are $(\gamma^{a_1}, \dots, \gamma^{a_k})$*

and the positive ends of u_- are $(\gamma^{b_1}, \dots, \gamma^{b_l})$, where γ is a simple elliptic Reeb orbit in Y. Assume further that we have $\sum_{i=1}^k a_i = \sum_{j=1}^l b_j$ and (5-3-1) holds. Then for any R sufficiently large there is a simple immersed regular J_R–holomorphic curve u in $\widehat{X}_R$ with positive asymptotics agreeing with those of u_+, and negative asymptotics agreeing with those of u_-.

Remark 5.3.9 If X is a compact symplectic cobordism and C is an asymptotically cylindrical J–holomorphic curve in X which is simple and has index zero, then C is automatically immersed provided that $J \in \mathcal{J}(X)$ is generic; cf [25, Theorem 4.1].

5.4 Curves with many positive ends via induction

We now seek to apply Theorem 5.3.8 in order to produce genus-zero pseudoholomorphic curves in $X \setminus E_{\mathrm{sk}}$ with one negative end, building on the main construction of [29].

Recall that E_{sk} denotes the ellipsoid $E(\epsilon, \epsilon s)$ with $s > 1$ sufficiently large and $\epsilon > 0$ sufficiently small, and by slight abuse we also use the same notation to denote its image under any symplectic embedding $\iota \colon E_{\mathrm{sk}} \overset{s}{\hookrightarrow} X$. Here the role of ϵ is just to ensure the existence of a symplectic embedding of $E(\epsilon, \epsilon s)$ into X, while s is the "skinniness" factor. More precisely, in the following context of curves in $X \setminus E(\epsilon, \epsilon s)$ with one negative end asymptotic to η_k (the k–fold cover of the short simple Reeb orbit in $\partial E(\epsilon, \epsilon s)$), we will say that $E(\epsilon, \epsilon s)$ is k*–skinny* (or simply *skinny*) if $s > k$. In this case we have $\mathrm{CZ}_{\tau_{\mathrm{ex}}}(\eta_i) = 2i + 1$ for $i = 1, \dots, k$, and hence, at least for the purposes of index computations, we can treat s as being arbitrarily large. On the other hand, note that for $k < s < k+1$, $E(\epsilon, \epsilon s)$ is k–skinny but not $(k+1)$–skinny, a fact which we will exploit in the proof of Lemma 5.4.2 given below.

Before proving the aforementioned lemma, we must deal with the following point. We showed in [30, Proposition 3.1.5] that if X is closed then the number of index-zero curves with fixed top end and a single negative end on E_{sk} is independent of the choice of ι, ϵ and s. However, in our situation with $\partial X \neq \varnothing$ we must be a little careful since in general — for example, if C is not formally perturbation invariant as in Proposition 3.7.1 — there may not be a well-defined count of curves of the given type. Therefore, our arguments only establish that there is a generic J on $X \setminus E_{\mathrm{sk}}$ for which certain curves exist.

Lemma 5.4.1 *Let X be a four-dimensional Liouville domain with nondegenerate contact boundary, and suppose that for some generic $J \in \mathcal{J}(X \setminus E_{\mathrm{sk}})$, there is a simple immersed index-zero J–holomorphic curve C in $X \setminus E_{\mathrm{sk}}$ with negative end asymptotic to η_k. Then, given any $s > k$, we may take $E_{\mathrm{sk}} = \iota(E(\epsilon, \epsilon s))$ for some $\epsilon > 0$ and some $\iota \colon E(\epsilon, \epsilon s) \overset{s}{\hookrightarrow} X$.*

Proof Let $E' = \epsilon' \cdot E(1, s)$, where $\epsilon' > 0$ is so small that we can identify E' with a subset of E_{sk}. Let $J_{X \setminus E'} \in \mathcal{J}(X \setminus E')$ be a generic admissible almost complex structure satisfying $J_{X \setminus E'}|_{X \setminus E} = J$, and put $J_{E \setminus E'} := J_{X \setminus E'}|_{E \setminus E'} \in \mathcal{J}(E \setminus E')$. By, for instance, [18, Theorem 2], there is a regular $J_{E_{\mathrm{sk}} \setminus E'}$–holomorphic cylinder Z in $E_{\mathrm{sk}} \setminus E'$ with positive end on η_1 and negative end on η'_1, and its k–fold cover is regular. We can glue (in the ordinary SFT sense) C to Z along cylindrical ends to produce a simple

$J'_{X\setminus E'}$–holomorphic curve C' in $X \setminus E'$. Here $J'_{X\setminus E'} \in \mathcal{J}(X \setminus E')$ corresponds to the concatenation of $J_{E_{\mathrm{sk}}\setminus E'}$ and $J_{X\setminus E_{\mathrm{sk}}}$ after inserting a sufficiently long neck region in between and reidentifying the resulting compact symplectic cobordism with $X \setminus E'$. Note that we can assume without loss of generality that J' is generic, since the curve C' will persist under small perturbations of J'. □

Lemma 5.4.2 *Let X be a four-dimensional Liouville domain with nondegenerate contact boundary, let $J \in \mathcal{J}(X \setminus E_{\mathrm{sk}})$ be generic, and let C_1 and C_2 be simple immersed index-zero J–holomorphic curves in $X \setminus E_{\mathrm{sk}}$ that have distinct images. For $i = 1, 2$, assume that C_i has positive ends Γ_i and a single negative end η_{k_i}. Then there exists a generic $J' \in \mathcal{J}(X \setminus E_{\mathrm{sk}})$ with $J'|_{\partial X} = J|_{\partial X} \in \mathcal{J}(\partial X)$, and a simple immersed index-zero J'–holomorphic curve C in $X \setminus E_{\mathrm{sk}}$ which has positive ends $\Gamma_1 \cup \Gamma_2$ and a single negative end $\eta_{k_1+k_2+1}$.*

Proof By Lemma 5.4.1, we may suppose that C_1 and C_2 lie in $X \setminus E'$ where $E' = \iota(E(\epsilon, \epsilon s))$ for $s = k_1 + k_2 - 1 + \delta'$, where $0 < \delta' < 1$.

Next put $E'' := \epsilon'' \cdot E(1, k_1 + k_2 + 1 + \delta'')$ for $\epsilon'', \delta'' > 0$ sufficiently small, and choose ι'' and ϵ'' so that we have $E'' \subset E'$. Let $J_{X\setminus E''} \in \mathcal{J}(X \setminus E'')$ be a generic admissible almost complex structure satisfying $J_{X\setminus E''}|_{X\setminus E'} = J'_{X\setminus E'}$, and put $J_{E'\setminus E''} := J_{X\setminus E''}|_{E'\setminus E''} \in \mathcal{J}(E' \setminus E'')$. Again by [18, Theorem 2] there is a (necessarily simple) $J_{E'\setminus E''}$–holomorphic cylinder Z in $E' \setminus E''$ with positive end $\eta_{k_1+k_2}$ in $\partial E'$ and negative end $\eta_{k_1+k_2+1}$ in $\partial E''$. The bottom ellipsoid E'' is skinny, since $k_1 + k_2 + 1 + \delta'' > k_1 + k_2 + 1$. However, the top ellipsoid is not, since $s < k_1 + k_2$. In fact, if we choose the split trivialization τ_{sp} of the contact distribution on $\partial E(1, x)$ as in [30, Section 3.2], then the monodromy angle of the short orbit is $1/x$, which implies that the cylinder Z has Fredholm index

$$2\left(k_1 + k_2 + \left\lfloor \frac{k_1 + k_2}{x} \right\rfloor\right) - 2(k_1 + k_2 + 1) = 0.$$

We now apply Theorem 5.3.8 with $u_+ := C'_1 \cup C'_2$ in $X \setminus E'$ and $u_- := Z$ in $E' \setminus E''$, in other words we glue in the neck $\mathbb{R} \times \partial E'$. Note that (5-3-2) holds since in $\partial E' = \partial(\epsilon \cdot E(1, s'))$, the monodromy angle is $1/s'$, where $k_i < s' < k_1 + k_2$, so that

$$\left\lceil \frac{k_1}{s} \right\rceil + \left\lceil \frac{k_2}{s} \right\rceil = 2 = \left\lceil \frac{k_1 + k_2}{s} \right\rceil.$$

Therefore, there is a curve C as stated. □

Lemma 5.4.2 suggests a natural inductive strategy for constructing curves. Fix a generic $J_{\partial \widetilde{X}_\Omega} \in \mathcal{J}(\partial \widetilde{X}_\Omega)$ as in Lemma 5.1.3. As before, $w_{\min}$ denotes the weakly permissible word with $\widetilde{\mathcal{A}}_\Omega$ minimal subject to $\mathrm{ind}(w_{\min}) = 2k$. We prove the following lemmas in the next subsection.

Lemma 5.4.3 *Let $J \in \mathcal{J}^{J_{\partial \widetilde{X}_\Omega}}(\widetilde{X}_\Omega \setminus E_{\mathrm{sk}})$ be generic. Consider an elliptic orbit $e_{i,j}$ in $\partial \widetilde{X}_\Omega$ such that either $i = 1$ or $j = 1$ (or both). Then there is a J–holomorphic cylinder in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ which is positively asymptotic to $e_{i,j}$ and negatively asymptotic to η_{i+j}.*

Lemma 5.4.4 *Let $J \in \mathcal{J}^{J_{\partial \widetilde{X}_\Omega}}(\widetilde{X}_\Omega \setminus E_{\mathrm{sk}})$ be generic. There is a J–holomorphic pair of pants in $\widetilde{X}_\Omega$ which is positively asymptotic to $e_{1,1} \times e_{1,1}$ and negatively asymptotic to η_5.*

Proposition 5.4.5 *Fix $k \in \mathbb{Z}_{\geq 1}$, and assume that $w_{\min} \neq e_{0,k}$ if $k \geq 2$. Then there exists $J \in \mathcal{J}^{J_{\partial \widetilde{X}_\Omega}}(\widetilde{X}_\Omega)$ for which the moduli space $\mathcal{M}^J_{\widetilde{X}_\Omega}(w_{\min})\langle\!\langle \mathcal{T}^{(m)} p\rangle\!\rangle$ is regular with nonzero signed count.*

Proof Let C be the formal curve in $\widetilde{X}_\Omega$ with positive ends $w_{\min}$ and constraint $\langle\!\langle \mathcal{T}^{(k)} p\rangle\!\rangle$. Recall that by Proposition 5.1.4, C is formally perturbation invariant with respect to $J_{\partial \widetilde{X}_\Omega}$. We explained in the introduction to this section that curves in this moduli space are robust and always count positively. Hence at this point it suffices to find a J–holomorphic curve in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ with positive asymptotics $w_{\min}$ and negative asymptotic η_k for some $J \in \mathcal{J}^{J_{\partial \widetilde{X}_\Omega}}(\widetilde{X}_\Omega)$.

We proceed to construct the desired curve, whose positive asymptotics $w_{\min}$ take one of the forms (1)–(3) in Proposition 4.3.8 or possibly $e_{0,1}$, by iteratively applying Lemma 5.4.2. Firstly, observe that by Lemma 5.4.3, we can construct any cylinder whose positive asymptotic is one of the Reeb orbits $e_{0,1}$, $e_{1,1}$, $e_{1,s}$, $e_{1,0}$ appearing in Proposition 4.3.8. Similarly, by Lemma 5.4.4 we can construct a pair of pants with positive asymptotics $e_{1,1} \times e_{1,1}$. We now iteratively construct curves with two or more positive ends by applying Lemma 5.4.2, with C_1 a previously constructed curve in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ and C_2 a cylinder in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ guaranteed by Lemma 5.4.3. Here we need C_1 and C_2 to have distinct images, and since neither is a multiple cover this is automatic as long as C_1 is not a cylinder with the same positive asymptotic Reeb orbit as C_2. In particular, the curve we seek with positive asymptotics $w_{\min}$ is readily constructed by this iterative construction. □

5.5 Existence of cylinders and pairs of pants

It remains to prove Lemmas 5.4.3 and 5.4.4. For this, we will use various results from the ECH literature, roughly as follows. Firstly, we use the computation of the ECH of $\widetilde{X}_\Omega$ from [22; 6], together with the holomorphic curve axiom for the ECH cobordism map, to establish the existence of a broken current in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ whose positive ends represent the same orbit set as our desired curve. We then argue that this broken pseudoholomorphic current must in fact be a genuine somewhere injective curve C of Fredholm index zero, but possibly of higher genus, whose ends satisfy the ECH partition conditions. Using this, we conclude that in specified situations C must have one negative end, the maximal possible number of positive ends, and genus zero.

Here are the details. Recall that an *orbit set* is a finite set of simple Reeb orbits, along with a choice of positive integer multiplicity for each. In the following we will view a word of Reeb orbits as an orbit set by remembering only the total multiplicity of each underlying simple orbit and forgetting the corresponding partition into iterates. This association from words to orbit sets is evidently not one-to-one, for instance the words η_3, $\eta_2 \times \eta_1$ and $\eta_1 \times \eta_1 \times \eta_1$ of Reeb orbits in ∂E_{sk} all define the same orbit set.

Similarly, a *pseudoholomorphic current* is a finite set of simple pseudoholomorphic curves (each modulo biholomorphic reparametrizations as usual), along with a choice of positive integer multiplicity for each.

We refer the reader to, say, [21, Section 3.4] for the definition of the ECH index $I(C)$. Since the first and second cohomology groups of $\widetilde{X}_\Omega$ and E_{sk} vanish, their ECH chain complexes (over $\mathbb{Z}/2$ for simplicity) have natural $\mathbb{Z}$–gradings, denoted again by I, such that $I(C) = I(\alpha) - I(\beta)$ if C is a holomorphic current which is positively asymptotic to the orbit set α and negatively asymptotic to the orbit set β. Also, the compact symplectic cobordism $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ induces a grading-preserving cobordism map from the ECH of $\partial \widetilde{X}_\Omega$ to that of ∂E_{sk}. If w is an elliptic word,[12] it is shown in [22, Lemma 5.4] that $I(w) = 2(\mathscr{L}(R) - 1) - h(R)$, where

- R denotes the lattice polygon in $\mathbb{R}^2_{\geq 0}$ defined in Section 4.2, and
- $\mathscr{L}(R)$ denotes the number of integer lattice points in the interior and boundary of R.

Lemma 5.5.1 *Let w be a word of elliptic Reeb orbits in $\partial \widetilde{X}_\Omega$, each of which is simple, and let J in $\mathcal{J}^{J_{\partial \widetilde{X}_\Omega}}(\widetilde{X}_\Omega \setminus E_{\mathrm{sk}})$ be generic. Then there is a curve C in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$, possibly of higher genus, with $\operatorname{ind}(C) = I(C) = 0$ and with positive asymptotics w and negative asymptote η_m with $m := \frac{1}{2} I(w)$.*

Proof By [22, Proposition A.4] (which assumes [22, Conjecture A.3], proved in [6]), the word w, when viewed as a generator of the ECH chain complex, represents a nontrivial homology class in the ECH of $\partial \widetilde{X}_\Omega$. Let β denote its image in the ECH of ∂E_{sk} under the ECH cobordism map Φ induced by $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$, and note that β is necessarily nontrivial since Φ is an isomorphism. Recall (see [21, Section 3.7]) that the ECH chain complex of an irrational four-dimensional ellipsoid has trivial differential, and the orbit set with k^{th} largest action has $I = 2k$. Then β is uniquely represented by the orbit set of η_m with $m := I(w)$.

Recall that the ECH cobordism map is defined via the isomorphism with Seiberg–Witten Floer homology, yet it is known to satisfy a "holomorphic curve axiom", which states that a coefficient can only be nonzero if it is represented by an ECH index-zero *broken pseudoholomorphic current*, ie the analogue of a stable pseudoholomorphic building but with each level a pseudoholomorphic current. As a result, we obtain a broken pseudoholomorphic current in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ with positive orbit set w and negative orbit set η_m. By [21, Proposition 3.7], each symplectization level has nonnegative ECH index, with ECH index zero if and only if it is a union of trivial cylinders with multiplicities. By Lemma 5.5.2 below, the main level in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ also has nonnegative ECH index. Using the SFT compactness stability condition (recall Section 2.1.4) and the fact that the total ECH index is zero, we conclude that there is only a single level D, which is a current $(\overline{D}, \kappa)$ in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$, where $\overline{D}$ is simple and $\kappa \in \mathbb{Z}_{\geq 1}$ represents its multiplicity, and we have $I(D) = 0$. By Lemma 5.5.2 again, we also have $I(\overline{D}) = 0$.

By [19, Theorem 4.15], we must have $\operatorname{ind}(\overline{D}) = 0$, and $\overline{D}$ satisfies the positive and negative partition conditions. Since the monodromy angle is positive and very small for each acceptable elliptic orbit, the positive partition conditions stipulate that each positive asymptotic orbit of $\overline{D}$ is simple, ie the positive ends are "as spread out as possible". Meanwhile, the negative partition condition implies that $\overline{D}$ has a single negative end. Finally, the desired curve C is given by taking a κ–fold cover of $\overline{C}$ which is fully ramified at the negative end and unramified at the each of the positive ends. □

[12] There is a more general formula computing I for ECH generators involving hyperbolic orbits, but we will not need this.

Lemma 5.5.2 *If $C = (\overline{C}, \kappa)$ is a J–holomorphic current in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ with $J \in \mathcal{J}(\widetilde{X}_\Omega \setminus E_{\mathrm{sk}})$ generic, we have $I(C) \geq 0$, with equality only if $I(\overline{C}) = 0$.*

Proof As in the proof of [22, Theorem 1.19], we can assume that the cobordism $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ is "L–tame" with L sufficiently large, whence the result follows immediately by [22, Proposition 4.6]. □

Proof of Lemma 5.4.3 By Lemma 5.5.1, there is a J–holomorphic curve C in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$, possibly of higher genus, with $\mathrm{ind}(C) = 0$, with positive asymptotics $e_{i,j}$ and negative asymptotics η_m for $m := \frac{1}{2} I(e_{i,j})$. As explained above, we have $I(e_{i,j}) = 2(\mathscr{L}(R) - 1)$, where R is the lattice triangle with vertices $(0,0)$, $(0,i)$, $(j,0)$, and $\mathscr{L}(R)$ denotes the number of integer lattice points in the interior or boundary of R. By our assumption that $i = 1$ or $j = 1$, we have $\mathscr{L}(R) = i + j + 1$ and hence $m = i + j$. It now follows immediately using the index formula (4-1-3) and $\mathrm{ind}(C) = 0$ that C has genus zero. □

Proof of Lemma 5.4.4 This is similar to the above proof. In this case Lemma 5.5.1 produces a J–holomorphic curve C in $\widetilde{X}_\Omega \setminus E_{\mathrm{sk}}$ with $\mathrm{ind}(C) = 0$ and with positive ends $e_{1,1} \times e_{1,1}$ and negative end η_m for $m := \frac{1}{2} I(e_{1,1} \times e_{1,1}) = 5$. The condition $\mathrm{ind}(C) = 0$ then forces the genus to be zero. □

Remark 5.5.3 For $e_{i,j}$ with $i, j \geq 2$, or $e_{1,1}^{\times k}$ with $k \geq 3$, the curve C coming from Lemma 5.5.1 will typically be forced to have higher genus.

5.6 Comparison with Gutt–Hutchings capacities

The following result likely holds in any dimension; for concreteness we give the proof in dimension four.

Proposition 5.6.1 *For X_Ω any four-dimensional convex toric domain, we have*

$$\widetilde{\mathfrak{g}}_k^{\leq 1}(X_\Omega) = c_k^{\mathrm{GH}}(X_\Omega) = \min_{\substack{(i,j) \in \mathbb{Z}^2_{\geq 0} \\ i+j=k}} \|(i,j)\|^*_\Omega. \tag{5-6-1}$$

Proof The second equality is [16, Theorem 1.6]. In order to compute $\widetilde{\mathfrak{g}}_k^{\leq 1}(X_\Omega)$, we can replace X_Ω with its full rounding $\widetilde{X}_\Omega$ as in Section 4.1. As shorthand put $X := X_\Omega$ and $\widetilde{X} := \widetilde{X}_\Omega$. Fix a generic almost complex structure $J_{\partial \widetilde{X}} \in \mathcal{J}(\partial \widetilde{X})$ as in Lemma 5.1.3, and a generic extension $J_{\widetilde{X}} \in \mathcal{J}^{J_{\partial \widetilde{X}}}(\widetilde{X}; D)$.

To prove that $\widetilde{\mathfrak{g}}_k^{\leq 1}(X) \geq c_k^{\mathrm{GH}}(X_\Omega)$, observe that by definition we can find a $J_{\widetilde{X}}$–holomorphic plane C in $\widetilde{X}$ satisfying the local tangency constraint $\langle\!\langle \mathcal{T}^{(k)} p \rangle\!\rangle$ and having $E(C) \leq \widetilde{\mathfrak{g}}_k^{\leq 1}(\widetilde{X})$. Let γ denote the asymptotic Reeb orbit of C, which we can take to be $e_{i,j}$ or $h_{i,j}$ for some i, j. If C is simple then by genericity it must be regular and hence satisfy $\mathrm{ind}(C) \geq 0$, and inspection of the index formula shows that this is also true if C is a multiple cover. In particular, we must have $i + j \geq k$, from which it follows that $\mathcal{A}(\gamma) = \|(i,j)\|^*_\Omega$ is greater than or equal to the right-hand side of (5-6-1). Since $E(C) = \widetilde{\mathcal{A}}(\gamma)$ is arbitrarily close to $\mathcal{A}(\gamma)$, this gives the desired lower bound.

To establish the upper bound for $\widetilde{\mathfrak{g}}_k^{\leq 1}(X)$, let (i,j) be a minimizer for the right-hand side of (5-6-1). We can assume that there are no common divisors of i, j, k, and we will then show that $\widetilde{\mathfrak{g}}_k^{\leq 1}(X) \leq \|(i,j)\|_\Omega^*$. Indeed, if there is a greatest common divisor $q \geq 2$ of i, j, k, then after putting $i' := i/q$, $j' := j/q$ and $k' := k/q$, it will follow that we have $\widetilde{\mathfrak{g}}_{k'}^{\leq 1}(X) \leq \|(i',j')\|_\Omega^*$, whence we have

$$\widetilde{\mathfrak{g}}_k^{\leq 1}(X) \leq q\widetilde{\mathfrak{g}}_{k'}^{\leq 1}(X) \leq q\|(i',j')\|_\Omega^* = \|(i,j)\|_\Omega^*.$$

Now let C be the (necessarily simple by the above) formal plane in $\widetilde{X}$ with positive end $e_{i,j}$ and carrying the constraint $\lessdot\mathcal{T}^{(k)}p\gtrdot$. By an argument paralleling the proof of Proposition 5.1.4, we find that C is formally perturbation invariant with respect to $J_{\partial\widetilde{X}}$. In particular, C cannot be represented by any nontrivial stable $J_{\widetilde{X}}$–holomorphic building. We claim that the signed count $\#\mathcal{M}_{\widetilde{X}}^{J_{\widetilde{X}}}(C)$ is nonzero, from which it follows that we have

$$\widetilde{\mathfrak{g}}_k^{\leq 1}(\widetilde{X}) \leq E(C) = \widetilde{\mathcal{A}}(e_{i,j}) \approx \|(i,j)\|_\Omega^*.$$

To justify the claim, note that we can use Proposition 2.2.3 to trade the local tangency constraint $\lessdot\mathcal{T}^{(k)}p\gtrdot$ for a skinny ellipsoidal constraint $\lessdot(k)\gtrdot_E$. Namely, letting $E_{\mathrm{sk}} = E(\epsilon, \epsilon x) \subset \widetilde{X}$ denote an ellipsoid with $x > k$ and $\epsilon > 0$ sufficiently small, it suffices to show that the moduli space of pseudoholomorphic cylinders in $\widetilde{X} \setminus E_{\mathrm{sk}}$ with positive end $e_{i,j}$ and negative end η_k has nonzero signed count. By slight abuse of notation we will denote the corresponding formal cylinder again by C. Recall that by Proposition 5.2.2 it suffices to show that this moduli space is nonempty. For this we invoke linearized contact homology as in [32], similar to the proof of [18, Theorem 2]. Indeed, observe that $e_{i,j}$ is necessarily a cycle with respect to the linearized contact homology differential thanks to Lemma 5.1.3 and the fact that any orbit $h_{i',j'}$ with $(i',j') \neq (i,j)$ necessarily has greater action by Lemma 4.1.1. Since the cobordism map on linearized contact homology induced by $\widetilde{X} \setminus E_{\mathrm{sk}}$ is an isomorphism, it follows that there is a stable pseudoholomorphic cylindrical building representing C, and by formal perturbation invariance this must be an honest pseudoholomorphic cylinder in $\widetilde{X} \setminus E_{\mathrm{sk}}$. □

Remark 5.6.2 As mentioned earlier, there is a natural higher-dimensional analogue of the fully rounding procedure, but for concreteness we have kept our discussion in Section 4.1 to dimension four and hence restrict Proposition 5.6.1 to dimension four. In order to extend the above argument to higher dimensions, one first ought to show that the higher-dimensional the analogue of C is formally perturbation invariant. Since the results in [32] hold in arbitrary dimension, one can then still invoke the cobordism map on linearized contact homology in higher dimensions in order to produce cylindrical buildings.

We also refer the reader to [33, Theorem 7.6.4] for the analogous statement $\mathfrak{g}_k^{\leq 1}(X) = c_k^{\mathrm{GH}}(X)$ for any Liouville domain X satisfying $\pi(X) = 2c_1(TX) = 0$.

Remark 5.6.3 We expect that the methods in this paper could be extended to compute $\widetilde{\mathfrak{g}}_k^{\leq l}(X_\Omega)$ for all $k, l \in \mathbb{Z}_{\geq 1}$, and it is an interesting question whether the entire family $\{\widetilde{\mathfrak{g}}_k^{\leq l}\}$ sometimes give stronger embedding obstructions than the sequence $\widetilde{\mathfrak{g}}_1, \widetilde{\mathfrak{g}}_2, \widetilde{\mathfrak{g}}_3, \ldots$ alone. A natural guess is that Theorem 1.2.8 generalizes to a formula for $\widetilde{\mathfrak{g}}_k^{\leq l}(X_\Omega)$ by requiring $q \leq l$ in the minimization.

6 Ellipsoids, polydisks, and more

In this section we apply our formalism to several examples, proving the remaining three theorems from the introduction. In each case, using Theorem 1.2.8 and the specific form of $\|-\|^*_\Omega$, it reduces to a purely combinatorial optimization problem. The latter is tractable thanks to Corollary 1.3.1, which implies that we can look for a minimizer taking one of the following forms:

(1) $(0,1)^{\times i} \times (1,1)^{\times j}$ for $i \geq 0,\ j \geq 1$.

(2) $(0,1)^{\times i} \times (1,s)$ for $i \geq 0$ and $s \geq 2$.

(3) $(0,1)^{\times i} \times (1,0)$ for $i \geq 1$.

(4) $(0,s)$ for $s \geq 1$.

Proof of Theorem 1.3.2 We consider $E(a,1)$, and by continuity we can assume $a > 1$ is irrational. Let Ω be the triangle with vertices $(0,0), (a,0), (0,1)$. Observe that for $\vec{v} = (v_x, v_y) \in \mathbb{R}^2_{\geq 0}$ we have

$$\|\vec{v}\|^*_\Omega = \max_{\vec{w} \in \Omega} \langle \vec{v}, \vec{w} \rangle = \max(v_x a, v_y).$$

We can ignore case (3), since we have

$$\|(1,2)\|^*_\Omega = \max(a,2) < 1 + a = \|(0,1)\|^*_\Omega + \|(1,0)\|^*_\Omega,$$

and hence $(0,1)^{\times i} \times (1,0)$ with $i \geq 1$ cannot be a minimizer.

Suppose first that $a > \frac{3}{2}$. Then we have

$$\|(1,2)\|^*_\Omega + \|(0,1)\|^*_\Omega = \max(a,2) + 1 < 2a = 2\|(1,1)\|^*_\Omega,$$

and hence $(0,1)^{\times i} \times (1,1)^{\times j}$ with $i \geq 0,\ j \geq 1$ can only be a minimizer if $j = 1$. If $s > a+1$, then

$$\|(0,1)\|^*_\Omega + \|(1,s-2)\|^*_\Omega = 1 + \max(a, s-2) < \max(a,s) < \|(1,s)\|^*_\Omega,$$

and therefore $(0,1)^{\times i} \times (1,s)$ with $i \geq 0$, $s \geq 2$ can only be a minimizer if $s \leq a+1$. Similarly, if $s < a-1$ then we have

$$\|(0,s+1)\|^*_\Omega = s + 1 < \max(a,s) = \|(1,s)\|^*_\Omega,$$
$$\|(1,s+2)\|^*_\Omega = \max(a,s+2) < 1 + \max(a,s) = \|(0,1)\|^*_\Omega + \|(1,s)\|^*_\Omega,$$

and therefore $(0,1)^{\times i} \times (1,s)$ with $i \geq 0$, $s \geq 1$ can only be a minimizer if $s \geq a-1$.

Since a is irrational, we have $[a-1, a+1] \cap \mathbb{Z} = \{\lfloor a \rfloor, \lfloor a \rfloor + 1\}$. Therefore, there must be a minimizer taking one of the forms

- $(0,1)^{\times i} \times (1, \lfloor a \rfloor)$ for $i \geq 0$,
- $(0,1)^{\times i} \times (1, \lfloor a \rfloor + 1)$ for $i \geq 0$,
- $(0,s)$ for $s \geq 1$,

from which (1-3-2) readily follows.

Now suppose that we have $a < \frac{3}{2}$. For $s \geq 3$ we have

$$\|(1,1)\|^*_\Omega + \|(0,s-2)\|^*_\Omega = a + s - 2 < s = \|(1,s)\|^*_\Omega,$$

and hence $(0,1)^{\times i} \times (1,s)$ with $i \geq 0$, $s \geq 3$ cannot be a minimizer. We have also

$$2\|(1,1)\|^*_\Omega = 2a < 3 = 3\|(0,1)\|^*_\Omega,$$

and hence $(0,1)^{\times i} \times (1,1)^{\times j}$ for $i \geq 0$, $j \geq 1$ can only be a minimizer if $i \in \{0,1,2\}$.

Therefore, there must be a minimizer taking one of the following forms:

- $(0,1)^{\times i} \times (1,1)^{\times j}$ for $i \in \{0,1,2\}$ and $j \geq 1$.
- $(0,1)^{\times i} \times (1,2)$ for $i \geq 0$.
- $(0,s)$ for $s \geq 1$.

Since $2\|(0,1)\|^*_\Omega = 2 = \|(1,2)\|^*_\Omega$, we can effectively ignore the second bullet by artificially allowing $j = 0$ in the first bullet. For $s \geq 2$ we have

$$\|(1,s-1)\|^*_\Omega = \max(a, s-1) < s = \|(0,s)\|^*_\Omega,$$

and hence $(0,s)$ can only be minimal if $s = 1$, so we can also effectively ignore the third bullet. Therefore we have

$$\tilde{\mathfrak{g}}_k(E(a,1)) = i\|(0,1)\|^*_\Omega + j\|(1,1)\|^*_\Omega = i + ja$$

for $i \in \{0,1,2\}$ and $j \geq 0$ satisfying $2i + 3j - 1 = k$. Note that i and j are uniquely determined via $i \equiv -k-1 \pmod 3$ and $j = \frac{1}{3}(j+1-2i)$, and (1-3-1) follows. □

Proof of Theorem 1.3.4 This is similar to the previous proof. We consider $P(a,1)$ with $a > 1$ irrational, and we take Ω to be the rectangle with vertices $(0,0)$, $(a,0)$, $(0,1)$, $(a,1)$. For $\vec{v} = (v_x, v_x) \in \mathbb{R}^2_{\geq 0}$ we then have

$$\|\vec{v}\|^*_\Omega = \langle \vec{v}, (a,1)\rangle = av_x + v_y.$$

For $s \geq 2$ we have

$$\|(1,0)\|^*_\Omega + \|(0,s-1)\|^*_\Omega = a + s - 1 < a + s = \|(1,s)\|^*_\Omega,$$

and hence case (2) in Corollary 1.3.1 cannot occur as a minimizer. For $i \geq 0$, $j \geq 1$ we have

$$2\|(0,1)\|^*_\Omega + \|(1,0)\|^*_\Omega = 2 + a < 2 + 2a = 2\|(1,1)\|^*_\Omega,$$

so case (1) can only occur if $j = 1$. Therefore, there must be a minimizer from the list

- $(0,1)^{\times i} \times (1,1)$ for $i \geq 0$,
- $(0,1)^{\times i} \times (1,0)$ for $i \geq 1$,
- $(0,s)$ for $s \geq 1$,

from which (1-3-3) follows. □

Proof of Theorem 1.3.7 The polygon $\Omega := Q(a,b,c) \subset \mathbb{R}^2_{\geq 0}$ has vertices $(0,0)$, $(c,0)$, (a,b), $(0,1)$. For $\vec{v} = (v_x, v_y) \in \mathbb{R}^2_{\geq 0}$, we have

$$\|\vec{v}\|^*_\Omega = \max_{\vec{w} \in \Omega} \langle \vec{v}, \vec{w} \rangle = \max(cv_x, av_x + bv_y, v_y).$$

Recall that by assumption we have $c \geq 1$, $a \leq c$, $b \leq 1$, $a + bc \geq c$, and $M := \max(a+b, c) \leq 2$.

For $j \geq 1$, we have

$$\|(1,j)\|^*_\Omega = \max(c, a + jb, j),$$

and, in particular,

$$\|(1,1)\|^*_\Omega = \max(c, a+b, 1) = \max(c, a+b) = M.$$

By the above, we have

$$\tilde{\mathfrak{g}}_2(X) = \min(\|(1,1)\|^*_\Omega, \|(0,2)\|^*_\Omega) = \min(M, 2) = M.$$

Next, because $c < 2$, we have

$$\|(0,3)\|^*_\Omega > 3 > \|(1,0)\|^*_\Omega + \|(0,1)\|^*_\Omega = 1 + c,$$

so that

$$\begin{aligned}\tilde{\mathfrak{g}}_3(X) = \min\bigl(\|(1,2)\|^*_\Omega, \|(1,0)\|^*_\Omega + \|(0,1)\|^*_\Omega\bigr) &= \min(\max(2, a+2b, c), 1+c) \\ &= \min(\max(2, a+2b), 1+c).\end{aligned}$$

Note that $\tilde{\mathfrak{g}}_3(X) = 2$ if $a + 2b < 2$ and otherwise $= \min(a+2b, 1+c) < 3$. In particular, if $a + 2b > 2$ the minimum could be represented by either orbit set.

We next claim that

$$\|(1,j)\|^*_\Omega > \|(0,1)\|^*_\Omega + \|(1,j-2)\|^*_\Omega \quad \text{for } j \geq 3.$$

If $b > \frac{1}{2}$, we must check that

$$\max(j, a + jb) > 1 + \max(j-2, a + (j-2)b) = \max(j-1, a + jb - 2b + 1),$$

which holds because $2b > 1$. If $b < \frac{1}{2}$ and $j \geq 3$ then $a + jb < 2 + \frac{1}{2}(j-1) \leq j$ for $j \geq 3$, so that

$$\|(1,j)\|^*_\Omega = j > \|(1,j-2)\|^*_\Omega + \|(0,1)\|^*_\Omega \quad \text{for } j \geq 3.$$

Thus in all cases, $(1,j)$ with $j \geq 3$ does not occur in a minimal orbit set. Further $(0,k)$ with $k \geq 2$ is never minimal since it can be replaced by $(1,1) \cup (0,1)^{\times k/2}$ for even k or $(1,0) \cup (0,1)^{\times (k-1)/2}$ for odd k.

Therefore, taking into account the discussion of $\tilde{\mathfrak{g}}_3$, we find that minimizers must take one of the forms

- $(0,1)^{\times i} \times (1,1)^{\times j}$, where $j = 0$ only if $i = 1$;
- $(0,1)^{\times i} \times (1,2)$ or $(0,1)^{\times i} \times (1,0)$ (but not both).

In particular,

$$\begin{aligned}\tilde{\mathfrak{g}}_4(X) &= \|(0,1)\|^*_\Omega + \|(1,1)\|^*_\Omega = 1 + M < 3, \\ \tilde{\mathfrak{g}}_6(X) &= \|(0,1)^{\times 2}\|^*_\Omega + \|(1,1)\|^*_\Omega = 2 + M < 4.\end{aligned}$$

On the other hand, $\widetilde{\mathfrak{g}}_5(X)$ might be represented by $(0,1)\times(1,2)$, $(0,1)^{\times 2}\times(1,0)$ or $(1,1)\times(1,1)$ and so is given by

$$\widetilde{\mathfrak{g}}_5(X)=\min(\max(3,1+a+2b,c),2+c,2M).$$

If $M<\frac{3}{2}$, then because the first two terms above are ≥ 3, we find that $\widetilde{\mathfrak{g}}_5(X)=2M$. However, if $\frac{3}{2}<M<2$ then any of these three terms might be minimal.

For $k>6$ it is again useful to consider the cases $M<\frac{3}{2}$ and $M>\frac{3}{2}$ separately. In the former case, it is more efficient to increase the index by adding copies of $(1,1)$ so that minimal orbit sets always have $i\leq 2$. In particular, orbit sets of the form $(0,1)^{\times i}\times(1,2)$ or $(0,1)^{\times i}\times(1,0)$ are not minimal when $i>2$, and so can only affect the capacities $\widetilde{\mathfrak{g}}_k$ for $k\leq 7$. Moreover when $M<\frac{3}{2}$,

$$\|(0,1)^{\times 2}\times(1,2)\|^*_\Omega=2+\max(2,a+2b,c)>4>1+2M=\|(0,1)\times(1,1)^{\times 2}\|^*_\Omega.$$

Therefore the capacities for $k\geq 6$ are given by the orbit sets

$$(0,1)^{\times 2}\times(1,1)^{\times j},\quad (0,1)\times(1,1)^{\times j+1},\quad (1,1)^{\times j+2},\quad \text{where } j\geq 1,\ M<\tfrac{3}{2}.$$

The claims in (i) follow readily.

If $M>\frac{3}{2}$, minimal orbit sets always have $j\leq 2$ since it is more efficient to use $(0,1)^{\times 3}\times(1,1)^{\times j-2}$ instead of $(1,1)^{\times j}$. Which of $(0,1)^{\times i}\times(1,2)$ or $(0,1)^{\times i}\times(1,0)$ is more efficient is determined by the value of $\widetilde{\mathfrak{g}}_3$, while the value of $\widetilde{\mathfrak{g}}_5$ determines whether it is in fact best to use $(1,1)^{\times 2}$ when representing elements of odd index ≥ 5. Thus the odd capacities for $k\geq 5$ are determined by $\widetilde{\mathfrak{g}}_5$, while the even capacities are more straightforward since they are always calculated by orbit sets of the form $(0,1)^{\times i}\times(1,1)$. This proves (ii). □

Remark 6.0.1 When $2\leq n<c<n+1$ one can check that $\widetilde{\mathfrak{g}}_k=k$ for $k\leq n$, represented by the orbit $e_{0,k}$. In this case, the $\widetilde{\mathfrak{g}}_k$ again limit on a period two cycle. However, the precise values in this cycle depend on b. To see this, note for example that if $n=2l$ is even, then

$$\widetilde{\mathfrak{g}}_{k+1}=\min_{i\leq l}\mathcal{A}(e^i_{0,1}\cup e_{1,2(l-i)})=\min_{i\leq l}\bigl(i+\max(2(l-i),\ a+2(l-i)b,c)\bigr),$$

and which orbit set gives the minimax depends on whether $b>\frac{1}{2}$ or $b<\frac{1}{2}$. For example, if we assume that $a<c$ are both very close to n then the minimax is determined by the minimum value of $i+a+2(l-i)b=a+2lb+i(1-2b)$. Thus if $b<\frac{1}{2}$ one should take $i=l$, while if $b<\frac{1}{2}$ one should take $i=0$.

Proof of Theorem 1.3.8 For $\Omega:=\Omega_p$, recall that we have $\mathcal{A}_\Omega(e_{i,j})=\|(i,j)\|^*_{\Omega_p}=\|(i,j)\|_q$. We can ignore case (4) in Corollary 4.3.9 for $s\geq 2$, since for $s=2$ we have

$$\|(1,1)\|_q=2^{1/q}\leq 2=\|(0,2)\|_q$$

and for $s\geq 3$ we have

$$\|(1,0)\|_q+\|(0,s-2)\|_q=s-1\leq s=\|(0,s)\|_q.$$

Similarly, we can ignore case (2), since we have

$$\|(1,0)\|_q + \|(0,s-1)\|_q = s < \|(1,s)\|_q.$$

Noting that $\|(0,1)\|_q = \|(1,0)\|_q$, we can also effectively ignore case (3) by relaxing the condition $j \geq 1$ in case (1). In other words, we have that $\widetilde{\mathfrak{g}}_k(X_{\Omega_p})$ is the minimal quantity of the form

$$i\,\|(0,1)\|_q + j\,\|(1,1)\|_q = i + j2^{1/q},$$

subject to $2i + 3j - 1 = k$ for $i, j \in \mathbb{Z}_{\geq 0}$.

We have $2\|(1,1)\|_q \leq 3\|(0,1)\|_q$ if and only if $2^{1/q} \leq \frac{3}{2}$, ie if and only if $q \geq \ln(2)/\ln\left(\frac{3}{2}\right)$, or equivalently $p \leq \ln(2)/\ln\left(\frac{4}{3}\right)$. In this case we can assume $i \in \{0,1,2\}$, and the value of i is then determined by looking at the equation $2i + 3j - 1 = k$ modulo 3, from which (1-3-6) immediately follows. Similarly, in the case $p > \ln(2)/\ln\left(\frac{4}{3}\right)$ we can assume $j \in \{0,1\}$, and the value of j is then determined by looking at the equation $2i + 3j - 1 = k$ modulo 2, which immediately gives (1-3-7). □

Appendix Regularity after stabilization

In this appendix we give a self-contained proof that regularity persists after dimensional stabilization. We also refer the reader to [33, Section 7.4] for a related approach.

Let X be a Liouville domain, and let $W := X \mathbin{\widehat{\times}} B^2(c)$ be a smoothing of $X \times B^2(c)$ for some $c > 0$, as in Lemma 3.6.2. Let D be a local symplectic divisor in X near a point $p \in X$, and let $\widetilde{D} = D \times B^2(\epsilon)$ for $\epsilon > 0$ small be a corresponding local symplectic divisor in W near $\widetilde{p} := (p, p_0)$ for $p_0 := 0 \in B^2(c)$. Let J be an admissible almost complex structure on $\widehat{X}$ which is integrable near p and preserves D, and let $\widetilde{J}$ be an admissible almost complex structure on $\widehat{W}$ which is integrable near $\widetilde{p}$, preserves $\widetilde{D}$, and restricts to J along $\widetilde{X} \times \{0\} \approx \widetilde{X}$ (so that, in particular, $\widehat{X} \times \{0\}$ is $\widetilde{J}$–holomorphic).

Our main goal is to prove:

Proposition A.1 *Let u be an asymptotically cylindrical J–holomorphic punctured sphere in $\widehat{X}$ satisfying the constraint $\lll \mathcal{T}_D^{(m)} p \ggg$ for some $m \in \mathbb{Z}_{\geq 1}$, and such that each asymptotic Reeb orbit is nondegenerate with normal Conley–Zehnder index one. Assume that u is regular and has index zero* (*taking into account the constraint $\lll \mathcal{T}_D^{(m)} p \ggg$*). *Let $\widetilde{u}$ denote the curve in $\widehat{W}$ given by the composition of u with the inclusion $\widehat{X} \subset \widehat{W}$. Then $\widetilde{u}$ is also regular* (*taking into account the constraint $\lll \mathcal{T}_{\widetilde{D}}^{(m)} p \ggg$*).

Note that in formulating the index and regularity of u and $\widetilde{u}$ we are as usual also allowing for arbitrary variations of the conformal structure of the domain. Recall that the normal Conley–Zehnder index is defined for a Reeb orbit in ∂X by taking into account the Reeb flow in the direction normal to $\widehat{X} \times \{0\}$ in $\widehat{W}$, and we are implicitly using trivializations coming from the natural trivialization of the normal bundle of $\widehat{X} \subset \widehat{W}$ as in Section 3.6.

Let $\Sigma = S^2 \setminus \{z_1, \dots, z_l\}$ denote the domain of u, where $z_1, \dots, z_l$ are the punctures, and let $z_0 \in \Sigma$ denote the marked point which realizes the local tangency constraint.

Regularity of u is equivalent to surjectivity of the linearized Cauchy–Riemann operator

$$D\overline{\partial}_J(u,j)\colon T_u\mathcal{B}\oplus T_j\mathcal{T}\to\mathcal{E}_{(u,j)},$$

where:

- $T_u\mathcal{B}=\mathcal{W}^{k,p,\delta}_{\ll\mathcal{T}_D^{(m)}p\gg}(u^*T\widehat{X})\oplus V$.

- $\mathcal{W}^{k,p,\delta}(u^*T\widehat{X})$ denotes the Banach space of sections ξ of $u^*T\widehat{X}$ of weighted Sobolev class $\mathcal{W}^{k,p,\delta}$ (cf [40, Section 7.2]), where we assume $k\ge m$ and $(k-m)p>2$ (so that ξ is C^m), and

$$\mathcal{W}^{k,p,\delta}_{\ll\mathcal{T}_D^{(m)}p\gg}(u^*T\widehat{X})\subset\mathcal{W}^{k,p,\delta}(u^*T\widehat{X})$$

denotes the subspace consisting of sections whose m–jet at z_0 lies in D as in [7, Section 6]. (In particular, $\mathcal{W}^{k,p,\delta}_{\ll p\gg}(u^*T\widehat{X})$ is the subspace such that ξ vanishes at z_0.)

- $V\subset\mathcal{W}^{k,p}_{\mathrm{loc}}(u^*T\widehat{X})$ is a $2l$–dimensional subspace as in [39, Section 3.1], consisting of smooth sections which are supported near the punctures and asymptotic to constant (in suitable trivializations) linear combinations of vector fields tangent to the trivial cylinders over the asymptotic Reeb orbits of u — this is needed to address the possibility of rotating and translating the asymptotic ends of u, as these deformations do not exponentially decay along the cylindrical ends.

- $\mathcal{T}\subset\mathcal{J}(\Sigma)$ is a Teichmüller slice through j as in [39, Section 3.1], which is in particular a smooth manifold containing j and having (in the stable case) dimension $2(l+1)-6$. By $T_j\mathcal{T}\subset\Gamma(\overline{\mathrm{End}}_{\mathbb{C}}(T\Sigma))$ we denote its tangent space at j.

- The space

$$\mathcal{E}_{(u,j)}=\mathcal{W}^{k-1,p,\delta}_{\ll\mathcal{T}_D^{(m-1)}p\gg}(\overline{\mathrm{Hom}}_{\mathbb{C}}(T\Sigma,u^*T\widehat{W}))$$

consists of bundle homomorphisms from $T\Sigma$ to $u^*T\widehat{W}$ over Σ which are (j,J)–antilinear and whose $(m-1)$–jet at z_0 lies in D.

Moreover, after choosing any symmetric connection ∇ on $T\widehat{X}$, for $\xi\in T_u\mathcal{B}$ and $y\in T_j\mathcal{T}$, the linearized Cauchy–Riemann operator $D\overline{\partial}_J(u,j)$ takes the explicit form

$$D\overline{\partial}_J(u,j)(\xi,y)=D_u\xi+G_uy,$$

where

- $D_u\colon T_u\mathcal{B}\to\mathcal{E}_{(u,j)}$ is given by

$$D_u\xi=\nabla\xi+J\circ(\nabla\xi)\circ j+\nabla_\xi J\circ du\circ j,$$

- $G_u\colon T_j\mathcal{T}\to\mathcal{E}_{(u,j)}$ is given by

$$G_uy=J\circ du\circ y.$$

Similarly, regularity of $\widetilde{u}$ is equivalent to surjectivity of the operator

$$D\overline{\partial}_J(\widetilde{u}, j)\colon T_{\widetilde{u}}\widetilde{\mathcal{B}} \oplus T_j\mathcal{T} \to \mathcal{E}_{(\widetilde{u},j)},$$

where

- $T_{\widetilde{u}}\widetilde{\mathcal{B}} = \mathcal{W}^{k,p,\delta}_{\lessdot\mathcal{T}^{(m)}_{\widetilde{D}}\widetilde{p}\gtrdot}(\widetilde{u}^*T\widehat{W}) \oplus V$,
- $\mathcal{E}_{(\widetilde{u},j)} = \mathcal{W}^{k-1,p,\delta}_{\lessdot\mathcal{T}^{(m-1)}_{\widetilde{D}}p\gtrdot}(\overline{\mathrm{Hom}}_{\mathbb{C}}(T\Sigma, \widetilde{u}^*T\widehat{W}))$,

and for $\xi \in T_{\widetilde{u}}\widetilde{\mathcal{B}}$ and $y \in T_j\mathcal{T}$ we have

$$D\overline{\partial}_J(\widetilde{u}, j)(\xi, y) = D_{\widetilde{u}}\xi + G_{\widetilde{u}}y,$$

where

- $D_{\widetilde{u}}\colon T_{\widetilde{u}}\widetilde{\mathcal{B}} \to \mathcal{E}_{(\widetilde{u},j)}$ is given by
$$D_{\widetilde{u}}\xi = \widetilde{\nabla}\xi + \widetilde{J}\circ(\widetilde{\nabla}\xi)\circ j + \widetilde{\nabla}_\xi\widetilde{J}\circ d\widetilde{u}\circ j,$$
- $G_{\widetilde{u}}\colon T_j\mathcal{T} \to \mathcal{E}_{(\widetilde{u},j)}$ is given by
$$G_{\widetilde{u}}y = \widetilde{J}\circ d\widetilde{u}\circ y,$$

where $\widetilde{\nabla}$ is any symmetric connection on $T\widehat{W}$.

Note that the embedding $W \hookrightarrow X \times B^2(c)$ naturally extends to a diffeomorphism $\widehat{W} \cong \widehat{X} \times \widehat{B^2(c)}$, and we get a corresponding splitting of the tangent bundle of $\widehat{W}$:

$$T\widehat{W} \cong T^{\mathrm{ver}}\widehat{W} \oplus T^{\mathrm{hor}}\widehat{W}.$$

Under the identification $T^{\mathrm{ver}}\widehat{W}|_{\widehat{X}\times\{0\}} \approx T\widehat{X}$, this induces natural splittings

$$T_{\widetilde{u}}\widetilde{\mathcal{B}} \oplus T_j\mathcal{T} \cong \underbrace{(\mathcal{W}^{k,p,\delta}_{\lessdot\mathcal{T}^{(m)}_D p\gtrdot}(u^*T\widehat{X}) \oplus V \oplus T_j\mathcal{T})}_{A_1} \oplus \underbrace{(\mathcal{W}^{k,p,\delta}_{\lessdot p_0\gtrdot}(\widetilde{u}^*T^{\mathrm{hor}}\widehat{W}))}_{A_2},$$

$$\mathcal{E}_{(\widetilde{u},j)} \cong \underbrace{\big(\mathcal{W}^{k-1,p,\delta}_{\lessdot\mathcal{T}^{(m-1)}_D p\gtrdot}(\overline{\mathrm{Hom}}_{\mathbb{C}}(T\Sigma, u^*T\widehat{X}))\big)}_{B_1} \oplus \underbrace{\big(\mathcal{W}^{k-1,p,\delta}(\overline{\mathrm{Hom}}_{\mathbb{C}}(T\Sigma, \widetilde{u}^*T^{\mathrm{hor}}\widehat{W}))\big)}_{B_2}.$$

From now on, we assume that the connection $\widetilde{\nabla}$ preserves this splitting and restricts to ∇ under the identification $T\widehat{X}$. The above splitting induces a block matrix decomposition

$$D\overline{\partial}_J(\widetilde{u}, j) = \begin{pmatrix} M_{1,1} = D\overline{\partial}_J(u, j) & M_{1,2} \\ M_{2,1} & M_{2,2} \end{pmatrix}. \tag{A-0-1}$$

Lemma A.2 *We have $M_{2,1} = 0$.*

Proof We need to show that the image of $D\overline{\partial}_J(\widetilde{u}, j)|_{A_1}$ lies in B_1. Note that for $y \in T_j\mathcal{T}$ we have

$$\widetilde{J}\circ d\widetilde{u}\circ y \in \Gamma(\overline{\mathrm{Hom}}_{\mathbb{C}}(T\Sigma, \widetilde{u}^*T\widehat{X})), \tag{A-0-2}$$

since $\widetilde{J}$ preserves $T^{\mathrm{ver}}\widehat{W}|_{\widehat{X}}$, and hence $G_{\widetilde{u}}y \in B_1$. It therefore suffices to show that for any ξ^{ver} in $\Gamma(\widetilde{u}^*T^{\mathrm{ver}}\widehat{W})$, the image $D_{\widetilde{u}}\xi^{\mathrm{ver}}$ lands in $\Gamma(\overline{\mathrm{Hom}}_{\mathbb{C}}(T\Sigma, \widetilde{u}^*T^{\mathrm{ver}}\widehat{W}))$. For $v \in \Gamma(T\Sigma)$, we have

$$(D_{\widetilde{u}}\xi^{\mathrm{ver}})(v) = \widetilde{\nabla}_v\xi^{\mathrm{ver}} + \widetilde{J}\widetilde{\nabla}_{jv}\xi^{\mathrm{ver}} + (\widetilde{\nabla}_{\xi^{\mathrm{ver}}}\widetilde{J})(q)$$

for $q := (d\widetilde{u})(jv) \in \Gamma(u^*T\widehat{X})$. Since $\widetilde{\nabla}$ and $\widetilde{J}$ respect the splitting $T\widehat{W}|_{\widehat{X}} = T^{\mathrm{ver}}\widehat{W}|_{\widehat{X}} \oplus T^{\mathrm{hor}}\widehat{W}|_{\widehat{X}}$, we have

$$\widetilde{\nabla}_v\xi^{\mathrm{ver}} + \widetilde{J}\widetilde{\nabla}_{jv}\xi^{\mathrm{ver}} \in \Gamma(\widetilde{u}^*T^{\mathrm{ver}}\widehat{W}).$$

Therefore it remains to show that $(\widetilde{\nabla}_{\xi^{\mathrm{ver}}}\widetilde{J})(q) \in \Gamma(\widetilde{u}^*T^{\mathrm{ver}}\widehat{W})$. For this, it suffices to establish

$$(\widetilde{\nabla}_a\widetilde{J})(b) \in \Gamma(T^{\mathrm{ver}}\widehat{W}|_{\widehat{X}}) \quad \text{for any } a, b \in \Gamma(T^{\mathrm{ver}}\widehat{W}|_{\widehat{X}}).$$

Recall that the term $\widetilde{\nabla}_a\widetilde{J} \in \mathrm{End}(T\widehat{W}|_{\widehat{X}})$ corresponds to applying the connection induced by $\widetilde{\nabla}$ — which we again denote by $\widetilde{\nabla}$ — on the endomorphism bundle, and by its definition we have

$$(\widetilde{\nabla}_a\widetilde{J})(b) = \widetilde{\nabla}_a(\widetilde{J}b) - \widetilde{J}(\widetilde{\nabla}_a b).$$

Similar to above, it is immediate that these last two terms lie in $\Gamma(T^{\mathrm{ver}}\widehat{W}|_{\widehat{X}})$. □

Lemma A.3 *The operator $M_{2,2}$ is surjective.*

Proof If we ignore the constraint $\lessdot p_0 \gtrdot$, the corresponding ($\mathbb{R}$–linear) Cauchy–Riemann type operator

$$\mathcal{W}^{k,p,\delta}(\widetilde{u}^*T^{\mathrm{hor}}\widehat{W}) \to \mathcal{W}^{k-1,p,\delta}(\overline{\mathrm{Hom}}_{\mathbb{C}}(T\Sigma, \widetilde{u}^*T^{\mathrm{hor}}\widehat{W}))$$

is Fredholm, and by a version of Riemann–Roch with its index is easily computed to be 2; see for instance [39, Section 2.1]. It follows that $M_{2,2}$ is also Fredholm, with index 0, and hence to prove its surjectivity it suffices to establish $\ker M_{2,2} = \{0\}$. Suppose by contradiction that η is a nonzero element in $\ker M_{2,2}$. By elliptic regularity we can assume that η is smooth, and its count $Z(\eta)$ of zeros is nonnegative (this follows by the similarity principle [40, Theorem 2.32]), and in fact strictly positive since η necessarily vanishes at the marked point z_0. On the other hand, in the notation of [39, Section 2.1], each puncture z_i of $\widetilde{u}$ has normal Conley–Zehnder index 1 and hence extremal winding number $\alpha_-(\mathbb{A}_{z_i}) = 0$, and therefore using [39, Equation 2.7] we have

$$1 \le Z(\eta) + Z_\infty(\eta) = c_1(\widetilde{u}^*T^{\mathrm{hor}}\widehat{W}) + \sum_{i=1}^{l} \alpha_-(\mathbb{A}_{z_i}) = 0,$$

a contradiction. □

Proof of Proposition A.1 This follows immediately from the decomposition (A-0-1) and Lemmas A.2 and A.3. □

Now suppose that J is an admissible almost complex structure on the symplectization of ∂X, and let $\widetilde{J}$ be an admissible almost complex structure on the symplectization of ∂W which restricts to J on $\mathbb{R} \times (\partial X \times \{0\})$. An argument nearly identical to the above proves:

Proposition A.4 *Let u be an asymptotically cylindrical J–holomorphic punctured sphere in $\mathbb{R} \times \partial X$ such that each asymptotic Reeb orbit is nondegenerate with normal Conley–Zehnder index one. Assume that u is regular and has index zero. Let $\widetilde{u}$ denote the curve given by the composition of u with the inclusion $\mathbb{R} \times \partial X \subset \mathbb{R} \times \partial W$. Then $\widetilde{u}$ is also regular.*

References

[1] **F Bourgeois**, *A Morse–Bott approach to contact homology*, from "Symplectic and contact topology: interactions and perspectives" (Y Eliashberg, B Khesin, F Lalonde, editors), Fields Inst. Commun. 35, Amer. Math. Soc., Providence, RI (2003) 55–77 MR Zbl

[2] **F Bourgeois**, **Y Eliashberg**, **H Hofer**, **K Wysocki**, **E Zehnder**, *Compactness results in symplectic field theory*, Geom. Topol. 7 (2003) 799–888 MR Zbl

[3] **F Bourgeois**, **K Mohnke**, *Coherent orientations in symplectic field theory*, Math. Z. 248 (2004) 123–146 MR Zbl

[4] **F Bourgeois**, **A Oancea**, *Symplectic homology, autonomous Hamiltonians, and Morse–Bott moduli spaces*, Duke Math. J. 146 (2009) 71–174 MR Zbl

[5] **J Chaidez**, **B Wormleighton**, *Lattice formulas for rational SFT capacities*, preprint (2021) arXiv 2106.07920

[6] **K Choi**, *Combinatorial embedded contact homology for toric contact manifolds*, preprint (2016) arXiv 1608.07988

[7] **K Cieliebak**, **K Mohnke**, *Symplectic hypersurfaces and transversality in Gromov–Witten theory*, J. Symplectic Geom. 5 (2007) 281–356 MR Zbl

[8] **K Cieliebak**, **K Mohnke**, *Punctured holomorphic curves and Lagrangian embeddings*, Invent. Math. 212 (2018) 213–295 MR Zbl

[9] **D Cristofaro-Gardiner**, **R Hind**, *Symplectic embeddings of products*, Comment. Math. Helv. 93 (2018) 1–32 MR Zbl

[10] **D Cristofaro-Gardiner**, **R Hind**, **D McDuff**, *The ghost stairs stabilize to sharp symplectic embedding obstructions*, J. Topol. 11 (2018) 309–378 MR Zbl

[11] **D Cristofaro-Gardiner**, **R Hind**, **K Siegel**, *Higher symplectic capacities and the stabilized embedding problem for integral elllipsoids [sic]*, J. Fixed Point Theory Appl. 24 (2022) art. id. 49 MR Zbl

[12] **I Ekeland**, **H Hofer**, *Symplectic topology and Hamiltonian dynamics*, Math. Z. 200 (1989) 355–378 MR Zbl

[13] **I Ekeland**, **H Hofer**, *Symplectic topology and Hamiltonian dynamics, II*, Math. Z. 203 (1990) 553–567 MR Zbl

[14] **M Gromov**, *Pseudo holomorphic curves in symplectic manifolds*, Invent. Math. 82 (1985) 307–347 MR Zbl

[15] **M Gromov**, *Soft and hard symplectic geometry*, from "Proceedings of the International Congress of Mathematicians, I" (A M Gleason, editor), Amer. Math. Soc., Providence, RI (1987) 81–98 MR Zbl

[16] **J Gutt**, **M Hutchings**, *Symplectic capacities from positive S^1–equivariant symplectic homology*, Algebr. Geom. Topol. 18 (2018) 3537–3600 MR Zbl

[17] **R Hind**, **E Kerman**, *New obstructions to symplectic embeddings*, Invent. Math. 196 (2014) 383–452 MR Zbl

[18] **R K Hind**, **E Kerman**, *J–holomorphic cylinders between ellipsoids in dimension four*, J. Symplectic Geom. 18 (2020) 1221–1245 MR Zbl

[19] **M Hutchings**, *The embedded contact homology index revisited*, from "New perspectives and challenges in symplectic field theory" (M Abreu, F Lalonde, L Polterovich, editors), CRM Proc. Lecture Notes 49, Amer. Math. Soc., Providence, RI (2009) 263–297 MR Zbl

[20] **M Hutchings**, *Quantitative embedded contact homology*, J. Differential Geom. 88 (2011) 231–266 MR Zbl

[21] **M Hutchings**, *Lecture notes on embedded contact homology*, from "Contact and symplectic topology" (F Bourgeois, V Colin, A Stipsicz, editors), Bolyai Soc. Math. Stud. 26, Bolyai Math. Soc., Budapest (2014) 389–484 MR Zbl

[22] **M Hutchings**, *Beyond ECH capacities*, Geom. Topol. 20 (2016) 1085–1126 MR Zbl

[23] **M Hutchings**, *An elementary alternative to ECH capacities*, Proc. Natl. Acad. Sci. USA 119 (2022) art. id. e2203090119

[24] **M Hutchings**, **C H Taubes**, *Gluing pseudoholomorphic curves along branched covered cylinders, I*, J. Symplectic Geom. 5 (2007) 43–137 MR Zbl

[25] **M Hutchings**, **C H Taubes**, *Gluing pseudoholomorphic curves along branched covered cylinders, II*, J. Symplectic Geom. 7 (2009) 29–133 MR Zbl

[26] **D Irvine**, *The stabilized symplectic embedding problem for polydiscs*, preprint (2019) arXiv 1907.13159

[27] **E Kerman**, **Y Liang**, *On symplectic capacities and their blind spots*, J. Topol. Anal. (online publication June 2023)

[28] **M Landry**, **M McMillan**, **E Tsukerman**, *On symplectic capacities of toric domains*, Involve 8 (2015) 665–676 MR Zbl

[29] **D McDuff**, *A remark on the stabilized symplectic embedding problem for ellipsoids*, Eur. J. Math. 4 (2018) 356–371 MR Zbl

[30] **D McDuff**, **K Siegel**, *Counting curves with local tangency constraints*, J. Topol. 14 (2021) 1176–1242 MR Zbl

[31] **A Moreno**, **R Siefring**, *Holomorphic curves in the presence of holomorphic hypersurface foliations*, preprint (2019) arXiv 1902.02700

[32] **J Pardon**, *Contact homology and virtual fundamental cycles*, J. Amer. Math. Soc. 32 (2019) 825–919 MR Zbl

[33] **M B Pereira**, *Equivariant symplectic homology, linearized contact homology and the Lagrangian capacity*, PhD thesis, Universität Augsburg (2022) arXiv 2205.13381

[34] **M Schwarz**, *Morse homology*, Progr. Math. 111, Birkhäuser, Basel (1993) MR Zbl

[35] **R Siefring**, *Intersection theory of punctured pseudoholomorphic curves*, Geom. Topol. 15 (2011) 2351–2457 MR Zbl

[36] **K Siegel**, *Computing higher symplectic capacities, II*, in preparation

[37] **K Siegel**, *Higher symplectic capacities*, preprint (2019) arXiv 1902.01490

[38] **K Siegel**, *Computing higher symplectic capacities, I*, Int. Math. Res. Not. 2022 (2022) 12402–12461 MR Zbl

[39] **C Wendl**, *Automatic transversality and orbifolds of punctured holomorphic curves in dimension four*, Comment. Math. Helv. 85 (2010) 347–407 MR Zbl

[40] **C Wendl**, *Lectures on symplectic field theory*, lecture notes (2020) Available at `https://tinyurl.com/SFTnotes`

Mathematics Department, Columbia University
New York, NY, United States

Department of Mathematics, University of Southern California
Los Angeles, CA, United States

`dusa@math.columbia.edu`, `kyler.siegel@usc.edu`

Proposed: Leonid Polterovich
Seconded: Paul Seidel, David Fisher

Received: 31 October 2021
Revised: 17 July 2022

Geometry & Topology **28**:3 (2024) 1287–1339
DOI: 10.2140/gt.2024.28.1287
Published: 10 May 2024

Quadric bundles and hyperbolic equivalence

ALEXANDER KUZNETSOV

We introduce the notion of hyperbolic equivalence for quadric bundles and quadratic forms on vector bundles and show that hyperbolic equivalent quadric bundles share many important properties: they have the same Brauer data; moreover, if they have the same dimension over the base, they are birational over the base and have equal classes in the Grothendieck ring of varieties.

Furthermore, when the base is a projective space we show that two quadratic forms are hyperbolic equivalent if and only if their cokernel sheaves are isomorphic up to twist, their fibers over a fixed point of the base are Witt equivalent, and, in some cases, certain quadratic forms on intermediate cohomology groups of the underlying vector bundles are Witt equivalent. For this we show that any quadratic form over $\mathbb{P}^n$ is hyperbolic equivalent to a quadratic form whose underlying vector bundle has many cohomology vanishings; this class of bundles, called VLC bundles in the paper, is interesting by itself.

14D06

1 Introduction

Let $Q \to X$ be a *quadric bundle*, that is, a proper morphism which can be presented as a composition $Q \hookrightarrow \mathbb{P}_X(\mathcal{E}) \to X$, where $\mathbb{P}_X(\mathcal{E}) \to X$ is the projectivization of a vector bundle $\mathcal{E}$ and $Q \hookrightarrow \mathbb{P}_X(\mathcal{E})$ is a divisorial embedding of relative degree 2 over X. A quadric bundle is determined by a *quadratic form* $q\colon \operatorname{Sym}^2 \mathcal{E} \to \mathcal{L}^\vee$ with values in a line bundle $\mathcal{L}^\vee$, or, equivalently, by a *self-dual morphism*

$$q\colon \mathcal{E} \otimes \mathcal{L} \to \mathcal{E}^\vee. \tag{1-1}$$

Conversely, the quadratic form q is determined by Q up to rescaling and a twist transformation

$$\mathcal{E} \mapsto \mathcal{E} \otimes \mathcal{M}, \quad \mathcal{L} \mapsto \mathcal{L} \otimes \mathcal{M}^{-2},$$

where $\mathcal{M}$ is a line bundle on X.

Furthermore, with a quadric bundle one associates the coherent sheaf

$$\mathcal{C}(Q) = \mathcal{C}(q) := \operatorname{Coker}(q\colon \mathcal{E} \otimes \mathcal{L} \to \mathcal{E}^\vee) \tag{1-2}$$

on X, which we call its *cokernel sheaf* and which is determined by Q up to a line bundle twist. We will usually assume that X is integral and the general fiber of $Q \to X$ is nondegenerate, or equivalently, that q is an isomorphism at the general point of X, so that $\operatorname{Ker}(q) = 0$ and $\mathcal{C}(q)$ is a torsion sheaf on X. Then the sheaf $\mathcal{C}(q)$ is endowed with a "shifted" self-dual isomorphism

$$\bar{q}\colon \mathcal{C}(q) \xrightarrow{\simeq} \mathcal{E}xt^1(\mathcal{C}(q), \mathcal{L}) = \mathcal{C}(q)^\vee \otimes \mathcal{L}[1], \tag{1-3}$$

where $\mathcal{C}(q)^\vee$ is the *derived* dual of $\mathcal{C}(q)$ and $[1]$ is the shift in the derived category; see Section 4.1 for a discussion of sheaves enjoying this property.

The main question addressed in this paper is: what properties of quadric bundles are determined by their cokernel sheaves? (We restate this question below in a more precise form as Question 1.2.) A priori it is hard to expect that the cokernel sheaf determines a lot; for instance because it is supported only on the discriminant divisor of Q/X. However, the main result of this paper is that, in the case where X is a projective space and some mild numerical conditions discussed below are satisfied, the cokernel sheaf determines the quadric bundle up to a natural equivalence relation, which we call *hyperbolic equivalence*, and which itself preserves the most important geometric properties of quadric bundles.

Hyperbolic equivalence is generated by operations of *hyperbolic reduction* and *hyperbolic extension*. The simplest instance of a hyperbolic reduction (over the trivial base) is the operation that takes a quadric $Q \subset \mathbb{P}^r$ and a smooth point $p \in Q$ and associates to it the fundamental locus of the linear projection $\operatorname{Bl}_p(Q) \to \mathbb{P}^{r-1}$, which is a quadric $Q_- \subset \mathbb{P}^{r-2} \subset \mathbb{P}^{r-1}$ of dimension $\dim(Q) - 2$. From the above geometric perspective it is clear that the hyperbolic reduction procedure is invertible: the inverse operation, which we call a hyperbolic extension, takes a quadric $Q \subset \mathbb{P}^r$ and a hyperplane embedding $\mathbb{P}^r \hookrightarrow \mathbb{P}^{r+1}$ and associates to it the quadric $Q_+ \subset \mathbb{P}^{r+2}$ obtained by blowing up $Q \subset \mathbb{P}^{r+1}$ and then contracting the strict transform of $\mathbb{P}^r \subset \mathbb{P}^{r+1}$.

The operations of hyperbolic reduction and extension can be defined in relative setting, ie for quadric bundles $Q \subset \mathbb{P}_X(\mathcal{E}) \to X$ over any base X, and, moreover, can be lifted to operations on quadratic forms. For the reduction a smooth point is replaced by a section $X \to Q$ that does not pass through singular points of fibers, or more generally, by a *regular isotropic subbundle* $\mathcal{F} \subset \mathcal{E}$, and for the extension a hyperplane embedding is replaced by an embedding $\mathcal{E} \hookrightarrow \mathcal{E}'$ of vector bundles of arbitrary corank. We define these operations for quadratic forms and quadric bundles in Sections 2.1 and 2.2 and say that quadratic forms $(\mathcal{E}_1, q_1)$ and $(\mathcal{E}_2, q_2)$ or quadric bundles Q_1 and Q_2 over X are *hyperbolic equivalent* if they can be connected by a chain of hyperbolic reductions and extensions.

While the construction of hyperbolic reduction is quite straightforward in the general case, this is far from true for hyperbolic extension. In fact, when we start with an extension $0 \to \mathcal{E} \to \mathcal{E}' \to \mathcal{G} \to 0$ of vector bundles, where the bundle $\mathcal{G}$ has rank *greater than one*, this operation does not have a simple geometric description (as in the rank-one case); moreover, the set $\mathrm{HE}(\mathcal{E}, q, \varepsilon)$ of all hyperbolic extensions of $(\mathcal{E}, q)$ with respect to an extension class $\varepsilon \in \mathrm{Ext}^1(\mathcal{G}, \mathcal{E})$ is empty unless a certain obstruction class

$$q(\varepsilon, \varepsilon) \in \mathrm{Ext}^2\left(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\right)$$

vanishes, and when the obstruction is zero, $\mathrm{HE}(\mathcal{E}, q, \varepsilon)$ is a principal homogeneous space under the natural action of the group $\mathrm{Ext}^1\left(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\right)$. This can be seen even in the simplest case where the extension is split, ie $\mathcal{E}' = \mathcal{E} \oplus \mathcal{G}$ — in this case the obstruction vanishes and the corresponding hyperbolic extensions have the form $\mathcal{E}_+ = \mathcal{E} \oplus \mathcal{G}_+$, where $\mathcal{G}_+$ is an arbitrary extension of $\mathcal{G}$ by $\mathcal{L}^\vee \otimes \mathcal{G}^\vee$ with the class in the subspace $\mathrm{Ext}^1\left(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\right) \subset \mathrm{Ext}^1(\mathcal{G}, \mathcal{L}^\vee \otimes \mathcal{G}^\vee)$. For a discussion of a slightly more complicated situation, see Remark 2.10. In general the situation is similar but even more complicated. The construction of hyperbolic extension explained in Section 2.2 (see Theorem 2.9) is the first main result of this paper.

As we mentioned above, hyperbolic equivalence does not change the basic invariants of a quadratic form. In Section 2.3 we prove the following result (for the definition of the Clifford algebra $\mathrm{Cliff}_0(\mathcal{E}, q)$ we refer to our earlier paper [11]).

Proposition 1.1 *Let $(\mathcal{E}, q)$ and $(\mathcal{E}', q')$ be hyperbolic equivalent generically nondegenerate quadratic forms over X, and let $Q \to X$ and $Q' \to X$ be the corresponding hyperbolic equivalent quadric bundles, where X is a scheme over a field $\Bbbk$ of characteristic not equal to 2. Then*:

- (0) *One has* $\dim(Q/X) \equiv \dim(Q'/X) \bmod 2$.
- (1) *The cokernel sheaves $\mathcal{C}(Q) = \mathcal{C}(q)$ and $\mathcal{C}(Q') = \mathcal{C}(q')$ are isomorphic up to twist by a line bundle on X, and their isomorphism is compatible with the shifted quadratic forms (1-3).*
- (2) *The discriminant divisors $\mathrm{Disc}_{Q/X} \subset X$ and $\mathrm{Disc}_{Q'/X} \subset X$ of Q and Q' coincide.*
- (3) *The even parts of Clifford algebras $\mathrm{Cliff}_0(\mathcal{E}, q)$ and $\mathrm{Cliff}_0(\mathcal{E}', q')$ on X are Morita equivalent.*
- (4) *If $\dim(Q/X) = \dim(Q'/X)$, then $[Q] = [Q']$ in the Grothendieck ring of varieties $\mathrm{K}_0(\mathrm{Var}/\Bbbk)$.*
- (5) *If the base scheme X is integral, the classes of general fibers $q_{\mathrm{K}(X)}$ and $q'_{\mathrm{K}(X)}$ in the Witt group of quadratic forms over the field of rational functions $\mathrm{K}(X)$ on X are equal. If, moreover, $\dim(Q/X) = \dim(Q'/X)$, then $Q_{\mathrm{K}(X)} \cong Q'_{\mathrm{K}(X)}$, and Q is birational to Q' over X.*

In the rest of the paper we explore whether the converse of Proposition 1.1(1) is true. More precisely, we discuss the following:

Question 1.2 *Does the cokernel sheaf endowed with its shifted quadratic form (1-3) determine the hyperbolic equivalence class of quadratic forms?*

At this point it makes sense to explain the relation of hyperbolic equivalence to Witt groups. Recall that the Witt group $W(K)$ of a field K is defined as the quotient of the monoid of isomorphism classes of *nondegenerate* quadratic forms (V, q), where V is a K–vector space and $q \in \mathrm{Sym}^2 V^\vee$ is a nondegenerate quadratic form, by the class of the hyperbolic plane $\left(K^{\oplus 2}, \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\right)$. Similarly, the Witt group $W(X)$ of a scheme X is defined as the quotient of the monoid of isomorphism classes of *unimodular*, ie everywhere nondegenerate quadratic forms $(\mathcal{E}, q)$, where $\mathcal{E}$ is a vector bundle on X and $q \in \operatorname{Hom}(\mathcal{O}_X, \operatorname{Sym}^2 \mathcal{E}^\vee)$ is everywhere nondegenerate, by the classes of *metabolic forms* $\left(\mathcal{F} \oplus \mathcal{F}^\vee, \left(\begin{smallmatrix} 0 & 1 \\ 1 & q' \end{smallmatrix}\right)\right)$; see Knebusch [10]. As explained in the survey by Balmer [4], modifying the standard duality operation on the category of vector bundles on X one can define the Witt group $\boldsymbol{W}(X, \mathcal{L})$ that classifies classes of line bundle valued nondegenerate quadratic forms $q\colon \operatorname{Sym}^2 \mathcal{E} \to \mathcal{L}^\vee$. Moreover, a trick described in Bayer-Fluckiger and Fainsilber [5] allows one to define the Witt group $\boldsymbol{W}_{\mathrm{nu}}(X, \mathcal{L})$ of *nonunimodular quadratic forms* (ie forms that are allowed to be degenerate) as the usual Witt group of the category of morphisms of vector bundle. Thus, quadratic forms (1-1) define elements of $\boldsymbol{W}_{\mathrm{nu}}(X, \mathcal{L})$.

It is well known that hyperbolic reduction (as defined above) does not change the class of a quadratic form $(\mathcal{E}, q)$ in the Witt group $\boldsymbol{W}_{\mathrm{nu}}(X, \mathcal{L})$; see eg [4, Section 1.1.5], where it is called *sublagrangian reduction*. On the other hand, Witt equivalence may change the cokernel sheaf of a quadratic form, eg for any morphism $\varphi\colon \mathcal{E}_1 \to \mathcal{E}_2$ of vector bundles the class of the quadratic form

$$\left(\mathcal{E}_1 \oplus \mathcal{E}_2^\vee, \begin{pmatrix} 0 & \varphi \\ \varphi^\vee & 0 \end{pmatrix}\right)$$

in the Witt group $\boldsymbol{W}_{\mathrm{nu}}(X, \mathcal{O}_X)$ is zero, but the corresponding cokernel sheaf $\mathcal{C} \cong \operatorname{Coker}(\varphi) \oplus \operatorname{Coker}(\varphi^\vee)$ is nontrivial unless φ is an isomorphism. Therefore, Question 1.2 does not reduce to a question about Witt groups.

To answer Question 1.2 (in the case $X = \mathbb{P}^n$) we define the following two basic hyperbolic equivalence invariants of quadratic forms that take values in the nonunimodular Witt group $\boldsymbol{W}_{\mathrm{nu}}(\mathsf{k})$ of the base field k. Here and everywhere below we assume that the characteristic of k is not equal to 2.

To define the first invariant, assume X is a k–scheme with a k–point $x \in X(\mathsf{k})$. We fix a trivialization of $\mathcal{L}_x$ and define

$$\mathrm{w}_x(\mathcal{E}, q) := [(\mathcal{E}_x, q_x)] \in \boldsymbol{W}_{\mathrm{nu}}(\mathsf{k}) \tag{1-4}$$

to be the class of the quadratic form q_x obtained as the composition $\operatorname{Sym}^2 \mathcal{E}_x \xrightarrow{q} \mathcal{L}_x^\vee \cong \mathsf{k}$, where the second arrow is given by the trivialization of $\mathcal{L}_x$. (We could also define $\mathrm{w}_x(\mathcal{E}, q)$ to be the class of the quotient of $(\mathcal{E}_x, q_x)$ by the kernel; then it would take values in $\boldsymbol{W}(\mathsf{k})$.) The class $\mathrm{w}_x(\mathcal{E}, q)$ depends on the choice of trivialization, but this is not a problem for our purposes. If the scheme X has no k–points, we could take x to be a k'–point for any field extension k'/k and define $\mathrm{w}_x(\mathcal{E}, q) \in \boldsymbol{W}_{\mathrm{nu}}(\mathsf{k}')$ in the same way.

For the second invariant, assume X is smooth, connected and proper k–scheme, $n = \dim(X)$ is even, and $\mathcal{L} \otimes \omega_X \cong \mathcal{M}^2$ for a line bundle $\mathcal{M}$ on X, where ω_X is the canonical line bundle of X. Then we define the bilinear form

$$H^{n/2}(X, \mathcal{E} \otimes \mathcal{M}) \otimes H^{n/2}(X, \mathcal{E} \otimes \mathcal{M}) \xrightarrow{q} H^n(X, \mathcal{L}^\vee \otimes \mathcal{M} \otimes \mathcal{M}) \cong H^n(X, \omega_X) = \mathsf{k} \tag{1-5}$$

on the cohomology group $H^{n/2}(X, \mathcal{E} \otimes \mathcal{M})$, which we denote by $H^{n/2}(q)$ or $H^{n/2}(Q)$. This form, of course, depends on the choice of the line bundle $\mathcal{M}$ (if $\operatorname{Pic}(X)$ has 2–torsion, there may be several choices), but we suppress this in the notation. The bilinear form $H^{n/2}(q)$ is symmetric if $\frac{1}{2}n$ is even (and skew-symmetric otherwise) and possibly degenerate. Anyway, if n is divisible by 4, we denote its class in the nonunimodular Witt group by

$$\mathrm{hw}(\mathcal{E},q) := [H^{n/2}(X,\mathcal{E}\otimes\mathcal{M}), H^{n/2}(q)] \in \boldsymbol{W}_{\mathrm{nu}}(\mathsf{k}). \tag{1-6}$$

(Again, we could define $\mathrm{hw}(\mathcal{E},q)$ to be the class of the quotient of $H^{n/2}(q)$ by its kernel; then it would take values in $\boldsymbol{W}(\mathsf{k})$.) As before, the class $\mathrm{hw}(\mathcal{E},q)$ depends on the choice of isomorphism $\mathcal{L}\otimes\omega_X \cong \mathcal{M}^2$, but this is still not a problem.

Note that when k is algebraically closed, $\boldsymbol{W}(\mathsf{k}) \cong \mathbb{Z}/2$ and so, if the corresponding forms are nondegenerate, the invariants $\mathrm{w}_x(\mathcal{E},q)$ and $\mathrm{hw}(\mathcal{E},q)$ take values in $\mathbb{Z}/2$, and do not depend on extra choices. In this case $\mathrm{w}_x(\mathcal{E},q)$ is just the parity of the rank of $\mathcal{E}$, and $\mathrm{hw}(\mathcal{E},q)$ is the parity of the rank of $H^{n/2}(q)$.

The second main result of this paper is the affirmative answer to Question 1.2 in the case $X=\mathbb{P}^n$. Recall that $\mathrm{Pic}(\mathbb{P}^n)=\mathbb{Z}$; hence any line bundle $\mathcal{L}$ has the form $\mathcal{L}=\mathcal{O}(-m)$ for some $m\in\mathbb{Z}$. We need to define the following two "standard" types of unimodular quadratic forms with values in $\mathcal{O}(m)$:

$$(\mathcal{E},q) \cong \bigoplus_{i\equiv m \bmod 2} W^i \otimes \mathcal{O}\big(\tfrac{1}{2}(m+i)\big), \quad \text{or} \tag{1-7}$$

$$(\mathcal{E},q) \cong \bigoplus_{i\equiv m+n+1 \bmod 2} W^i \otimes \Omega^{n/2}\big(\tfrac{1}{2}(m+n+1+i)\big) \quad \text{if } n \text{ is even,} \tag{1-8}$$

where $\{W^i\}$ is a collection of vector spaces and q is the sum of tensor products of the natural pairings

$$\mathcal{O}\big(\tfrac{1}{2}(m-i)\big)\otimes\mathcal{O}\big(\tfrac{1}{2}(m+i)\big) \longrightarrow \mathcal{O}(m),$$
$$\Omega^{n/2}\big(\tfrac{1}{2}(m+n+1-i)\big)\otimes\Omega^{n/2}\big(\tfrac{1}{2}(m+n+1+i)\big) \xrightarrow{\wedge} \Omega^n(m+n+1)\cong\mathcal{O}(m)$$

(the second is given by wedge product, hence it is symmetric if $\frac{1}{2}n$ is even and skew-symmetric if $\frac{1}{2}n$ is odd) and of nondegenerate bilinear forms $q_{W^i}\colon W^{-i}\otimes W^i \to \mathsf{k}$ which for $i=0$ are symmetric in the case (1-7) and (1-8) with $\frac{1}{2}n$ even and skew-symmetric in the case (1-8) with $\frac{1}{2}n$ odd.

Recall that the cokernel sheaf $\mathcal{C}(q)$ of a quadratic form $(\mathcal{E},q)$ is endowed with the shifted self-duality isomorphism $\bar{q}$; see (1-3). In conditions (1) and (2) of the theorem we use the same trivialization of $\mathcal{O}(-m)_x$ and the same isomorphism $\mathcal{O}(-m)\otimes\omega_{\mathbb{P}^n}\cong\mathcal{M}^2$ for $(\mathcal{E}_1,q_1)$ and $(\mathcal{E}_2,q_2)$.

Theorem 1.3 *Let k be a field of characteristic not equal to 2 and let $X=\mathbb{P}^n$ be a projective space over k. Let $\mathcal{E}_1(-m)\xrightarrow{q_1}\mathcal{E}_1^\vee$ and $\mathcal{E}_2(-m)\xrightarrow{q_2}\mathcal{E}_2^\vee$ be generically nondegenerate self-dual morphisms over $\mathbb{P}^n$. Assume there is an isomorphism of sheaves $\mathcal{C}(q_1)\cong\mathcal{C}(q_2)$ compatible with the quadratic forms $\bar{q}_1$ and $\bar{q}_2$. Then $(\mathcal{E}_1,q_1)$ is hyperbolic equivalent to the direct sum of $(\mathcal{E}_2,q_2)$ and one of the standard quadratic forms (1-7) or (1-8), where $W^i=0$ for $i\neq 0$, and q_{W^0} is anisotropic.*

If, moreover, the following conditions hold true:

1. *if m is even then $\mathrm{w}_x(\mathcal{E}_1,q_1)=\mathrm{w}_x(\mathcal{E}_2,q_2)\in\boldsymbol{W}_{\mathrm{nu}}(\mathsf{k})$ for some k–point $x\in\mathbb{P}^n$;*
2. *if m is odd and n is divisible by 4 then $\mathrm{hw}(\mathcal{E}_1,q_1)=\mathrm{hw}(\mathcal{E}_2,q_2)\in\boldsymbol{W}_{\mathrm{nu}}(\mathsf{k})$;*

then $(\mathcal{E}_1,q_1)$ is hyperbolic equivalent to $(\mathcal{E}_2,q_2)$.

If k is algebraically closed and x is chosen away from the support of $\mathcal{C}(q_i)$, condition (1) in the theorem just amounts to $\mathcal{E}_1$ and $\mathcal{E}_2$ having ranks of the same parity. Similarly, condition (2) amounts to the forms $\mathrm{hw}(\mathcal{E}_i, q_i)$ having ranks of the same parity.

Adding a standard summand of type (1-7) with $W^i = 0$ for $i \neq 0$ and $\dim(W^0) = 1$ corresponds geometrically to replacing a quadric bundle $Q \subset \mathbb{P}_{\mathbb{P}^n}(\mathcal{E}) \to \mathbb{P}^n$ by the quadric bundle $\widetilde{Q} \to \mathbb{P}^n$, where $\widetilde{Q} \to \mathbb{P}_{\mathbb{P}^n}(\mathcal{E})$ is the double covering branched along Q — note that this operation changes the parity of the rank of $\mathcal{E}$. The geometric meaning of adding a trivial summand of type (1-8) is not so obvious.

Remark 1.4 The condition of compatibility of an isomorphism $\mathcal{C}(q_1) \cong \mathcal{C}(q_2)$ with the shifted quadratic forms $\overline{q}_1$ and $\overline{q}_2$ may seem subtle, but in many applications it is easy to verify. For instance, if the sheaves $\mathcal{C}(q_i)$ are *simple*, ie $\mathrm{End}(\mathcal{C}(q_i)) \cong \mathsf{k}$, then a nondegenerate shifted quadratic form on $\mathcal{C}(q_i)$ is unique up to scalar, so if k is quadratically closed then any isomorphism of $\mathcal{C}(q_i)$, after appropriate rescaling, is compatible with the shifted quadratic forms.

To prove Theorem 1.3 we develop, in Section 3, the theory of what we call *VHC morphisms* (here VHC stands for *vanishing of half cohomology*). These are morphisms of vector bundles $\mathcal{E}_{\mathrm{L}} \to \mathcal{E}_{\mathrm{U}}$ on $\mathbb{P}^n$ such that

$$H^p(\mathbb{P}^n, \mathcal{E}_{\mathrm{L}}(t)) = 0 \quad \text{for } 1 \le p \le \left\lfloor \tfrac{1}{2}n \right\rfloor \quad \text{and all } t \in \mathbb{Z},$$
$$H^p(\mathbb{P}^n, \mathcal{E}_{\mathrm{U}}(t)) = 0 \quad \text{for } \left\lceil \tfrac{1}{2}n \right\rceil \le p \le n-1 \quad \text{and all } t \in \mathbb{Z}.$$

(We say then that $\mathcal{E}_{\mathrm{L}}$ is VLC as its *lower* intermediate cohomology vanishes, and $\mathcal{E}_{\mathrm{U}}$ is VUC as its *upper* intermediate cohomology vanishes.) The main results of this section are Theorem 3.15, in which we prove the uniqueness (under appropriate assumptions) of VHC resolutions, and Corollary 3.18, proving the existence of VHC resolutions for any sheaf of projective dimension one.

In Section 4 we apply this technique to the case of resolutions of *symmetric sheaves*; see Definition 4.1. Any cokernel sheaf $\mathcal{C}(q)$ is symmetric, and conversely, if $X = \mathbb{P}^n$ then under a mild technical assumption any symmetric sheaf is isomorphic to $\mathcal{C}(q)$ for some self-dual morphism $q\colon \mathcal{E}(-m) \to \mathcal{E}^\vee$; see [7] or Theorem 4.8 and Remark 4.9 in Section 4.

Our main technical result here is the modification theorem (Theorem 4.17), in which we show that any self-dual morphism over $\mathbb{P}^n$ is hyperbolic equivalent to the sum of a self-dual VHC morphism and a standard unimodular self-dual morphism of type (1-7) or (1-8). This implies Theorem 1.3; see Section 4.4 for the proof.

Combining Theorem 1.3 with Proposition 1.1 we obtain the following corollary, which for simplicity we state over an algebraically closed ground field.

Corollary 1.5 *Let k be an algebraically closed field of characteristic not equal to 2. Let $Q \to \mathbb{P}^n$ and $Q' \to \mathbb{P}^n$ be generically smooth quadric bundles such that there is an isomorphism of the cokernel sheaves $\mathcal{C}(Q) \cong \mathcal{C}(Q')$ compatible with their shifted quadratic forms. If n is divisible by 4 and m is odd, assume also that $\mathrm{rk}(H^{n/2}(Q)) \equiv \mathrm{rk}(H^{n/2}(Q')) \bmod 2$, where the quadratic forms $H^{n/2}(Q)$ and $H^{n/2}(Q')$ are defined by (1-5). Then:*

(1) *If* $\dim(Q/\mathbb{P}^n)$ *and* $\dim(Q'/\mathbb{P}^n)$ *are even, then the corresponding discriminant double covers* $S \to \mathbb{P}^n$ *and* $S' \to \mathbb{P}^n$ *are isomorphic over* $\mathbb{P}^n$, *and also the Brauer classes* $\beta_S \in \mathrm{Br}(S_{\le 1})$ *and* $\beta'_S \in \mathrm{Br}(S'_{\le 1})$ *on the corank* ≤ 1 *loci inside* S *and* S' *are equal.*

(2) *If* $\dim(Q/\mathbb{P}^n)$ *and* $\dim(Q'/\mathbb{P}^n)$ *are odd, then the corresponding discriminant root stacks* $S \to \mathbb{P}^n$ *and* $S' \to \mathbb{P}^n$ *are isomorphic over* $\mathbb{P}^n$, *and also the Brauer classes* $\beta_S \in \mathrm{Br}(S_{\le 1})$ *and* $\beta'_S \in \mathrm{Br}(S'_{\le 1})$ *on the corank* ≤ 1 *loci inside* S *and* S' *are equal.*

(3) *If* $\dim(Q/\mathbb{P}^n) = \dim(Q'/\mathbb{P}^n)$, *then* $[Q] = [Q']$ *in the Grothendieck ring of varieties* $\mathrm{K}_0(\mathrm{Var}/\mathsf{k})$.

(4) *If* $\dim(Q/\mathbb{P}^n) = \dim(Q'/\mathbb{P}^n)$, *then there is a birational isomorphism* $Q \sim Q'$ *over* $\mathbb{P}^n$.

To finish the introduction it should be said that this paper was inspired by the recent paper of Bini, Kapustka and Kapustka [6], where similar questions were discussed. In particular, assertions (1) and (4) of Corollary 1.5 in the case $n = 2$ were proved there. We refer to [6] for various geometric applications of these results.

On the other hand, we want to stress that the approach of the present paper is completely different: the results of [6] are based on an explicit computation of the Brauer class of a quadric bundle using the technique developed by Ingalls, Obus, Ozman and Viray in [9]. It is unclear whether these methods can be effectively generalized to higher dimensions.

It also makes sense to mention that the technique of hyperbolic extensions and VHC resolutions developed in this paper can be used for other questions related to quadric bundles over arbitrary schemes and vector bundles on projective spaces.

Convention Throughout the paper we work over an arbitrary field k of characteristic not equal to 2.

Acknowledgements This paper owes its very existence to [6], so I am very grateful to its authors for inspiration and useful discussions. I would also like to thank Alexey Ananyevskiy for a suggestion that allowed me to improve significantly the results of Proposition 1.1(4) and Corollary 1.5(3), and the referee for many useful comments about the first version of the paper.

This work is supported by the Russian Science Foundation grant 19-11-00164, https://rscf.ru/project/19-11-00164/.

2 Quadric bundles and hyperbolic equivalence

Recall from the introduction the definition of a quadric bundle, of its associated quadratic form and self-dual morphism (1-1), which we assume to be generically nondegenerate, of the cokernel sheaf (1-2) and of its shifted self-duality (1-3). Conversely, we denote by

$$Q(\mathcal{E}, q) \subset \mathbb{P}_X(\mathcal{E})$$

the quadric bundle associated with a quadratic form $(\mathcal{E}, q)$ or a morphism (1-1).

2.1 Hyperbolic reduction

We start with the notion of hyperbolic reduction, which is well known; see [3; 13]. For the reader's convenience we remind the definition in a slightly different form.

Let (1-1) be a self-dual morphism of vector bundles on a scheme X. We will say that a vector subbundle $\phi\colon \mathcal{F} \hookrightarrow \mathcal{E}$ is *regular isotropic* if the composition

$$\mathcal{E} \xrightarrow{q} \mathcal{E}^\vee \otimes \mathcal{L}^\vee \xrightarrow{\phi^\vee} \mathcal{F}^\vee \otimes \mathcal{L}^\vee$$

is surjective and vanishes on the subbundle $\mathcal{F} \subset \mathcal{E}$, ie $\mathcal{F}$ is contained in the subbundle

$$\mathcal{F}^\perp := \operatorname{Ker}(\mathcal{E} \twoheadrightarrow \mathcal{F}^\vee \otimes \mathcal{L}^\vee) \subset \mathcal{E}. \tag{2-1}$$

If $\mathcal{F}$ is regular isotropic, the restriction of q to $\mathcal{F}^\perp$ contains $\mathcal{F}$ in the kernel, hence induces a quadratic form on $\mathcal{F}^\perp/\mathcal{F}$. We summarize these observations in the following.

Lemma 2.1 *Let (1-1) be a self-dual morphism of vector bundles on a scheme X. Let $\phi\colon \mathcal{F} \hookrightarrow \mathcal{E}$ be a regular isotropic subbundle. Write*

$$\mathcal{E}_- := \mathcal{F}^\perp/\mathcal{F}.$$

The restriction of q to $\mathcal{F}^\perp$ induces a self-dual morphism $q_-\colon \mathcal{E}_- \otimes \mathcal{L} \to \mathcal{E}_-^\vee$ such that there is an isomorphism $\mathcal{C}(q_-) \cong \mathcal{C}(q)$ of the cokernel bundles compatible with their shifted self-dualities $\overline{q}$ and $\overline{q}_-$.

Proof The result follows from the argument of [13, Lemma 2.4]. Indeed, it is explained in loc. cit. that the cokernel sheaf $\mathcal{C}(q_-)$ is isomorphic to the cohomology of the bicomplex (cf [13, equation (2)])

$$\begin{tikzcd} \mathcal{F} \otimes \mathcal{L} \arrow[r, "\phi"] \arrow[d, "\mathrm{id}"'] & \mathcal{E} \otimes \mathcal{L} \arrow[r, "\phi^\vee \circ q"] \arrow[d, "q"] & \mathcal{F}^\vee \arrow[d, "\mathrm{id}"] \\ \mathcal{F} \otimes \mathcal{L} \arrow[r, "q \circ \phi"] & \mathcal{E}^\vee \arrow[r, "\phi^\vee"] & \mathcal{F}^\vee \end{tikzcd} \tag{2-2}$$

Its left and right columns are acyclic, while the middle one coincides with (1-1), hence $\mathcal{C}(q_-) \cong \mathcal{C}(q)$. Furthermore, using the self-duality of q, we see that the dual of (2-2) twisted by $\mathcal{L}$ is isomorphic to (2-2), and moreover, this isomorphism is compatible with the isomorphism of the dual of (1-1) twisted by $\mathcal{L}$ with (1-1). This means that the isomorphism of the cokernel sheaves $\mathcal{C}(q_-) \cong \mathcal{C}(q)$ is compatible with their shifted self-dualities. □

The operation

$$(\mathcal{E}, q) \mapsto (\mathcal{E}_-, q_-) \quad \text{or} \quad Q(\mathcal{E}, q) \mapsto Q(\mathcal{E}_-, q_-)$$

defined in Lemma 2.1 is called *hyperbolic reduction* of a quadratic form (resp. of a quadric bundle) with respect to the subbundle $\mathcal{F}$. As explained in [13, Proposition 2.5], this operation can be interpreted geometrically in terms of the linear projection of $Q \subset \mathbb{P}_X(\mathcal{E})$ from the linear subbundle $\mathbb{P}_X(\mathcal{F}) \subset Q \subset \mathbb{P}_X(\mathcal{E})$.

The next simple lemma motivates the terminology.

Lemma 2.2 *Assume X is integral and $\mathrm{K}(X)$ is the field of rational functions on X. If Q/X is a generically nondegenerate quadric bundle and Q_-/X is its hyperbolic reduction, then the quadratic forms $q_{\mathrm{K}(X)}$ and $(q_-)_{\mathrm{K}(X)}$ corresponding to their general fibers are equal in the Witt group $\boldsymbol{W}(\mathrm{K}(X))$ of $\mathrm{K}(X)$.*

Proof Hyperbolic reduction commutes with base change, so the question reduces to the case where the base is the spectrum of $\mathrm{K}(X)$, ie to the case of hyperbolic reduction of a quadric $Q_{\mathrm{K}(X)} \subset \mathbb{P}(E_{\mathrm{K}(X)})$ with respect to a linear subspace $F_{\mathrm{K}(X)} \subset E_{\mathrm{K}(X)}$. In this case q_- is the induced quadratic form on $F^{\perp}_{\mathrm{K}(X)}/F_{\mathrm{K}(X)}$ (the orthogonal is taken with respect to the quadratic form q). It is easy to see that the quadratic form q is isomorphic to the orthogonal sum $q_- \perp q_0$ of q_- with the hyperbolic form

$$q_0 = \begin{pmatrix} 0 & 1_{\dim(F)} \\ 1_{\dim(F)} & 0 \end{pmatrix},$$

hence $q = q_-$ in the Witt group $\boldsymbol{W}(\mathrm{K}(X))$. □

The following obvious lemma shows that hyperbolic reduction is transitive.

Lemma 2.3 *Let $(\mathcal{E}_-, q_-)$ be the hyperbolic reduction of $(\mathcal{E}, q)$ with respect to a regular isotropic subbundle $\mathcal{F} \hookrightarrow \mathcal{E}$ and let $(\mathcal{E}_{--}, q_{--})$ be the hyperbolic reduction of $(\mathcal{E}_-, q_-)$ with respect to a regular isotropic subbundle $\mathcal{F}_- \hookrightarrow \mathcal{E}_-$. Then $(\mathcal{E}_{--}, q_{--})$ is a hyperbolic reduction of $(\mathcal{E}, q)$.*

Proof Let $\widetilde{\mathcal{F}} \subset \mathcal{F}^{\perp}$ be the preimage of $\mathcal{F}_- \subset \mathcal{E}_-$ under the map $\mathcal{F}^{\perp} \twoheadrightarrow \mathcal{F}^{\perp}/\mathcal{F} = \mathcal{E}_-$, so that there is an exact sequence $0 \to \mathcal{F} \to \widetilde{\mathcal{F}} \to \mathcal{F}_- \to 0$ and an embedding $\widetilde{\mathcal{F}} \hookrightarrow \mathcal{E}$. Then $\widetilde{\mathcal{F}}$ is regular isotropic and the hyperbolic reduction of $(\mathcal{E}, q)$ with respect to $\widetilde{\mathcal{F}}$ is isomorphic to $(\mathcal{E}_{--}, q_{--})$. □

In the next subsection we will describe a construction inverse to hyperbolic reduction, and in the rest of this subsection we introduce the input data for that construction.

Assume $\mathcal{F} \subset \mathcal{E}$ is a regular isotropic subbundle with respect to a quadratic form q and let $(\mathcal{E}_-, q_-)$ be the hyperbolic reduction of $(\mathcal{E}, q)$ with respect to $\mathcal{F}$. Consider the length 3 filtration

$$0 \hookrightarrow \mathcal{F} \hookrightarrow \mathcal{F}^{\perp} \hookrightarrow \mathcal{E}. \tag{2-3}$$

Its associated graded is $\mathrm{gr}^{\bullet}(\mathcal{E}) = \mathcal{F} \oplus \mathcal{E}_- \oplus (\mathcal{F}^{\vee} \otimes \mathcal{L}^{\vee})$. In particular, we have two exact sequences

$$0 \to \mathcal{F} \to \mathcal{F}^{\perp} \to \mathcal{E}_- \to 0, \tag{2-4}$$

$$0 \to \mathcal{E}_- \to \mathcal{E}/\mathcal{F} \to \mathcal{F}^{\vee} \otimes \mathcal{L}^{\vee} \to 0. \tag{2-5}$$

The next lemma describes a relation between their extension classes.

Lemma 2.4 *Let $\varepsilon \in \mathrm{Ext}^1(\mathcal{F}^{\vee} \otimes \mathcal{L}^{\vee}, \mathcal{E}_-)$ be the extension class of (2-5). Then the extension class of (2-4) is equal to $q_-(\varepsilon)$, the Yoneda product of ε with the map $q_- \colon \mathcal{E}_- \to \mathcal{E}_-^{\vee} \otimes \mathcal{L}^{\vee}$, so that*

$$q_-(\varepsilon) \in \mathrm{Ext}^1(\mathcal{F}^{\vee} \otimes \mathcal{L}^{\vee}, \mathcal{E}_-^{\vee} \otimes \mathcal{L}^{\vee}) \cong \mathrm{Ext}^1(\mathcal{E}_-, \mathcal{F}).$$

Moreover, the Yoneda product $q_-(\varepsilon, \varepsilon) := q_-(\varepsilon) \circ \varepsilon \in \mathrm{Ext}^2(\mathcal{F}^{\vee} \otimes \mathcal{L}^{\vee}, \mathcal{F})$ vanishes.

Proof Tensoring diagram (2-2) by $\mathcal{L}^\vee$ and taking quotients by $\mathcal{F}$ we obtain a morphism of exact sequences

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{E}_- \arrow[r] \arrow[d, "q_-"'] & \mathcal{E}/\mathcal{F} \arrow[r] \arrow[d, "q"'] & \mathcal{F}^\vee \otimes \mathcal{L}^\vee \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathcal{E}_-^\vee \otimes \mathcal{L}^\vee \arrow[r] & (\mathcal{E}^\vee \otimes \mathcal{L}^\vee)/\mathcal{F} \arrow[r] & \mathcal{F}^\vee \otimes \mathcal{L}^\vee \arrow[r] & 0
\end{tikzcd}$$

This is a pushout diagram and the extension class of the top row is ε, hence the extension class of the bottom row is $q_-(\varepsilon)$. It remains to note that the bottom row is the twisted dual of (2-4).

Since the sequences (2-4) and (2-5) come from a length 3 filtration of $\mathcal{E}$, the Yoneda product of their extension classes vanishes. $\square$

We axiomatize the property of the class ε observed in Lemma 2.4 as follows; recall that for $s \in \mathbb{Z}$ we denote by $[s]$ the shift by s in the derived category.

Definition 2.5 Let (1-1) be a self-dual morphism, let $\mathcal{G}$ be a vector bundle on X, and let $\varepsilon \in \operatorname{Ext}^1(\mathcal{G}, \mathcal{E})$ be an extension class. We define the classes $q(\varepsilon) \in \operatorname{Ext}^1(\mathcal{E}, \mathcal{G}^\vee \otimes \mathcal{L}^\vee)$ and $q(\varepsilon, \varepsilon) \in \operatorname{Ext}^2(\mathcal{G}, \mathcal{G}^\vee \otimes \mathcal{L}^\vee)$ as the Yoneda products

$$q(\varepsilon)\colon \mathcal{E} \xrightarrow{q} \mathcal{E}^\vee \otimes \mathcal{L}^\vee \xrightarrow{\varepsilon} \mathcal{G}^\vee \otimes \mathcal{L}^\vee[1] \quad \text{and} \quad q(\varepsilon,\varepsilon)\colon \mathcal{G} \xrightarrow{\varepsilon} \mathcal{E}[1] \xrightarrow{q(\varepsilon)} \mathcal{G}^\vee \otimes \mathcal{L}^\vee[2].$$

We say that ε is *q–isotropic* if $q(\varepsilon, \varepsilon) = 0$.

Using this terminology we can reformulate Lemma 2.4 by saying that the class of (2-5) is q_-–isotropic.

Remark 2.6 It is easy to see that $q(\varepsilon, \varepsilon) \in \operatorname{Ext}^2\left(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\right) \subset \operatorname{Ext}^2(\mathcal{G} \otimes \mathcal{G}, \mathcal{L}^\vee) = \operatorname{Ext}^2(\mathcal{G}, \mathcal{G}^\vee \otimes \mathcal{L}^\vee)$. Indeed, the morphism $q(\varepsilon, \varepsilon) = \varepsilon \circ q \circ \varepsilon$ is symmetric because q is, hence it defines a morphism $\operatorname{Sym}^2(\mathcal{G}[-1]) \to \mathcal{L}^\vee$, and it remains to note that $\operatorname{Sym}^2(\mathcal{G}[-1]) \cong \bigwedge^2 \mathcal{G}[-2]$.

2.2 Hyperbolic extension

The following definition is central for this section.

Definition 2.7 Given a self-dual morphism (1-1) and a q–isotropic extension class $\varepsilon \in \operatorname{Ext}^1(\mathcal{G}, \mathcal{E})$ we say that $(\mathcal{E}_+, q_+)$ is a *hyperbolic extension of* $(\mathcal{E}, q)$ *with respect to* ε if there is a regular isotropic embedding $\mathcal{L}^\vee \otimes \mathcal{G}^\vee \hookrightarrow \mathcal{E}_+$ such that the hyperbolic reduction of $(\mathcal{E}_+, q_+)$ with respect to $\mathcal{L}^\vee \otimes \mathcal{G}^\vee$ is isomorphic to $(\mathcal{E}, q)$ and the induced extension $0 \to \mathcal{E} \to \mathcal{E}_+/(\mathcal{L}^\vee \otimes \mathcal{G}^\vee) \to \mathcal{G} \to 0$ has class ε.

We denote by $\operatorname{HE}(\mathcal{E}, q, \varepsilon)$ the set of isomorphism classes of all hyperbolic extensions of $(\mathcal{E}, q)$ with respect to a q–isotropic extension class ε. The main goal of this section is to show that $\operatorname{HE}(\mathcal{E}, q, \varepsilon)$ is nonempty; we will moreover see that this set may be quite big.

We start, however, with a simpler case, where the set $\operatorname{HE}(\mathcal{E}, q, \varepsilon)$ consists of a single element.

Proposition 2.8 *Let (1-1) be a self-dual morphism of vector bundles. If $\mathcal{G}$ is a line bundle, then for any extension class $\varepsilon \in \operatorname{Ext}^1(\mathcal{G}, \mathcal{E})$ there exists a unique (up to isomorphism) hyperbolic extension of $(\mathcal{E}, q)$ with respect to ε.*

Proof We start by proving the existence of a hyperbolic extension. The construction described below is an algebraic version of the geometric construction sketched in the introduction.

Let

$$0 \to \mathcal{E} \to \mathcal{E}' \to \mathcal{G} \to 0 \tag{2-6}$$

be an extension of class ε and consider its symmetric square $0 \to \operatorname{Sym}^2 \mathcal{E} \to \operatorname{Sym}^2 \mathcal{E}' \to \mathcal{E}' \otimes \mathcal{G} \to 0$, its tensor product with $\mathcal{G}^\vee$, and its pushout along the map $\operatorname{Sym}^2 \mathcal{E} \otimes \mathcal{G}^\vee \xrightarrow{q} \mathcal{L}^\vee \otimes \mathcal{G}^\vee$,

$$\begin{tikzcd} 0 \arrow[r] & \operatorname{Sym}^2 \mathcal{E} \otimes \mathcal{G}^\vee \arrow[r] \arrow[d, "q"'] & \operatorname{Sym}^2 \mathcal{E}' \otimes \mathcal{G}^\vee \arrow[r] \arrow[d, "\phi"'] & \mathcal{E}' \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & \mathcal{L}^\vee \otimes \mathcal{G}^\vee \arrow[r] & \mathcal{E}_+ \arrow[r] & \mathcal{E}' \arrow[r] & 0 \end{tikzcd} \tag{2-7}$$

defining a vector bundle $\mathcal{E}_+$ and a morphism ϕ. We will show that $\mathcal{E}_+$ comes with a natural quadratic form q_+ such that the embedding $\mathcal{L}^\vee \otimes \mathcal{G}^\vee \hookrightarrow \mathcal{E}_+$ in the bottom row of (2-7) is regular isotropic and the corresponding hyperbolic reduction is isomorphic to $(\mathcal{E}, q)$. For this we consider a component of the symmetric square of ϕ:

$$\operatorname{Sym}^2(\phi) \colon \operatorname{Sym}^4 \mathcal{E}' \otimes (\mathcal{G}^\vee)^{\otimes 2} \to \operatorname{Sym}^2 \mathcal{E}_+. \tag{2-8}$$

We will show that its cokernel is canonically isomorphic to $\mathcal{L}^\vee$, and we will take the cokernel morphism $\operatorname{Sym}^2 \mathcal{E}_+ \to \mathcal{L}^\vee$ as the definition of the quadratic form q_+.

Indeed, considering (2-6) as a length 2 filtration on $\mathcal{E}'$ and taking its fourth symmetric power we obtain a length 5 filtration on $\operatorname{Sym}^4 \mathcal{E}' \otimes (\mathcal{G}^\vee)^{\otimes 2}$ with factors

$$\operatorname{Sym}^4 \mathcal{E} \otimes (\mathcal{G}^\vee)^{\otimes 2}, \quad \operatorname{Sym}^3 \mathcal{E} \otimes \mathcal{G}^\vee, \quad \operatorname{Sym}^2 \mathcal{E}, \quad \mathcal{E} \otimes \mathcal{G}, \quad \mathcal{G}^{\otimes 2}. \tag{2-9}$$

Similarly, the combination of the bottom row of (2-7) with (2-6) provides $\mathcal{E}_+$ with a length 3 filtration, which induces a length 5 filtration on $\operatorname{Sym}^2 \mathcal{E}_+$ with factors

$$(\mathcal{L}^\vee)^{\otimes 2} \otimes (\mathcal{G}^\vee)^{\otimes 2}, \quad \mathcal{E} \otimes \mathcal{L}^\vee \otimes \mathcal{G}^\vee, \quad \mathcal{L}^\vee \oplus \operatorname{Sym}^2 \mathcal{E}, \quad \mathcal{E} \otimes \mathcal{G}, \quad \mathcal{G}^{\otimes 2}. \tag{2-10}$$

It is easy to check that the morphism (2-8) is compatible with the filtrations, induces isomorphisms of the last two factors, epimorphisms on the first two factors, and the morphism

$$\operatorname{Sym}^2 \mathcal{E} \xrightarrow{(q,\mathrm{id})} \mathcal{L}^\vee \oplus \operatorname{Sym}^2 \mathcal{E}$$

on the middle factors. Therefore, the cokernel of (2-8) is canonically isomorphic to $\operatorname{Coker}(q, \mathrm{id}) \cong \mathcal{L}^\vee$. This induces a canonical morphism $q_+ \colon \operatorname{Sym}^2 \mathcal{E}_+ \to \mathcal{L}^\vee$, which vanishes on the first two factors of (2-10) and restricts to the morphism $(-\mathrm{id}, q)$ on the middle factor.

Since the morphism q_+ vanishes on the first factor $(\mathcal{L}^\vee)^{\otimes 2} \otimes (\mathcal{G}^\vee)^{\otimes 2} \cong (\mathcal{G}^\vee \otimes \mathcal{L}^\vee)^{\otimes 2}$ of (2-10), the subbundle $\mathcal{G}^\vee \otimes \mathcal{L}^\vee \hookrightarrow \mathcal{E}_+$ is q_+–isotropic. Similarly, since the morphism q_+ vanishes on the second factor of (2-10) and nowhere vanishes on the summand $\mathcal{L}^\vee \cong (\mathcal{G}^\vee \otimes \mathcal{L}^\vee) \otimes \mathcal{G}$ of the third factor, the subbundle $\mathcal{G}^\vee \otimes \mathcal{L}^\vee \hookrightarrow \mathcal{E}_+$ is regular isotropic, the underlying vector bundle of the hyperbolic reduction of $(\mathcal{E}_+, q_+)$ is isomorphic to $\mathcal{E}$, and the induced extension of $\mathcal{G}$ by $\mathcal{E}$ coincides with (2-6). Finally, since the restriction of q_+ to the summand $\operatorname{Sym}^2 \mathcal{E}$ of the middle factor of (2-10) equals q, the induced quadratic form on $\mathcal{E}$ is equal to q. Thus, $(\mathcal{E}_+, q_+)$ is a hyperbolic extension of $(\mathcal{E}, q)$ with respect to ε.

Now we prove that the constructed hyperbolic extension is unique. For this it is enough to show that for any hyperbolic extension $(\mathcal{E}_+, q_+)$ of $(\mathcal{E}, q)$ with respect to ε there is a diagram (2-7) such that q_+ is the cokernel of $\operatorname{Sym}^2(\phi)$.

First, consider the morphism

$$\phi_+ \colon \operatorname{Sym}^2 \mathcal{E}_+ \otimes \mathcal{G}^\vee \to \mathcal{E}_+, \quad e_1 e_2 \otimes f \mapsto q_+(e_1, f) e_2 + q_+(e_2, f) e_1 - q_+(e_1, e_2) f,$$

where e_i are sections of $\mathcal{E}_+$ and f is a section of $\mathcal{G}^\vee$ that we consider as a subbundle in $\mathcal{E}_+ \otimes \mathcal{L}$. The symmetric square of the exact sequence $0 \to \mathcal{L}^\vee \otimes \mathcal{G}^\vee \to \mathcal{E}_+ \to \mathcal{E}' \to 0$ tensored with $\mathcal{G}^\vee$ takes the form

$$0 \to \mathcal{E}_+ \otimes \mathcal{L}^\vee \otimes \mathcal{G}^\vee \otimes \mathcal{G}^\vee \to \operatorname{Sym}^2 \mathcal{E}_+ \otimes \mathcal{G}^\vee \to \operatorname{Sym}^2 \mathcal{E}' \otimes \mathcal{G}^\vee \to 0,$$

where the first map takes $e \otimes f_1 \otimes f_2$ to $e f_1 \otimes f_2$. The composition of this map with ϕ_+ acts as

$$e \otimes f_1 \otimes f_2 \mapsto \phi_+(e f_1 \otimes f_2) = q_+(e, f_2) f_1 + q_+(f_1, f_2) e - q_+(e, f_1) f_2.$$

The second summand is zero because $\mathcal{L}^\vee \otimes \mathcal{G}^\vee \subset \mathcal{E}_+$ is isotropic and the first summand cancels with the last because the rank of $\mathcal{G}$ is 1, hence f_1 and f_2 are proportional. Therefore, the map ϕ_+ factors through a map $\phi \colon \operatorname{Sym}^2 \mathcal{E}' \otimes \mathcal{G}^\vee \to \mathcal{E}_+$. Moreover, it is easy to see that this map fits into the diagram (2-7). Finally, it is straightforward (but tedious) to check that the composition

$$\operatorname{Sym}^4 \mathcal{E}' \otimes (\mathcal{G}^\vee)^{\otimes 2} \xrightarrow{\operatorname{Sym}^2(\phi)} \operatorname{Sym}^2 \mathcal{E}_+ \xrightarrow{q_+} \mathcal{L}^\vee$$

vanishes, and since q_+ is a hyperbolic extension of q, it vanishes on the first two factors of (2-10) and induces the morphism $\mathcal{L}^\vee \oplus \operatorname{Sym}^2 \mathcal{E} \to \mathcal{L}^\vee$ of the third factor, which is equal to q on $\operatorname{Sym}^2 \mathcal{E}$, hence equal to $(-\mathrm{id}, q)$ on this third factor, and thus coincides with the canonical cokernel of $\operatorname{Sym}^2(\phi)$. □

Note that the general case (where the rank of $\mathcal{G}$ is greater than 1) does not immediately reduce to a rank 1 case, because a general vector bundle does not admit a filtration by line bundles. Besides, even if such a filtration exists, it is hard to trace what happens with the obstructions and to see how the nontrivial space of extensions shows up. So, in the proof of the theorem below, we use the projective bundle trick.

Theorem 2.9 *For any self-dual morphism (1-1) and a q–isotropic extension class $\varepsilon \in \operatorname{Ext}^1(\mathcal{G}, \mathcal{E})$, the set $\operatorname{HE}(\mathcal{E}, q, \varepsilon)$ of hyperbolic extensions of $(\mathcal{E}, q)$ with respect to ε is nonempty and is a principal homogeneous variety under an action of the group $\operatorname{Ext}^1\big(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\big)$.*

The action of the group $\operatorname{Ext}^1\big(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\big)$ on the set $\operatorname{HE}(\mathcal{E}, q, \varepsilon)$ will be constructed in course of the proof.

Proof Consider the projectivization $\pi : \mathbb{P}_X(\mathcal{G}) \to X$ and the tautological line subbundle $\mathcal{O}(-1) \hookrightarrow \pi^*\mathcal{G}$. Note that the quotient bundle $\pi^*\mathcal{G}/\mathcal{O}(-1)$ can be identified with $\mathcal{T}_\pi(-1)$, where $\mathcal{T}_\pi$ is the relative tangent bundle for the morphism π. We denote by $\gamma \in \operatorname{Ext}^1(\mathcal{T}_\pi(-1), \mathcal{O}(-1))$ the extension class of the tautological sequence

$$0 \to \mathcal{O}(-1) \to \pi^*\mathcal{G} \to \mathcal{T}_\pi(-1) \to 0. \tag{2-11}$$

Pulling back the class $\pi^*\varepsilon \in \operatorname{Ext}^1(\pi^*\mathcal{G}, \pi^*\mathcal{E})$ along the embedding $\mathcal{O}(-1) \hookrightarrow \pi^*\mathcal{G}$ we obtain the extension

$$0 \to \pi^*\mathcal{E} \to \widetilde{\mathcal{E}}' \to \mathcal{O}(-1) \to 0 \tag{2-12}$$

on $\mathbb{P}_X(\mathcal{G})$; we denote its extension class by $\tilde{\varepsilon} \in \operatorname{Ext}^1(\mathcal{O}(-1), \pi^*\mathcal{E})$.

By Proposition 2.8 there is a unique hyperbolic extension of $(\pi^*\mathcal{E}, \pi^*q)$ with respect to $\tilde{\varepsilon}$, which is given by an extension of vector bundles

$$0 \to \pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1) \to \widetilde{\mathcal{E}}_+ \to \widetilde{\mathcal{E}}' \to 0$$

and a quadratic form $\tilde{q}_+ : \pi^*\mathcal{L} \to \operatorname{Sym}^2 \widetilde{\mathcal{E}}_+^\vee$. We denote the extension class of the above sequence by

$$\tilde{\varepsilon}' \in \operatorname{Ext}^1(\widetilde{\mathcal{E}}', \pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1)).$$

Note that by Lemma 2.4 the restriction of $\tilde{\varepsilon}'$ to $\pi^*\mathcal{E} \subset \widetilde{\mathcal{E}}'$ is $\pi^*q(\tilde{\varepsilon})$; in particular, $\widetilde{\mathcal{E}}_+$ has a length 3 filtration with

$$\operatorname{gr}^\bullet(\widetilde{\mathcal{E}}_+) = (\pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1)) \oplus \pi^*\mathcal{E} \oplus \mathcal{O}(-1),$$

and the extension classes linking its factors are $(\pi^*q)(\tilde{\varepsilon})$ and $\tilde{\varepsilon}$, respectively.

It would be natural at this point to consider a hyperbolic extension of $(\widetilde{\mathcal{E}}_+, \tilde{q}_+)$ by $\mathcal{T}_\pi(-1)$ (note that the rank of $\mathcal{T}_\pi(-1)$ is less than $\mathcal{G}$) and then show that the result descends to a self-dual morphism on X. However, it turns out to be more convenient to use a simpler construction by "adding" the (twisted) dual bundle $\pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)$ to the kernel space of $\tilde{q}$ and then applying another version of descent.

Consider the product of extension classes (recall that γ is the extension class of (2-11))

$$\widetilde{\mathcal{E}}' \xrightarrow{\tilde{\varepsilon}'} \pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1)[1] \xrightarrow{\gamma} \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)[2],$$

where $\Omega_\pi = \mathcal{T}_\pi^\vee$ is the relative sheaf of Kähler differentials. We claim that $\gamma \circ \tilde{\varepsilon}' = 0$. Indeed, using (2-12) and taking into account isomorphisms $\boldsymbol{R}\pi_*(\Omega_\pi(1)) = 0$ and $\boldsymbol{R}\pi_*(\Omega_\pi(2)) \cong \bigwedge^2 \mathcal{G}^\vee$, we obtain

$$\operatorname{Ext}^p(\widetilde{\mathcal{E}}', \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) \cong \operatorname{Ext}^p(\mathcal{O}(-1), \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) \cong \operatorname{Ext}^p\left(\textstyle\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\right),$$

for all $p \in \mathbb{Z}$, and note that under this isomorphism the product $\gamma \circ \tilde{\varepsilon}' \in \operatorname{Ext}^2(\widetilde{\mathcal{E}}', \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1))$ coincides with the obstruction class $q(\varepsilon, \varepsilon) \in \operatorname{Ext}^2\left(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\right)$, and hence vanishes as ε is assumed to be q–isotropic.

Consider the tensor product of the dual sequence of (2-11) with $\pi^*\mathcal{L}^\vee$:

$$0 \to \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1) \to \pi^*\mathcal{L}^\vee \otimes \pi^*\mathcal{G}^\vee \to \pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1) \to 0,$$

its extension class is also γ. The vanishing of the product $\gamma \circ \tilde{\varepsilon}'$ implies that the class $\tilde{\varepsilon}'$ lifts to a class in $\operatorname{Ext}^1(\widetilde{\mathcal{E}}', \pi^*\mathcal{L}^\vee \otimes \pi^*\mathcal{G}^\vee)$, or, equivalently, that the class $\gamma \in \operatorname{Ext}^1(\pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1), \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1))$ lifts to a class in $\mathrm{Ext}^1(\widetilde{\mathcal{E}}_+, \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1))$. Moreover, $\mathrm{Hom}(\pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1), \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) = 0$, hence we have an exact sequence

$$0 \to \mathrm{Ext}^1(\widetilde{\mathcal{E}}', \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) \to \mathrm{Ext}^1(\widetilde{\mathcal{E}}_+, \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) \to \mathrm{Ext}^1(\pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1), \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)),$$

which shows that such a lift of γ is unique up to the natural free action of the group

$$\mathrm{Ext}^1(\widetilde{\mathcal{E}}', \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) \cong \mathrm{Ext}^1(\mathcal{O}(-1), \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) \cong \mathrm{Ext}^1\big(\textstyle\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\big).$$

In other words, the set of such lifts is a principal homogeneous space under an action of $\mathrm{Ext}^1\big(\bigwedge^2 \mathcal{G}, \mathcal{L}^\vee\big)$.

The lifted classes define a vector bundle $\widehat{\mathcal{E}}_+$ that fits into two exact sequences

$$0 \to \pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1) \to \widehat{\mathcal{E}}_+ \to \widetilde{\mathcal{E}}_+ \to 0 \quad \text{and} \quad 0 \to \pi^*\mathcal{L}^\vee \otimes \pi^*\mathcal{G}^\vee \to \widehat{\mathcal{E}}_+ \to \widetilde{\mathcal{E}}' \to 0. \tag{2-13}$$

We consider the quadratic form on $\widehat{\mathcal{E}}_+$ defined by the composition

$$\hat{q}_+ \colon \pi^*\mathcal{L} \xrightarrow{\tilde{q}_+} \mathrm{Sym}^2 \widetilde{\mathcal{E}}_+^\vee \hookrightarrow \mathrm{Sym}^2 \widehat{\mathcal{E}}_+^\vee,$$

where the latter embedding is induced by the surjection $\widehat{\mathcal{E}}_+ \twoheadrightarrow \widetilde{\mathcal{E}}_+$ from (2-13). By construction $\widehat{\mathcal{E}}_+$ has a length 4 filtration with

$$\mathrm{gr}^\bullet(\widehat{\mathcal{E}}_+) = (\pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)) \oplus (\pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1)) \oplus \pi^*\mathcal{E} \oplus \mathcal{O}(-1),$$

and the extension classes linking its adjacent factors are γ, $(\pi^*q)(\tilde{\varepsilon})$ and $\tilde{\varepsilon}$, respectively. Furthermore, the subbundle $\pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1) \subset \widehat{\mathcal{E}}_+$ is contained in the kernel of the quadratic form $\hat{q}_+$. Now we explain how to descend the quadratic form $(\widehat{\mathcal{E}}_+, \hat{q}_+)$ over $\mathbb{P}_X(\mathcal{G})$ to a quadratic form $(\mathcal{E}_+, q_+)$ over X.

Consider the subbundle $\mathrm{Ker}(\widehat{\mathcal{E}}_+ \to \mathcal{O}(-1)) \subset \widehat{\mathcal{E}}_+$ generated by the first three factors of the filtration. Since the first two factors are linked by the class γ of the twisted dual of (2-11), this bundle is an extension of $\pi^*\mathcal{E}$ by $\pi^*(\mathcal{L}^\vee \otimes \mathcal{G}^\vee)$. Since the functor π^* is fully faithful on the derived category of coherent sheaves, its extension class is a pullback, hence there exists a vector bundle $\mathcal{E}''$ on X and exact sequences

$$0 \to \pi^*\mathcal{L}^\vee \otimes \pi^*\mathcal{G}^\vee \to \pi^*\mathcal{E}'' \to \pi^*\mathcal{E} \to 0, \tag{2-14}$$

$$0 \to \pi^*\mathcal{E}'' \to \widehat{\mathcal{E}}_+ \to \mathcal{O}(-1) \to 0. \tag{2-15}$$

Since $\mathrm{Ext}^1(\mathcal{O}(-1), \pi^*\mathcal{E}'') \cong \mathrm{Ext}^1(\mathcal{G}, \mathcal{E}'')$, there is an extension $0 \to \mathcal{E}'' \to \mathcal{E}_+ \to \mathcal{G} \to 0$ on X and a pullback diagram

$$\begin{tikzcd}
0 \arrow[r] & \pi^*\mathcal{E}'' \arrow[r] \arrow[d, equal] & \widehat{\mathcal{E}}_+ \arrow[r] \arrow[d] & \mathcal{O}(-1) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \pi^*\mathcal{E}'' \arrow[r] & \pi^*\mathcal{E}_+ \arrow[r] & \pi^*\mathcal{G} \arrow[r] & 0
\end{tikzcd} \tag{2-16}$$

where the right vertical arrow is the tautological embedding. The embedding of bundles $\widehat{\mathcal{E}}_+ \hookrightarrow \pi^*\mathcal{E}_+$ in the middle column is identical on the subbundle $\pi^*\mathcal{E}''$; hence, the induced morphism

$$\rho \colon \mathbb{P}_{\mathbb{P}_X(\mathcal{G})}(\widehat{\mathcal{E}}_+) \to \mathbb{P}_X(\mathcal{E}_+)$$

is the blowup with center $\mathbb{P}_X(\mathcal{E}'') \subset \mathbb{P}_X(\mathcal{E}_+)$, ie we have $\mathbb{P}_{\mathbb{P}_X(\mathcal{G})}(\widehat{\mathcal{E}}_+) \cong \mathrm{Bl}_{\mathbb{P}_X(\mathcal{E}'')}(\mathbb{P}_X(\mathcal{E}_+))$. Note that $\mathbb{P}_X(\mathcal{E}'') \subset \mathbb{P}_X(\mathcal{E}_+)$ is a locally complete intersection, hence

$$\boldsymbol{R}\rho_*\mathcal{O}_{\mathbb{P}_{\mathbb{P}_X(\mathcal{G})}(\widehat{\mathcal{E}}_+)} \cong \boldsymbol{R}\rho_*\mathcal{O}_{\mathrm{Bl}_{\mathbb{P}_X(\mathcal{E}'')}(\mathbb{P}_X(\mathcal{E}_+))} \cong \mathcal{O}_{\mathbb{P}_X(\mathcal{E}_+)},$$

and therefore the derived pullback functor ρ^* is fully faithful.

Let $\pi_+ \colon \mathbb{P}_X(\mathcal{E}_+) \to X$ and $\hat{\pi}_+ \colon \mathbb{P}_{\mathbb{P}_X(\mathcal{G})}(\widehat{\mathcal{E}}_+) \to X$ be the projections, so that $\hat{\pi}_+ = \pi_+ \circ \rho$, and we have a commutative diagram

$$\begin{tikzcd} & \mathbb{P}_{\mathbb{P}_X(\mathcal{G})}(\widehat{\mathcal{E}}_+) \arrow[dl] \arrow[dd, "\hat{\pi}_+"'] \arrow[dr, "\rho"] & \\ \mathbb{P}_X(\mathcal{G}) \arrow[dr, "\pi"'] & & \mathbb{P}_X(\mathcal{E}_+) \arrow[dl, "\pi_+"] \\ & X & \end{tikzcd}$$

Furthermore, let H_+ and $\widehat{H}_+$ be the relative hyperplane classes of $\mathbb{P}_X(\mathcal{E}_+)$ and $\mathbb{P}_{\mathbb{P}_X(\mathcal{G})}(\widehat{\mathcal{E}}_+)$, respectively, so that $\rho^*\mathcal{O}(H_+) \cong \mathcal{O}(\widehat{H}_+)$. Note that the quadratic form $\hat{q}_+$ can be represented by a section of the line bundle $\hat{\pi}_+^*\mathcal{L}^\vee \otimes \mathcal{O}(2\widehat{H}_+)$ on $\mathbb{P}_{\mathbb{P}_X(\mathcal{G})}(\widehat{\mathcal{E}}_+)$. Using full faithfulness of ρ^* we compute

$$\mathrm{Hom}(\hat{\pi}_+^*\mathcal{L}, \mathcal{O}(2\widehat{H}_+)) = \mathrm{Hom}(\rho^*\pi_+^*\mathcal{L}, \rho^*\mathcal{O}(2H_+)) \cong \mathrm{Hom}(\pi_+^*\mathcal{L}, \mathcal{O}(2H_+)).$$

Thus, $\hat{q}_+$ is (in a unique way) the pullback of a section q_+ of the line bundle $\pi_+^*\mathcal{L}^\vee \otimes \mathcal{O}(2H_+)$ on $\mathbb{P}_X(\mathcal{E}_+)$, ie $\hat{q}_+ = \rho^*(q_+)$. Furthermore, q_+ induces a morphism

$$q_+ \colon \mathcal{L} \to \mathrm{Sym}^2 \mathcal{E}_+^\vee$$

on X. It remains to show that $(\mathcal{E}, q)$ is a hyperbolic reduction of $(\mathcal{E}_+, q_+)$.

First, note that a combination of (2-14) and the second row of (2-16) shows that $\mathcal{E}_+$ has a filtration

$$0 \hookrightarrow \mathcal{L}^\vee \otimes \mathcal{G}^\vee \hookrightarrow \mathcal{E}'' \hookrightarrow \mathcal{E}_+$$

with factors $\mathcal{L}^\vee \otimes \mathcal{G}^\vee$, $\mathcal{E}$ and $\mathcal{G}$, respectively. In particular, there is an exact sequence

$$0 \to \mathcal{E} \to \mathcal{E}_+/(\mathcal{L}^\vee \otimes \mathcal{G}^\vee) \to \mathcal{G} \to 0,$$

and the diagram (2-16) implies that the sequence (2-12) is its pullback. Using the natural isomorphism $\mathrm{Ext}^1(\mathcal{O}(-1), \pi^*\mathcal{E}) \cong \mathrm{Ext}^1(\mathcal{G}, \mathcal{E})$ and the definition of (2-12) we conclude that the extension class of the above sequence is ε. So we only need to show that the subbundle $\mathcal{L}^\vee \otimes \mathcal{G}^\vee \hookrightarrow \mathcal{E}_+$ is regular isotropic and that the induced quadratic form on $\mathcal{E}$ coincides with q.

The first follows immediately from the fact that $\pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1) \subset \widehat{\mathcal{E}}_+$ is contained in the kernel of the quadratic form $\hat{q}_+$ (as was mentioned above) and that the subbundle $\pi^*\mathcal{L}^\vee \otimes \mathcal{O}(1) \subset \widetilde{\mathcal{E}}_+$ is isotropic for the quadratic form $\tilde{q}_+$ (because $(\widetilde{\mathcal{E}}_+, \tilde{q}_+)$ is a hyperbolic extension). Moreover, by the same reason the induced quadratic form on $\pi^*\mathcal{E}$ coincides with π^*q.

To finish the proof of the theorem we must check that any hyperbolic extension of $(\mathcal{E}, q)$ comes from the above construction. So, assume that $(\mathcal{E}_+, q_+)$ is a hyperbolic extension of $(\mathcal{E}, q)$ with respect to ε. Define the bundle $\widehat{\mathcal{E}}_+$ from the diagram (2-16), consider the blowup morphism ρ as above, and the pullback $\widehat{q}_+ = \rho^*(q_+)$ of the quadratic form q_+. It defines a quadratic form on $\widehat{\mathcal{E}}_+$ over $\mathbb{P}_X(\mathcal{G})$. It is easy to see that $\pi^*\mathcal{L}^\vee \otimes \Omega_\pi(1)$ is contained in the kernel of $\widehat{q}_+$ and that the quotient $(\widetilde{\mathcal{E}}_+, \widetilde{q}_+)$, where $\widetilde{\mathcal{E}}_+$ is defined by the first sequence in (2-13), is a hyperbolic extension of $\pi^*\mathcal{E}$ with respect to (2-12). Therefore, by the uniqueness result in Proposition 2.8, this quadratic form coincides with the one constructed in the proof and the rest of the construction shows that $(\mathcal{E}_+, q_+)$ coincides with one of the hyperbolic extensions of the theorem. □

The nontriviality of the construction of hyperbolic extension is demonstrated by the following.

Remark 2.10 If $q(\varepsilon) \in \operatorname{Ext}^1(\mathcal{G}, \mathcal{L}^\vee \otimes \mathcal{E}^\vee)$ is zero, then $\mathcal{E}_+$ is an extension of $\mathcal{G}$ by $(\mathcal{L}^\vee \otimes \mathcal{G}^\vee) \oplus \mathcal{E}$. The component of its extension class in $\operatorname{Ext}^1(\mathcal{G}, \mathcal{E})$ equals ε, and the component in $\operatorname{Ext}^1(\mathcal{G}, \mathcal{L}^\vee \otimes \mathcal{G}^\vee)$ is in general nontrivial; one can identify it with the Massey product $\mu(\varepsilon, q, \varepsilon)$.

The operation of hyperbolic extension is transitive in the following sense.

Lemma 2.11 *Let $(\mathcal{E}_+, q_+)$ be a hyperbolic extension of $(\mathcal{E}, q)$ with respect to a q–isotropic extension class $\varepsilon \in \operatorname{Ext}^1(\mathcal{G}, \mathcal{E})$, and let $(\mathcal{E}_{++}, q_{++})$ be a hyperbolic extension of $(\mathcal{E}_+, q_+)$ with respect to a q_+–isotropic extension class $\varepsilon_+ \in \operatorname{Ext}^1(\mathcal{G}_+, \mathcal{E}_+)$. Then $(\mathcal{E}_{++}, q_{++})$ is a hyperbolic extension of $(\mathcal{E}, q)$.*

Proof By definition, the hyperbolic reduction of $(\mathcal{E}_{++}, q_{++})$ with respect to $\mathcal{L}^\vee \otimes \mathcal{G}_+^\vee \hookrightarrow \mathcal{E}_{++}$ is $(\mathcal{E}_+, q_+)$ and the hyperbolic reduction of $(\mathcal{E}_+, q_+)$ with respect to $\mathcal{L}^\vee \otimes \mathcal{G}^\vee \hookrightarrow \mathcal{E}_+$ is $(\mathcal{E}, q)$. Therefore, by Lemma 2.3 we see that $(\mathcal{E}, q)$ is a hyperbolic reduction of $(\mathcal{E}_{++}, q_{++})$, hence by definition we conclude that $(\mathcal{E}_{++}, q_{++})$ is a hyperbolic extension of $(\mathcal{E}, q)$. □

2.3 Hyperbolic equivalence

We combine the notions of hyperbolic reduction and extension defined in the previous sections into the notion of hyperbolic equivalence.

Definition 2.12 We say that two quadratic forms $q_1 \colon \mathcal{E}_1 \otimes \mathcal{L} \to \mathcal{E}_1^\vee$ and $q_2 \colon \mathcal{E}_2 \otimes \mathcal{L} \to \mathcal{E}_2^\vee$ or two quadric bundles $Q_1 \to X$ and $Q_2 \to X$ are *hyperbolically equivalent* if they can be connected by a chain of hyperbolic reductions and hyperbolic extensions.

Since the operations of hyperbolic reduction and hyperbolic extension are mutually inverse by definition, this is an equivalence relation. In this subsection we discuss *hyperbolic invariants*, ie invariants of quadratic forms and quadric bundles with respect to hyperbolic equivalence.

Recall the invariants (1-4) and (1-6) with values in the (nonunimodular) Witt group $W_{\mathrm{nu}}(\mathsf{k})$ defined in the introduction. The hyperbolic invariance of (1-4) is obvious.

Lemma 2.13 *For any* k*–point* $x \in X$ *and a fixed trivialization of the fiber* $\mathcal{L}_x$ *of the line bundle* $\mathcal{L}$, *the class* $\mathrm{w}_x(\mathcal{E}.q) = [(\mathcal{E}_x, q_x)] \in \boldsymbol{W}_{\mathrm{nu}}(\mathsf{k})$ *is hyperbolic invariant. In particular, the parity of* $\mathrm{rk}(\mathcal{E})$ *is hyperbolic invariant.*

Proof This follows immediately from the fact that if $(\mathcal{E}_-, q_-)$ is the hyperbolic reduction of $(\mathcal{E}, q)$ with respect to a regular isotropic subbundle $\mathcal{F}$, then $(\mathcal{E}_{-,x}, q_{-,x})$ is the sublagrangian reduction of $(\mathcal{E}_x, q_x)$ with respect to the subspace $\mathcal{F}_x \subset \mathcal{E}_x$; see [4, Section 1.1.5].

Applying the rank parity homomorphism $\boldsymbol{W}_{\mathrm{nu}}(\mathsf{k}) \to \mathbb{Z}/2$ we deduce the invariance of the parity of $\mathrm{rk}(\mathcal{E})$ from that of $\mathrm{w}_x(\mathcal{E}, q)$; alternatively, this invariance can be seen directly from the construction. □

The hyperbolic invariance of (1-6) requires a bit more work.

Lemma 2.14 *If* X *is smooth and proper,* $\mathcal{L} \otimes \omega_X$ *is a square in* $\mathrm{Pic}(X)$, *and* $n = \dim(X)$ *is divisible by* 4, *then the class* $\mathrm{hw}(\mathcal{E}, q) \in \boldsymbol{W}_{\mathrm{nu}}(\mathsf{k})$ *is hyperbolic invariant. In particular, the parity of the rank of the form* $H^{n/2}(q)$ *defined by* (1-5) *is hyperbolic invariant.*

Proof Let $\mathcal{M}$ be a square root of $\mathcal{L} \otimes \omega_X$. By Serre duality we have

$$H^{n/2}(X, \mathcal{E} \otimes \mathcal{M})^\vee = H^{n/2}(X, \mathcal{E}^\vee \otimes \mathcal{M}^\vee \otimes \omega_X) \cong H^{n/2}(X, \mathcal{E}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}).$$

Therefore, the pairing (1-5) can be rewritten as the composition of the morphism

$$H^{n/2}(X, \mathcal{E} \otimes \mathcal{M}) \xrightarrow{q} H^{n/2}(X, \mathcal{E}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}) \tag{2-17}$$

and the Serre duality pairing.

Now assume that $\mathcal{F} \hookrightarrow \mathcal{E}$ is a regular isotropic subbundle and $(\mathcal{E}_-, q_-)$ is the hyperbolic reduction. It is enough to check that $\mathrm{hw}(\mathcal{E}, q) = \mathrm{hw}(\mathcal{E}_-, q_-)$. Note that $\mathcal{E}_- \otimes \mathcal{M}$ and $\mathcal{E}_-^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}$ by definition are the cohomology bundles (in the middle terms) of the complexes

$$\{\mathcal{F} \otimes \mathcal{M} \hookrightarrow \mathcal{E} \otimes \mathcal{M} \twoheadrightarrow \mathcal{F}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}\} \quad \text{and} \quad \{\mathcal{F} \otimes \mathcal{M} \hookrightarrow \mathcal{E}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M} \twoheadrightarrow \mathcal{F}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}\} \tag{2-18}$$

and the morphism $q_- \colon \mathcal{E}_- \otimes \mathcal{M} \to \mathcal{E}_-^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}$ is induced by the morphism of complexes

$$\begin{tikzcd}
\mathcal{F} \otimes \mathcal{M} \arrow[r] \arrow[d, "\mathrm{id}"] & \mathcal{E} \otimes \mathcal{M} \arrow[r] \arrow[d, "q"] & \mathcal{F}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M} \arrow[d, "\mathrm{id}"] \\
\mathcal{F} \otimes \mathcal{M} \arrow[r] & \mathcal{E}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M} \arrow[r] & \mathcal{F}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}
\end{tikzcd}$$

Therefore, the morphism of cohomology $H^{n/2}(X, \mathcal{E}_- \otimes \mathcal{M}) \xrightarrow{q} H^{n/2}(X, \mathcal{E}_-^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M})$ is computed by the morphism of the spectral sequences whose first pages look like

$$\boldsymbol{E}_1^{\bullet,\bullet}(\mathcal{E}_- \otimes \mathcal{M}) = \left\{ \begin{array}{l} H^{n/2+1}(X, \mathcal{F} \otimes \mathcal{M}) \longrightarrow H^{n/2+1}(X, \mathcal{E} \otimes \mathcal{M}) \longrightarrow H^{n/2+1}(X, \mathcal{F}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}) \\ H^{n/2}(X, \mathcal{F} \otimes \mathcal{M}) \longrightarrow H^{n/2}(X, \mathcal{E} \otimes \mathcal{M}) \longrightarrow H^{n/2}(X, \mathcal{F}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}) \\ H^{n/2-1}(X, \mathcal{F} \otimes \mathcal{M}) \longrightarrow H^{n/2-1}(X, \mathcal{E} \otimes \mathcal{M}) \longrightarrow H^{n/2-1}(X, \mathcal{F}^\vee \otimes \mathcal{L}^\vee \otimes \mathcal{M}) \end{array} \right\}$$

(dotted arrows show the directions of the only higher differentials $\boldsymbol{d}_2$), and

$$\boldsymbol{E}_1^{\bullet,\bullet}(\mathcal{E}_-^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M})$$
$$=\left\{\begin{array}{lll} H^{n/2+1}(X,\mathcal{F}\otimes\mathcal{M}) \longrightarrow & H^{n/2+1}(X,\mathcal{E}^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M}) \longrightarrow & H^{n/2+1}(X,\mathcal{F}^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M}) \\ H^{n/2}\quad(X,\mathcal{F}\otimes\mathcal{M}) \longrightarrow & H^{n/2}\quad(X,\mathcal{E}^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M}) \longrightarrow & H^{n/2}\quad(X,\mathcal{F}^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M}) \\ H^{n/2-1}(X,\mathcal{F}\otimes\mathcal{M}) \longrightarrow & H^{n/2-1}(X,\mathcal{E}^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M}) \longrightarrow & H^{n/2-1}(X,\mathcal{F}^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M}) \end{array}\right\}.$$

Moreover, the morphism of spectral sequences is equal to the identity on the first and last columns, and is induced by q on the middle column. On the other hand, by Serre duality

$$H^i(X,\mathcal{F}\otimes\mathcal{M})^\vee = H^{n-i}(X,\mathcal{F}^\vee\otimes\mathcal{M}^\vee\otimes\omega_X)\cong H^{n-i}(X,\mathcal{F}^\vee\otimes\mathcal{L}^\vee\otimes\mathcal{M}),$$

hence the morphism of spectral sequences is self-dual.

It follows that $(H^{n/2}(X,\mathcal{E}_-\otimes\mathcal{M}), H^{n/2}(q_-))$ is obtained from $(H^{n/2}(X,\mathcal{E}\otimes\mathcal{M}), H^{n/2}(q))$ by a composition of the hyperbolic reduction with respect to the regular isotropic subspace

$$\operatorname{Im}(\boldsymbol{E}_1^{-1,n/2}(\mathcal{E}_-\otimes\mathcal{M})\to\boldsymbol{E}_1^{0,n/2}(\mathcal{E}_-\otimes\mathcal{M})) = \operatorname{Im}(H^{n/2}(X,\mathcal{F}\otimes\mathcal{M})\to H^{n/2}(X,\mathcal{E}\otimes\mathcal{M}))$$

followed by a hyperbolic extension with respect to the space

$$\boldsymbol{E}_3^{1,n/2-1}(\mathcal{E}_-\otimes\mathcal{M}) = \operatorname{Coker}\big(\boldsymbol{E}_2^{-1,n/2}(\mathcal{E}_-\otimes\mathcal{M})\xrightarrow{\boldsymbol{d}_2}\boldsymbol{E}_2^{1,n/2-1}(\mathcal{E}_-\otimes\mathcal{M})\big).$$

Therefore, we have the required equality $\operatorname{hw}(\mathcal{E}_-,q_-)=\operatorname{hw}(\mathcal{E},q)$ in the Witt group $\boldsymbol{W}_{\mathrm{nu}}(\mathsf{k})$.

Applying the rank parity homomorphism $\boldsymbol{W}_{\mathrm{nu}}(\mathsf{k})\to\mathbb{Z}/2$ we deduce the invariance of the parity of the rank of $H^{n/2}(q)$ from that of $\operatorname{hw}(\mathcal{E},q)$. □

Other hyperbolic invariants of quadric bundles have been listed in Proposition 1.1. We are ready now to prove this proposition.

Proof of Proposition 1.1 Since assertion (0) is clear from the definition (or follows from Lemma 2.13), it is enough to prove assertions (1)–(5) of the proposition. Moreover, in most cases it is enough to prove the assertions for a single hyperbolic reduction. So assume that (1-1) is a self-dual morphism and $(\mathcal{E}_-,q_-)$ is its hyperbolic reduction with respect to a regular isotropic subbundle $\mathcal{F}\hookrightarrow\mathcal{E}$.

By Lemma 2.1 we have $\mathcal{C}(q)\cong\mathcal{C}(q_-)$, an isomorphism compatible with the shifted quadratic forms; this proves assertion (1). Furthermore, the equality of the discriminant divisors

$$D=\operatorname{supp}(\mathcal{C}(Q))=\operatorname{supp}(\mathcal{C}(Q'))=D'$$

follows as well, and proves (2). Similarly, (5) follows from Lemma 2.2 and Witt's cancellation theorem.

Now we prove (3). We refer to [11] for generalities about sheaves of Clifford algebras and modules. Here we just recall that for a vector bundle $\mathcal{E}$ with a quadratic form $q\colon\mathcal{L}\to\operatorname{Sym}^2\mathcal{E}^\vee$ we write

$$\operatorname{Cliff}_0(\mathcal{E},q)=\mathcal{O}\oplus\big(\textstyle\bigwedge^2\mathcal{E}\otimes\mathcal{L}\big)\oplus\big(\textstyle\bigwedge^4\mathcal{E}\otimes\mathcal{L}^{\otimes 2}\big)\oplus\cdots,$$
$$\operatorname{Cliff}_1(\mathcal{E},q)=\mathcal{E}\oplus\big(\textstyle\bigwedge^3\mathcal{E}\otimes\mathcal{L}\big)\oplus\big(\textstyle\bigwedge^5\mathcal{E}\otimes\mathcal{L}^{\otimes 2}\big)\oplus\cdots,$$

and set $\mathrm{Cliff}_{i+2}(\mathcal{E},q)=\mathcal{L}^\vee\otimes\mathrm{Cliff}_i(\mathcal{E},q)$. The Clifford multiplication (see [12, Section 3])

$$\mathrm{Cliff}_i(\mathcal{E},q)\otimes\mathrm{Cliff}_j(\mathcal{E},q)\to\mathrm{Cliff}_{i+j}(\mathcal{E},q),$$

induced by q and the wedge product on $\bigwedge^\bullet\mathcal{E}$, provides $\mathrm{Cliff}_0(\mathcal{E},q)$ with the structure of an $\mathcal{O}_X$–algebra (called the *sheaf of even parts of Clifford algebras*), and each $\mathrm{Cliff}_i(\mathcal{E},q)$ with the structure of $\mathrm{Cliff}_0(\mathcal{E},q)$–bimodule. In the case where the line bundle $\mathcal{L}$ is trivial, the sum

$$\mathrm{Cliff}(\mathcal{E},q)=\mathrm{Cliff}_0(\mathcal{E},q)\oplus\mathrm{Cliff}_1(\mathcal{E},q)$$

also acquires a structure of $\mathcal{O}_X$–algebra (called the *total Clifford algebra*), which is naturally $\mathbb{Z}/2$–graded.

Now consider the subbundle $\mathcal{F}^\perp\subset\mathcal{E}$ defined by (2-1). It comes with the quadratic form $q_{\mathcal{F}^\perp}$, the restriction of the form q, so that the subbundle $\mathcal{F}\subset\mathcal{F}^\perp$ is contained in the kernel of $q_{\mathcal{F}^\perp}$ and the induced quadratic form on the quotient $\mathcal{F}^\perp/\mathcal{F}=\mathcal{E}_-$ coincides with q_-. Thus, the maps $\mathcal{F}^\perp\hookrightarrow\mathcal{E}$ and $\mathcal{F}^\perp\twoheadrightarrow\mathcal{E}_-$ are morphisms of quadratic spaces. Therefore, they are compatible with the Clifford multiplications and induce $\mathcal{O}_X$–algebra morphisms of sheaves of even parts of Clifford algebras

$$\mathrm{Cliff}_0(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})\hookrightarrow\mathrm{Cliff}_0(\mathcal{E},q)\quad\text{and}\quad\mathrm{Cliff}_0(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})\twoheadrightarrow\mathrm{Cliff}_0(\mathcal{E}_-,q_-).$$

The kernel of the second morphism is the two-sided ideal

$$\mathcal{R}:=\mathrm{Im}(\mathcal{F}\otimes\mathrm{Cliff}_{-1}(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})\to\mathrm{Cliff}_0(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})),$$

where the arrow is the natural morphism induced by the embedding $\mathcal{F}\hookrightarrow\mathcal{F}^\perp\hookrightarrow\mathrm{Cliff}_1(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})$ and the Clifford multiplication.

Now we write $k=\mathrm{rk}(\mathcal{F})$ and consider the right ideal in $\mathrm{Cliff}_0(\mathcal{E},q)$ defined as

$$\mathcal{P}:=\mathrm{Im}\big(\textstyle\bigwedge^k\mathcal{F}\otimes\mathrm{Cliff}_{-k}(\mathcal{E},q)\to\mathrm{Cliff}_0(\mathcal{E},q)\big).$$

Since $\mathcal{F}^\perp$ is orthogonal to $\mathcal{F}$ with respect to q, the subalgebra $\mathrm{Cliff}_0(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})\subset\mathrm{Cliff}_0(\mathcal{E},q)$ anticommutes with $\bigwedge^k\mathcal{F}\subset\mathrm{Cliff}_k(\mathcal{E},q)$, hence $\mathcal{P}$ is invariant under the left action of $\mathrm{Cliff}_0(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})$ on $\mathrm{Cliff}_0(\mathcal{E},q)$. Furthermore, since $\mathcal{F}$ is isotropic, the Clifford multiplication vanishes on $\mathcal{F}\otimes\bigwedge^k\mathcal{F}$, hence the ideal $\mathcal{R}$ annihilates $\mathcal{P}$. Therefore, $\mathcal{P}$ has the structure of a left module over the algebra

$$\mathrm{Cliff}_0(\mathcal{F}^\perp,q_{\mathcal{F}^\perp})/\mathcal{R}\cong\mathrm{Cliff}_0(\mathcal{E}_-,q_-).$$

This structure obviously commutes with the right $\mathrm{Cliff}_0(\mathcal{E},q)$–module structure, hence $\mathcal{P}$ is naturally a $(\mathrm{Cliff}_0(\mathcal{E}_-,q_-),\mathrm{Cliff}_0(\mathcal{E},q))$–bimodule. We show below that $\mathcal{P}$ defines the required Morita equivalence.

The question now is local over X, so we may assume that $\mathcal{L}=\mathcal{O}_X$ and there is an orthogonal direct sum decomposition

$$\mathcal{E}=\mathcal{E}_-\oplus\mathcal{E}_0,\quad q=q_-\perp q_0,\tag{2-19}$$

where $\mathcal{E}_0=\mathcal{F}\oplus\mathcal{F}^\vee$ and the quadratic form q_0 is given by the natural pairing $\mathcal{F}\otimes\mathcal{F}^\vee\to\mathcal{O}_X$. Furthermore, as $\mathcal{L}=\mathcal{O}_X$, we can consider the total $\mathbb{Z}/2$–graded Clifford algebras.

On the one hand, the orthogonal direct sum decomposition (2-19) implies the natural isomorphism

$$\mathrm{Cliff}(\mathcal{E}, q) \cong \mathrm{Cliff}(\mathcal{E}_-, q_-) \otimes \mathrm{Cliff}(\mathcal{E}_0, q_0)$$

(where the right-hand side is the tensor product in the category of $\mathbb{Z}/2$–graded algebras), compatible with the gradings. On the other hand, since $\mathcal{F} \subset \mathcal{E}_0$ is Lagrangian, the algebra

$$\mathrm{Cliff}(\mathcal{E}_0, q_0) \cong \mathcal{E}nd\big(\textstyle\bigwedge^{\bullet} \mathcal{F}\big)$$

is Morita trivial, and its $\mathbb{Z}/2$–grading is induced by the natural $\mathbb{Z}/2$–grading of $\bigwedge^{\bullet} \mathcal{F}$. It follows that the $(\mathrm{Cliff}(\mathcal{E}_-, q_-), \mathrm{Cliff}(\mathcal{E}, q))$–bimodule

$$\widetilde{\mathcal{P}} = \mathrm{Cliff}(\mathcal{E}_-, q_-) \otimes \textstyle\bigwedge^{\bullet} \mathcal{F}$$

defines a Morita equivalence of $\mathrm{Cliff}(\mathcal{E}_-, q_-)$ and $\mathrm{Cliff}(\mathcal{E}, q)$, compatible with the grading. Therefore, the even part of $\widetilde{\mathcal{P}}$,

$$\widetilde{\mathcal{P}}_0 = \big(\mathrm{Cliff}_0(\mathcal{E}_-, q_-) \otimes \textstyle\bigwedge^{\mathrm{even}} \mathcal{F}\big) \oplus \big(\mathrm{Cliff}_1(\mathcal{E}_-, q_-) \otimes \textstyle\bigwedge^{\mathrm{odd}} \mathcal{F}\big),$$

defines a Morita equivalence between the even Clifford algebras $\mathrm{Cliff}_0(\mathcal{E}_-, q_-)$ and $\mathrm{Cliff}_0(\mathcal{E}, q)$. Finally, a simple computation shows that the globally defined bimodule $\mathcal{P}$ is locally isomorphic to the bimodule $\widetilde{\mathcal{P}}_0$, hence it defines a global Morita equivalence.

In conclusion we prove (4). To show that $[Q] = [Q']$ we will first show that for any point $x \in X$ there is a Zariski neighborhood $x \in U \subset X$ such that $Q_U \cong Q'_U$ (where $Q_U = Q \times_X U$ and $Q'_U = Q' \times_X U$), hence a fortiori $[Q_U] = [Q'_U]$, and after that we will use this local equality to deduce the global one.

Since we are going to work locally, we may assume that the line bundle $\mathcal{L}$ is trivial and the base is affine. Then two things happen with hyperbolic extension: first, any extension class $\varepsilon \in \mathrm{Ext}^1(\mathcal{G}, \mathcal{E})$ vanishes (in particular, any such class is q–isotropic), and second, the group $\mathrm{Ext}^1\big(\bigwedge^2 \mathcal{G}, \mathcal{E}\big)$ vanishes as well, so that the result of hyperbolic extension becomes unambiguous. Moreover, it is clear that this result becomes isomorphic to $\mathcal{E}_+ = \mathcal{E} \oplus (\mathcal{G} \oplus \mathcal{G}^\vee)$, the orthogonal direct sum of $\mathcal{E}$ and $\mathcal{G} \oplus \mathcal{G}^\vee$, with the quadratic form on $\mathcal{G} \oplus \mathcal{G}^\vee$ induced by duality. Similarly, hyperbolic reduction reduces to splitting off an orthogonal summand $\mathcal{F} \oplus \mathcal{F}^\vee$. Thus, locally, hyperbolic equivalence turns into Witt equivalence (in the nonunimodular Witt ring of the base scheme). Therefore, a hyperbolic equivalence between Q and Q' locally can be realized by a single quadric bundle $\widehat{Q}$ such that both Q and Q' are obtained from $\widehat{Q}$ by hyperbolic reduction. In other words, we may assume that the quadrics Q and Q' correspond to quadratic forms obtained from a single quadratic form $(\widehat{\mathcal{E}}, \widehat{q})$ by isotropic reduction with respect to regular isotropic subbundles $\mathcal{F} \subset \widehat{\mathcal{E}}$ and $\mathcal{F}' \subset \widehat{\mathcal{E}}$ of the same rank. Below we prove isomorphism of Q and Q' in a neighborhood of x by induction on the rank of $\mathcal{F}$ and $\mathcal{F}'$.

First assume that the rank of $\mathcal{F}$ and $\mathcal{F}'$ is 1 and $\widehat{q}(\mathcal{F}, \mathcal{F}') \neq 0$ at x (hence also in a neighborhood of x). Since $\mathcal{F}$ and $\mathcal{F}'$ are isotropic, the restriction of $\widehat{q}$ to $\mathcal{F} \oplus \mathcal{F}'$ is nondegenerate, hence there is an orthogonal direct sum decomposition

$$\widehat{\mathcal{E}} = \overline{\mathcal{E}} \oplus (\mathcal{F} \oplus \mathcal{F}').$$

Then obviously $\mathcal{F}^\perp = \overline{\mathcal{E}} \oplus \mathcal{F}$, hence the hyperbolic reduction of $(\widehat{\mathcal{E}}, \widehat{q})$ with respect to $\mathcal{F}$ is isomorphic to $(\overline{\mathcal{E}}, q|_{\overline{\mathcal{E}}})$. Similarly, the hyperbolic reduction of $(\widehat{\mathcal{E}}, \widehat{q})$ with respect to $\mathcal{F}'$ is isomorphic to $(\overline{\mathcal{E}}, q|_{\overline{\mathcal{E}}})$ as well. In particular, the two hyperbolic reductions are isomorphic.

On the other hand, assume that the rank of $\mathcal{F}$ and $\mathcal{F}'$ is 1 and $\widehat{q}(\mathcal{F}, \mathcal{F}')$ vanishes at x. Then we find (locally) yet another regular isotropic subbundle $\mathcal{F}'' \subset \widehat{\mathcal{E}}$ such that $\widehat{q}(\mathcal{F}, \mathcal{F}'') \neq 0$ and $\widehat{q}(\mathcal{F}', \mathcal{F}'') \neq 0$ at x. Let $v, v' \in \widehat{\mathcal{E}}_x$ be the points corresponding to $\mathcal{F}$ and $\mathcal{F}'$. Let $v'' \in \widehat{\mathcal{E}}_x$ be a point such that $\widehat{q}_x(v, v'') \neq 0$ and $\widehat{q}_x(v', v'') \neq 0$. The existence of a regular subbundle $\mathcal{F}$ implies rationality of $\widehat{Q}$ over X, hence (maybe over a smaller neighborhood of x) there exists a regular isotropic subbundle $\mathcal{F}''$ corresponding to the point v''. Now, when we have such $\mathcal{F}''$, we apply the previous argument and conclude that the hyperbolic reduction of $(\widehat{\mathcal{E}}, \widehat{q})$ with respect to $\mathcal{F}''$ is isomorphic to the hyperbolic reductions with respect to $\mathcal{F}$ and $\mathcal{F}'$; hence the latter two reductions are mutually isomorphic.

Now assume the rank of $\mathcal{F}$ and $\mathcal{F}'$ is bigger than 1. Shrinking the neighborhood of x if necessary, we may split $\mathcal{F} = \mathcal{F}_1 \oplus \mathcal{F}_2$ and $\mathcal{F}' = \mathcal{F}'_1 \oplus \mathcal{F}'_2$, where the rank of $\mathcal{F}_1$ and $\mathcal{F}'_1$ is 1. The above argument shows that the isotropic reductions of $(\widehat{\mathcal{E}}, \widehat{q})$ with respect to $\mathcal{F}_1$ and $\mathcal{F}'_1$ are isomorphic. Hence Q and Q' correspond to hyperbolic reductions of the same quadratic form with respect to regular isotropic subbundles $\mathcal{F}_2$ and $\mathcal{F}'_2$, which have smaller rank than $\mathcal{F}$ and $\mathcal{F}'$, and therefore by induction Q and Q' are isomorphic.

Finally, we deduce the global result from the local results obtained above. Indeed, the argument above and quasicompactness of X imply that X has a finite open covering $\{U_i\}$ such that over each U_i we have an isomorphism $Q_{U_i} \cong Q'_{U_i}$, hence an equality $[Q_{U_i}] = [Q'_{U_i}]$ in the Grothendieck ring of varieties. For any finite set I of indices set $U_I = \cap_{i \in I} U_i$. Then inclusion–exclusion gives

$$[Q] = \sum_{|I| \geq 1} (-1)^{|I|-1} [Q_{U_I}] \quad \text{and} \quad [Q'] = \sum_{|I| \geq 1} (-1)^{|I|-1} [Q'_{U_I}],$$

and since by base change we have isomorphisms $Q_{U_I} \cong Q'_{U_I}$, hence equalities $[Q_{U_I}] = [Q'_{U_I}]$ for each I with $|I| \geq 1$, the equality $[Q] = [Q']$ follows. □

Remark 2.15 The same technique proves the more general formula

$$[Q] = [Q'] \mathbb{L}^d + [X][\mathbb{P}^{d-1}](1 + \mathbb{L}^{n-d+1}) \tag{2-20}$$

for any hyperbolic equivalent quadric bundles Q/X and Q'/X, where $n = \dim(Q/X)$ and we assume that it is greater or equal than $\dim(Q'/X)$, which we write in the form $\dim(Q'/X) = n - 2d$. Indeed, first (2-20) can be proved over a small neighborhood of any point of X; for this the same argument reduces everything to the case where Q' is a hyperbolic reduction of Q, in which case the formula is proved in [13, Corollary 2.7]. After that the inclusion–exclusion trick proves (2-20) in general.

3 VHC resolutions on projective spaces

From now on we consider the case $X = \mathbb{P}^n$. This section serves as preparation for the next one. Here we introduce a class of locally free resolutions (which we call VHC resolutions), which plays the main role in Section 4, and we show that on $\mathbb{P}^n$ any sheaf of projective dimension 1 has a (essentially unique) VHC resolution; see Corollary 3.18 for existence and Theorem 3.15 for uniqueness.

3.1 Complexes of split bundles

For each coherent sheaf $\mathcal{F}$ on $\mathbb{P}^n = \mathbb{P}(V)$ (and more generally, for any object of the bounded derived category $\boldsymbol{D}(\mathbb{P}^n)$ of coherent sheaves) and each integer p, we write

$$\boldsymbol{H}^p_*(\mathcal{F}) := \bigoplus_{t=-\infty}^{\infty} H^p(\mathbb{P}^n, \mathcal{F}(t)). \tag{3-1}$$

This is a graded module over the homogeneous coordinate ring

$$\mathbb{S} = \boldsymbol{H}^0_*(\mathcal{O}_{\mathbb{P}^n}) \cong \operatorname{Sym}^\bullet(V^\vee) = \mathsf{k} \oplus V^\vee \oplus \operatorname{Sym}^2 V^\vee \oplus \cdots . \tag{3-2}$$

For a sheaf $\mathcal{F}$ we will often consider the $\mathbb{S}$–module of *intermediate cohomology*

$$\bigoplus_{p=1}^{n-1} \boldsymbol{H}^p_*(\mathcal{F}) = \bigoplus_{p=1}^{n-1} \Bigl(\bigoplus_{t=-\infty}^{\infty} H^p(\mathbb{P}^n, \mathcal{F}(t)) \Bigr)$$

as a bigraded $\mathbb{S}$–module; with index p corresponding to the *homological* and index t corresponding to the *internal* grading. We will use the notation $[p]$ and (t) for the corresponding shifts of grading.

Recall the following well-known result.

Lemma 3.1 *Let $\mathcal{F}$ be a coherent sheaf, so that the $\mathbb{S}$–module $\boldsymbol{H}^0_*(\mathcal{F})$ is finitely generated. For any epimorphism $\bigoplus \mathbb{S}(t_i) \to \boldsymbol{H}^0_*(\mathcal{F})$ of graded $\mathbb{S}$–modules there is an epimorphism $\bigoplus \mathcal{O}(t_i) \to \mathcal{F}$ such that the induced morphism $\bigoplus \mathbb{S}(t_i) = \boldsymbol{H}^0_*\bigl(\bigoplus \mathcal{O}(t_i)\bigr) \to \boldsymbol{H}^0_*(\mathcal{F})$ coincides with the original epimorphism.*

Proof Any morphism of graded $\mathbb{S}$–modules $\mathbb{S}(t) \to \boldsymbol{H}^0_*(\mathcal{F})$ is given by an element in the graded component $H^0(\mathbb{P}^n, \mathcal{F}(-t)) = \operatorname{Hom}(\mathcal{O}(t), \mathcal{F})$ of $\boldsymbol{H}^0_*(\mathcal{F})$, hence the epimorphism $\bigoplus \mathbb{S}(t_i) \to \boldsymbol{H}^0_*(\mathcal{F})$ corresponds to a morphism $\bigoplus \mathcal{O}(t_i) \to \mathcal{F}$. The only nontrivial statement here is the surjectivity of this morphism. To prove it let $\mathcal{K}$ and $\mathcal{C}$ denote its kernel and cokernel, so that we have an exact sequence

$$0 \to \mathcal{K} \to \bigoplus \mathcal{O}(t_i) \to \mathcal{F} \to \mathcal{C} \to 0.$$

We need to show $\mathcal{C} = 0$. When twisted by $\mathcal{O}(t)$ with $t \gg 0$ all sheaves above have no higher cohomology, therefore there is an exact sequence

$$0 \to H^0(\mathbb{P}^n, \mathcal{K}(t)) \to H^0\bigl(\mathbb{P}^n, \bigoplus \mathcal{O}(t_i + t)\bigr) \to H^0(\mathbb{P}^n, \mathcal{F}(t)) \to H^0(\mathbb{P}^n, \mathcal{C}(t)) \to 0.$$

By assumption the middle arrow is surjective, hence $H^0(\mathbb{P}^n, \mathcal{C}(t)) = 0$ for $t \gg 0$. Therefore, $\mathcal{C} = 0$. $\square$

We will say that $\mathcal{E}$ is a *split bundle* if it is isomorphic to a direct sum of line bundles; note that by Horrocks' theorem a vector bundle $\mathcal{E}$ is split if and only if $\boldsymbol{H}^p_*(\mathcal{E}) = 0$ for all $1 \le p \le n-1$.

We will need the following simple generalization of the Horrocks' theorem.

Lemma 3.2 *Let $\mathcal{E}$ be a vector bundle on $\mathbb{P}^n$ and let $0 \le \ell \le n-1$. Then $\boldsymbol{H}^p_*(\mathcal{E}) = 0$ for all p such that $1 \le p \le n-\ell-1$ if and only if $\mathcal{E}$ has a resolution of length ℓ*

$$0 \to \mathcal{L}_\ell \to \cdots \to \mathcal{L}_1 \to \mathcal{L}_0 \to \mathcal{E} \to 0$$

by split bundles.

Proof We use induction on ℓ. If $\ell = 0$ the result follows from the Horrocks' theorem. Assume $\ell > 0$. Choose an epimorphism $\mathcal{L}_0 \twoheadrightarrow \mathcal{E}$ from a split bundle $\mathcal{L}_0$ which is surjective on $\boldsymbol{H}^0_*$ (it exists by Lemma 3.1) and let $\mathcal{E}'$ be its kernel. The cohomology exact sequence

$$\cdots \to \boldsymbol{H}^{p-1}_*(\mathcal{L}_0) \to \boldsymbol{H}^{p-1}_*(\mathcal{E}) \to \boldsymbol{H}^p_*(\mathcal{E}') \to \boldsymbol{H}^p_*(\mathcal{L}_0) \to \boldsymbol{H}^p_*(\mathcal{E}) \to \cdots$$

implies that $\boldsymbol{H}^p_*(\mathcal{E}') = 0$ for $1 \le p \le n-\ell$. By the induction hypothesis $\mathcal{E}'$ has a resolution of length $\ell-1$

$$0 \to \mathcal{L}_\ell \to \cdots \to \mathcal{L}_1 \to \mathcal{E}' \to 0$$

by split bundles. It follows that the complex $\mathcal{L}_\ell \to \cdots \to \mathcal{L}_1 \to \mathcal{L}_0$ (where the morphism $\mathcal{L}_1 \to \mathcal{L}_0$ is defined as the composition $\mathcal{L}_1 \twoheadrightarrow \mathcal{E}' \hookrightarrow \mathcal{L}_0$) is a resolution of $\mathcal{E}$ of length ℓ by split bundles.

The converse statement follows immediately from the hypercohomology spectral sequence applied to the resolution since the intermediate cohomology of split bundles vanishes. □

The following obvious observation about complexes is quite useful.

Lemma 3.3 *Let $\mathcal{L}_\bullet$ be a complex of coherent sheaves such that $\mathcal{L}_i = \mathcal{O}(t) \oplus \mathcal{L}'_i$ and $\mathcal{L}_{i-1} = \mathcal{O}(t) \oplus \mathcal{L}'_{i-1}$ for some $i \in \mathbb{Z}$ and $t \in \mathbb{Z}$, and the differential $\mathrm{d}_i \colon \mathcal{L}_i \to \mathcal{L}_{i-1}$ of $\mathcal{L}_\bullet$ takes the summand $\mathcal{O}(t)$ of $\mathcal{L}_i$ isomorphically to the summand $\mathcal{O}(t)$ of $\mathcal{L}_{i-1}$. Then there is an isomorphism of complexes*

$$\mathcal{L}_\bullet \cong \mathcal{L}'_\bullet \oplus (\mathcal{O}(t) \xrightarrow{\mathrm{id}} \mathcal{O}(t))[i]. \tag{3-3}$$

Proof By assumption the differential d_i can be written in the form

$$\mathcal{O}(t) \oplus \mathcal{L}'_i \xrightarrow{\left(\begin{smallmatrix} 1 & f \\ 0 & \mathrm{d}'_i \end{smallmatrix}\right)} \mathcal{O}(t) \oplus \mathcal{L}'_{i-1}$$

for some $f \in \mathrm{Hom}(\mathcal{L}'_i, \mathcal{O}(t))$ and $\mathrm{d}'_i \in \mathrm{Hom}(\mathcal{L}'_i, \mathcal{L}'_{i-1})$. After the modification of the direct sum decomposition of $\mathcal{L}_i$ by the automorphism $\left(\begin{smallmatrix} 1 & f \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{End}(\mathcal{O}(t) \oplus \mathcal{L}'_i)$ this differential takes the form $\left(\begin{smallmatrix} 1 & 0 \\ 0 & \mathrm{d}'_i \end{smallmatrix}\right)$. Then the equalities $\mathrm{d}_i \circ \mathrm{d}_{i+1} = 0$ and $\mathrm{d}_{i-1} \circ \mathrm{d}_i = 0$ imply that

$$\mathrm{d}_{i+1} = \begin{pmatrix} 0 \\ \mathrm{d}'_{i+1} \end{pmatrix} \in \mathrm{Hom}(\mathcal{L}_{i+1}, \mathcal{O}(t) \oplus \mathcal{L}'_i) \quad \text{and} \quad \mathrm{d}_{i-1} = (0 \;\; \mathrm{d}'_{i-1}) \in \mathrm{Hom}(\mathcal{O}(t) \oplus \mathcal{L}'_{i-1}, \mathcal{L}'_{i-2}),$$

which implies (3-3), where $\mathcal{L}'_j = \mathcal{L}_j$ for $j \notin \{i, i-1\}$. □

Now let

$$\mathcal{L}_\bullet := \{\mathcal{L}_\ell \to \mathcal{L}_{\ell-1} \to \cdots \to \mathcal{L}_1 \to \mathcal{L}_0\}$$

be a complex of split bundles on $\mathbb{P}^n$. Since split bundles have no intermediate cohomology, the first page of the hypercohomology spectral sequence of $\mathcal{L}_\bullet$ has only two nontrivial rows:

$$\begin{tikzcd}
H^n(\mathcal{L}_\ell) \arrow[r] \arrow[ddddrr, dashed] & H^n(\mathcal{L}_{\ell-1}) \arrow[r] \arrow[ddddrr, dashed] & \cdots \arrow[r] & \cdots \arrow[r] & H^n(\mathcal{L}_{n+1}) \arrow[r] \arrow[ddddrr, dashed] & \cdots \arrow[r] & H^n(\mathcal{L}_0) \\
0 & 0 & \cdots & \cdots & 0 & \cdots & 0 \\
\vdots & & & & & & \vdots \\
0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
H^0(\mathcal{L}_\ell) \arrow[r] & \cdots \arrow[r] & H^0(\mathcal{L}_{\ell-n-1}) \arrow[r] & H^0(\mathcal{L}_{\ell-n-2}) \arrow[r] & \cdots \arrow[r] & \cdots \arrow[r] & H^0(\mathcal{L}_0)
\end{tikzcd}$$

one formed by $H^0(\mathbb{P}^n, \mathcal{L}_i)$ and the other by $H^n(\mathbb{P}^n, \mathcal{L}_i)$. The dashed arrows show the only nontrivial higher differentials $\boldsymbol{d}_{n+1}$ — these differentials are directed n steps down and $n+1$ steps to the right. Therefore, if $\ell \le n$ there are no higher differentials, and if $\ell = n+1$ there is exactly one, which acts from $H^{\mathrm{top}}(\mathbb{P}^n, \mathcal{L}_\bullet)$ to $H^{\mathrm{bot}}(\mathbb{P}^n, \mathcal{L}_\bullet)$, where we define

$$\begin{aligned} H^{\mathrm{top}}(\mathbb{P}^n, \mathcal{L}_\bullet) &:= \operatorname{Ker}(H^n(\mathbb{P}^n, \mathcal{L}_\ell) \to H^n(\mathbb{P}^n, \mathcal{L}_{\ell-1})), \\ H^{\mathrm{bot}}(\mathbb{P}^n, \mathcal{L}_\bullet) &:= \operatorname{Coker}(H^0(\mathbb{P}^n, \mathcal{L}_1) \to H^0(\mathbb{P}^n, \mathcal{L}_0)). \end{aligned} \tag{3-4}$$

We also set $\boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}_\bullet) = \bigoplus_t H^{\mathrm{top}}(\mathbb{P}^n, \mathcal{L}_\bullet(t))$ and $\boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet) = \bigoplus_t H^{\mathrm{bot}}(\mathbb{P}^n, \mathcal{L}_\bullet(t))$.

Lemma 3.4 *If a complex $\mathcal{L}_\bullet$ of split bundles quasiisomorphic to an object $\mathcal{F}$ of the derived category $\boldsymbol{D}(\mathbb{P}^n)$ has length $\ell = n$, then there is a canonical exact sequence*

$$0 \to \boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet) \to \boldsymbol{H}^0_*(\mathcal{F}) \to \boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}_\bullet) \to 0.$$

Proof This follows immediately from the hypercohomology spectral sequence. □

The next two lemmas are crucial for the rest of the paper.

Lemma 3.5 *If an acyclic complex $\mathcal{L}_\bullet$ of split bundles has length $\ell = n+1$, then the following conditions are equivalent*:

(1) $\boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet) = 0$.

(2) $\boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}_\bullet) = 0$.

(3) *The canonical morphism $\boldsymbol{d}_{n+1} \colon \boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}_\bullet) \to \boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet)$ is zero.*

(4) *The complex $\mathcal{L}_\bullet$ is isomorphic to a direct sum of shifts of trivial complexes $\mathcal{O}(t) \xrightarrow{\mathrm{id}} \mathcal{O}(t)$.*

Proof Since $\mathcal{L}_\bullet$ is acyclic, its hypercohomology spectral sequence converges to zero, hence the canonical morphism $\boldsymbol{d}_{n+1} \colon \boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}_\bullet) \to \boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet)$ is an isomorphism. It follows that (1), (2), and (3) are equivalent.

Now we prove (3) $\Longrightarrow$ (4). So, assume (3) holds. Then for each t the hypercohomology spectral sequence of $\mathcal{L}_\bullet(-t)$ degenerates on the second page; in particular the bottom row of the first page is exact. Let t be the maximal integer such that $\mathcal{O}(t)$ appears as one of summands of one of the split bundles $\mathcal{L}_i$. Then the bottom row of the first page of the hypercohomology spectral sequence of $\mathcal{L}_\bullet(-t)$ is nonzero, and takes the form

$$H^0(\mathbb{P}^n, \mathcal{L}_\bullet(-t)) = \{\mathsf{k}^{m_\ell} \to \mathsf{k}^{m_{\ell-1}} \to \cdots \to \mathsf{k}^{m_1} \to \mathsf{k}^{m_0}\},$$

where m_i is the multiplicity of $\mathcal{O}(t)$ in $\mathcal{L}_i$. Since this complex is exact, it is a direct sum of shifts of trivial complexes $\mathsf{k} \xrightarrow{\mathrm{id}} \mathsf{k}$. Since $\mathrm{Hom}(\mathcal{O}(t), \mathcal{O}(t')) = 0$ for all $t' < t$, it follows that $\mathcal{L}_\bullet$ contains the subcomplex $H^0(\mathbb{P}^n, \mathcal{L}_\bullet(-t)) \otimes \mathcal{O}(t)$; this subcomplex is isomorphic to a direct sum of shifts of trivial complexes $\mathcal{O}(t) \xrightarrow{\mathrm{id}} \mathcal{O}(t)$, and each of its terms is a direct summand of the corresponding term of $\mathcal{L}_\bullet$. Applying Lemma 3.3 to one of these trivial subcomplexes we obtain the direct sum decomposition (3-3). The condition (3) holds for $\mathcal{L}'_\bullet$ (because it is a direct summand of $\mathcal{L}_\bullet$), hence by induction $\mathcal{L}'_\bullet$ is the sum of shifts of trivial complexes, and hence the same is true for $\mathcal{L}_\bullet$; which means that (4) holds.

The implication (4) $\Longrightarrow$ (3) is evident. □

Lemma 3.6 *Assume objects $\mathcal{F}$ and $\mathcal{F}'$ in $\boldsymbol{D}(\mathbb{P}^n)$ are quasiisomorphic to complexes $\mathcal{L}_\bullet$ and $\mathcal{L}'_\bullet$ of split bundles of length ℓ. If $\ell < n$, then any morphism $\varphi\colon \mathcal{F} \to \mathcal{F}'$ is induced by a morphism of complexes*

$$\begin{tikzcd}
\mathcal{L}_\ell \arrow[r] \arrow[d, "\varphi_\ell"'] & \mathcal{L}_{\ell-1} \arrow[r] \arrow[d, "\varphi_{\ell-1}"'] & \cdots \arrow[r] & \mathcal{L}_1 \arrow[r] \arrow[d, "\varphi_1"'] & \mathcal{L}_0 \arrow[d, "\varphi_0"'] \\
\mathcal{L}'_\ell \arrow[r] & \mathcal{L}'_{\ell-1} \arrow[r] & \cdots \arrow[r] & \mathcal{L}'_1 \arrow[r] & \mathcal{L}'_0
\end{tikzcd}$$

If $\ell = n$, the same is true for a morphism $\varphi\colon \mathcal{F} \to \mathcal{F}'$ if and only if the composition

$$\boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet) \hookrightarrow \boldsymbol{H}^0_*(\mathcal{F}) \xrightarrow{\boldsymbol{H}^0_*(\varphi)} \boldsymbol{H}^0_*(\mathcal{F}') \twoheadrightarrow \boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}'_\bullet) \tag{3-5}$$

vanishes, where the first and last morphisms are defined in Lemma 3.4. Moreover, in both cases a morphism of complexes $\varphi_\bullet$ inducing a morphism φ as above is unique up to a homotopy $h_\bullet\colon \mathcal{L}_\bullet \to \mathcal{L}'_{\bullet+1}$.

Proof Obviously, the first page of the spectral sequence

$$\boldsymbol{E}_1^{p,q} = \bigoplus_i \mathrm{Ext}^q(\mathcal{L}_i, \mathcal{L}'_{i-p}) \Rightarrow \mathrm{Ext}^{p+q}(\mathcal{F}, \mathcal{F}')$$

is nonzero only when $-\ell \le p \le \ell$ and $q \in \{0, n\}$. Consequently, we have an exact sequence

$$0 \to \boldsymbol{E}_\infty^{0,0} \to \mathrm{Hom}(\mathcal{F}, \mathcal{F}') \to \boldsymbol{E}_\infty^{-n,n} \to 0$$

and (under the assumption $\ell \le n$) the last term is nonzero only if $\ell = n$. Furthermore, we have

$$\begin{aligned}
\boldsymbol{E}_\infty^{0,0} &= \boldsymbol{E}_2^{0,0} \\
&= \mathrm{Ker}\Big(\bigoplus_i \mathrm{Hom}(\mathcal{L}_i, \mathcal{L}'_i) \to \bigoplus_i \mathrm{Hom}(\mathcal{L}_i, \mathcal{L}'_{i-1})\Big) \Big/ \mathrm{Im}\Big(\bigoplus_i \mathrm{Hom}(\mathcal{L}_i, \mathcal{L}'_{i+1}) \to \bigoplus_i \mathrm{Hom}(\mathcal{L}_i, \mathcal{L}'_i)\Big),
\end{aligned}$$

hence a morphism $\varphi\colon \mathcal{F} \to \mathcal{F}'$ can be represented by a morphism of complexes $\varphi_\bullet\colon \mathcal{L}_\bullet \to \mathcal{L}'_\bullet$ if and only if it comes from $\boldsymbol{E}_\infty^{0,0}$. In particular, this holds true for $\ell < n$ since in this case $\boldsymbol{E}_\infty^{-n,n} = 0$.

Now assume $\ell = n$. We have $\boldsymbol{E}_\infty^{-n,n} \subset \operatorname{Ext}^n(\mathcal{L}_0, \mathcal{L}'_n)$ and it is easy to see that if $\partial\varphi \in \operatorname{Ext}^n(\mathcal{L}_0, \mathcal{L}'_n)$ is the image of φ under the composition

$$\operatorname{Hom}(\mathcal{F}, \mathcal{F}') \to \boldsymbol{E}_\infty^{-n,n} \to \operatorname{Ext}^n(\mathcal{L}_0, \mathcal{L}'_n),$$

then the composition

$$\boldsymbol{H}_*^0(\mathcal{L}_0) \twoheadrightarrow \boldsymbol{H}_*^{\mathrm{bot}}(\mathcal{L}_\bullet) \to \boldsymbol{H}_*^{\mathrm{top}}(\mathcal{L}'_\bullet) \hookrightarrow \boldsymbol{H}_*^n(\mathcal{L}'_n),$$

where the middle arrow is (3-5), is given by $\partial\varphi$. Thus, if (3-5) vanishes then $\partial\varphi = 0$, and it follows that φ is in the image of $\boldsymbol{E}_\infty^{0,0}$, hence is induced by a morphism of complexes.

Conversely, if φ is given by a morphism of complexes $\varphi_\bullet$, the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \boldsymbol{H}_*^{\mathrm{bot}}(\mathcal{L}_\bullet) \arrow[r] \arrow[d, "\boldsymbol{H}_*^0(\varphi_0)"'] & \boldsymbol{H}_*^0(\mathcal{F}) \arrow[r] \arrow[d, "\boldsymbol{H}_*^0(\varphi)"'] & \boldsymbol{H}_*^{\mathrm{top}}(\mathcal{L}_\bullet) \arrow[r] \arrow[d, "\boldsymbol{H}_*^n(\varphi_n)"'] & 0 \\
0 \arrow[r] & \boldsymbol{H}_*^{\mathrm{bot}}(\mathcal{L}'_\bullet) \arrow[r] & \boldsymbol{H}_*^0(\mathcal{F}') \arrow[r] & \boldsymbol{H}_*^{\mathrm{top}}(\mathcal{L}'_\bullet) \arrow[r] & 0
\end{tikzcd}$$

where the rows are the exact sequences of Lemma 3.4, shows that (3-5) is zero.

The uniqueness up to homotopy of $\varphi_\bullet$ in both cases follows from the above formula for $\boldsymbol{E}_\infty^{0,0}$. □

3.2 VHC resolutions and uniqueness

The notion of a VHC resolution is based on the following.

Definition 3.7 We will say that a vector bundle $\mathcal{E}$ on $\mathbb{P}^n$ has

- the *vanishing lower cohomology* property if
$$\boldsymbol{H}_*^p(\mathcal{E}) = 0 \quad \text{for } 1 \le p \le \left\lfloor \tfrac{1}{2}n \right\rfloor;$$
- the *vanishing upper cohomology* property if
$$\boldsymbol{H}_*^p(\mathcal{E}) = 0 \quad \text{for } \left\lceil \tfrac{1}{2}n \right\rceil \le p \le n-1.$$

We will abbreviate these properties by *VLC* and *VUC*, respectively.

Example 3.8 Every split bundle is both VLC and VUC. Moreover:

- Every vector bundle on $\mathbb{P}^1$ is both VLC and VUC since the conditions are void.
- A vector bundle on $\mathbb{P}^2$ is VLC if and only if it is VUC if and only if it is split.
- If $1 \le p, q \le n-1$ and $t \in \mathbb{Z}$, we have

(3-6) $$H^q(\mathbb{P}^n, \Omega^p(t)) = \begin{cases} \mathsf{k} & \text{if } q = p \text{ and } t = 0, \\ 0 & \text{otherwise.} \end{cases}$$

  Thus, $\Omega^p(t)$ is VLC if and only if $p > \left\lfloor \tfrac{1}{2}n \right\rfloor$ and it is VUC if and only if $p < \left\lceil \tfrac{1}{2}n \right\rceil$.

Note that for even n, the bundle $\Omega^{n/2}(t)$ is neither VLC nor VUC.

Lemma 3.9 *The properties VLC and VUC are invariant under twists, direct sums, and passing to direct summands. Moreover, a vector bundle $\mathcal{E}$ is VLC if and only if $\mathcal{E}^\vee$ is VUC. Finally, if a bundle $\mathcal{E}$ is VLC and VUC at the same time, it is split.*

Proof Follows from the definition, Serre duality, and Horrock's theorem. □

Below we give a characterization of VLC and VUC bundles in terms of resolutions by split bundles.

Lemma 3.10 *A vector bundle $\mathcal{E}$ on $\mathbb{P}^n$ is VLC if and only if there is an exact sequence*

$$0 \to \mathcal{L}_{\lfloor (n-1)/2 \rfloor} \to \cdots \to \mathcal{L}_0 \to \mathcal{E} \to 0,$$

where $\mathcal{L}_i$ are split bundles.

A vector bundle $\mathcal{E}$ on $\mathbb{P}^n$ is VUC if and only if there is an exact sequence

$$0 \to \mathcal{E} \to \mathcal{L}_0 \to \cdots \to \mathcal{L}_{\lfloor (n-1)/2 \rfloor} \to 0,$$

where $\mathcal{L}_i$ are split bundles.

Proof First assume that $\mathcal{E}$ is a VLC vector bundle.

- If $n = 2k$, then $\boldsymbol{H}^p_*(\mathcal{E}) = 0$ for $1 \le p \le k = n-k$; by Lemma 3.2 this is equivalent to the existence of a resolution of length $\ell = k-1 = \lfloor \frac{1}{2}(n-1) \rfloor$ by split bundles.
- Similarly, if $n = 2k+1$, then $\boldsymbol{H}^p_*(\mathcal{E}) = 0$ for $1 \le p \le k = n-(k+1)$; by Lemma 3.2 this is equivalent to the existence of a resolution of length $\ell = k = \lfloor \frac{1}{2}(n-1) \rfloor$ by split bundles.

The case of a VUC bundle follows from this and Lemma 3.9 by duality. □

Definition 3.11 We will say that a locally free resolution $0 \to \mathcal{E}_{\mathrm{L}} \to \mathcal{E}_{\mathrm{U}} \to \mathcal{F} \to 0$ of a sheaf $\mathcal{F}$ has the *VHC* (*vanishing of half cohomology*) property (or simply is a *VHC resolution*) if $\mathcal{E}_{\mathrm{L}}$ is a VLC vector bundle and $\mathcal{E}_{\mathrm{U}}$ is a VUC vector bundle; see Definition 3.7.

The cohomology of bundles constituting a VHC resolution of a sheaf $\mathcal{F}$ are related to the cohomology of $\mathcal{F}$ as follows.

Lemma 3.12 *Let $0 \to \mathcal{E}_{\mathrm{L}} \to \mathcal{E}_{\mathrm{U}} \to \mathcal{F} \to 0$ be a VHC resolution on $\mathbb{P}^n$ and assume that $1 \le p \le n-1$. If $n = 2k$, then*

$$\boldsymbol{H}^p_*(\mathcal{E}_{\mathrm{L}}) = \begin{cases} 0 & \text{if } 1 \le p \le k, \\ \boldsymbol{H}^{p-1}_*(\mathcal{F}) & \text{if } k+1 \le p \le n-1, \end{cases} \qquad \boldsymbol{H}^p_*(\mathcal{E}_{\mathrm{U}}) = \begin{cases} \boldsymbol{H}^p_*(\mathcal{F}) & \text{if } 1 \le p \le k-1, \\ 0 & \text{if } k \le p \le n-1. \end{cases}$$

If $n = 2k+1$, then

$$\boldsymbol{H}^p_*(\mathcal{E}_{\mathrm{L}}) = \begin{cases} 0 & \text{if } 1 \le p \le k, \\ \boldsymbol{H}^{p-1}_*(\mathcal{F}) & \text{if } k+2 \le p \le n-1, \end{cases} \qquad \boldsymbol{H}^p_*(\mathcal{E}_{\mathrm{U}}) = \begin{cases} \boldsymbol{H}^p_*(\mathcal{F}) & \text{if } 1 \le p \le k-1, \\ 0 & \text{if } k+1 \le p \le n-1, \end{cases}$$

while $\boldsymbol{H}^{k+1}_*(\mathcal{E}_{\mathrm{L}})$ and $\boldsymbol{H}^k_*(\mathcal{E}_{\mathrm{U}})$ fit into an exact sequence of graded $\mathbb{S}$–modules

$$0 \to \boldsymbol{H}^k_*(\mathcal{E}_{\mathrm{U}}) \to \boldsymbol{H}^k_*(\mathcal{F}) \to \boldsymbol{H}^{k+1}_*(\mathcal{E}_{\mathrm{L}}) \to 0. \tag{3-7}$$

Proof Follows immediately from the long exact sequences of cohomology groups and the vanishings in the definition of VLC and VUC bundles. $\square$

If $n = 2k+1$, we will often use sequence (3-7) to identify $\boldsymbol{H}^k_*(\mathcal{E}_\mathrm{U})$ with an $\mathbb{S}$–submodule of $\boldsymbol{H}^k_*(\mathcal{F})$.

Lemma 3.13 *Let $0 \to \mathcal{E}_\mathrm{L} \to \mathcal{E}_\mathrm{U} \to \mathcal{F} \to 0$ be a VHC resolution on $\mathbb{P}^n$. Set $k = \lfloor \frac{1}{2}(n-1) \rfloor$. Then the object $\mathcal{F}[k] \in \boldsymbol{D}(\mathbb{P}^n)$ is quasiisomorphic to a complex of split bundles*

$$\{\mathcal{L}_{2k+1} \to \mathcal{L}_{2k} \to \cdots \to \mathcal{L}_1 \to \mathcal{L}_0\}$$

such that its first half $\{\mathcal{L}_{2k+1} \to \cdots \to \mathcal{L}_{k+1}\}$ is a resolution of $\mathcal{E}_\mathrm{L}$ and its second half $\{\mathcal{L}_k \to \cdots \to \mathcal{L}_0\}$ is a resolution of $\mathcal{E}_\mathrm{U}$.

Moreover, if $n = 2k+1$ then $\boldsymbol{H}^k_*(\mathcal{E}_\mathrm{U}) = \boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet)$, $\boldsymbol{H}^{k+1}_*(\mathcal{E}_\mathrm{L}) = \boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}_\bullet)$, and the exact sequence (3-7) coincides with the exact sequence of Lemma 3.4.

Proof By Lemma 3.10, the sheaves $\mathcal{E}_\mathrm{L}$ and $\mathcal{E}_\mathrm{U}$ have resolutions of length k by split bundles, which we can write in the form

$$0 \to \mathcal{L}_{2k+1} \to \cdots \to \mathcal{L}_{k+1} \to \mathcal{E}_\mathrm{L} \to 0 \quad \text{and} \quad 0 \to \mathcal{E}_\mathrm{U} \to \mathcal{L}_k \to \cdots \to \mathcal{L}_0 \to 0. \tag{3-8}$$

The morphism $\mathcal{E}_\mathrm{L} \to \mathcal{E}_\mathrm{U}$ gives a morphism $\mathcal{L}_{k+1} \to \mathcal{L}_k$, which allows us to concatenate the resolutions into a single complex

$$\{\mathcal{L}_{2k+1} \to \mathcal{L}_{2k} \to \cdots \to \mathcal{L}_1 \to \mathcal{L}_0\}$$

of split bundles quasiisomorphic to $\mathrm{Cone}(\mathcal{E}_\mathrm{L} \to \mathcal{E}_\mathrm{U})[k] \cong \mathcal{F}[k]$. If $n = 2k+1$ the hypercohomology spectral sequences of (3-8) show that $\boldsymbol{H}^k_*(\mathcal{E}_\mathrm{U}) = \boldsymbol{H}^{\mathrm{bot}}_*(\mathcal{L}_\bullet)$ and $\boldsymbol{H}^{k+1}_*(\mathcal{E}_\mathrm{L}) = \boldsymbol{H}^{\mathrm{top}}_*(\mathcal{L}_\bullet)$, and allow us to identify the exact sequences of Lemmas 3.12 and 3.4. $\square$

For the uniqueness result stated below we need the following technical notion.

Definition 3.14 A VHC resolution is *linearly minimal* if it has no trivial complex $\mathcal{O}(t) \xrightarrow{\mathrm{id}} \mathcal{O}(t)$ as a direct summand. In other words, if $f : \mathcal{E}_\mathrm{L} \to \mathcal{E}_\mathrm{U}$ is not isomorphic to $\mathrm{id} \oplus f' : \mathcal{O}(t) \oplus \mathcal{E}'_\mathrm{L} \to \mathcal{O}(t) \oplus \mathcal{E}'_\mathrm{U}$.

Clearly, any VHC resolution is isomorphic to the direct sum of a linearly minimal VHC resolution and several trivial complexes $\mathcal{O}(t_i) \xrightarrow{\mathrm{id}} \mathcal{O}(t_i)$.

Theorem 3.15 *Let $0 \to \mathcal{E}_\mathrm{L} \xrightarrow{f} \mathcal{E}_\mathrm{U} \to \mathcal{F} \to 0$ and $0 \to \mathcal{E}'_\mathrm{L} \xrightarrow{f'} \mathcal{E}'_\mathrm{U} \to \mathcal{F} \to 0$ be linearly minimal VHC resolutions of the same sheaf $\mathcal{F}$. If $n = 2k+1$ assume also we have an equality $\boldsymbol{H}^k_*(\mathcal{E}_\mathrm{U}) = \boldsymbol{H}^k_*(\mathcal{E}'_\mathrm{U})$ of $\mathbb{S}$–submodules in $\boldsymbol{H}^k_*(\mathcal{F})$ with respect to the embeddings given by (3-7). Then the resolutions are isomorphic, ie there is a commutative diagram*

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{E}_\mathrm{L} \arrow[r, "f"] \arrow[d, "\varphi_\mathrm{L}"'] & \mathcal{E}_\mathrm{U} \arrow[r] \arrow[d, "\varphi_\mathrm{U}"] \arrow[dl, dotted, "h"'] & \mathcal{F} \arrow[r] \arrow[d, "\mathrm{id}"] & 0 \\
0 \arrow[r] & \mathcal{E}'_\mathrm{L} \arrow[r, "f'"] & \mathcal{E}'_\mathrm{U} \arrow[r] & \mathcal{F} \arrow[r] & 0
\end{tikzcd}$$

where φ_{L} and φ_{U} are isomorphisms. Moreover, an isomorphism $(\varphi_{\mathrm{L}},\varphi_{\mathrm{U}})$ of resolutions inducing the identity morphism of $\mathcal{F}$ is unique up to a homotopy $h\colon \mathcal{E}_{\mathrm{U}}\to\mathcal{E}'_{\mathrm{L}}$. Finally, the endomorphisms $\varphi_{\mathrm{L}}^{-1}\circ h\circ f$ and $\varphi_{\mathrm{U}}^{-1}\circ f'\circ h$ of $\mathcal{E}_{\mathrm{L}}$ and $\mathcal{E}_{\mathrm{U}}$ induced by any homotopy h are nilpotent.

Proof Let $k=\lfloor\frac{1}{2}(n-1)\rfloor$, so that $n=2k+1$ or $n=2k+2$. By Lemma 3.13, the object $\mathcal{F}[k]$ is quasiisomorphic to complexes of split bundles

$$\{\mathcal{L}_{2k+1}\to\mathcal{L}_{2k}\to\cdots\to\mathcal{L}_1\to\mathcal{L}_0\}\quad\text{and}\quad\{\mathcal{L}'_{2k+1}\to\mathcal{L}'_{2k}\to\cdots\to\mathcal{L}'_1\to\mathcal{L}'_0\}$$

corresponding to the resolutions $\mathcal{E}_{\mathrm{L}}\to\mathcal{E}_{\mathrm{U}}$ and $\mathcal{E}'_{\mathrm{L}}\to\mathcal{E}'_{\mathrm{U}}$, respectively. Using linear minimality we can assume that each of these complexes has no trivial complex $\mathcal{O}(t)\xrightarrow{\mathrm{id}}\mathcal{O}(t)$ as a direct summand.

If $n=2k+2$, the lengths of the resolutions are less than n, hence the first part of Lemma 3.6 ensures that the identity morphism $\mathcal{F}\to\mathcal{F}$ is induced by a morphism of complexes. If $n=2k+1$, we use the second part of Lemma 3.6 — the composition (3-5) vanishes due to the assumption $\boldsymbol{H}^k_*(\mathcal{E}_{\mathrm{U}})=\boldsymbol{H}^k_*(\mathcal{E}'_{\mathrm{U}})$ and Lemma 3.13 — and obtain the same conclusion. Thus, we obtain a quasiisomorphism of complexes of split bundles

$$\begin{tikzcd}
\mathcal{L}_{2k+1} \arrow[r] \arrow[d, "\varphi_{2k+1}"'] & \mathcal{L}_{2k} \arrow[r] \arrow[d, "\varphi_{2k}"'] & \cdots \arrow[r] & \mathcal{L}_1 \arrow[r] \arrow[d, "\varphi_1"'] & \mathcal{L}_0 \arrow[d, "\varphi_0"'] \\
\mathcal{L}'_{2k+1} \arrow[r] & \mathcal{L}'_{2k-1} \arrow[r] & \cdots \arrow[r] & \mathcal{L}'_1 \arrow[r] & \mathcal{L}'_0
\end{tikzcd} \tag{3-9}$$

We prove below that it is necessarily an isomorphism, ie that each φ_i is an isomorphism. For this we use the induction on the sum of ranks of the bundles $\mathcal{L}_i$.

The base of the induction follows from Lemma 3.5. Indeed, if $\mathcal{L}_\bullet=0$ then $\mathcal{L}'_\bullet$ is acyclic, hence is the sum of trivial complexes. But by assumption it has no trivial summands, hence $\mathcal{L}'_\bullet=0$.

Now assume that $\mathcal{L}_\bullet\neq0$. The totalization of (3-9) is the acyclic complex

$$\mathcal{L}_{2k+1}\to\mathcal{L}_{2k}\oplus\mathcal{L}'_{2k+1}\to\cdots\to\mathcal{L}_0\oplus\mathcal{L}'_1\to\mathcal{L}'_0 \tag{3-10}$$

of split bundles of length $2k+2$. If $n=2k+2$ we can formally add the zero term on the right and obtain an acyclic complex of length $\ell=n+1$ of split bundles for which the condition (1) of Lemma 3.5 holds true. If $n=2k+1$, the condition (1) of Lemma 3.5 follows from the assumption $\boldsymbol{H}^k_*(\mathcal{E}_{\mathrm{U}})=\boldsymbol{H}^k_*(\mathcal{E}'_{\mathrm{U}})$. In both cases Lemma 3.5 implies that (3-10) is isomorphic to a direct sum of shifts of trivial complexes.

To make this direct sum decomposition more precise, we consider as in the proof of Lemma 3.5 the maximal integer t such that $\mathcal{O}(t)$ appears as one of summands of one of the split bundles $\mathcal{L}_i$ or $\mathcal{L}'_i$. Twisting (3-9) by $\mathcal{O}(-t)$ and applying the functor $H^0(\mathbb{P}^n,-)$ we obtain a nonzero bicomplex

$$\begin{tikzcd}
\mathsf{k}^{m_{2k+1}} \arrow[r] \arrow[d] & \mathsf{k}^{m_{2k}} \arrow[r] \arrow[d] & \cdots \arrow[r] & \mathsf{k}^{m_1} \arrow[r] \arrow[d] & \mathsf{k}^{m_0} \arrow[d] \\
\mathsf{k}^{m'_{2k+1}} \arrow[r] & \mathsf{k}^{m'_{2k}} \arrow[r] & \cdots \arrow[r] & \mathsf{k}^{m'_1} \arrow[r] & \mathsf{k}^{m'_0}
\end{tikzcd}$$

(as before, m_i and m'_i are the multiplicities of $\mathcal{O}(t)$ in $\mathcal{L}_i$ and $\mathcal{L}'_i$, respectively) with acyclic totalization.

If any of the horizontal arrows in this bicomplex is nontrivial, Lemma 3.3 implies that the trivial complex $\mathcal{O}(t) \to \mathcal{O}(t)$ is a direct summand of either $\mathcal{L}_\bullet$ or $\mathcal{L}'_\bullet$, which contradicts the linear minimality assumption. Therefore, the horizontal arrows are zero, and hence the vertical arrows are all isomorphisms.

This means that $m_i = m'_i$ for all i and we can write

$$\mathcal{L}_i = \mathcal{O}(t)^{\oplus m_i} \oplus \overline{\mathcal{L}}_i, \quad \mathcal{L}'_i = \mathcal{O}(t)^{\oplus m_i} \oplus \overline{\mathcal{L}}'_i, \quad \varphi_i = \begin{pmatrix} 1 & \psi_i \\ 0 & \overline{\varphi}_i \end{pmatrix},$$

and that $\overline{\varphi}_\bullet \colon \overline{\mathcal{L}}_\bullet \to \overline{\mathcal{L}}'_\bullet$ is a quasiisomorphism of complexes of split bundles which have no trivial summands. Moreover, we have $\sum \mathrm{rk}(\overline{\mathcal{L}}_i) < \sum \mathrm{rk}(\mathcal{L}_i)$. By induction, we deduce that $\overline{\varphi}_i$ is an isomorphism for each i, hence so is φ_i.

Since $\varphi_\bullet$ is an isomorphism of complexes, it induces an isomorphism of resolutions of $\mathcal{E}_{\mathrm{L}}$ and $\mathcal{E}'_{\mathrm{L}}$ and of $\mathcal{E}_{\mathrm{U}}$ and $\mathcal{E}'_{\mathrm{U}}$, compatible with the maps $\mathcal{E}_{\mathrm{L}} \to \mathcal{E}_{\mathrm{U}}$ and $\mathcal{E}'_{\mathrm{L}} \to \mathcal{E}'_{\mathrm{U}}$, hence an isomorphism $(\varphi_{\mathrm{L}}, \varphi_{\mathrm{U}})$ of the original VHC resolutions. This proves the first part of the theorem.

Further, recall that by Lemma 3.6 the morphism $\varphi_\bullet$ in (3-9) inducing the identity of $\mathcal{F}$ is unique up to a homotopy $h_\bullet \colon \mathcal{L}_\bullet \to \mathcal{L}'_{\bullet+1}$. Note that the first part $(h_i)_{0 \le i \le k-1}$ of such a homotopy replaces the morphism $(\varphi_i)_{0 \le i \le k}$ of the right resolutions of $\mathcal{E}_{\mathrm{U}}$ and $\mathcal{E}'_{\mathrm{U}}$ by a homotopy equivalent morphism, hence it does not change φ_{U}, and a fortiori does not change φ_{L}. Similarly, the last part $(h_i)_{k+1 \le i \le 2k}$ of a homotopy does not change $(\varphi_{\mathrm{L}}, \varphi_{\mathrm{U}})$. Finally, it is clear that the middle component $h_k \colon \mathcal{L}_k \to \mathcal{L}'_{k+1}$ of a homotopy modifies $(\varphi_{\mathrm{L}}, \varphi_{\mathrm{U}})$ by the homotopy

$$\mathcal{E}_{\mathrm{U}} \hookrightarrow \mathcal{L}_k \xrightarrow{h_k} \mathcal{L}'_{k+1} \twoheadrightarrow \mathcal{E}'_{\mathrm{L}}$$

of the VHC resolutions. This proves the second part of the theorem.

So, it only remains to check the nilpotence of the induced endomorphisms of $\mathcal{E}_{\mathrm{L}}$ and $\mathcal{E}_{\mathrm{U}}$. For this let us write

$$\mathcal{L}_{k+1} = \bigoplus \mathcal{O}(a_i), \quad \mathcal{L}_k = \bigoplus \mathcal{O}(b_i), \quad \mathcal{L}'_{k+1} = \bigoplus \mathcal{O}(a'_i), \quad \mathcal{L}'_k = \bigoplus \mathcal{O}(b'_i),$$

and for each $c \in \mathbb{Z}$ define finite filtrations of these bundles by

$$\mathrm{F}_{\ge c}\mathcal{L}_{k+1} = \bigoplus_{a_i \ge c} \mathcal{O}(a_i), \quad \mathrm{F}_{\ge c}\mathcal{L}_k = \bigoplus_{b_i \ge c} \mathcal{O}(b_i), \quad \mathrm{F}_{\ge c}\mathcal{L}'_{k+1} = \bigoplus_{a'_i \ge c} \mathcal{O}(a'_i), \quad \mathrm{F}_{\ge c}\mathcal{L}'_k = \bigoplus_{b'_i \ge c} \mathcal{O}(b'_i).$$

Then the morphism $\mathcal{L}_{k+1} \to \mathcal{L}_k$ induced by f takes $\mathrm{F}_{\ge c}\mathcal{L}_{k+1}$ to $\mathrm{F}_{\ge c+1}\mathcal{L}_k$ (because f is assumed to be linearly minimal) and obviously any morphism $h \colon \mathcal{L}_k \to \mathcal{L}'_{k+1}$ takes $\mathrm{F}_{\ge c+1}\mathcal{L}_k$ to $\mathrm{F}_{\ge c+1}\mathcal{L}'_{k+1}$. Since φ_{L} is an isomorphism, we conclude that the composition $\varphi_{\mathrm{L}}^{-1} \circ h \circ f$ is induced by an endomorphism of $\mathcal{L}_{k+1}$ that takes $\mathrm{F}_{\ge c}\mathcal{L}_{k+1}$ to $\mathrm{F}_{\ge c+1}\mathcal{L}_{k+1}$, hence is nilpotent. A similar argument works for $\varphi_{\mathrm{U}}^{-1} \circ f' \circ h$. $\square$

3.3 Existence of VHC resolutions

The results of this subsection are not necessary for Section 4, but the technique used in their proofs is similar.

Definition 3.16 Let $\mathcal{F}$ be a coherent sheaf and $1 \le k \le n-1$. We will say that a graded $\mathbb{S}$–submodule $\mathrm{A}^k \subset \boldsymbol{H}^k_*(\mathcal{F})$ is *shadowless* if for any $t_0 \in \mathbb{Z}$ such that $\mathrm{A}^k_{t_0} \neq 0$ we have $\mathrm{A}^k_t = H^k(\mathbb{P}^n, \mathcal{F}(t))$ for any $t > t_0$. Similarly, for any $1 \le p_0 \le n-1$ and any $t_0 \in \mathbb{Z}$ we define the *shadow* of (p_0, t_0) as the set

$$\mathbf{Sh}(p_0, t_0) = \{(p, t) \mid 1 \le p \le p_0 \text{ and } t > t_0\}, \tag{3-11}$$

and say that a bigraded $\mathbb{S}$–submodule $\mathrm{A} \subset \bigoplus_{p=1}^{n-1} \boldsymbol{H}^p_*(\mathcal{F})$ is *shadowless* if for any (p_0, t_0) such that $\mathrm{A}^{p_0}_{t_0} \neq 0$ we have $\mathrm{A}^p_t = H^p(\mathbb{P}^n, \mathcal{F}(t))$ for any $(p, t) \in \mathbf{Sh}(p_0, t_0)$.

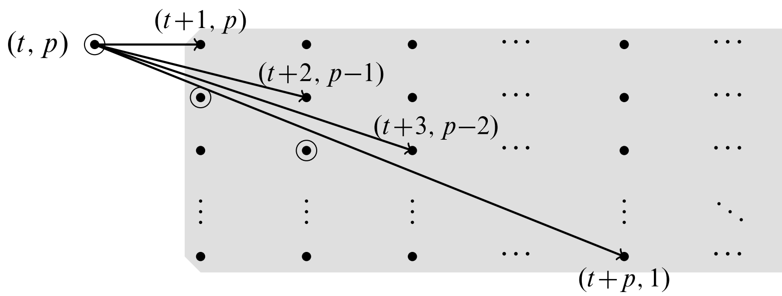
To understand the meaning of this notion observe the following. Let $\mathcal{T}$ be the tangent bundle of $\mathbb{P}^n$. Recall the Koszul resolution of its exterior power,

$$0 \to \mathcal{O} \to V \otimes \mathcal{O}(1) \to \cdots \to \textstyle\bigwedge^s V \otimes \mathcal{O}(s) \to \bigwedge^s \mathcal{T} \to 0, \tag{3-12}$$

where V is a vector space such that $\mathbb{P}^n = \mathbb{P}(V)$. If $\mathcal{F}$ is a sheaf on $\mathbb{P}^n$, tensoring (3-12) by $\mathcal{F}(t)$ we obtain the hypercohomology spectral sequence

$$\boldsymbol{E}_1^{i,j} = \textstyle\bigwedge^i V \otimes H^j(\mathbb{P}^n, \mathcal{F}(i+t)) \Rightarrow H^{i+j-s}\bigl(\mathbb{P}^n, \bigwedge^s \mathcal{T} \otimes \mathcal{F}(t)\bigr).$$

The following picture shows the arrows $\boldsymbol{d}_r$, for $1 \le r \le p$, of the spectral sequence with source at the terms $\boldsymbol{E}_r^{t,p}$, as well as the terms that in the limit compute the filtration on $H^{p-s}\bigl(\mathbb{P}^n, \bigwedge^s \mathcal{T} \otimes \mathcal{F}(t)\bigr)$ (these terms are circled), and the shadow of (p, t):



It is important that the arrows $\boldsymbol{d}_r$, for $1 \le r \le p$, applied to the terms $\boldsymbol{E}_r^{t,p}$ of the spectral sequence land in its shadow. This property will be used in Propositions 3.17 and 4.10 below.

Proposition 3.17 *For any coherent sheaf $\mathcal{F}$ on $\mathbb{P}^n$ and any finite-dimensional shadowless $\mathbb{S}$–submodule*

$$\mathrm{A} \subset \bigoplus_{p=1}^{n-1} \boldsymbol{H}^p_*(\mathcal{F})$$

there exists a vector bundle $\mathcal{E}_{\mathrm{A}}$ and an epimorphism $\pi_{\mathrm{A}} \colon \mathcal{E}_{\mathrm{A}} \twoheadrightarrow \mathcal{F}$ such that

- *the map $\boldsymbol{H}^0_*(\mathcal{E}_A) \xrightarrow{\boldsymbol{H}^0_*(\pi_{\mathrm{A}})} \boldsymbol{H}^0_*(\mathcal{F})$ is surjective, and*
- *the map $\bigoplus_{p=1}^{n-1} \boldsymbol{H}^p_*(\mathcal{E}_A) \xrightarrow{\oplus \boldsymbol{H}^p_*(\pi_{\mathrm{A}})} \bigoplus_{p=1}^{n-1} \boldsymbol{H}^p_*(\mathcal{F})$ is an isomorphism onto* A.

Note that the assumption $\dim(\mathrm{A}) < \infty$ in the proposition is necessary because $\boldsymbol{H}_*^p(\mathcal{E})$ is finite-dimensional for any vector bundle $\mathcal{E}$ if $1 \le p \le n-1$. On the other hand, if $\mathcal{F}$ is a coherent sheaf with p–dimensional support then $\boldsymbol{H}_*^p(\mathcal{F})$ is not finite-dimensional, so a priori A could have infinite dimension.

Proof We argue by induction on $\dim(\mathrm{A})$. If $\mathrm{A} = 0$ we take $\mathcal{E}_{\mathrm{A}}$ to be the split bundle that corresponds to a free $\mathbb{S}$–module surjecting onto $\boldsymbol{H}_*^0(\mathcal{F})$ as in Lemma 3.1. The desired condition is tautologically true.

Assume $\dim(\mathrm{A}) > 0$. Let

$$p_0 = \min\{p \ge 1 \mid \mathrm{A}^p \ne 0\} \quad \text{and} \quad t_0 = \max\{t \mid \mathrm{A}_t^{p_0} \ne 0\}.$$

Since A is shadowless we have $H^p(\mathbb{P}^n, \mathcal{F}(t)) = \mathrm{A}_t^p = 0$ for all $(p,t) \in \mathbf{Sh}(p_0, t_0)$ — the first equality holds because A is shadowless and the second follows from the above definition of (p_0, t_0). In particular, the subspace $\mathrm{A}_{t_0}^{p_0} \subset H^{p_0}(\mathbb{P}^n, \mathcal{F}(t_0))$ sits in the kernels of differentials $\boldsymbol{d}_1, \dots, \boldsymbol{d}_{p_0-1}$ of the hypercohomology spectral sequence of $\mathcal{F}(t_0)$ tensored with the Koszul complex (3-12) for $s = p_0 - 1$. Moreover, $H^{p_0}(\mathbb{P}^n, \mathcal{F}(t_0))$ is the only nonzero subspace on the diagonal of the spectral sequence that in the limit computes the filtration on $H^1\bigl(\mathbb{P}^n, \bigwedge^{p_0-1}\mathcal{T} \otimes \mathcal{F}(t_0)\bigr)$. Therefore, we obtain an inclusion

$$\mathrm{A}_{t_0}^{p_0} \subset H^{p_0}(\mathbb{P}^n, \mathcal{F}(t_0)) = H^1\bigl(\mathbb{P}^n, \textstyle\bigwedge^{p_0-1}\mathcal{T} \otimes \mathcal{F}(t_0)\bigr) = \operatorname{Ext}^1(\Omega^{p_0-1}(-t_0), \mathcal{F}),$$

which induces an extension

$$0 \to \mathcal{F} \to \mathcal{F}' \to \mathrm{A}_{t_0}^{p_0} \otimes \Omega^{p_0-1}(-t_0) \to 0$$

such that the connecting morphism $\mathrm{A}_{t_0}^{p_0} = H^{p_0-1}(\mathbb{P}^n, \mathrm{A}_{t_0}^{p_0} \otimes \Omega^{p_0-1}) \to H^{p_0}(\mathbb{P}^n, \mathcal{F}(t_0))$ is the natural embedding (the first identification uses (3-6)). Now the cohomology exact sequence implies that

$$\bigoplus_{p=1}^{n-1} \boldsymbol{H}_*^p(\mathcal{F}') = \Bigl(\bigoplus_{p=1}^{n-1} \boldsymbol{H}_*^p(\mathcal{F})\Bigr) \big/ \mathrm{A}_{t_0}^{p_0},$$

hence the quotient $\mathbb{S}$–module $\mathrm{A}' := \mathrm{A}/\mathrm{A}_{t_0}^{p_0}$ is an $\mathbb{S}$–submodule in $\bigoplus \boldsymbol{H}_*^p(\mathcal{F}')$. Clearly $\dim(\mathrm{A}') < \dim(\mathrm{A})$ and it is straightforward to check that A' is shadowless. By the induction hypothesis there is a vector bundle $\mathcal{E}_{\mathrm{A}'}$ and an epimorphism $\pi_{\mathrm{A}'} : \mathcal{E}_{\mathrm{A}'} \twoheadrightarrow \mathcal{F}'$ inducing surjection on $\boldsymbol{H}_*^0$ and the natural embedding of A' into the intermediate cohomology of $\mathcal{F}'$. We define $\mathcal{E}_{\mathrm{A}}$ as the kernel of the composition of epimorphisms

$$\mathcal{E}_{\mathrm{A}'} \twoheadrightarrow \mathcal{F}' \twoheadrightarrow \mathrm{A}_{t_0}^{p_0} \otimes \Omega^{p_0-1}(-t_0).$$

By construction the map $\pi_{\mathrm{A}'}$ lifts to a map π_{A} that fits into a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{E}_{\mathrm{A}} \arrow[r] \arrow[d, "\pi_{\mathrm{A}}"'] & \mathcal{E}_{\mathrm{A}'} \arrow[r] \arrow[d, two heads, "\pi_{\mathrm{A}'}"'] & \mathrm{A}_{t_0}^{p_0} \otimes \Omega^{p_0-1}(-t_0) \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathcal{F} \arrow[r] & \mathcal{F}' \arrow[r] & \mathrm{A}_{t_0}^{p_0} \otimes \Omega^{p_0-1}(-t_0) \arrow[r] & 0
\end{tikzcd}$$

The surjectivity of $\pi_{\mathrm{A}'}$ and $\boldsymbol{H}_*^0(\pi_{\mathrm{A}'})$ implies that of π_{A} and $\boldsymbol{H}_*^0(\pi_{\mathrm{A}})$. Similarly, it follows that $\bigoplus \boldsymbol{H}_*^p(\pi_{\mathrm{A}})$ is an isomorphism onto A. Thus, the required result holds for A. $\square$

Corollary 3.18 *If $\mathcal{F}$ is a sheaf on $\mathbb{P}^n$ of projective dimension at most one, then $\mathcal{F}$ has a VHC resolution.*

Proof Since the projective dimension of $\mathcal{F}$ is at most one, there exists a locally free resolution

$$0 \to \mathcal{E}_1 \to \mathcal{E}_0 \to \mathcal{F} \to 0.$$

Since $\boldsymbol{H}^p_*(\mathcal{E}_i)$ is finite-dimensional for $1 \le p \le n-1$, the cohomology exact sequence

$$\cdots \to \boldsymbol{H}^p_*(\mathcal{E}_0) \to \boldsymbol{H}^p_*(\mathcal{F}) \to \boldsymbol{H}^{p+1}_*(\mathcal{E}_1) \to \cdots$$

implies that $\boldsymbol{H}^p_*(\mathcal{F})$ is finite-dimensional for $1 \le p \le n-2$. Let

$$\mathrm{A} := \begin{cases} \bigoplus_{p=1}^{k-1} \boldsymbol{H}^p_*(\mathcal{F}) & \text{if } n = 2k, \\ \bigoplus_{p=1}^{k-1} \boldsymbol{H}^p_*(\mathcal{F}) \oplus \mathrm{A}^k & \text{if } n = 2k+1, \end{cases}$$

where $\mathrm{A}^k \subset \boldsymbol{H}^k_*(\mathcal{F})$ is any finite-dimensional shadowless $\mathbb{S}$–submodule, for instance $\mathrm{A}^k = 0$. Note that we have $k-1 \le n-2$ as soon as $n \ge 1$, hence A is finite-dimensional. Moreover, A is shadowless by construction.

Let $\pi_{\mathrm{A}} \colon \mathcal{E}_{\mathrm{A}} \to \mathcal{F}$ be the epimorphism constructed in Proposition 3.17 and let $\mathcal{K}_{\mathrm{A}} = \mathrm{Ker}(\pi_{\mathrm{A}})$, so that

$$0 \to \mathcal{K}_{\mathrm{A}} \to \mathcal{E}_{\mathrm{A}} \to \mathcal{F} \to 0$$

is an exact sequence. First, $\boldsymbol{H}^p_*(\mathcal{E}_{\mathrm{A}}) = \mathrm{A}^p = 0$ for $\lceil \frac{1}{2}n \rceil \le p \le n-1$ by definition, hence $\mathcal{E}_{\mathrm{A}}$ is VUC. Furthermore, $\mathcal{K}_{\mathrm{A}}$ is locally free because the projective dimension of $\mathcal{F}$ is at most one. Finally, the cohomology exact sequence implies that $\boldsymbol{H}^p_*(\mathcal{K}_{\mathrm{A}}) = 0$ for $1 \le p \le k = \lfloor \frac{1}{2}n \rfloor$, hence $\mathcal{K}_{\mathrm{A}}$ is VLC. □

4 Hyperbolic equivalence on projective spaces

In this section we prove Theorem 4.17 on VHC modifications of quadratic forms and deduce from it Theorem 1.3 and Corollary 1.5 from the introduction. In Section 4.1 we recall a characterization of cokernel sheaves of quadratic forms (*symmetric sheaves*), in Section 4.2 we define elementary modifications of quadratic forms with respect to some intermediate cohomology classes, and in Section 4.3 we state the modification theorem (Theorem 4.17) and prove it by applying an appropriate sequence of elementary modifications. Finally, in Section 4.4 we combine these results to prove Theorem 1.3 and Corollary 1.5.

4.1 Reminder on symmetric sheaves

For a scheme Y and an object $\mathcal{C} \in \boldsymbol{D}(Y)$ we write

$$\mathcal{C}^\vee := \mathrm{R}\mathcal{H}om(\mathcal{C}, \mathcal{O}_Y)$$

for the derived dual of $\mathcal{C}$. Note that the cohomology sheaves $\mathcal{H}^i(\mathcal{C}^\vee)$ of $\mathcal{C}^\vee$ are isomorphic to the local Ext–sheaves $\mathcal{E}xt^i(\mathcal{C}, \mathcal{O}_Y)$.

Definition 4.1 (cf [7, Definition 0.2]) We say that a coherent sheaf $\mathcal{C}$ on $\mathbb{P}^n$ is (d,δ)*–symmetric* if $\mathcal{C} \cong i_*\mathcal{R}$, where $i : D \hookrightarrow \mathbb{P}^n$ is the embedding of a degree d hypersurface and $\mathcal{R}$ is a coherent sheaf on D endowed with a symmetric morphism

$$\mathcal{R} \otimes \mathcal{R} \to \mathcal{O}_D(\delta)$$

such that the induced morphism $\mathcal{R}(-\delta) \to \mathcal{R}^\vee$ (where the duality is applied on D) is an isomorphism.

Note that d, δ, D and $\mathcal{R}$ in Definition 4.1 are *not* determined by the sheaf $\mathcal{C}$; see Remark 4.4.

The goal of this subsection is to relate symmetric sheaves to cokernel sheaves of quadratic forms. Most of these results are well-known and not really necessary for the rest of the paper, but useful for the context.

Lemma 4.2 *If $\mathcal{C}$ is a (d,δ)–symmetric coherent sheaf on $\mathbb{P}^n$, there is a self-dual isomorphism*

$$\mathcal{C}^\vee \cong \mathcal{C}(m)[-1],$$

where $m = d - \delta$. In particular, the sheaf $\mathcal{C}$ has projective dimension one.

Proof Let $\mathcal{C} = i_*\mathcal{R}$. Using the definitions and Grothendieck duality we deduce

$$\begin{aligned}\mathcal{C}^\vee = \mathrm{R}\mathcal{H}om(i_*\mathcal{R}, \mathcal{O}_{\mathbb{P}^n}) &\cong i_*\,\mathrm{R}\mathcal{H}om(\mathcal{R}, i^!\mathcal{O}_{\mathbb{P}^n}) \cong i_*\,\mathrm{R}\mathcal{H}om(\mathcal{R}, \mathcal{O}_D(d)[-1]) \\ &\cong i_*\mathcal{R}^\vee(d)[-1] \cong i_*\mathcal{R}(d-\delta)[-1] = \mathcal{C}(m)[-1].\end{aligned}$$

This proves the required isomorphism. Moreover, it follows that this isomorphism is self-dual because so is the isomorphism $\mathcal{R}(-\delta) \cong \mathcal{R}^\vee$. Finally, it follows that $\mathcal{E}xt^1(\mathcal{C}, \mathcal{O}_{\mathbb{P}^n}) \cong \mathcal{C}(m)$ and $\mathcal{E}xt^i(\mathcal{C}, \mathcal{O}_{\mathbb{P}^n}) = 0$ for $i \ge 2$, which means that the projective dimension of $\mathcal{C}$ is one. □

The above lemma implies that symmetric sheaves can be understood as quadratic spaces in the derived category $\boldsymbol{D}(\mathbb{P}^n)$, and define classes in the *shifted Witt group* $\boldsymbol{W}^1(\boldsymbol{D}(\mathbb{P}^n), \mathcal{O}(-m))$ in the sense of [4, Section 1.4].

The following well-known lemma shows that cokernel sheaves of generically nondegenerate quadratic forms are symmetric. For the reader's convenience we provide a proof.

Lemma 4.3 *If a nonzero sheaf $\mathcal{C}$ on $\mathbb{P}^n$ has a self-dual locally free resolution*

$$0 \to \mathcal{E}(-m) \xrightarrow{q} \mathcal{E}^\vee \to \mathcal{C} \to 0, \tag{4-1}$$

then $\mathcal{C}$ is a (d,δ)–symmetric sheaf, where $d = 2\mathrm{c}_1(\mathcal{E}^\vee) + m\,\mathrm{rk}(\mathcal{E})$ and $\delta = d - m$.

Proof Applying the functor $\mathrm{R}\mathcal{H}om(-, \mathcal{O}_{\mathbb{P}^n})$ to the resolution (4-1) of $\mathcal{C}$ we obtain a distinguished triangle

$$\mathcal{C}^\vee \to \mathcal{E} \xrightarrow{q^\vee} \mathcal{E}^\vee(m)$$

in $\boldsymbol{D}(\mathbb{P}^n)$. Since q is self-dual, we have $q = q^\vee$. In particular, it is generically an isomorphism, hence $\mathcal{C}^\vee[1]$ is a pure sheaf and, moreover, $\mathcal{C}^\vee[1] \cong \mathcal{C}(m)$.

Let $\det(q)$ be the determinant of q, which we understand as a global section of the line bundle

$$\det(\mathcal{E}(-m))^\vee \otimes \det(\mathcal{E}^\vee) \cong \det(\mathcal{E}^\vee)^{\otimes 2} \otimes \mathcal{O}(rm),$$

where r is the rank of $\mathcal{E}$. Let $D \subset \mathbb{P}^n$ be the zero locus of $\det(q)$ and set $d := \deg(D)$, so that the line bundle above is $\mathcal{O}(d)$. Consider also the adjugate morphism $q' := \bigwedge^{r-1} q \colon \bigwedge^{r-1}(\mathcal{E}(-m)) \to \bigwedge^{r-1} \mathcal{E}^\vee$; twisting it by $\det(\mathcal{E}) \otimes \mathcal{O}(d-m)$ we obtain a morphism $\mathcal{E}^\vee \xrightarrow{q'} \mathcal{E}(d-m)$. Note that

$$q' \circ q = \det(q) \otimes \mathrm{id}_{\mathcal{E}} \quad \text{and} \quad q \circ q' = \det(q) \otimes \mathrm{id}_{\mathcal{E}^\vee}.$$

It follows that $\mathcal{C} = \mathrm{Coker}(q)$ is supported on D, ie $\mathcal{C} \cong i_*\mathcal{R}$, where $i \colon D \hookrightarrow \mathbb{P}^n$ is the embedding.

Inverting the computation of Lemma 4.2 and using the fact that the functor i_* is exact and fully faithful on coherent sheaves we deduce that $\mathcal{R}^\vee \cong \mathcal{R}(-\delta)$. So, it remains to show that this isomorphism is induced by a symmetric morphism $\mathcal{R} \otimes \mathcal{R} \to \mathcal{O}_D(\delta)$. For this we consider the diagram

(4-2)
$$\begin{tikzcd}
 & (\mathcal{E}(-m) \otimes \mathcal{E}^\vee) \oplus (\mathcal{E}^\vee \otimes \mathcal{E}(-m)) \arrow[r, "q \otimes 1 + 1 \otimes q"] \arrow[d, "(\mathrm{Tr},\mathrm{Tr})"'] & \mathcal{E}^\vee \otimes \mathcal{E}^\vee \arrow[r] \arrow[d, "q'"'] & i_*(\mathcal{R} \otimes \mathcal{R}) \arrow[r] \arrow[d, dashed] & 0 \\
0 \arrow[r] & \mathcal{O}(-m) \arrow[r, "\det(q)"] & \mathcal{O}(\delta) \arrow[r] & i_*\mathcal{O}_D(\delta) \arrow[r] & 0
\end{tikzcd}$$

where the top row is the tensor square of resolution (4-1) of $\mathcal{C} = i_*\mathcal{R}$, and $\mathrm{Tr} \colon \mathcal{E} \otimes \mathcal{E}^\vee \to \mathcal{O}$ is the trace map. It is easy to check that the left square commutes. Therefore, there exists a unique dashed arrow on the right such that the right square commutes. Since q' is symmetric, the dashed arrow is symmetric as well. Now it is easy to see that it induces the isomorphism $\mathcal{R} \to \mathcal{R}^\vee(\delta)$ constructed above. □

Remark 4.4 If $\mathcal{C}$ is a sheaf as in Lemma 4.3, it is not in general true that the presentation of $\mathcal{C}$ as a symmetric sheaf is unique. For instance, if $\mathcal{E} = \mathcal{O} \oplus \mathcal{O}$ and $q = \mathrm{diag}(f, f)$ for a homogeneous polynomial f, we have $\mathcal{C} \cong \mathcal{O}_{D(f)} \oplus \mathcal{O}_{D(f)}$, where $D(f) \subset \mathbb{P}^n$ is the divisor of f; however, the construction of the lemma represents $\mathcal{C}$ as a symmetric sheaf on the nonreduced hypersurface $D = D(f^2)$.

As is explained in Theorem 4.8 (see also Example 4.7), the converse of Lemma 4.3 is not always true. Below we explain the obstruction.

Let $n = 2k+1$. Recall the graded ring $\mathbb{S}$ defined in (3-2). Let $\mathcal{C}$ be a (d,δ)–symmetric sheaf on $\mathbb{P}^n$. A combination of the self-dual isomorphism $\mathcal{C}^\vee \cong \mathcal{C}(m)[-1]$ of Lemma 4.2 with Serre duality endows the $\mathbb{S}$–module $\boldsymbol{H}^k_*(\mathcal{C})$ with a perfect $\mathbb{S}$–bilinear pairing

(4-3)
$$\boldsymbol{H}^k_*(\mathcal{C}) \otimes \boldsymbol{H}^k_*(\mathcal{C}) \to \mathbb{S}^\vee(n+1-m) \twoheadrightarrow \mathsf{k}(n+1-m),$$

which is symmetric when k is even, and skew-symmetric when k is odd.

Lemma 4.5 *Assume $n = 2k+1$. If (4-1) is a self-dual resolution of a symmetric sheaf $\mathcal{C}$, then the $\mathbb{S}$–submodule $\mathrm{Im}(\boldsymbol{H}^k_*(\mathcal{E}^\vee) \to \boldsymbol{H}^k_*(\mathcal{C}))$ is Lagrangian for the pairing (4-3). In particular, $\dim(\boldsymbol{H}^k_*(\mathcal{C}))$ is even.*

Proof First, we need to check that the subspace $\mathrm{Im}(\boldsymbol{H}_*^k(\mathcal{E}^\vee) \to \boldsymbol{H}_*^k(\mathcal{C}))$ in $\boldsymbol{H}_*^k(\mathcal{C})$ is isotropic. For this we note that commutativity of the right square in (4-2) implies that the restriction of the pairing (4-3) to this subspace factors as the composition

$$\boldsymbol{H}_*^k(\mathcal{E}^\vee) \otimes \boldsymbol{H}_*^k(\mathcal{E}^\vee) \xrightarrow{q'} \boldsymbol{H}_*^{2k}(\mathcal{O}(\delta)) \to \boldsymbol{H}_*^{2k}(\mathcal{O}_D(\delta)) \to \boldsymbol{H}_*^{2k+1}(\mathcal{O}(-m)) = \mathbb{S}^\vee(n+1-m),$$

and it follows that it is zero since the composition of the two middle arrows is.

On the other hand, by Serre duality the maps $\boldsymbol{H}_*^k(\mathcal{E}) \to \boldsymbol{H}_*^k(\mathcal{E}^\vee)$ and $\boldsymbol{H}_*^{k+1}(\mathcal{E}) \to \boldsymbol{H}_*^{k+1}(\mathcal{E}^\vee)$ in the cohomology exact sequence

$$\cdots \to \boldsymbol{H}_*^k(\mathcal{E}) \xrightarrow{q} \boldsymbol{H}_*^k(\mathcal{E}^\vee) \to \boldsymbol{H}_*^k(\mathcal{C}) \to \boldsymbol{H}_*^{k+1}(\mathcal{E}) \xrightarrow{q} \boldsymbol{H}_*^{k+1}(\mathcal{E}^\vee) \to \cdots$$

are mutually dual (up to shift of internal grading), so we conclude that

$$\begin{aligned}\dim \mathrm{Im}(\boldsymbol{H}_*^k(\mathcal{E}^\vee) \to \boldsymbol{H}_*^k(\mathcal{C})) &= \dim \mathrm{Coker}(\boldsymbol{H}_*^k(\mathcal{E}) \to \boldsymbol{H}_*^k(\mathcal{E}^\vee)) \\ &= \dim \mathrm{Ker}(\boldsymbol{H}_*^{k+1}(\mathcal{E}) \to \boldsymbol{H}_*^{k+1}(\mathcal{E}^\vee)) = \dim \mathrm{Coker}(\boldsymbol{H}_*^k(\mathcal{E}^\vee) \to \boldsymbol{H}_*^k(\mathcal{C})),\end{aligned}$$

hence

$$\dim\bigl(\mathrm{Im}(\boldsymbol{H}_*^k(\mathcal{E}^\vee) \to \boldsymbol{H}_*^k(\mathcal{C}))\bigr) = \tfrac{1}{2}\dim(\boldsymbol{H}_*^k(\mathcal{C})),$$

and hence $\mathrm{Im}(\boldsymbol{H}_*^k(\mathcal{E}^\vee) \to \boldsymbol{H}_*^k(\mathcal{C}))$ is Lagrangian. □

Remark 4.6 Lemma 4.5 gives an important obstruction to the existence of a self-dual resolution for a symmetric sheaf $\mathcal{C}$: if $n = 2k+1$ and k is even the class of the quadratic space $\boldsymbol{H}_*^k(\mathcal{C})$ in the Witt group $\boldsymbol{W}(\mathsf{k})$ must be trivial; in particular, the dimension of the space $\boldsymbol{H}_*^k(\mathcal{C})$ must be even. Note also that the latter condition is sufficient for the existence of a Lagrangian $\mathbb{S}$–submodule $\mathrm{A}^k \subset \boldsymbol{H}_*^k(\mathcal{C})$. Indeed, taking into account the twist in (4-3) we see that the subspace

$$\mathrm{A}^k = \begin{cases} \bigoplus_{t>(m-n-1)/2} H^k(\mathbb{P}^n, \mathcal{C}(t)) & \subset \boldsymbol{H}_*^k(\mathcal{C}) \quad \text{if } m-n-1 \text{ is odd}, \\ \bigoplus_{t>(m-n-1)/2} H^k(\mathbb{P}^n, \mathcal{C}(t)) \oplus \mathrm{A}^k_{(m-n-1)/2} & \subset \boldsymbol{H}_*^k(\mathcal{C}) \quad \text{if } m-n-1 \text{ is even}, \end{cases} \tag{4-4}$$

is an $\mathbb{S}$–submodule of $\boldsymbol{H}_*^k(\mathcal{C})$ and it is Lagrangian as soon as

$$\mathrm{A}^k_{(m-n-1)/2} \subset H^k\bigl(\mathbb{P}^n, \mathcal{C}\bigl(\tfrac{1}{2}(m-n-1)\bigr)\bigr)$$

is a Lagrangian subspace for the restriction of the pairing (4-3) (if $m-n-1$ is odd the Witt class of $\boldsymbol{H}_*^k(\mathcal{C})$ is trivial, and if $m-n-1$ is even, this class equals the class of the space $H^k\bigl(\mathbb{P}^n, \mathcal{C}\bigl(\tfrac{1}{2}(m-n-1)\bigr)\bigr)$, hence the latter has a Lagrangian subspace as soon as the Witt class of the former is trivial). Note also that the submodule $\mathrm{A}^k \subset \boldsymbol{H}_*^k(\mathcal{C})$ defined by (4-4) is shadowless in the sense of Definition 3.16.

The obstruction of Lemma 4.5 is well known to be nontrivial.

Example 4.7 Let $i \colon D \hookrightarrow \mathbb{P}^5$ be the so-called EPW sextic; see [8, Example 9.3]. Then there is a sheaf $\mathcal{R}$ on D such that $\mathcal{C} = i_*\mathcal{R}$ is symmetric, but $\dim(\boldsymbol{H}_*^2(\mathcal{C})) = 1$. Consequently, $\mathcal{C}$ does not admit a self-dual resolution.

The following fundamental result has been proved by Casnati and Catanese.

Theorem 4.8 [7, Theorem 0.3], [8, Theorem 9.1] *Let $\mathcal{C}$ be a (d,δ)–symmetric sheaf on $\mathbb{P}^n$. If $n = 2k+1$, and k and $m = d-\delta$ are even, assume that the class of the space $H^k\big(\mathbb{P}^n, \mathcal{C}\big(\tfrac{1}{2}(m-n-1)\big)\big)$ endowed with the quadratic form (4-3) is trivial in the Witt group $\boldsymbol{W}(\mathsf{k})$. Let $\mathrm{A}^k \subset \boldsymbol{H}^k_*(\mathcal{C})$ be any Lagrangian $\mathbb{S}$–submodule defined as in (4-4). Then there is a symmetric resolution (4-1) such that* $\mathrm{Im}(\boldsymbol{H}^k_*(\mathcal{E}^\vee) \to \boldsymbol{H}^k_*(\mathcal{C})) = \mathrm{A}^k$.

Remark 4.9 In fact, Theorem 4.8 has been proved in [7] for $n = 3$ and over an algebraically closed field, but as pointed out in [7, Remark 2.2], the proof applies to any n as soon as a Lagrangian subspace in $H^k\big(\mathbb{P}^n, \mathcal{C}\big(\tfrac{1}{2}(m-n-1)\big)\big)$ exists. For $n = 3$ this condition is automatically satisfied because the pairing (4-3) is skew-symmetric, but for $n = 2k+1$ with k even this becomes a nontrivial obstruction.

4.2 Elementary modifications

Throughout this section we fix a generically nondegenerate quadratic form $(\mathcal{E}, q)$ with its associated self-dual morphism $q\colon \mathcal{E}(-m) \to \mathcal{E}^\vee$.

We will need the following auxiliary result. Recall the exact sequence (3-12) and note that it can be considered as concatenation of short exact sequences

$$0 \to \textstyle\bigwedge^{p-1} \mathcal{T} \to \bigwedge^p V \otimes \mathcal{O}(p) \to \bigwedge^p \mathcal{T} \to 0. \tag{4-5}$$

We denote by $\tau_p \in \mathrm{Ext}^1\big(\bigwedge^p \mathcal{T}, \bigwedge^{p-1}\mathcal{T}\big) = \mathrm{Ext}^1(\Omega^{p-1}, \Omega^p)$ the extension class of (4-5).

The following observation is used to translate higher cohomology of $\mathcal{E}$ to hyperbolic extension classes. Recall from (3-11) the definition of the shadow $\mathbf{Sh}(p,t)$.

Proposition 4.10 *Let $\mathcal{E}$ be a vector bundle on $\mathbb{P}^n$. Let $1 \le p \le n-1$ and let $0 \ne \varepsilon_p \in H^p(\mathbb{P}^n, \mathcal{E}(t))$ be a cohomology class such that*

$$H^{p'}(\mathbb{P}^n, \mathcal{E}(t')) = 0 \quad \text{for any } (p', t') \in \mathbf{Sh}(p,t). \tag{4-6}$$

Then for $0 \le i \le p$ there exists a sequence of classes $\varepsilon_i \in H^i\big(\mathbb{P}^n, \mathcal{E}(t) \otimes \bigwedge^{p-i}\mathcal{T}\big) = \mathrm{Hom}(\Omega^{p-i}[-i], \mathcal{E}(t))$ that fit into a commutative diagram

$$\begin{tikzcd}
\mathcal{O}[-p] \arrow[r, "\tau_1"] \arrow[d, "\varepsilon_p"'] & \Omega^1[1-p] \arrow[r, "\tau_2"] \arrow[dl, "\varepsilon_{p-1}"'] & \cdots \arrow[r, "\tau_{p-1}"] \arrow[dll, phantom, "\cdots"] & \Omega^{p-1}[-1] \arrow[r, "\tau_p"] \arrow[dlll, "\varepsilon_1"'] & \Omega^p \arrow[r, hook] \arrow[dllll, bend left=10, "\varepsilon_0"] & \bigwedge^p V \otimes \mathcal{O}(-p) \arrow[dlllll, dotted, bend left=10] \\
\mathcal{E}(t)
\end{tikzcd}$$

where τ_i are the extension classes of the complexes (4-5); in other words,

$$\varepsilon_p = \varepsilon_i \circ \tau_{p-i} \circ \cdots \circ \tau_1 \tag{4-7}$$

for each $0 \le i \le p$. Moreover, for $i \ge 1$ such ε_i are unique, while ε_0 is unique up to a composition

$$\Omega^p \hookrightarrow \textstyle\bigwedge^p V \otimes \mathcal{O}(-p) \to \mathcal{E}(t),$$

where the first arrow is the canonical embedding.

Finally, if one of the following conditions is satisfied:

$$2p \le n, \quad \textit{or} \tag{4-8}$$

$$2p = n+1 \quad \textit{and} \quad 2t+m+n+1 \ge 0, \tag{4-9}$$

then $q(\varepsilon_p, \varepsilon_p) = 0$ *implies* $q(\varepsilon_{p-i}, \varepsilon_{p-i}) = 0$ *for each* $1 \le i \le p-1$, *where*

$$q(\varepsilon_{p-i}, \varepsilon_{p-i}) \in \operatorname{Ext}^{2(p-i)}\bigl(\Omega^i(-t-m), \textstyle\bigwedge^i \mathcal{T}(t)\bigr)$$

is defined as the composition $\Omega^i(-t-m)[i-p] \xrightarrow{\varepsilon_{p-i}} \mathcal{E}(-m) \xrightarrow{q} \mathcal{E}^\vee \xrightarrow{\varepsilon_{p-i}} \bigwedge^i \mathcal{T}(t)[p-i]$.

Proof The existence of ε_i satisfying (4-7) and their uniqueness follow by descending induction from the cohomology exact sequences of complexes (4-5) tensored with $\mathcal{E}(t)$ in view of the vanishing (4-6).

For the second assertion we also induct on i. Assume $1 \le i \le p-1$. We have

$$q(\varepsilon_{p-i}, \varepsilon_{p-i}) \in \operatorname{Ext}^{2(p-i)}(\Omega^i(-t-m), \textstyle\bigwedge^i \mathcal{T}(t)) = H^{2(p-i)}\bigl(\mathbb{P}^n, \bigwedge^i \mathcal{T} \otimes \bigwedge^i \mathcal{T}(2t+m)\bigr).$$

Consider the tensor square of (4-5):

$$0 \to \textstyle\bigwedge^{i-1} \mathcal{T} \otimes \bigwedge^{i-1} \mathcal{T} \to \bigl(\bigwedge^i V \otimes \bigwedge^{i-1} \mathcal{T}(i)\bigr)^{\oplus 2} \to \bigwedge^i V \otimes \bigwedge^i V \otimes \mathcal{O}(2i) \to \bigwedge^i \mathcal{T} \otimes \bigwedge^i \mathcal{T} \to 0. \tag{4-10}$$

Note that its extension class is $\tau_i \otimes \tau_i \in \operatorname{Ext}^2\bigl(\bigwedge^i \mathcal{T} \otimes \bigwedge^i \mathcal{T}, \bigwedge^{i-1} \mathcal{T} \otimes \bigwedge^{i-1} \mathcal{T}\bigr)$. Furthermore, we note that

$$\begin{aligned} H^{2(p-i)+1}\bigl(\mathbb{P}^n, \textstyle\bigwedge^{i-1} \mathcal{T}_{\mathbb{P}^n}(2t+m+i)\bigr) &= H^{2(p-i)+1}\bigl(\mathbb{P}^n, \Omega^{n-i+1}(2t+m+i+n+1)\bigr) = 0, \\ H^{2(p-i)}\bigl(\mathbb{P}^n, \mathcal{O}(2t+m+2i)\bigr) &= 0. \end{aligned}$$

Indeed, if (4-8) holds we use $1 \le 2(p-i)+1 < n-i+1 \le n$ together with (3-6) for the first vanishing and $1 \le 2(p-i) < n$ for the second. Similarly, if (4-9) holds and $i \ge 2$, the same arguments prove the vanishings. Finally, if (4-9) holds and $i = 1$, the same arguments prove the second vanishing, while the first cohomology space is equal to $H^n(\mathbb{P}^n, \mathcal{O}(2t+m+1))$, hence vanishes since $2t+m+1 > -n-1$.

The cohomology vanishings that we just established imply that the morphism

$$\tau_i \otimes \tau_i \colon H^{2(p-i)}\bigl(\mathbb{P}^n, \textstyle\bigwedge^i \mathcal{T} \otimes \bigwedge^i \mathcal{T}(2t+m)\bigr) \to H^{2(p-i+1)}\bigl(\mathbb{P}^n, \bigwedge^{i-1} \mathcal{T} \otimes \bigwedge^{i-1} \mathcal{T}(2t+m)\bigr)$$

is injective, and hence the condition

$$0 = q(\varepsilon_{p-i+1}, \varepsilon_{p-i+1}) = q(\varepsilon_{p-i} \circ \tau_i, \varepsilon_{p-i+1} \circ \tau_i) = (\tau_i \otimes \tau_i)(q(\varepsilon_{p-i}, \varepsilon_{p-i}))$$

implies $q(\varepsilon_{p-i}, \varepsilon_{p-i}) = 0$. ◻

The following *elementary modification* procedure allows us to kill an isotropic cohomology class of a quadratic form by a hyperbolic extension; see Section 2.2. For a cohomology class $\varepsilon_p \in H^p(\mathbb{P}^n, \mathcal{E}(t))$ we denote by $q(\varepsilon_p) \in H^p(\mathbb{P}^n, \mathcal{E}^\vee(m+t))$ the image of ε_p under the map

$$H^p(\mathbb{P}^n, \mathcal{E}(t)) \xrightarrow{q} H^p(\mathbb{P}^n, \mathcal{E}^\vee(m+t)).$$

Using the class $q(\varepsilon_p)$ we consider the map

$$\bigoplus_{i=1}^{n-1} \boldsymbol{H}^i_*(\mathcal{E}) \twoheadrightarrow H^{n-p}(\mathbb{P}^n, \mathcal{E}(-m-t-n-1)) \xrightarrow{q(\varepsilon_p)} H^n(\mathbb{P}^n, \mathcal{O}(-n-1)) = \mathsf{k}, \tag{4-11}$$

where the first arrow is the projection to a direct summand. We denote by $q(\varepsilon_p)^\perp \subset \bigoplus_{i=1}^{n-1} \boldsymbol{H}^i_*(\mathcal{E})$ the kernel of (4-11).

Proposition 4.11 *Let $0 \neq \varepsilon_p \in H^p(\mathbb{P}^n, \mathcal{E}(t))$ be a cohomology class such that $q(\varepsilon_p, \varepsilon_p) = 0$, and assume that the condition (4-6) holds and either (4-8) or (4-9) is satisfied. Let $\varepsilon_1 \in \operatorname{Ext}^1(\Omega^{p-1}(-t), \mathcal{E})$ be the extension class defined in Proposition 4.10. Then ε_1 is q–isotropic, and for any hyperbolic extension $(\mathcal{E}_+, q_+)$ of $(\mathcal{E}, q)$ with respect to ε_1, we have*

$$\bigoplus_{i=1}^{n-1} \boldsymbol{H}^i_*(\mathcal{E}_+) = \begin{cases} \left(q(\varepsilon_p)^\perp \cap \bigoplus_{i=1}^{n-1} \boldsymbol{H}^i_*(\mathcal{E})\right)/\mathsf{k}\varepsilon_p & \text{if } q(\varepsilon_p) \neq 0, \\ \mathsf{k}\varepsilon_+ \oplus \left(\bigoplus_{i=1}^{n-1} \boldsymbol{H}^i_*(\mathcal{E})\right)/\mathsf{k}\varepsilon_p & \text{if } q(\varepsilon_p) = 0, \end{cases} \tag{4-12}$$

where, in the second line, $\varepsilon_+ \in H^{n-p+1}(\mathbb{P}^n, \mathcal{E}_+(t+m+n+1))$ is a nonzero cohomology class that depends on the choice of $(\mathcal{E}_+, q_+)$.

Proof By Proposition 4.10, we have $q(\varepsilon_1, \varepsilon_1) = 0$, hence the extension class ε_1 is q–isotropic and a hyperbolic extension $(\mathcal{E}_+, q_+)$ exists by Theorem 2.9. By Lemma 2.4, its underlying bundle $\mathcal{E}_+$ has a length 3 filtration with the factors

$$\textstyle\bigwedge^{p-1} \mathcal{T}(t+m), \quad \mathcal{E}, \quad \Omega^{p-1}(-t)$$

linked by the classes $q(\varepsilon_1) \in \operatorname{Ext}^1\left(\mathcal{E}, \bigwedge^{p-1} \mathcal{T}(t+m)\right)$ and $\varepsilon_1 \in \operatorname{Ext}^1(\Omega^{p-1}(-t), \mathcal{E})$, respectively. Recall that τ_i denote the extension classes of complexes (4-5).

Consider the spectral sequence of a filtered complex that computes the cohomology of (twists of) $\mathcal{E}_+$; the terms of its first page $\boldsymbol{E}_1^{\bullet,\bullet}$ which compute intermediate cohomology look like

$$\begin{aligned} \boldsymbol{E}_1^{-1,i} &= \boldsymbol{H}^{i-1}_*(\Omega^{p-1}(-t)) = \begin{cases} \mathsf{k}(-t) & \text{if } i = p, \\ 0 & \text{otherwise, for } 2 \le i \le n, \end{cases} \\ \boldsymbol{E}_1^{0,i} &= \boldsymbol{H}^i_*(\mathcal{E}), \\ \boldsymbol{E}_1^{1,i} &= \boldsymbol{H}^{i+1}_*\left(\textstyle\bigwedge^{p-1} \mathcal{T}(t+m)\right) = \begin{cases} \mathsf{k}(t+m+n+1) & \text{if } i = n-p, \\ 0 & \text{otherwise, for } 0 \le i \le n-2, \end{cases} \end{aligned}$$

and the first differentials are given by $\varepsilon_1 \colon \boldsymbol{E}_1^{-1,i} \to \boldsymbol{E}_1^{0,i}$ and $q(\varepsilon_1) \colon \boldsymbol{E}_1^{0,i} \to \boldsymbol{E}_1^{1,i}$, respectively. In particular, there are only two possibly nontrivial differentials here:

$$\begin{gathered} \mathsf{k} \xrightarrow[\simeq]{\tau_{p-1} \circ \cdots \circ \tau_1} H^{p-1}(\mathbb{P}^n, \Omega^{p-1}) \xrightarrow{\varepsilon_1} H^p(\mathbb{P}^n, \mathcal{E}(t)), \\ H^{n-p}(\mathbb{P}^n, \mathcal{E}(-m-t-n-1)) \xrightarrow{q(\varepsilon_1)} H^{n-p+1}\left(\mathbb{P}^n, \textstyle\bigwedge^{p-1} \mathcal{T}(-n-1)\right) \xrightarrow[\simeq]{\tau_{p-1} \circ \cdots \circ \tau_1} \mathsf{k}. \end{gathered}$$

Since the spectral sequence is supported in three columns, the differential $\boldsymbol{d}_2$ acts as $\boldsymbol{E}_2^{-1,i} \to \boldsymbol{E}_2^{1,i-1}$, and using (3-6) we see that its source is nonzero only for $i \in \{1, n+1\}$ (note that $\varepsilon_1 \neq 0$), while its target is nonzero only in $i \in \{0, n+1-p, n\}$, hence $\boldsymbol{d}_2 = 0$. The further differentials a fortiori vanish, so that $\boldsymbol{E}_\infty^{\bullet,\bullet} = \boldsymbol{E}_2^{\bullet,\bullet}$. On the other hand, by (4-7) the image of the first map is $\Bbbk\varepsilon_p$ and the second map coincides with the map $q(\varepsilon_p)$ defined in (4-11). Therefore, the totalization of $\boldsymbol{E}_2^{\bullet,\bullet}$ takes the form of the right-hand side of (4-12), where in the case $q(\varepsilon_p) = 0$ the class ε_+ comes from $\boldsymbol{E}_2^{1,n-p}$, which survives exactly in this case. $\square$

As explained in Theorem 2.9 the construction of a hyperbolic extension might be ambiguous. In the situation described in Proposition 4.11 this happens precisely when the space $\operatorname{Ext}^1\big(\bigwedge^2 \Omega^{p-1}, \mathcal{O}(2t+m)\big)$ is nonzero. In the next lemma we determine when this happens.

Lemma 4.12 *Assume* $2 \le 2p \le n+1$. *Then the space* $\operatorname{Ext}^1\big(\bigwedge^2 \Omega^{p-1}, \mathcal{O}(s)\big)$ *is nonzero if and only if* $n = 2k+1$, $p = k+1$, $k \ge 1$ *is odd and* $s = -n-1$, *in which case* $\dim\big(\operatorname{Ext}^1\big(\bigwedge^2 \Omega^{p-1}, \mathcal{O}(s)\big)\big) = 1$.

Proof Set $k = p-1$, so that $2k \le n-1$. We have $\operatorname{Ext}^1\big(\bigwedge^2 \Omega^{p-1}, \mathcal{O}(s)\big) = H^1\big(\mathbb{P}^n, \bigwedge^2\big(\bigwedge^k \mathcal{T}\big) \otimes \mathcal{O}(s)\big)$. Taking the exterior square of (3-12) we see that $\bigwedge^2\big(\bigwedge^k \mathcal{T}\big)$ is quasiisomorphic to the complex of split bundles of length $2k$ if k is odd and $2k-1$ if k is even. Since split bundles have no intermediate cohomology and since $2k \le n-1$, the hypercohomology spectral sequence shows that $H^1\big(\mathbb{P}^n, \bigwedge^2\big(\bigwedge^k \mathcal{T}\big) \otimes \mathcal{O}(s)\big) = 0$ unless k is odd and $n = 2k+1$, and in the latter case we have

$$H^1\big(\mathbb{P}^n, \textstyle\bigwedge^2\big(\bigwedge^k \mathcal{T}\big) \otimes \mathcal{O}(s)\big) = \operatorname{Ker}\big(H^n(\mathbb{P}^n, \mathcal{O}(s)) \to H^n(\mathbb{P}^n, V \otimes \mathcal{O}(s+1))\big),$$

where the morphism in the right side is induced by the tautological embedding $\mathcal{O} \hookrightarrow V \otimes \mathcal{O}(1)$. Now it is easy to see that this space is zero unless $s = -n-1$, in which case it is one-dimensional. $\square$

Now let $\mathcal{C}$ be the cokernel sheaf of a generically nondegenerate quadratic form $(\mathcal{E}, q)$. The next result shows that in the case where the construction of an elementary modification of Proposition 4.11 is ambiguous, ie $\operatorname{Ext}^1\big(\bigwedge^2 \Omega^{p-1}, \mathcal{O}(2t+m)\big) \ne 0$, one can choose one such modification $(\mathcal{E}_+, q_+)$, which has an additional nice property, namely, it has a prescribed image of $\boldsymbol{H}_*^k(\mathcal{E}_+^\vee)$ in $\boldsymbol{H}_*^k(\mathcal{C})$. For our purposes it will be enough to consider the case where the bundle $\mathcal{E}$ is VLC; see Definition 3.7.

So, assume $n = 2k+1$ and the bundle $\mathcal{E}$ in (4-1) is VLC. Note that $\mathcal{E}^\vee$ is VUC by Lemma 3.9. By Lemma 3.12 we have an exact sequence of graded $\mathbb{S}$–modules

$$0 \to \boldsymbol{H}_*^k(\mathcal{E}^\vee) \to \boldsymbol{H}_*^k(\mathcal{C}) \to \boldsymbol{H}_*^{k+1}(\mathcal{E}) \to 0; \tag{4-13}$$

see (3-7). Recall also that the space $\boldsymbol{H}_*^k(\mathcal{C})$ is endowed with the perfect pairing (4-3) and that the subspace $\boldsymbol{H}_*^k(\mathcal{E}^\vee) \subset \boldsymbol{H}_*^k(\mathcal{C})$ is Lagrangian; see Lemma 4.5. In particular, the pairing induces an isomorphism

$$\boldsymbol{H}_*^{k+1}(\mathcal{E}) \cong \boldsymbol{H}_*^k(\mathcal{E}^\vee)^\vee$$

Using this isomorphism, any class $\varepsilon_{k+1} \in H^{k+1}(\mathbb{P}^n, \mathcal{E}(t))$ can be considered as a homogeneous linear function on $\boldsymbol{H}_*^k(\mathcal{E}^\vee)$; we denote by $\varepsilon_{k+1}^\perp \subset \boldsymbol{H}_*^k(\mathcal{E}^\vee)$ its kernel. Note that $\varepsilon_{k+1}^\perp$ is a graded isotropic $\mathbb{S}$–submodule in $\boldsymbol{H}_*^k(\mathcal{E}^\vee)$ and hence also in $\boldsymbol{H}_*^k(\mathcal{C})$.

Proposition 4.13 *Assume $k \ge 1$, $n = 2k+1$, and the bundle $\mathcal{E}$ in (4-1) is VLC. Let t be the maximal integer such that $H^{k+1}(\mathbb{P}^n, \mathcal{E}(t)) \neq 0$, let $\varepsilon_{k+1} \in H^{k+1}(\mathbb{P}^n, \mathcal{E}(t))$ be a cohomology nonzero class, and let $\varepsilon_{k+1}^{\perp} \subset \boldsymbol{H}_*^k(\mathcal{E}^\vee)$ be the corresponding graded isotropic $\mathbb{S}$–submodule. For each graded Lagrangian $\mathbb{S}$–submodule $\mathrm{A} \subset \boldsymbol{H}_*^k(\mathcal{C})$ such that $\mathrm{A} \neq \boldsymbol{H}_*^k(\mathcal{E}^\vee)$ and*

$$\varepsilon_{k+1}^{\perp} \subset \mathrm{A} \cap \boldsymbol{H}_*^k(\mathcal{E}^\vee), \tag{4-14}$$

there is a unique elementary modification $(\mathcal{E}_+, q_+)$ of $(\mathcal{E}, q)$ with respect to ε_{k+1} such that

$$\boldsymbol{H}_*^k(\mathcal{E}_+^\vee) = \mathrm{A}. \tag{4-15}$$

Proof Let $\varepsilon_1 \in \operatorname{Ext}^1(\Omega^k(-t), \mathcal{E})$ be the extension class constructed from ε_{k+1} in Proposition 4.10 and consider the variety $\mathrm{HE}(\mathcal{E}, q, \varepsilon_1)$ of all hyperbolic extensions of $(\mathcal{E}, q)$ with respect to ε_1, ie the set of all elementary modifications of $(\mathcal{E}, q)$ with respect to ε_{k+1}. By (4-12) every $(\mathcal{E}_+, q_+) \in \mathrm{HE}(\mathcal{E}, q, \varepsilon_1)$ is a VLC bundle, hence $\boldsymbol{H}_*^k(\mathcal{E}_+^\vee)$ is a graded Lagrangian $\mathbb{S}$–submodule in $\boldsymbol{H}_*^k(\mathcal{C})$ by Lemma 4.5. Moreover, the equality (4-12) also implies that $\varepsilon_{k+1}^{\perp} \subset \boldsymbol{H}_*^k(\mathcal{E}_+^\vee)$, ie $\mathrm{A} = \boldsymbol{H}_*^k(\mathcal{E}_+^\vee)$ satisfies (4-14). Therefore, there is a morphism

$$\lambda\colon \mathrm{HE}(\mathcal{E}, q, \varepsilon_1) \to \mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C})), \quad (\mathcal{E}_+, q_+) \mapsto [\boldsymbol{H}_*^k(\mathcal{E}_+^\vee)], \tag{4-16}$$

where $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C}))$ is the variety of all graded Lagrangian $\mathbb{S}$–submodules $\mathrm{A} \subset \boldsymbol{H}_*^k(\mathcal{C})$ satisfying (4-14). We will show that λ is an isomorphism onto the complement of the point $[\boldsymbol{H}_*^k(\mathcal{E}^\vee)]$ in $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C}))$.

First we check that the image of λ is contained in the complement of $[\boldsymbol{H}_*^k(\mathcal{E}^\vee)]$ in $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C}))$. Recall that $\varepsilon_1 \in \operatorname{Ext}^1(\Omega^k(-t), \mathcal{E})$ denotes the extension class constructed from ε_{k+1} in Proposition 4.10 and let, as usual, $q(\varepsilon_1) \in \operatorname{Ext}^1(\mathcal{E}, \Omega^{k+1}(t+m+n+1))$ be the class obtained from it by the application of q. Let $(\mathcal{E}_+, q_+)$ be any hyperbolic extension of $(\mathcal{E}, q)$ with respect to ε_1, so that $(\mathcal{E}, q)$ is the hyperbolic reduction of $(\mathcal{E}_+, q_+)$ with respect to an embedding $\Omega^{k+1}(t+m+n+1) \hookrightarrow \mathcal{E}_+$. Then we have the commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
 & \Omega^{k+1}(2t+m+n+1) \arrow[r, equal] \arrow[d] & \Omega^{k+1}(2t+m+n+1) \arrow[d] & & \\
0 \arrow[r] & \mathcal{E}''(t) \arrow[r] \arrow[d] & \mathcal{E}_+(t) \arrow[r] \arrow[d] & \Omega^k \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathcal{E}(t) \arrow[r] \arrow[d] & \mathcal{E}'(t) \arrow[r] \arrow[d] & \Omega^k \arrow[r] & 0 \\
 & 0 & 0 & &
\end{tikzcd} \tag{4-17}$$

with the extension class of the bottom row being ε_1 and that of the left column being $q(\varepsilon_1)$. Note that the cohomology exact sequence of the bottom row and the nontriviality of ε_1 imply that $\mathcal{E}'$ is VLC, hence $\mathcal{E}'^\vee$ is VUC. Similarly, the cohomology exact sequence of the left column implies that $\mathcal{E}''$ is VLC. We will use these observations below.

Consider the dual of the diagram (4-17) and the induced cohomology exact sequences

$$\begin{tikzcd} & & \mathsf{k}(-t-m) \arrow[r, equal] & \mathsf{k}(-t-m) \\ (4\text{-}18) & \mathsf{k}(t+n+1) \arrow[d, equal] & \boldsymbol{H}^k_*(\mathcal{E}''^\vee) \arrow[l, "\varepsilon''"'] \arrow[u, "\iota"] & \boldsymbol{H}^k_*(\mathcal{E}^\vee_+) \arrow[l] \arrow[u] \\ & \mathsf{k}(t+n+1) & \boldsymbol{H}^k_*(\mathcal{E}^\vee) \arrow[l, "\varepsilon_1"'] \arrow[u] & \boldsymbol{H}^k_*(\mathcal{E}'^\vee) \arrow[l] \arrow[u] \end{tikzcd}$$

(the map ι is induced by the embedding $\Omega^{k+1}(2t+m+n+1) \to \mathcal{E}''(t)$ in the left column of (4-17)). Since $\mathcal{E}'^\vee$ is VUC, the upper arrow in the right column of (4-18) is surjective. From the commutativity of the diagram we conclude that the composition

$$\boldsymbol{H}^k_*(\mathcal{E}^\vee_+) \to \boldsymbol{H}^k_*(\mathcal{E}''^\vee) \xrightarrow{\iota} \mathsf{k}$$

(of the right arrow in the middle row and the upper arrow in the middle column) is nontrivial, while

$$\boldsymbol{H}^k_*(\mathcal{E}^\vee) \to \boldsymbol{H}^k_*(\mathcal{E}''^\vee) \xrightarrow{\iota} \mathsf{k}$$

(the composition of arrows in the middle column) is zero. This proves that the images of $\boldsymbol{H}^k_*(\mathcal{E}^\vee_+)$ and $\boldsymbol{H}^k_*(\mathcal{E}^\vee)$ in $\boldsymbol{H}^k_*(\mathcal{E}''^\vee)$ are distinct. On the other hand, we have an obvious commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & \mathcal{E}(-m) \arrow[r, "q"] \arrow[d, hook] & \mathcal{E}^\vee \arrow[r] \arrow[d, hook] & \mathcal{C} \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & \mathcal{E}'(-m) \arrow[r] & \mathcal{E}''^\vee \arrow[r] & \mathcal{C} \arrow[r] & 0 \\ 0 \arrow[r] & \mathcal{E}_+(-m) \arrow[r, "q_+"] \arrow[u, two heads] & \mathcal{E}^\vee_+ \arrow[r] \arrow[u, two heads] & \mathcal{C} \arrow[r] \arrow[u, equal] & 0 \end{tikzcd}$$

and since $\mathcal{E}'$ is VLC, the map $\boldsymbol{H}^k_*(\mathcal{E}''^\vee) \to \boldsymbol{H}^k_*(\mathcal{C})$ is injective, hence

$$[\boldsymbol{H}^k_*(\mathcal{E}^\vee_+)] \neq [\boldsymbol{H}^k_*(\mathcal{E}^\vee)] \in \mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}^k_*(\mathcal{C})),$$

hence the image of λ is contained in the complement of $[\boldsymbol{H}^k_*(\mathcal{E}^\vee)]$.

Now we separate the following cases:

(a) k is even.

(b) $2t+m+n+1 \neq 0$.

(c) $2t+m+n+1 = 0$ and k is odd.

First, we describe the variety $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}^k_*(\mathcal{C})) \setminus [\boldsymbol{H}^k_*(\mathcal{E}^\vee)]$ in each case.

In case (a) the bilinear form (4-3) is symmetric, hence there exists exactly two Lagrangian subspaces in $\boldsymbol{H}^k_*(\mathcal{C})$ containing $\varepsilon_{k+1}^\perp$, hence $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}^k_*(\mathcal{C})) \setminus [\boldsymbol{H}^k_*(\mathcal{E}^\vee)]$ is a single point.

In case (b) the two-dimensional space $(\varepsilon_{k+1}^{\perp})^{\perp}/\varepsilon_{k+1}^{\perp}$ lives in two distinct degrees, hence it has exactly two *graded* Lagrangian subspaces, hence $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C})) \setminus [\boldsymbol{H}_*^k(\mathcal{E}^\vee)]$ is again a single point.

Finally, in case (c) the two-dimensional space $(\varepsilon_{k+1}^{\perp})^{\perp}/\varepsilon_{k+1}^{\perp}$ is symplectic and lives in a single degree, hence $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C})) \cong \mathbb{P}^1$ and $\mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C})) \setminus [\boldsymbol{H}_*^k(\mathcal{E}^\vee)] \cong \mathbb{A}^1$.

Now we see that in cases (a) and (b) the map λ is a map between two one-point sets, hence it is an isomorphism. Finally, in case (c) it is a map $\mathbb{A}^1 \to \mathbb{A}^1$, and to show it is an isomorphism (and thus to complete the proof of the proposition), it is enough to check its injectivity. So, for the rest of the proof we assume k is odd and $2t+m+n+1=0$ and prove that λ is injective.

First, we note that since the extension class $q(\varepsilon_1)$ of the left column of (4-17) does not depend on any choice, the hyperbolic extension $(\mathcal{E}_+, q_+)$ is determined by the class $\varepsilon'' \in \mathrm{Ext}^1(\Omega^k, \mathcal{E}''(t))$ of the middle row in (4-17), which is a lift of $\varepsilon_1 \in \mathrm{Ext}^1(\Omega^k, \mathcal{E}(t))$ with respect to the exact sequence

$$\text{(4-19)}\quad \mathrm{Hom}(\Omega^k, \mathcal{E}(t)) \to \mathrm{Ext}^1(\Omega^k, \Omega^{k+1}(2t+m+n+1)) \to \mathrm{Ext}^1(\Omega^k, \mathcal{E}''(t)) \to \mathrm{Ext}^1(\Omega^k, \mathcal{E}(t)).$$

Since the left arrow in the middle row of (4-18) is determined by the class ε'' and since $\lambda((\mathcal{E}_+, q_+))$ is its kernel, we conclude that the morphism λ factors as a composition

$$\mathrm{HE}(\mathcal{E}, q, \varepsilon_1) \xrightarrow{\lambda_1} \mathrm{Ext}^1(\Omega^k, \mathcal{E}''(t)) \xrightarrow{\lambda_2} \mathrm{LGr}_{\varepsilon_{k+1}}(\boldsymbol{H}_*^k(\mathcal{C})),$$

where λ_1 takes a hyperbolic extension $(\mathcal{E}_+, q_+)$ to the extension class of the middle row in (4-17) and λ_2 takes an extension class $\varepsilon'' \in \mathrm{Ext}^1(\Omega^k, \mathcal{E}''(t))$ to the kernel of the map ε'' in (4-18). To check the injectivity of λ it is enough to check the injectivity of λ_1 and λ_2.

To prove the injectivity of λ_2 consider the composition

$$\mathrm{Ext}^1(\Omega^k, \mathcal{E}''(t)) = H^{k+1}(\mathbb{P}^n, \mathcal{E}''(t)) \xrightarrow{\sim} \mathrm{Hom}\big(H^k(\mathbb{P}^n, \mathcal{E}''^\vee(-t-n-1)), \mathsf{k}\big) \subset \mathrm{Hom}(\boldsymbol{H}_*^k(\mathcal{E}''^\vee), \mathsf{k}(t+n+1)).$$

The equality follows from Proposition 4.10 applied to $\mathcal{E}''$ (recall that the bundle $\mathcal{E}''$ is VLC and, moreover, we have $H^{k+1}(\mathbb{P}^n, \mathcal{E}''(s)) = 0$ for $s > t$ by definition of t and the assumption $2t+m+n+1=0$) and the middle arrow is an isomorphism by Serre duality. Therefore, the composition is injective, and since $\lambda_2(\varepsilon'')$ is determined by the image of ε'' under this composition, we conclude that λ_2 is injective.

Finally, we note that $\mathrm{HE}(\mathcal{E}, q, \varepsilon_1)$ comes with a transitive action of the group

$$\mathrm{Ext}^1\big(\textstyle\bigwedge^2 \Omega^k, \mathcal{O}(2t+m)\big) \subset \mathrm{Ext}^1(\Omega^k, \Omega^{k+1}(2t+m+n+1)).$$

Therefore, to check the injectivity of λ_1, it is enough to check the injectivity of the middle arrow in (4-19). And for this, it is enough to check that the first arrow in (4-19) vanishes. To prove this vanishing consider the commutative square

$$\begin{array}{ccc} \mathrm{Hom}\big(\bigwedge^k V \otimes \mathcal{O}(-k), \mathcal{E}(t)\big) & \longrightarrow & \mathrm{Hom}(\Omega^k, \mathcal{E}(t)) \\ \downarrow & & \downarrow \\ \mathrm{Ext}^1\big(\bigwedge^k V \otimes \mathcal{O}(-k), \Omega^{k+1}(2t+m+n+1)\big) & \longrightarrow & \mathrm{Ext}^1(\Omega^k, \Omega^{k+1}(2t+m+n+1)) \end{array}$$

where the vertical arrows are induced by the extension class $q(\varepsilon_1)$ of the left column of (4-17), and the horizontal arrows are induced by the morphisms in the dual of (3-12) with $s = k$. The space in the lower left corner is zero by (3-6) (recall that $k \geq 1$), hence the compositions of arrows are zero. On the other hand, the argument of Proposition 4.10 shows that the top horizontal arrow is surjective. Therefore the right vertical arrow is zero, and as we explained above, this implies the injectivity of λ_1, and hence of λ, and completes the proof of the proposition. □

The elementary modification $(\mathcal{E}_+, q_+)$ of $(\mathcal{E}, q)$ satisfying the properties of Proposition 4.13 for a given Lagrangian $\mathbb{S}$–submodule $\mathrm{A} \subset \boldsymbol{H}^k_*(\mathcal{C})$ will be referred to as *the refined elementary modification*.

4.3 Modification theorem

Recall that a quadratic form $(\mathcal{E}, q)$ is called *unimodular* if the corresponding cokernel sheaf $\mathcal{C}$ vanishes, ie if $q\colon \mathcal{E}(-m) \to \mathcal{E}^\vee$ is an isomorphism. Recall the definitions (1-7) and (1-8) of standard unimodular quadratic forms. We will say that a standard unimodular quadratic form is *anisotropic* if $W = W^0$ and the form q_{W^0} is symmetric and anisotropic.

To prove the main result of this section we need the following simple observations. Recall the notion of linear minimality; see Definition 3.14.

Lemma 4.14 *Assume $(\mathcal{E}, q)$ is a generically nondegenerate quadratic form such that $q\colon \mathcal{E}(-m) \to \mathcal{E}^\vee$ is not linearly minimal. Then $(\mathcal{E}, q)$ is isomorphic to the orthogonal direct sum $(\mathcal{E}_0, q_0) \oplus (\mathcal{E}_1, q_1)$, where the second summand is a standard unimodular quadratic form (1-7) of rank 1 or 2.*

Proof Since q is not linearly minimal, it can be written as a direct sum of morphisms $f\colon \mathcal{E}'(-m) \to \mathcal{E}''$ and $\mathrm{id}\colon \mathcal{O}(t-m) \to \mathcal{O}(t-m)$ for some $t \in \mathbb{Z}$; in particular, $\mathcal{E} \cong \mathcal{E}' \oplus \mathcal{O}(t)$, and the restriction of q to the summand $\mathcal{O}(t-m)$ of $\mathcal{E}(-m)$ is a split monomorphism. Consider the composition

$$\varphi\colon \mathcal{O}(t-m) \hookrightarrow \mathcal{E}'(-m) \oplus \mathcal{O}(t-m) = \mathcal{E}(-m) \xrightarrow{q} \mathcal{E}^\vee = \mathcal{E}'^\vee \oplus \mathcal{O}(-t).$$

Let $\varphi_0\colon \mathcal{O}(t-m) \to \mathcal{E}'^\vee$ and $\varphi_1\colon \mathcal{O}(t-m) \to \mathcal{O}(-t)$ be its components. Since φ is a split monomorphism, there is a map $\psi = (\psi_0, \psi_1)\colon \mathcal{E}'^\vee \oplus \mathcal{O}(-t) \to \mathcal{O}(t-m)$ such that

$$\psi \circ \varphi = \psi_0 \circ \varphi_0 + \psi_1 \circ \varphi_1 = 1.$$

We consider the summand $\psi_1 \circ \varphi_1\colon \mathcal{O}(t-m) \to \mathcal{O}(t-m)$.

First, assume $\psi_1 \circ \varphi_1 \neq 0$. Then it is an isomorphism, hence φ_1 is a split monomorphism, hence an isomorphism, hence $t - m = -t$ and so $m = 2t$. Furthermore, it follows that the restriction of q to the subbundle $\mathcal{E}_1 = \mathcal{O}(t)$ of $\mathcal{E}$ is unimodular. Taking $\mathcal{E}_0 = \mathcal{E}_1^\perp$ to be the orthogonal of $\mathcal{E}_1$ in $\mathcal{E}$, we obtain the required direct sum decomposition.

Next, assume $\psi_1 \circ \varphi_1 = 0$. Then it follows that $\psi_0 \circ \varphi_0 = 1$, hence φ_0 is a split monomorphism. Therefore, we have $\mathcal{E}' \cong \mathcal{E}'' \oplus \mathcal{O}(m-t)$, so that $\mathcal{E} = \mathcal{E}'' \oplus \mathcal{O}(m-t) \oplus \mathcal{O}(t)$. Furthermore, it follows that the restriction of q to the subbundle $\mathcal{E}_1 = \mathcal{O}(m-t) \oplus \mathcal{O}(t)$ of $\mathcal{E}$ is unimodular (the restriction to $\mathcal{O}(m-t)$ is zero and the pairing between $\mathcal{O}(m-t)$ and $\mathcal{O}(t)$ is a nonzero constant). Taking $\mathcal{E}_0 = \mathcal{E}_1^{\perp}$ to be the orthogonal of $\mathcal{E}_1$ in $\mathcal{E}$, we obtain the required direct sum decomposition. □

Corollary 4.15 *If $(\mathcal{E}, q)$ is a unimodular quadratic form and $\mathcal{E}$ is VLC, then $(\mathcal{E}, q)$ is isomorphic to a standard unimodular quadratic form (1-7); in particular, $\mathcal{E}$ is split.*

Proof Since q is unimodular, we have $\mathcal{E}(-m) \cong \mathcal{E}^\vee$, so if $\mathcal{E}$ is VLC, and hence $\mathcal{E}^\vee$ is VUC, then $\mathcal{E}$ is both VLC and VUC, hence it is split by Lemma 3.9. Furthermore, $q\colon \mathcal{E}(-m) \to \mathcal{E}^\vee$ is an isomorphism of split bundles, hence it is not linearly minimal. Applying Lemma 4.14 we obtain a direct sum decomposition $\mathcal{E} = \mathcal{E}_0 \oplus \mathcal{E}_1$, where $\mathcal{E}_1$ is standard unimodular of type (1-7) and $\mathcal{E}_0$, being a direct summand of a unimodular VLC quadratic form, is itself unimodular and VLC. Iterating the argument, we conclude that $\mathcal{E}_0$ is standard unimodular of type (1-7), hence so is $\mathcal{E}$. □

Lemma 4.16 *If $(\mathcal{E}, q)$ is a standard unimodular quadratic form of type (1-7) or (1-8), it is hyperbolic equivalent to one of the following:*

- $(W^0, q_{W^0}) \otimes \mathcal{O}\big(\tfrac{1}{2}m\big)$ *if m is even, or*
- $(W^0, q_{W^0}) \otimes \Omega^{n/2}\big(\tfrac{1}{2}(m+n+1)\big)$ *if m is odd and n is divisible by 4,*

where in each case (W^0, q_{W^0}) is an anisotropic quadratic space; or to zero, otherwise.

Proof By definition of standard unimodular quadratic forms for each $i \neq 0$ the summands

$$W^i \otimes \mathcal{O}\big(\tfrac{1}{2}(m+i)\big) \oplus W^{-i} \otimes \mathcal{O}\big(\tfrac{1}{2}(m-i)\big),$$
$$W^i \otimes \Omega^{n/2}\big(\tfrac{1}{2}(m+m+1+i)\big) \oplus W^{-i} \otimes \Omega^{n/2}\big(\tfrac{1}{2}(m+n+1-i)\big)$$

are hyperbolic equivalent to zero, hence any standard unimodular quadratic form is hyperbolic equivalent to the one with $W = W^0$. It remains to note that by the standard Witt theory the bilinear form (W^0, q_{W^0}) is hyperbolic equivalent to an anisotropic form. Finally, in the case where m is odd and $n \equiv 2 \bmod 4$ the form q_{W^0} is skew-symmetric, so if it is anisotropic, it is just zero. □

Now we are ready to prove the main result of this section. Recall Definition 3.16.

Theorem 4.17 *Any generically nondegenerate quadratic form $q\colon \mathcal{E}(-m) \to \mathcal{E}^\vee$ over $\mathbb{P}^n$ is hyperbolic equivalent to an orthogonal direct sum*

$$(\mathcal{E}_{\min}, q_{\min}) \oplus (\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}}), \tag{4-20}$$

where $\mathcal{E}_{\min}$ is a VLC bundle, $(\mathcal{E}_{\min}, q_{\min})$ has no unimodular direct summands, and $(\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}})$ is an anisotropic standard unimodular quadratic form which has type (1-7) if m is even, type (1-8) if m is odd and $n \equiv 0 \bmod 4$, and is zero otherwise.

Moreover, if $n = 2k+1$, $\mathcal{C} = \mathcal{C}(q)$ is the cokernel sheaf of $(\mathcal{E}, q)$, and $\mathrm{A}^k \subset \boldsymbol{H}^k_*(\mathcal{C})$ is any shadowless subspace which is Lagrangian with respect to the bilinear form (4-3), then the quadratic form $(\mathcal{E}_{\min}, q_{\min})$ in (4-20) can be chosen in such a way that there is an equality $\boldsymbol{H}^k_*(\mathcal{E}^\vee_{\min}) = \mathrm{A}^k$ of $\mathbb{S}$–submodules in $\boldsymbol{H}^k_*(\mathcal{C})$.

Proof We split the proof into a number of steps.

Step 1 First we show that q is hyperbolic equivalent to a quadratic form $(\mathcal{E}_1, q_1)$ such that $\boldsymbol{H}^i_*(\mathcal{E}_1) = 0$ for each $1 \le i \le \lfloor \frac{1}{2}(n-1) \rfloor$ (if n is odd this is equivalent to the VLC property, and if n is even this is a bit weaker). For this we use induction on the parameter

$$\ell_1(\mathcal{E}) := \sum_{i=1}^{\lfloor (n-1)/2 \rfloor} \dim \boldsymbol{H}^i_*(\mathcal{E}).$$

Note that $\ell_1(\mathcal{E}) < \infty$ for any vector bundle $\mathcal{E}$.

Assume $\ell_1(\mathcal{E}) > 0$. Let $1 \le p_0 \le \lfloor \frac{1}{2}(n-1) \rfloor$ be the minimal integer such that $\boldsymbol{H}^{p_0}_*(\mathcal{E}) \neq 0$ and let t_0 be the maximal integer such that $H^{p_0}(\mathbb{P}^n, \mathcal{E}(t_0)) \neq 0$. Choose a nonzero element $\varepsilon_{p_0} \in H^{p_0}(\mathbb{P}^n, \mathcal{E}(t_0))$. Note that the class $q(\varepsilon_{p_0}, \varepsilon_{p_0}) \in H^{2p_0}(\mathbb{P}^n, \mathcal{O}(2t_0 + m))$ vanishes because $2 \le 2p_0 \le n-1$. Note also that the conditions (4-6) and (4-8) are satisfied for ε_{p_0}. Let $(\mathcal{E}_+, q_+)$ be the elementary modification of $(\mathcal{E}, q)$ with respect to ε_{p_0} constructed in Proposition 4.11. Then $(\mathcal{E}_+, q_+)$ is hyperbolic equivalent to $(\mathcal{E}, q)$ and the formula (4-12) implies that $\ell_1(\mathcal{E}_+) = \ell_1(\mathcal{E}) - 1$. Indeed, $n - p_0 + 1 > \lfloor \frac{1}{2}(n-1) \rfloor$, so even if the extra cohomology class ε_+ appears in $\boldsymbol{H}_*(\mathcal{E}_+)$ it does not contribute to $\ell_1(\mathcal{E}_+)$. By the induction hypothesis, the quadratic form $(\mathcal{E}_+, q_+)$ is hyperbolic equivalent to a quadratic form $(\mathcal{E}_1, q_1)$ such that $\boldsymbol{H}^i_*(\mathcal{E}_1) = 0$ for each $1 \le i \le \lfloor \frac{1}{2}(n-1) \rfloor$, hence so is $(\mathcal{E}, q)$.

From now on we assume that $\ell_1(\mathcal{E}) = 0$ and discuss separately the cases of even and odd n.

Step 2 Assume that $n = 2k$. In this case $\lfloor \frac{1}{2}(n-1) \rfloor = k - 1 < k = \lfloor \frac{1}{2} n \rfloor$, hence by Step 1 the only nontrivial intermediate cohomology of $\mathcal{E}$ preventing it from being VLC is $\boldsymbol{H}^k_*(\mathcal{E})$ and it fits into the exact sequence

$$0 \to \boldsymbol{H}^{k-1}_*(\mathcal{E}^\vee) \to \boldsymbol{H}^{k-1}_*(\mathcal{C}) \to \boldsymbol{H}^k_*(\mathcal{E}) \xrightarrow{\boldsymbol{H}^k_*(q)} \boldsymbol{H}^k_*(\mathcal{E}^\vee) \to \boldsymbol{H}^k_*(\mathcal{C}) \to \boldsymbol{H}^{k+1}_*(\mathcal{E}) \to 0,$$

where $\boldsymbol{H}^k_*(q)$ is the map induced by q. Note also that the combination of the morphism $\boldsymbol{H}^k_*(q)$ with the Serre duality pairing is a graded $\mathbb{S}$–bilinear form

$$\boldsymbol{H}^k_*(q) \colon \boldsymbol{H}^k_*(\mathcal{E}) \otimes \boldsymbol{H}^k_*(\mathcal{E}) \to \mathsf{k}(m + n + 1),$$

which is symmetric if k is even and skew-symmetric if k is odd.

First, we show that $(\mathcal{E}, q)$ is hyperbolic equivalent to a quadratic form $(\mathcal{E}', q')$ such that $\ell_1(\mathcal{E}') = 0$, the form $\boldsymbol{H}^k_*(q')$ is nondegenerate, and $H^k(\mathbb{P}^n, \mathcal{E}(t)) = 0$ unless $t = -\frac{1}{2}(m + n + 1)$.

If $\operatorname{Ker}(\boldsymbol{H}^k_*(q)) \neq 0$ let t be the maximal integer such that $\operatorname{Ker}(\boldsymbol{H}^k_*(q)) \cap H^k(\mathbb{P}^n, \mathcal{E}(t)) \neq 0$ and let ε_k be any nonzero class in this space (note that $q(\varepsilon_k, \varepsilon_k) = 0$); otherwise let t be the maximal integer such that $H^k(\mathbb{P}^n, \mathcal{E}(t)) \neq 0$ and let ε_k be any nonzero class in this space such that $q(\varepsilon_k, \varepsilon_k) = 0$ (if it exists). As before conditions (4-6) and (4-8) are satisfied for ε_k. Applying the elementary modification of Proposition 4.11 we obtain a quadratic form $(\mathcal{E}_+, q_+)$ hyperbolic equivalent to $(\mathcal{E}, q)$ and such that

$$\boldsymbol{H}^k_*(\mathcal{E}_+) \subset \boldsymbol{H}^k_*(\mathcal{E}) / \mathsf{k}\varepsilon_k \qquad \text{and} \qquad \boldsymbol{H}^p_*(\mathcal{E}_+) = \boldsymbol{H}^p_*(\mathcal{E}) = 0 \quad \text{for } 1 \le p < k.$$

In particular, $\ell_1(\mathcal{E}_+) = 0$ and $\dim(\boldsymbol{H}^k_*(\mathcal{E}_+)) < \dim(\boldsymbol{H}^k_*(\mathcal{E}))$. Iterating this argument we eventually obtain a quadratic form $(\mathcal{E}', q')$ such that $\ell_1(\mathcal{E}') = 0$, the form $\boldsymbol{H}^k_*(q')$ is nondegenerate, and if t is the maximal integer such that $H^k(\mathbb{P}^n, \mathcal{E}'(t)) \neq 0$ then $q'(\varepsilon_k, \varepsilon_k) \neq 0$ for any $0 \neq \varepsilon_k \in H^k(\mathbb{P}^n, \mathcal{E}'(t))$.

If $\boldsymbol{H}^k_*(\mathcal{E}') = 0$ there is nothing to prove anymore. Otherwise, the condition $q'(\varepsilon_k, \varepsilon_k) \neq 0$ implies that

$$2t + m + n + 1 = 0.$$

It remains to note that $H^k(\mathbb{P}^n, \mathcal{E}'(s)) = 0$ for $s \neq t$. Indeed, for $s > t$ the vanishing holds by definition of t. On the other hand, we have

$$H^k(\mathbb{P}^n, \mathcal{E}'(s)) = H^k(\mathbb{P}^n, \mathcal{E}'^{\vee}(m+s)) = H^k(\mathbb{P}^n, \mathcal{E}'(-s-m-n-1))^{\vee}$$

(the first equality follows from nondegeneracy of $\boldsymbol{H}^k_*(q')$ and the second from Serre duality), and as the right-hand side vanishes for $-s-m-n-1 > t$, the left-hand side vanishes for $s < -t-m-n-1 = t$.

Now, replacing $(\mathcal{E}, q)$ by $(\mathcal{E}', q')$, we may assume that $\boldsymbol{H}^k_*(q)$ is nondegenerate and $H^k(\mathbb{P}^n, \mathcal{E}(t)) = 0$ unless $t = -\frac{1}{2}(m+n+1)$. So, we set $t := -\frac{1}{2}(m+n+1)$ and let

$$\boldsymbol{\varepsilon}_k \colon H^k(\mathbb{P}^n, \mathcal{E}(t)) \otimes \mathcal{O}(-t)[-k] = \operatorname{Ext}^k(\mathcal{O}(-t), \mathcal{E}) \otimes \mathcal{O}(-t)[-k] \to \mathcal{E}$$

be the evaluation morphism in the derived category. Let

$$\boldsymbol{\varepsilon}_0 \colon H^k(\mathbb{P}^n, \mathcal{E}(t)) \otimes \Omega^k(-t) = \operatorname{Hom}(\Omega^k(-t), \mathcal{E}) \otimes \Omega^k(-t) \to \mathcal{E}$$

be the morphism constructed from $\boldsymbol{\varepsilon}_k$ in Proposition 4.10. Consider the composition

$$H^k(\mathbb{P}^n, \mathcal{E}(t)) \otimes \Omega^k(-t) \xrightarrow{\boldsymbol{\varepsilon}_0} \mathcal{E} \xrightarrow{q} \mathcal{E}^{\vee}(m) \xrightarrow{\boldsymbol{\varepsilon}_0^{\vee}} H^k(\mathbb{P}^n, \mathcal{E}(t))^{\vee} \otimes \textstyle\bigwedge^k \mathcal{T}(t+m). \tag{4-21}$$

By Proposition 4.10 and Serre duality the first and last arrows in the composition

$$\boldsymbol{H}^k_*\big(H^k(\mathbb{P}^n, \mathcal{E}(t)) \otimes \Omega^k(-t)\big) \xrightarrow{\boldsymbol{\varepsilon}_0} \boldsymbol{H}^k_*(\mathcal{E}) \xrightarrow{\boldsymbol{H}^k_*(q)} \boldsymbol{H}^k_*(\mathcal{E}^{\vee}) \xrightarrow{\boldsymbol{\varepsilon}_0^{\vee}} \boldsymbol{H}^k_*\big(H^k(\mathbb{P}^n, \mathcal{E}(t))^{\vee} \otimes \textstyle\bigwedge^k \mathcal{T}(t+m)\big)$$

are isomorphisms, while the middle arrow is an isomorphism by nondegeneracy of $\boldsymbol{H}^k_*(q)$. It follows that the composition (4-21) is an isomorphism; note that $\bigwedge^k \mathcal{T}(t+m) \cong \Omega^k(t+m+n+1) \cong \Omega^k(-t)$ since $n = 2k$ and $2t+m+n+1 = 0$. This means that $\boldsymbol{\varepsilon}_0$ is a split monomorphism, ie

$$\mathcal{E} \cong \mathcal{E}_0 \oplus \mathcal{E}_1, \quad \text{where } \mathcal{E}_1 = H^k(\mathbb{P}^n, \mathcal{E}(t)) \otimes \Omega^k(-t),$$

so that $\mathcal{E}_1$ is a standard unimodular bundle of type (1-8) and $\mathcal{E}_0$ is the orthogonal of $\mathcal{E}_1$ with respect to the quadratic form q. From the direct sum decomposition it easily follows that $\mathcal{E}_0$ is VLC.

Step 3 Assume $n = 2k+1$. In this case $\lfloor \frac{1}{2}(n-1) \rfloor = k = \lfloor \frac{1}{2} n \rfloor$, hence by Step 1 the bundle $\mathcal{E}$ is already VLC. It remains to find a VLC quadratic form $(\mathcal{E}_0, q_0)$ hyperbolic equivalent to $(\mathcal{E}, q)$ with $\boldsymbol{H}^k_*(\mathcal{E}_0^{\vee}) = \mathrm{A}^k$, where recall that $\mathrm{A}^k \subset \boldsymbol{H}^k_*(\mathcal{C})$ is a given shadowless Lagrangian subspace. To construct $\mathcal{E}_0$ we induct on the parameter

$$\ell_{\mathrm{A}}(\mathcal{E}) := \dim\big(\operatorname{Im}(\mathrm{A}^k \to \boldsymbol{H}^{k+1}_*(\mathcal{E}))\big) = \operatorname{codim}_{\mathrm{A}^k}(\boldsymbol{H}^k_*(\mathcal{E}^{\vee}) \cap \mathrm{A}^k), \tag{4-22}$$

where the equality follows from the exact sequence (4-13).

Assume $\ell_{\mathrm{A}}(\mathcal{E}) > 0$, choose a nonzero homogeneous element $\varepsilon_{k+1} \in \operatorname{Im}(\mathrm{A}^k \to \boldsymbol{H}^{k+1}_*(\mathcal{E}))$ of maximal degree and let $\tilde{\varepsilon}_{k+1}$ be its arbitrary homogeneous lift to $\mathrm{A}^k \subset \boldsymbol{H}^k_*(\mathcal{C})$. Recall the definition of the hyperplane $\varepsilon_{k+1}^{\perp} \subset \boldsymbol{H}^k_*(\mathcal{E}^\vee)$ that was given before Proposition 4.13. Note that

$$\varepsilon_{k+1}^{\perp} = \boldsymbol{H}^k_*(\mathcal{E}^\vee) \cap \tilde{\varepsilon}_{k+1}^{\perp},$$

where $\tilde{\varepsilon}_{k+1}^{\perp} \subset \boldsymbol{H}^k_*(\mathcal{C})$ is the hyperplane orthogonal of $\tilde{\varepsilon}_{k+1}$ with respect to the perfect pairing (4-3); in particular $\varepsilon_{k+1}^{\perp}$ is isotropic and orthogonal to $\tilde{\varepsilon}_{k+1}$, hence the subspace

$$\mathrm{A}^k_+ := \varepsilon_{k+1}^{\perp} \oplus \mathsf{k}\tilde{\varepsilon}_{k+1} \subset \boldsymbol{H}^k_*(\mathcal{C})$$

is Lagrangian. Applying Proposition 4.13 we conclude that there exists a refined elementary modification $(\mathcal{E}_+, q_+)$ of $(\mathcal{E}, q)$ which is VLC and has the property $\boldsymbol{H}^k_*(\mathcal{E}_+^\vee) = \mathrm{A}^k_+$. Now it is easy to see that

$$\boldsymbol{H}^k_*(\mathcal{E}^\vee) \cap \mathrm{A}^k \subset \boldsymbol{H}^k_*(\mathcal{E}^\vee) \cap \tilde{\varepsilon}_{k+1}^{\perp} = \varepsilon_{k+1}^{\perp} \subset \mathrm{A}^k_+$$

(the first inclusion follows from $\tilde{\varepsilon}_{k+1} \in \mathrm{A}^k$ since A^k is Lagrangian) and

$$\tilde{\varepsilon}_{k+1} \in (\mathrm{A}^k_+ \cap \mathrm{A}^k) \setminus (\boldsymbol{H}^k_*(\mathcal{E}^\vee) \cap \mathrm{A}^k),$$

hence $\dim(\mathrm{A}^k_+ \cap \mathrm{A}^k) > \dim(\boldsymbol{H}^k_*(\mathcal{E}^\vee) \cap \mathrm{A}^k)$ and so $\ell_{\mathrm{A}}(\mathcal{E}_+) < \ell_{\mathrm{A}}(\mathcal{E})$. By the induction hypothesis the quadratic form $(\mathcal{E}_+, q_+)$ is hyperbolic equivalent to $(\mathcal{E}_0, q_0)$ such that $\boldsymbol{H}^k_*(\mathcal{E}_0^\vee) = \mathrm{A}^k$, hence so is $(\mathcal{E}, q)$.

Step 4 We already have proved that the quadratic form $(\mathcal{E}, q)$ is hyperbolic equivalent to an orthogonal direct sum $(\mathcal{E}_0, q_0) \oplus (\mathcal{E}_1, q_1)$, where $\mathcal{E}_0$ is VLC (with prescribed Lagrangian subspace $\boldsymbol{H}^k_*(\mathcal{E}_0^\vee) \subset \boldsymbol{H}^k_*(\mathcal{C})$ if $n = 2k+1$) and $(\mathcal{E}_1, q_1)$ is standard unimodular of type (1-8) (if $n = 2k$).

Obviously we can write $(\mathcal{E}_0, q_0) \cong (\mathcal{E}_{\mathrm{min}}, q_{\mathrm{min}}) \oplus (\mathcal{E}_2, q_2)$, where $(\mathcal{E}_{\mathrm{min}}, q_{\mathrm{min}})$ has no unimodular direct summands and $(\mathcal{E}_2, q_2)$ is unimodular. Then, defining

$$(\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}}) := (\mathcal{E}_1, q_1) \oplus (\mathcal{E}_2, q_2),$$

we obtain a decomposition of type (4-20), and it remains to modify it slightly.

First, note that since $\mathcal{E}_2$ is a direct summand of $\mathcal{E}_0$, it is VLC, hence $(\mathcal{E}_2, q_2)$ is standard of type (1-7) by Corollary 4.15. Second, by Lemma 4.16 we can replace the summands $(\mathcal{E}_1, q_1)$ and $(\mathcal{E}_2, q_2)$ above by summands of the same type with $W^i = 0$ for $i \neq 0$ and q_{W^0} anisotropic. It remains to note that we have $0 = i \equiv m \bmod 2$ for the summand of type (1-7) and $0 = i \equiv m + n + 1 \equiv m + 1 \bmod 2$ and n is even for the summand of type (1-8). Moreover, in the latter case, if n is not divisible by 4, the form q_{W^0} is skew-symmetric, and since it is also anisotropic, $W^0 = 0$. Thus, we obtain the required description of the summand $(\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}})$. □

In the following corollary we deduce from Theorem 4.17 a generalization of the result of Arason [2] about the untwisted unimodular Witt group $\boldsymbol{W}(\mathbb{P}^n, \mathcal{O})$ of a projective space to the case of the twisted unimodular group $\boldsymbol{W}(\mathbb{P}^n, \mathcal{O}(m))$; thus reproving a result of Walter [14] (see also [4, Theorem 1.5.28]) in the special case of trivial base.

Corollary 4.18 *If m is even, or if m is odd and n is divisible by 4, one has $\boldsymbol{W}(\mathbb{P}^n, \mathcal{O}(m)) \cong \boldsymbol{W}(\mathsf{k})$. Otherwise, $\boldsymbol{W}(\mathbb{P}^n, \mathcal{O}(m)) = 0$.*

Proof By Theorem 4.17 a unimodular quadratic form $(\mathcal{E}, q)$ on $\mathbb{P}^n$ is hyperbolic equivalent to a sum (4-20). The summand $(\mathcal{E}_{\min}, q_{\min})$ is unimodular (as a direct summand of a unimodular quadratic form) and has no unimodular summands by assumption, hence $(\mathcal{E}_{\min}, q_{\min}) = 0$ and $(\mathcal{E}, q) = (\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}})$ is an anisotropic standard unimodular quadratic form of type (1-7) or (1-8).

If m is even, $(\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}}) \cong (W^0, q_{W^0}) \otimes \mathcal{O}\big(\tfrac{1}{2}m\big)$ and $\mathrm{w}_x(\mathcal{E}, q) = \mathrm{w}_x(\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}}) = [(W^0, q_{W^0})]$ — the first equality follows from Lemma 2.13, and for the second to be true one has to choose the trivialization of $\mathcal{O}(m)_x$ to be induced by a trivialization of $\mathcal{O}\big(\tfrac{1}{2}m\big)_x$ — for any k–point $x \in \mathbb{P}^n$. Therefore, the group homomorphism

$$\mathrm{w}_x \colon \boldsymbol{W}(\mathbb{P}^n, \mathcal{O}(m)) \to \boldsymbol{W}(\mathsf{k})$$

is injective. On the other hand, it is obviously surjective, hence it is an isomorphism.

Next, assume m is odd and n is divisible by 4. Then $(\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}}) = (W^0, q_{W^0}) \otimes \Omega^{n/2}\big(\tfrac{1}{2}(m+n+1)\big)$, and $\mathrm{hw}(\mathcal{E}, q) = \mathrm{hw}(\mathcal{E}_{\mathrm{uni}}, q_{\mathrm{uni}}) = [(W^0, q_{W^0})]$ (the first equality follows from Lemma 2.14), hence the group homomorphism

$$\mathrm{hw} \colon \boldsymbol{W}(\mathbb{P}^n, \mathcal{O}(m)) \to \boldsymbol{W}(\mathsf{k})$$

is injective. On the other hand, it is obviously surjective, hence it is an isomorphism.

In the remaining cases, $\mathcal{E}_{\mathrm{uni}} = 0$ by Theorem 4.17, hence $\boldsymbol{W}(\mathbb{P}^n, \mathcal{O}(m)) = 0$. □

4.4 Proof of Theorem 1.3 and Corollary 1.5

In this final subsection we prove Theorem 1.3 and Corollary 1.5 from the introduction. The next proposition provides the crucial step.

Proposition 4.19 *Assume $(\mathcal{E}_1, q_1)$ and $(\mathcal{E}_2, q_2)$ are generically nondegenerate quadratic forms which have no unimodular direct summands and such that the bundles $\mathcal{E}_i$ are VLC. Let $\varphi \colon \mathcal{C}(q_1) \xrightarrow{\simeq} \mathcal{C}(q_2)$ be an isomorphism of their cokernel sheaves compatible with their induced shifted quadratic forms (1-3). If $n = 2k+1$, assume also that $\boldsymbol{H}^k_*(\varphi)$ identifies the Lagrangian subspaces $\boldsymbol{H}^k_*(\mathcal{E}_i^\vee) \subset \boldsymbol{H}^k_*(\mathcal{C}(q_i))$. Then φ is induced by a unique isomorphism $(\mathcal{E}_1, q_1) \cong (\mathcal{E}_2, q_2)$ of quadratic forms.*

Proof By Lemma 4.14 the VHC morphisms $\mathcal{E}_i(-m) \xrightarrow{q_i} \mathcal{E}_i^\vee$ are linearly minimal resolutions of the sheaves $\mathcal{C}(q_i)$, hence by Theorem 3.15 they are isomorphic, ie there is a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{E}_1(-m) \arrow[r, "q_1"] \arrow[d, "\varphi_{\mathrm{L}}"'] & \mathcal{E}_1^\vee \arrow[r] \arrow[d, "\varphi_{\mathrm{U}}"'] \arrow[dl, dotted] & \mathcal{C}(q_1) \arrow[r] \arrow[d, "\varphi"'] & 0 \\
0 \arrow[r] & \mathcal{E}_2(-m) \arrow[r, "q_2"] & \mathcal{E}_2^\vee \arrow[r] & \mathcal{C}(q_2) \arrow[r] & 0
\end{tikzcd}$$

where φ_{L} and φ_{U} are isomorphisms. Moreover, such diagram is unique up to a homotopy represented by the dotted arrow.

From now on we identify $\mathcal{E}_2$ with $\mathcal{E}_1$ by means of φ_{L}, so we assume $\mathcal{E}_1 = \mathcal{E}_2$ and $\varphi_{\mathrm{L}} = \mathrm{id}$. Now consider the dual diagram, and then invert its vertical arrows:

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{E}_1(-m) \arrow[r, "q_1"] \arrow[d, "(\varphi_{\mathrm{U}}^{\vee})^{-1}"'] & \mathcal{E}_1^{\vee} \arrow[r] \arrow[d, "\mathrm{id}"'] \arrow[dl, dotted, "h"'] & \mathcal{C}(q_1) \arrow[r] \arrow[d, "(\varphi^{\vee})^{-1}"'] & 0 \\
0 \arrow[r] & \mathcal{E}_2(-m) \arrow[r, "q_2"] & \mathcal{E}_2^{\vee} \arrow[r] & \mathcal{C}(q_2) \arrow[r] & 0
\end{tikzcd}$$

Since φ is compatible with the shifted quadratic forms on $\mathcal{C}(q_i)$, we have $(\varphi^{\vee})^{-1} = \varphi$, hence by the uniqueness property of the diagram, there is a homotopy h such that

$$\varphi_{\mathrm{U}} = \mathrm{id} + q_2 \circ h \quad \text{and} \quad \mathrm{id} = h \circ q_1 + (\varphi_{\mathrm{U}}^{\vee})^{-1}.$$

Now note that the endomorphism $q_2 \circ h$ of $\mathcal{E}_1^{\vee} = \mathcal{E}_2^{\vee}$ is nilpotent, again by Theorem 3.15. Therefore, φ_{U} is unipotent. On the other hand, commutativity of the first diagram (with the convention $\varphi_{\mathrm{L}} = \mathrm{id}$ taken into account) means that

$$q_2 = \varphi_{\mathrm{U}} \circ q_1,$$

and since q_2 is self-dual, it follows that φ_{U} is self-adjoint with respect to q_1.

Now note that if a unipotent operator over a field is self-adjoint with respect to a nondegenerate quadratic form, it is the identity. Indeed, to prove this we can pass to an algebraic closure of the field, then the operator can be diagonalized, and a diagonal operator is unipotent only if it is the identity.

The above argument thus shows that φ_{U} restricted to the generic point of $\mathbb{P}^n$ is the identity. Finally, since the bundle $\mathcal{E}_2 = \mathcal{E}_1$ is torsion free and φ_{U} is an automorphism, it follows that φ_{U} is the identity. Thus, $q_2 = q_1$, ie the quadratic forms $(\mathcal{E}_i, q_i)$ are isomorphic. □

Now we can deduce the theorem.

Proof of Theorem 1.3 Let $\mathcal{C} = \mathcal{C}(q_1) = \mathcal{C}(q_2)$. If $n = 2k+1$, define $\mathrm{A}^k \subset \boldsymbol{H}_*^k(\mathcal{C})$ by the formula (4-4), where, if $m-n-1$ is even,

$$\mathrm{A}^k_{(m-n-1)/2} := H^k\big(\mathbb{P}^n, \mathcal{E}_2^{\vee}\big(\tfrac{1}{2}(m-n-1)\big)\big) \subset H^k\big(\mathbb{P}^n, \mathcal{C}\big(\tfrac{1}{2}(m-n-1)\big)\big);$$

this is a shadowless Lagrangian $\mathbb{S}$–submodule as explained in Remark 4.6. By Theorem 4.17 the quadratic forms $(\mathcal{E}_i, q_i)$ are hyperbolic equivalent to orthogonal direct sums

$$(\mathcal{E}_{i,\min}, q_{i,\min}) \oplus (\mathcal{E}_{i,\mathrm{uni}}, q_{i,\mathrm{uni}}),$$

where $\mathcal{E}_{i,\min}$ are VLC bundles, $(\mathcal{E}_{i,\min}, q_{i,\min})$ have no unimodular direct summands, $\boldsymbol{H}_*^k(\mathcal{E}_{i,\min}^{\vee}) = \mathrm{A}^k$ if $n = 2k+1$, and $(\mathcal{E}_{i,\mathrm{uni}}, q_{i,\mathrm{uni}})$ are anisotropic standard unimodular quadratic forms (1-7) or (1-8).

Since the summands $(\mathcal{E}_{i,\mathrm{uni}}, q_{i,\mathrm{uni}})$ are unimodular, ie $\mathcal{C}(q_{i,\mathrm{uni}}) = 0$, we have

$$\mathcal{C}(q_{i,\min}) = \mathcal{C}(q_{i,\min} \oplus q_{i,\mathrm{uni}}) \cong \mathcal{C}(q_i) = \mathcal{C},$$

where the isomorphism is induced by the hyperbolic equivalence (see Proposition 1.1(1)); hence it is compatible with the shifted quadratic forms (1-3). Moreover, we have $\boldsymbol{H}^k_*(\mathcal{E}^\vee_{1,\min}) = \boldsymbol{H}^k_*(\mathcal{E}^\vee_{2,\min})$ if $n = 2k+1$. Applying Proposition 4.19, we conclude that

$$(\mathcal{E}_{1,\min}, q_{1,\min}) \cong (\mathcal{E}_{2,\min}, q_{2,\min}). \tag{4-23}$$

Next, we identify the unimodular summands $(\mathcal{E}_{1,\mathrm{uni}}, q_{1,\mathrm{uni}})$ and $(\mathcal{E}_{2,\mathrm{uni}}, q_{2,\mathrm{uni}})$.

If m is even, the unimodular summands have type (1-7), ie they can be written as

$$(\mathcal{E}_{i,\mathrm{uni}}, q_{i,\mathrm{uni}}) \cong (W_i^0, q_{W_i^0}) \otimes \mathcal{O}\big(\tfrac{1}{2}m\big),$$

where $(W_i^0, q_{W_i^0})$ are anisotropic. Moreover, we have

$$\mathrm{w}_X(\mathcal{E}_i, q_i) = \mathrm{w}_X(\mathcal{E}_{i,\min}, q_{i,\min}) + \mathrm{w}_X(\mathcal{E}_{i,\mathrm{uni}}, q_{i,\mathrm{uni}}) = \mathrm{w}_X(\mathcal{E}_{i,\min}, q_{i,\min}) + [(W_i^0, q_{W_i^0})],$$

hence (4-23) and the equality $\mathrm{w}_X(\mathcal{E}_1, q_1) = \mathrm{w}_X(\mathcal{E}_2, q_2)$ imply the equality $[(W_1^0, q_{W_1^0})] = [(W_2^0, q_{W_2^0})]$ of the Witt classes of the quadratic spaces $(W_i^0, q_{W_i^0})$. But since these quadratic spaces are anisotropic, they are isomorphic, hence we have $(\mathcal{E}_{1,\mathrm{uni}}, q_{1,\mathrm{uni}}) \cong (\mathcal{E}_{2,\mathrm{uni}}, q_{2,\mathrm{uni}})$.

If m is odd and n is divisible by 4, the unimodular summands have type (1-8), ie they can be written as

$$(\mathcal{E}_{i,\mathrm{uni}}, q_{i,\mathrm{uni}}) \cong (W_i^0, q_{W_i^0}) \otimes \Omega^{n/2}\big(\tfrac{1}{2}(m+n+1)\big),$$

where again $(W_i^0, q_{W_i^0})$ are anisotropic. Moreover, we have

$$\mathrm{hw}(\mathcal{E}_i, q_i) = \mathrm{hw}(\mathcal{E}_{i,\min}, q_{i,\min}) + \mathrm{hw}(\mathcal{E}_{i,\mathrm{uni}}, q_{i,\mathrm{uni}}) = \mathrm{hw}(\mathcal{E}_{i,\min}, q_{i,\min}) + [(W_i^0, q_{W_i^0})],$$

hence (4-23) and the equality $\mathrm{hw}(\mathcal{E}_1, q_1) = \mathrm{hw}(\mathcal{E}_2, q_2)$ imply the equality $[(W_1^0, q_{W_1^0})] = [(W_2^0, q_{W_2^0})]$ of the Witt classes of the quadratic spaces $(W_i^0, q_{W_i^0})$. Again, since these quadratic spaces are anisotropic, they are isomorphic, hence we have $(\mathcal{E}_{1,\mathrm{uni}}, q_{1,\mathrm{uni}}) \cong (\mathcal{E}_{2,\mathrm{uni}}, q_{2,\mathrm{uni}})$.

Finally, if m is odd and n is not divisible by 4, we have $(\mathcal{E}_{i,\min}, q_{i,\min}) = 0$ for $i = 1, 2$.

Thus, in all the cases we have $(\mathcal{E}_{1,\mathrm{uni}}, q_{1,\mathrm{uni}}) \cong (\mathcal{E}_{2,\mathrm{uni}}, q_{2,\mathrm{uni}})$. Combining this isomorphism with (4-23), we obtain a chain of hyperbolic equivalences and isomorphisms

$$(\mathcal{E}_1, q_1) \overset{\mathrm{he}}{\sim} (\mathcal{E}_{1,\min}, q_{1,\min}) \oplus (\mathcal{E}_{1,\mathrm{uni}}, q_{1,\mathrm{uni}}) \cong (\mathcal{E}_{2,\min}, q_{2,\min}) \oplus (\mathcal{E}_{2,\mathrm{uni}}, q_{2,\mathrm{uni}}) \overset{\mathrm{he}}{\sim} (\mathcal{E}_2, q_2),$$

and conclude that $(\mathcal{E}_1, q_1)$ is hyperbolic equivalent to $(\mathcal{E}_2, q_2)$. □

Before proving Corollary 1.5 let us recall the definitions of the discriminant double cover and root stack associated with a generically nondegenerate quadric bundle $Q \subset \mathbb{P}_X(\mathcal{E})$, and of the Brauer classes on their corank ≤ 1 loci.

First, assume that $\dim(Q/X) \equiv 0 \bmod 2$. The determinant of the morphism $q\colon \mathcal{E} \otimes \mathcal{L} \to \mathcal{E}^\vee$ is a nonzero global section $\det(q)$ of the line bundle $((\det \mathcal{E})^{\otimes 2} \otimes \mathcal{L}^{\otimes \mathrm{rk}(\mathcal{E})})^\vee$; it defines a $\mathbb{Z}/2$–graded commutative $\mathcal{O}_X$–algebra structure on the sheaf $\mathcal{O}_X \oplus \det \mathcal{E} \otimes \mathcal{L}^{\otimes \mathrm{rk}(\mathcal{E})/2}$, and the determinant double cover is defined as its relative spectrum

$$S = \mathrm{Spec}_X(\mathcal{O}_X \oplus \det \mathcal{E} \otimes \mathcal{L}^{\otimes \mathrm{rk}(\mathcal{E})/2}).$$

On the other hand, by [11, Section 3.5] the algebra $\mathcal{O}_X \oplus \det \mathcal{E} \otimes \mathcal{L}^{\otimes \mathrm{rk}(\mathcal{E})/2}$ is identified with the center of the even part of the Clifford algebra $\mathrm{Cliff}_0(\mathcal{E}, q)$, which therefore can be written as the pushforward of an $\mathcal{O}_S$–algebra $\mathcal{B}_0$ from S. Moreover, the restriction of $\mathcal{B}_0$ to the open subset $S_{\le 1} \subset S$, the preimage of the open subset $X_{\le 1} \subset X$ parametrizing quadrics of corank at most one in the quadric bundle $Q \to X$, is an Azumaya algebra. We denote by $\beta_S \in \mathrm{Br}(S_{\le 1})$ the Brauer class of $\mathcal{B}_0|_{S_{\le 1}}$.

Similarly, assume $\dim(Q/X) \equiv 1 \bmod 2$. Locally over X we can trivialize the line bundle $\mathcal{L}$; then using $\det(q)$ in the same way as above we can define local double covers of X, which do not necessarily glue into a global double cover, but whose quotient stacks by the covering involutions glue into a global stack $S \to X$. This is, in fact, the root stack

$$S = \sqrt[2]{\big(((\det \mathcal{E})^{\otimes 2} \otimes \mathcal{L}^{\otimes \mathrm{rk}(\mathcal{E})})^\vee, \det(q)\big)/X}$$

as defined in [1, Section B.2]. By [11, Section 3.6], the algebra $\mathrm{Cliff}_0(\mathcal{E}, q)$ can be written as the pushforward of an $\mathcal{O}_S$–algebra $\mathcal{B}_0$ from S. Moreover, the restriction of $\mathcal{B}_0$ to the open substack $S_{\le 1} \subset S$, the preimage of the open subset $X_{\le 1} \subset X$, is an Azumaya algebra. We denote by $\beta_S \in \mathrm{Br}(S_{\le 1})$ the Brauer class of $\mathcal{B}_0|_{S_{\le 1}}$.

Proof of Corollary 1.5 Since the field k is algebraically closed, we have $W(\mathsf{k}) = \mathbb{Z}/2$ via the dimension parity homomorphism, hence the assumptions $\mathrm{w}_x(\mathcal{E}, q) = \mathrm{w}_x(\mathcal{E}', q')$ and $\mathrm{hw}(\mathcal{E}, q) = \mathrm{hw}(\mathcal{E}', q')$ of Theorem 1.3 reduce, respectively, to the equality $\mathrm{rk}(\mathcal{E}) \equiv \mathrm{rk}(\mathcal{E}') \bmod 2$, which holds true in each part of the corollary, and to $\mathrm{rk}(H^{n/2}(q)) \equiv \mathrm{rk}(H^{n/2}(q')) \bmod 2$, which is one of the assumptions of the corollary. Therefore, by Theorem 1.3, the quadric bundles Q and Q' are hyperbolic equivalent. Now parts (1) and (2) of the corollary follow from the Morita equivalence of $\mathrm{Cliff}_0(\mathcal{E}, q)$ and $\mathrm{Cliff}_0(\mathcal{E}', q')$ proved in Proposition 1.1(3), part (3) follows from Proposition 1.1(4), and part (4) from Proposition 1.1(5). □

References

[1] **D Abramovich**, **T Graber**, **A Vistoli**, *Gromov–Witten theory of Deligne–Mumford stacks*, Amer. J. Math. 130 (2008) 1337–1398 MR Zbl

[2] **J K Arason**, *Der Wittring projektiver Räume*, Math. Ann. 253 (1980) 205–212 MR Zbl

[3] **A Auel**, **M Bernardara**, **M Bolognesi**, *Fibrations in complete intersections of quadrics, Clifford algebras, derived categories, and rationality problems*, J. Math. Pures Appl. 102 (2014) 249–291 MR Zbl

[4] **P Balmer**, *Witt groups*, from "Handbook of K–theory, II" (E M Friedlander, D R Grayson, editors), Springer (2005) 539–576 MR Zbl

[5] **E Bayer-Fluckiger**, **L Fainsilber**, *Non-unimodular Hermitian forms*, Invent. Math. 123 (1996) 233–240 MR Zbl

[6] **G Bini**, **G Kapustka**, **M Kapustka**, *Symmetric locally free resolutions and rationality problems*, Commun. Contemp. Math. 25 (2023) art. id. 2250033 MR Zbl

[7] **G Casnati**, **F Catanese**, *Even sets of nodes are bundle symmetric*, J. Differential Geom. 47 (1997) 237–256 MR Zbl

[8] **D Eisenbud**, **S Popescu**, **C Walter**, *Lagrangian subbundles and codimension 3 subcanonical subschemes*, Duke Math. J. 107 (2001) 427–467 MR Zbl

[9] **C Ingalls**, **A Obus**, **E Ozman**, **B Viray**, *Unramified Brauer classes on cyclic covers of the projective plane*, from "Brauer groups and obstruction problems" (A Auel, B Hassett, A Várilly-Alvarado, B Viray, editors), Progr. Math. 320, Birkhäuser, Cham (2017) 115–153 MR Zbl

[10] **M Knebusch**, *Symmetric bilinear forms over algebraic varieties*, from "Conference on quadratic forms" (G Orzech, editor), Queen's Papers in Pure and Appl. Math. 46, Queen's Univ., Kingston, ON (1977) 103–283 MR Zbl

[11] **A Kuznetsov**, *Derived categories of quadric fibrations and intersections of quadrics*, Adv. Math. 218 (2008) 1340–1369 MR Zbl

[12] **A Kuznetsov**, *Scheme of lines on a family of 2–dimensional quadrics: geometry and derived category*, Math. Z. 276 (2014) 655–672 MR Zbl

[13] **A Kuznetsov**, **E Shinder**, *Grothendieck ring of varieties, D– and L–equivalence, and families of quadrics*, Selecta Math. 24 (2018) 3475–3500 MR Zbl

[14] **C Walter**, *Grothendieck–Witt groups of projective bundles*, preprint (2003)

Algebraic Geometry Section, Steklov Mathematical Institute of Russian Academy of Sciences
Moscow, Russia

akuznet@mi-ras.ru

Proposed: Marc Levine — Received: 20 November 2021
Seconded: Mark Gross, Richard P Thomas — Revised: 12 August 2022

msp

Geometry & Topology **28**:3 (2024) 1341–1407
DOI: 10.2140/gt.2024.28.1341
Published: 10 May 2024

Categorical wall-crossing formula for Donaldson–Thomas theory on the resolved conifold

YUKINOBU TODA

We prove a wall-crossing formula for categorical Donaldson–Thomas invariants on the resolved conifold, which categorifies the Nagao–Nakajima wall-crossing formula for numerical DT invariants on it. The categorified Hall products are used to describe the wall-crossing formula as semiorthogonal decompositions. A successive application of the categorical wall-crossing formula yields semiorthogonal decompositions of categorical Pandharipande–Thomas stable pair invariants on the resolved conifold, which categorify the product expansion formula of the generating series of numerical PT invariants on it.

14N35; 18G80

1 Introduction

1.1 Background and summary of the paper

In this paper, we establish a wall-crossing formula for categorical Donaldson–Thomas invariants on the resolved conifold, and apply it to give a complete description of categorical Pandharipande–Thomas (PT) stable pair invariants on it.

The PT invariants count stable pairs on CY 3–folds, and were introduced by Pandharipande and Thomas [2009] in order to give a better formulation of the GW/DT correspondence conjecture of Maulik, Nekrasov, Okounkov and Pandharipande [Maulik et al. 2006]. They are special cases of Donaldson–Thomas (DT) type invariants counting stable objects in the derived category, and are now understood as fundamental enumerative invariants of curves on CY 3–folds as well as Gromov–Witten invariants and Gopakumar–Vafa invariants. Now by efforts from derived algebraic geometry due to Pantev, Toën, Vaquié and Vezzosi [Pantev et al. 2013] and Brav, Bussi and Joyce [Brav et al. 2019], the moduli spaces which define DT (in particular PT) invariants are known to be locally written as critical loci. In [Toda 2019], we proposed a study of categorical DT theory by gluing locally defined dg-categories of matrix factorizations on these moduli spaces. A definition of categorical DT invariants is introduced in the case of local surfaces in [Toda 2019] via Koszul duality and singular support quotients. We also proposed several conjectures on wall-crossing of categorical DT invariants on local surfaces, motivated by a d–critical analogue of the D/K equivalence conjecture of Bondal and Orlov [1995] and Kawamata [2002], and also categorifications of wall-crossing formulas of numerical DT invariants [Joyce and Song 2012; Kontsevich and Soibelman 2008]. In [Toda 2019], we also derived a wall-crossing formula of categorical PT invariants on local surfaces in the setting of simple wall-crossing (ie there are at most two Jordan–Hölder factors at the wall).

The purpose of this paper is to prove wall-crossing formula for categorical DT invariants on the resolved conifold, which categorifies the wall-crossing formula of Nagao and Nakajima [2011] for numerical DT invariants on it. In this case the relevant moduli spaces are global critical loci, so there is no issue with gluing dg-categories of matrix factorizations. However, the wall-crossing is not necessarily a simple wall-crossing, and the analysis of categorical wall-crossings is much harder. Our strategy is to use categorified Hall products for quivers with superpotentials introduced by Păduraru [2019; 2023]. A key observation is that, up to Knörrer periodicity, a wall-crossing diagram for the resolved conifold locally looks like a Grassmannian flip together with some superpotential (d-critical Grassmannian flip in the sense of d-critical birational geometry [Toda 2022]). We refine the result of Ballard, Chidambaram, Favero, McFaddin and Vandermolen [Ballard et al. 2021] on derived categories of Grassmannian flips via categorified Hall products, and compare them with more global categorified Hall products under the Knörrer periodicity. The above approach via categorified Hall products yields a desired categorical wall-crossing formula. A successive iteration of wall-crossing gives a semiorthogonal decomposition of categorical PT invariants on the resolved conifold, whose semiorthogonal summands are the simplest categories of matrix factorizations over a point. We emphasize that the result of this paper is a first instance where a categorical wall-crossing formula is obtained for nonsimple wall-crossing in the context of categorical DT theory.

1.2 Categorical PT stable pair theory on the resolved conifold

The *resolved conifold* X is defined by

$$X := \mathrm{Tot}_{\mathbb{P}^1}(\mathcal{O}_{\mathbb{P}^1}(-1)^{\oplus 2}),$$

which is also obtained as a crepant small resolution of the conifold singularity $\{xy+zw=0\} \subset \mathbb{C}^4$. The resolved conifold is a noncompact CY 3–fold, and an important toy model for enumerative geometry on CY 3–folds such as PT invariants.

For each $(\beta, n) \in \mathbb{Z}^2$, we denote by

$$P_n(X, \beta)$$

the moduli space of PT stable pairs (F, s) on X, ie F is a pure one-dimensional coherent sheaf on X and $s\colon \mathcal{O}_X \to F$ is surjective in dimension one, satisfying $[F] = \beta[C]$ and $\chi(F) = n$. Here $C \subset X$ is the zero section of the projection $X \to \mathbb{P}^1$, and $[F]$ is the fundamental one-cycle of F. The PT invariant $P_{n,\beta} \in \mathbb{Z}$ is defined by either taking the integration over the zero-dimensional virtual fundamental class on $P_n(X,\beta)$, or weighted Euler characteristic of the Behrend constructible function [Behrend 2009] on it. It is well-known that the generating series of PT invariants on X is given by the formula

$$\sum_{n,\beta} P_{n,\beta} q^n t^\beta = \prod_{m \geq 1} (1-(-q)^m t)^m. \tag{1-1}$$

The above formula is available in [Nagao and Nakajima 2011, Theorem 3.15], which is also obtained from the DT calculation in [Behrend and Bryan 2007] together with the DT/PT correspondence [Bridgeland 2011; Toda 2010; Stoppa and Thomas 2011].

The purpose of this paper is to give a categorification of the formula (1-1). In the case of the resolved conifold, the moduli space $P_n(X,\beta)$ is written as a global critical locus, ie there is a pair (M,w) where M is a smooth quasiprojective scheme and $w\colon M \to \mathbb{A}^1$ is a regular function such that $P_n(X,\beta)$ is isomorphic to the critical locus of w. A choice of (M,w) is not unique, and we take it using the noncommutative crepant resolution of X due to Van den Bergh [2004]; see Section 5.10. We define the *categorical PT invariant* on X to be the triangulated category of matrix factorizations of w,

$$\mathcal{DT}(P_n(X,\beta)) := \mathrm{MF}(M,w).$$

See Definition 5.23. The above triangulated category (or more precisely its dg-enhancement) recovers $P_{n,\beta}$ by taking the Euler characteristic of its periodic cyclic homology; see equation (5-59). The following is a consequence of the main result in this paper:

Theorem 1.1 (Corollary 5.24) *There exists a semiorthogonal decomposition of the form*

$$\mathcal{DT}(P_n(X,\beta)) = \langle a_{n,\beta} \textit{ copies of } \mathrm{MF}(\mathrm{Spec}\,\mathbb{C},0)\rangle. \tag{1-2}$$

Here $a_{n,\beta}$ is defined by

$$a_{n,\beta} := \sum_{\substack{l\colon \mathbb{Z}_{\geq 1} \to \mathbb{Z}_{\geq 0} \\ \sum_{m\geq 1} l(m)\cdot(m,1)=(n,\beta)}} \prod_{m\geq 1} \binom{m}{l(m)}. \tag{1-3}$$

Here $\mathrm{MF}(\mathrm{Spec}\,\mathbb{C},0)$ is the category of matrix factorizations of the zero superpotential over the point, which is equivalent to the $\mathbb{Z}/2$–periodic derived category of finite-dimensional $\mathbb{C}$–vector spaces. As the formula (1-1) is equivalent to $P_{n,\beta} = (-1)^{n+\beta} a_{n,\beta}$, by taking the periodic cyclic homologies of both sides and Euler characteristics, the result of Theorem 1.1 recovers the formula (1-1); see Remark 5.25.

1.3 Categorical wall-crossing formula

Nagao and Nakajima [2011, Theorem 3.15] derived the formula (1-1) by proving wall-crossing formula for stable perverse coherent systems on X. Under a derived equivalence of X with a noncommutative crepant

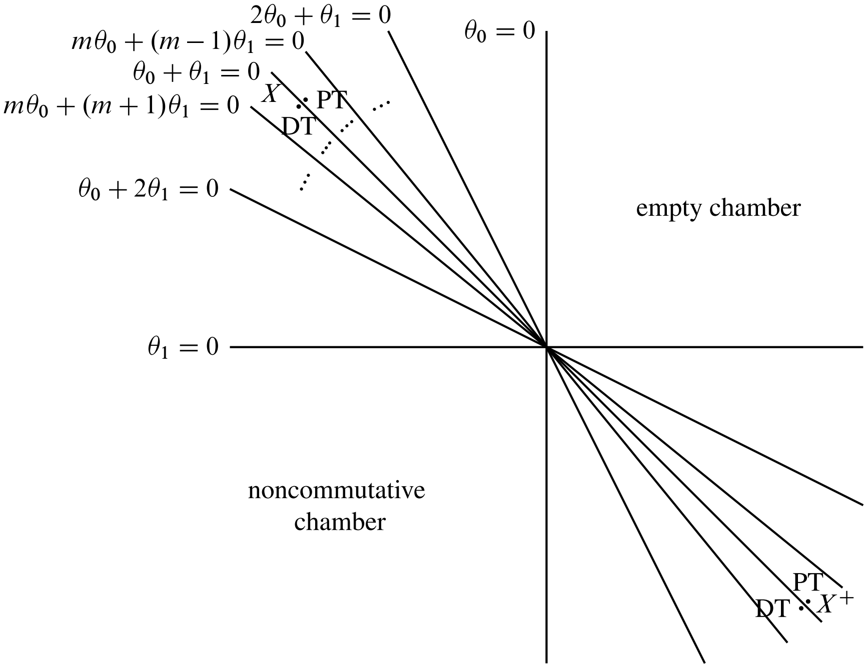


Figure 1: Wall-chamber structures.

resolution of the conifold due to Van den Bergh [2004], the category of perverse coherent systems on X is equivalent to the category of representations of the quiver with superpotential $(Q^\dagger, W)$, where

$$Q^\dagger = \begin{tikzcd} \bullet_\infty \arrow[d] & \\ \bullet_0 \arrow[r, bend left=50, "a_1"] \arrow[r, bend left=20, "a_2"] & \bullet_1 \arrow[l, bend left=20, "b_1"] \arrow[l, bend left=50, "b_2"] \end{tikzcd} \qquad \text{and} \qquad W = a_1b_1a_2b_2 - a_1b_2a_2b_1.$$

For $v = (v_0, v_1) \in \mathbb{Z}^2_{\geq 0}$, we denote by $\mathcal{M}^\dagger_Q(v)$ the $\mathbb{C}^*$–rigidified moduli stack of $Q^\dagger$–representations with dimension vector $(1, v_0, v_1)$, where 1 is the dimension vector at ∞. It is equipped with a superpotential

$$w = \mathrm{Tr}(W)\colon \mathcal{M}^\dagger_Q(v) \to \mathbb{A}^1,$$

whose critical locus is isomorphic to the moduli stack of $(Q^\dagger, W)$–representations with dimension vector $(1, v_0, v_1)$. There is also a stability parameter $\theta = (\theta_0, \theta_1) \in \mathbb{R}^2$ of $(Q^\dagger, W)$–representations, whose wall-chamber structure is pictured in Figure 1, taken from [Nagao and Nakajima 2011, Figure 1].

For $m \in \mathbb{Z}_{\geq 1}$, there is a wall in the second quadrant in Figure 1,

$$W_m := \mathbb{R}_{>0}(1-m, m) \subset \mathbb{R}^2.$$

We take a stability condition on the wall $\theta \in W_m$ and $\theta_\pm = \theta \pm (-\varepsilon, \varepsilon)$ for $\varepsilon > 0$ which lie on its adjacent chambers. Let $\mathrm{DT}^{\theta_\pm}(v_0, v_1) \in \mathbb{Z}$ be the DT invariant counting $\theta_\pm$–stable $(Q^\dagger, W)$–representations with dimension vector $(1, v_0, v_1)$. We have the wall-crossing formula

$$\sum_{(v_0,v_1)\in\mathbb{Z}^2_{\geq 0}} \mathrm{DT}^{\theta_+}(v_0,v_1)q_0^{v_0}q_1^{v_1} = \Bigg(\sum_{(v_0,v_1)\in\mathbb{Z}^2_{\geq 0}} \mathrm{DT}^{\theta_-}(v_0,v_1)q_0^{v_0}q_1^{v_1}\Bigg)\cdot(1+q_0^m(-q_1)^{m-1})^m \tag{1-4}$$

proved by Nagao and Nakajima [2011, Theorem 3.12]. The formula (1-1) is obtained from the above wall-crossing formula by applying it from $m = 1$ to $m \gg 0$ and noting that the PT invariants correspond to a chamber which is sufficiently close to (and above) the wall $\mathbb{R}_{>0}(-1, 1)$.

We prove Theorem 1.1 by giving a categorification of the formula (1-4). For $\theta \in \mathbb{R}^2$, we denote by

$$\mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v) \subset \mathcal{M}_Q^{\dagger}(v)$$

the open substack of θ–semistable $Q^\dagger$–representations. The following is the main result of this paper, which gives a categorification of the formula (1-4):

Theorem 1.2 (Corollary 5.18) *For $\theta \in W_m$, by setting $s_m = (m, m-1)$, there exists a semiorthogonal decomposition*

$$\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_+\text{–ss}}(v), w) = \Big\langle \binom{m}{l} \text{ copies of } \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_-\text{–ss}}(v - l s_m), w) : l \geq 0 \Big\rangle. \tag{1-5}$$

There is also a precisely defined order among semiorthogonal summands in (1-2); see Corollary 5.18 for the precise statement. Again by taking the periodic cyclic homologies and the Euler characteristics, the result of Theorem 1.2 recovers the Nagao–Nakajima formula (1-4); see Remark 5.19. The result of Theorem 1.1 follows by applying Theorem 1.2 from $m = 1$ to $m \gg 0$.

We also remark that the similar categorical wall-crossing formula holds at other walls except walls at $\{\theta_0 + \theta_1 = 0\}$, ie DT/PT wall on X or on its flop; see Remark 5.21. Note that the numerical DT/PT wall-crossing formula was not directly obtained in [Nagao and Nakajima 2011], but was proved in [Bridgeland 2011; Toda 2010; Stoppa and Thomas 2011] using the full machinery of motivic Hall algebras in [Joyce and Song 2012; Kontsevich and Soibelman 2008].

1.4 Outline of the proof of Theorem 1.2

The strategy of the proof of Theorem 1.2 is to use the following ingredients:

(i) The window subcategories for GIT quotient stacks developed by Halpern-Leistner [2015] and Ballard, Favero and Katzarkov [Ballard et al. 2019].

(ii) The categorified Hall products for quivers with superpotentials introduced and studied by Pădurariu [2023; 2024; 2019].

(iii) The descriptions of derived categories under Grassmannian flips by Ballard, Chidambaram, Favero, McFaddin and Vandermolen [Ballard et al. 2021], which itself relies on earlier work by Donovan and Segal [2014] for Grassmannian flops.

For $\theta \in W_m$, let $\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v) \to M_Q^{\dagger,\theta\text{-ss}}(v)$ be the good moduli space [Alper 2013]. We have the wall-crossing diagram

$$\begin{tikzcd} M_Q^{\dagger,\theta_+\text{-ss}}(v) \arrow[rr, dotted] \arrow[dr] & & M_Q^{\dagger,\theta_-\text{-ss}}(v) \arrow[dl] \\ & M_Q^{\dagger,\theta\text{-ss}}(v) & \end{tikzcd} \tag{1-6}$$

which is shown to be a flip of smooth quasiprojective varieties. The D/K principle by Bondal and Orlov [1995] and Kawamata [2002] predicts the existence of a fully faithful functor of their derived categories or categories of matrix factorizations.

The window subcategories have been used to investigate the D/K conjecture under variations of GIT quotients. In the above setting, there exist subcategories $\mathbb{W}_{\mathrm{glob}}^{\theta_\pm}(v) \subset \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w)$, called *window subcategories*, such that the compositions

$$\mathbb{W}_{\mathrm{glob}}^{\theta_\pm}(v) \hookrightarrow \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w) \twoheadrightarrow \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_\pm\text{-ss}}(v), w)$$

are equivalences. If we can show that $\mathbb{W}_{\mathrm{glob}}^{\theta_-}(v) \subset \mathbb{W}_{\mathrm{glob}}^{\theta_+}(v)$ for some choice of window subcategories, then we have a desired fully faithful functor

$$\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_-\text{-ss}}(v), w) \hookrightarrow \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_+\text{-ss}}(v), w). \tag{1-7}$$

In fact, the above argument is used in [Toda 2019, Theorem 4.3.5] to show the existence of a fully faithful functor (1-7).

We are interested in the semiorthogonal complement of the fully faithful functor (1-7). If the wall-crossing is enough simple, eg satisfying the DHT condition in [Ballard et al. 2019, Definition 4.1.4], then the above window subcategory argument also describes the semiorthogonal complement; see [Ballard et al. 2019, Theorem 4.2.1]. However our wall-crossing (1-6) does not necessary satisfy the DHT condition, and we cannot directly apply it. Instead we use categorified Hall products to describe the semiorthogonal complement of (1-7).

The categorified Hall product for quivers with superpotentials was introduced by Pădurariu [2023; 2024; 2019] in order to give a K–theoretic version of critical COHA, which was introduced in [Kontsevich and Soibelman 2011] and developed in [Davison 2017]. For $v = v_1 + v_2$ with $\theta(v_1) = 0$, it is a functor

$$*\colon \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v_1), w) \boxtimes \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v_2), w) \to \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w)$$

which is defined by the pullback/pushforward with respect to the stack of short exact sequences of $Q^\dagger$–representations. We will show that, for $l \geq 0$ and a sequence of integers $0 \leq j_1 \leq \cdots \leq j_l \leq m - l$, the categorified Hall product gives a fully faithful functor

$$\bigboxtimes_{i=1}^{l} \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(s_m), w)_{j_i+(2i-1)(m^2-m)} \boxtimes (\mathbb{W}_{\mathrm{glob}}^{\theta_-}(v - l s_m) \otimes \chi_0^{j_l+2l(m^2-m)}) \to \mathbb{W}_{\mathrm{glob}}^{\theta_+}(v), \tag{1-8}$$

whose essential images form a semiorthogonal decomposition. Here the subscript $j_i + (2i-1)(m^2-m)$ indicates the fixed $\mathbb{C}^*$–weight part, and χ_0 is some character regarded as a line bundle on $\mathcal{M}_Q^{\dagger}(v)$; see Theorem 5.17 for details. It follows that the categorified Hall products describe the semiorthogonal complement of (1-7), which lead to a proof of Theorem 1.2.

In order to show that the functor (1-8) is fully faithful and they form a semiorthogonal decomposition, we prove these statements formally locally on the good moduli space $M_Q^{\dagger,\theta\text{-ss}}(v)$ at any point p corresponding to a θ–polystable $(Q^{\dagger}, W)$–representation R. By the étale slice theorem, one can describe the formal fibers of the diagram (1-6) at p in terms of a wall-crossing diagram of the Ext quiver $Q_p^{\dagger}$ associated with R, which is much simpler than $Q^{\dagger}$. After removing a quadratic part of the superpotential, one observes that the wall-crossing diagram for $Q_p^{\dagger}$–representations is the product of a Grassmannian flip with some trivial part. Here a *Grassmannian flip* is a birational map

$$G_{a,b}^{+}(d) \dashrightarrow G_{a,b}^{-}(d)$$

given by two GIT stable loci of the quotient stack

$$\mathcal{G}_{a,b}(d) = [(\mathrm{Hom}(A, V) \oplus \mathrm{Hom}(V, B))/\mathrm{GL}(V)],$$

where $d = \dim V$, $a = \dim A$ and $b = \dim B$ with $a \geq b$.

Donovan and Segal [2014] proved a derived equivalence $D^b(G_{a,b}^{-}(d)) \simeq D^b(G_{a,b}^{+}(d))$ in the case of $a = b$ (ie Grassmannian flop) using window subcategories, and the same argument also applies to construct a fully faithful functor $D^b(G_{a,b}^{-}(d)) \hookrightarrow D^b(G_{a,b}^{+}(d))$. However it is in a rather recent work of Ballard, Chidambaram, Favero, McFaddin and Vandermolen [Ballard et al. 2021] where the semiorthogonal complement of the above fully faithful functor is considered. We will interpret the description of semiorthogonal complement in [Ballard et al. 2021] in terms of categorified Hall products, and refine it as a semiorthogonal decomposition

$$D^b(G_{a,b}^{+}(d)) = \left\langle \binom{a-b}{l} \text{ copies of } D^b(G_{a,b}^{-}(d-l)) : 0 \leq l \leq d \right\rangle. \tag{1-9}$$

See Corollary 4.19. The above semiorthogonal decomposition unifies Kapranov's exceptional collections of derived categories of Grassmannians, and also semiorthogonal decompositions of standard toric flips, so it may be of independent interest; see Remarks 4.20 and 4.21.

A semiorthogonal decomposition similar to (1-9) also holds for categories of factorizations of a superpotential of $\mathcal{G}_{a,b}(d)$. Under the Knörrer periodicity, we compare global categorified Hall products (1-8) with local categorified Hall products giving the semiorthogonal decomposition (1-9). By combining these arguments, we see that the functor (1-8) is fully faithful and they form a semiorthogonal decomposition formally locally on $M_Q^{\dagger,\theta\text{-ss}}(v)$, hence they also hold globally.

1.5 Related works

The wall-crossing formula (1-4) was proved by Nagao and Nakajima [2011] in order to give an understanding of the product expansion formula of noncommutative DT invariants of the conifold studied by Szendrői [2008]. The wall-crossing formula (1-4) was later extended to the case of a global flopping contraction by Toda [2013] and Calabrese [2016], to the motivic DT invariants by Morrison, Mozgovoy, Nagao and Szendrői [Morrison et al. 2012], and to the DT4 invariants by Cao and Toda [2023]. Recently Tasuki Kinjo [≥ 2024] has studied cohomological DT theory on the resolved conifold and proves a cohomological version of DT/PT correspondence in this case. It would be interesting to extend the argument in this paper and categorify his cohomological DT/PT correspondence.

As we already mentioned, the study of wall-crossing of categorical PT invariants was posed in [Toda 2019]. In the case of local surfaces, a categorical wall-crossing formula is conjectured in [Toda 2019, Conjecture 6.2.6] in the case of simple wall-crossing, and proved in some cases in [Toda 2019, Theorem 6.3.19] using Porta–Sala categorified Hall products for surfaces [Porta and Sala 2023]. The wall-crossing we consider in this paper is not necessary simple, so it is beyond the cases we considered in [Toda 2019, Conjecture 6.2.6]. A similar wall-crossing at $(-1,-1)$–curve is also considered in [Toda 2021, Section 7], but we only proved the existence of fully faithful functors and their semiorthogonal complements are not considered.

The categorified (K–theoretic) Hall algebras for quivers with superpotentials were introduced and studied by Tudor Pădurariu [2019; 2023]. He also proved the PBW theorem for K–theoretic Hall algebras [Pădurariu 2019;2024] via much more sophisticated combinatorial arguments (based on earlier works of Špenko and Van den Bergh [2017] and Halpern-Leistner and Sam [2020]). We expect that his arguments proving the K–theoretic PBW theorem can be applied to prove categorical (or K–theoretic) wall-crossing formula in a broader setting, including DT/PT wall-crossing in this paper.

Recently Qingyuan Jiang [2021] has studied derived categories of Quot schemes of locally free quotients, and proposed conjectural semiorthogonal decompositions of them; see [Jiang 2021, Conjecture A.5]. He proved the above conjecture in the case of rank-two quotients. His conjectural semiorthogonal decompositions resemble the one in Theorem 1.2. It would be interesting to see whether the technique in this paper can be applied to his conjecture.

Koseki Naoki [2021] has studied derived categories of moduli spaces of stable perverse coherent sheaves for a blow-up of a surface (studied by Nakajima and Yoshioka [2011]), and proved the existence of fully faithful functors under wall-crossing. As the situation is similar to the one in this paper, a similar argument to that in this paper may be applied to describe the semiorthogonal complements of his fully faithful functors.

1.6 Notation and conventions

In this paper, all the schemes or stacks are defined over $\mathbb{C}$. For an Artin stack $\mathcal{Y}$, we denote by $D^b(\mathcal{Y})$ the bounded derived category of coherent sheaves on $\mathcal{Y}$. For an algebraic group G and its representation V, we regard it as a vector bundle on BG. For a variety Y on which G acts, we denote by $V \otimes \mathcal{O}_{[Y/G]}$ the vector bundle given by the pullback of V by $[Y/G] \to BG$. For a morphism $\mathfrak{M} \to M$ from a stack $\mathfrak{M}$ to a scheme M and a closed point $y \in M$, the *formal fiber* at y is defined by

$$\widehat{\mathfrak{M}}_y := \mathfrak{M} \times_M \operatorname{Spec} \widehat{\mathcal{O}}_{M,y} \to \widehat{M}_y := \operatorname{Spec} \widehat{\mathcal{O}}_{M,y}.$$

For a triangulated category $\mathcal{D}$, its triangulated subcategory $\mathcal{D}' \subset \mathcal{D}$ is called *dense* if any object in $\mathcal{D}$ is a direct summand of an object in $\mathcal{D}'$.

Acknowledgements The author is grateful to Yalong Cao, Qingyuan Jiang, Naoki Koseki, Tudor Păduraru and Gufang Zhao for comments on the first version of this paper. The author also thanks the referee for careful reading of the paper and several valuable comments. The author is supported by World Premier International Research Center Initiative (WPI initiative), MEXT, Japan, and Grant in Aid for Scientific Research grant 19H01779 from MEXT, Japan.

2 Preliminaries

In this section, we review triangulated categories of factorizations, the window theorem for categories of factorizations over GIT quotient stacks, and the Knörrer periodicity.

2.1 The category of factorizations

Let $\mathcal{Y}$ be a noetherian algebraic stack over $\mathbb{C}$ and take $w \in \Gamma(\mathcal{O}_{\mathcal{Y}})$. A (coherent) factorization of w consists of

$$\begin{tikzcd} \mathcal{P}_0 \arrow[r, bend left, "\alpha_0"] & \mathcal{P}_1 \arrow[l, bend left, "\alpha_1"] \end{tikzcd}, \quad \text{where } \alpha_0 \circ \alpha_1 = \cdot w \text{ and } \alpha_1 \circ \alpha_0 = \cdot w,$$

where each $\mathcal{P}_i$ is a coherent sheaf on $\mathcal{Y}$, and the α_i are morphisms of coherent sheaves. The category of coherent factorizations naturally forms a dg-category, whose homotopy category is denoted by $\mathrm{HMF}(\mathcal{Y}, w)$. The subcategory of absolutely acyclic objects

$$\mathrm{Acy}^{\mathrm{abs}} \subset \mathrm{HMF}(\mathcal{Y}, w)$$

is defined to be the minimum thick triangulated subcategory which contains totalizations of short exact sequences of coherent factorizations of w. The triangulated category of factorizations of w is defined by

$$\mathrm{MF}(\mathcal{Y}, w) := \mathrm{HMF}(\mathcal{Y}, w)/\mathrm{Acy}^{\mathrm{abs}}.$$

(See [Orlov 2012; Efimov and Positselski 2015; Polishchuk and Vaintrob 2011].)

If $\mathcal{Y}$ is an affine scheme, then $\mathrm{MF}(\mathcal{Y}, w)$ is equivalent to Orlov's triangulated category [2009] of matrix factorizations of w. For two pairs $(\mathcal{Y}_i, w_i)$ for $i = 1, 2$, we use the notation

$$\mathrm{MF}(\mathcal{Y}_1, w_1) \boxtimes \mathrm{MF}(\mathcal{Y}_2, w_2) := \mathrm{MF}(\mathcal{Y}_1 \times \mathcal{Y}_2, w_1 + w_2).$$

For triangulated subcategories $\mathcal{C}_i \subset \mathrm{MF}(\mathcal{Y}_i, w_i)$, we denote by $\mathcal{C}_1 \boxtimes \mathcal{C}_2$ the smallest thick triangulated subcategory of $\mathrm{MF}(\mathcal{Y}_1, w_1) \boxtimes \mathrm{MF}(\mathcal{Y}_2, w_2)$ which contains $C_1 \boxtimes C_2$ for $C_i \in \mathcal{C}_i$.

It is well known that $\mathrm{MF}(\mathcal{Y}, w)$ only depends on an open neighborhood of $\mathrm{Crit}(w) \subset \mathcal{Y}$. Namely let $\mathcal{Y}' \subset \mathcal{Y}$ be an open substack such that $\mathrm{Crit}(w) \subset \mathcal{Y}'$. Then the restriction functor gives an equivalence

$$\mathrm{MF}(\mathcal{Y}, w) \xrightarrow{\sim} \mathrm{MF}(\mathcal{Y}', w|_{\mathcal{Y}'}). \tag{2-1}$$

(See [Polishchuk and Vaintrob 2011, Corollary 5.3] and [Halpern-Leistner and Sam 2020, Lemma 5.5].) Suppose that $\mathcal{Y} = [Y/G]$, where G is an algebraic group which acts on a scheme Y. Assume that $\mathbb{C}^* \subset G$ lies in the center of G, which acts on Y trivially. Then $\mathrm{MF}(\mathcal{Y}, w)$ decomposes into the direct sum

$$\mathrm{MF}(\mathcal{Y}, w) = \bigoplus_{j \in \mathbb{Z}} \mathrm{MF}(\mathcal{Y}, w)_j, \tag{2-2}$$

where each summand corresponds to the $\mathbb{C}^*$–weight j part.

2.2 Attracting loci

Let G be a reductive algebraic group with maximal torus T, which acts on a smooth affine scheme Y. We denote by M the character lattice of T and N the cocharacter lattice of T. There is a perfect pairing

$$\langle -, - \rangle \colon N \times M \to \mathbb{Z}.$$

For a one-parameter subgroup $\lambda \colon \mathbb{C}^* \to G$, let $Y^{\lambda \geq 0}$ and $Y^{\lambda = 0}$ be defined by

$$Y^{\lambda \geq 0} := \bigl\{ y \in Y : \lim_{t \to 0} \lambda(t)(y) \text{ exists} \bigr\},$$
$$Y^{\lambda = 0} := \{ y \in Y : \lambda(t)(y) = y \text{ for all } t \in \mathbb{C}^* \}.$$

The Levi subgroup and the parabolic subgroup

$$G^{\lambda = 0} \subset G^{\lambda \geq 0} \subset G$$

are also similarly defined by the conjugate G–action on G, ie $g \cdot (-) = g(-)g^{-1}$. The G–action on Y restricts to the $G^{\lambda \geq 0}$–action on $Y^{\lambda \geq 0}$, and the $G^{\lambda = 0}$–action on $Y^{\lambda = 0}$. We note that λ factors through $\lambda \colon \mathbb{C}^* \to G^{\lambda = 0}$, and it acts on $Y^{\lambda = 0}$ trivially. So we have the decomposition into λ–weight spaces

$$D^b([Y^{\lambda=0}/G^{\lambda=0}]) = \bigoplus_{j \in \mathbb{Z}} D^b([Y^{\lambda=0}/G^{\lambda=0}])_{\lambda\text{-wt}=j}.$$

We have the diagram of attracting loci

$$\begin{tikzcd} {[Y^{\lambda \geq 0}/G^{\lambda \geq 0}]} \arrow[r, "p_\lambda"] \arrow[d, "q_\lambda"] & {[Y/G]} \\ {[Y^{\lambda=0}/G^{\lambda=0}]} \arrow[u, bend left, "\sigma_\lambda"] & \end{tikzcd} \tag{2-3}$$

Here p_λ is induced by the inclusion $Y^{\lambda \geq 0} \subset Y$, and q_λ is given by taking the $t \to 0$ limit of the action of $\lambda(t)$ for $t \in \mathbb{C}^*$. The morphism σ_λ is a section of q_λ induced by inclusions $Y^{\lambda=0} \subset Y^{\lambda \geq 0}$ and $G^{\lambda=0} \subset G^{\lambda \geq 0}$. We will use the following lemma.

Lemma 2.1 [Halpern-Leistner 2015, Corollary 3.17, Amplification 3.18]

(i) *For* $\mathscr{E}_i \in D^b([Y^{\lambda\geq 0}/G^{\lambda\geq 0}])$ *with* $i = 1, 2$, *suppose that*

$$\sigma_\lambda^*\mathscr{E}_1 \in D^b([Y^{\lambda=0}/G^{\lambda=0}])_{\lambda\text{-wt}\geq j} \quad \text{and} \quad \sigma_\lambda^*\mathscr{E}_2 \in D^b([Y^{\lambda=0}/G^{\lambda=0}])_{\lambda\text{-wt}<j}$$

for some j. *Then* $\operatorname{Hom}(\mathscr{E}_1, \mathscr{E}_2) = 0$.

(ii) *For* $j \in \mathbb{Z}$, *the functor*

$$q_\lambda^* \colon D^b([Y^{\lambda=0}/G^{\lambda=0}])_{\lambda\text{-wt}=j} \to D^b([Y^{\lambda\geq 0}/G^{\lambda\geq 0}])$$

is fully faithful.

2.3 Kempf–Ness stratification

Here we review Kempf–Ness stratifications associated with GIT quotients of reductive algebraic groups, and the corresponding window theorem following the convention of [Halpern-Leistner 2015, Section 2.1]. Let Y and G be as in the previous subsection. For an element $l \in \operatorname{Pic}([Y/G])_{\mathbb{R}}$, we have the open subset of l–semistable points

$$Y^{l\text{-ss}} \subset Y$$

characterized by the set of points $y \in Y$ such that for any one-parameter subgroup $\lambda\colon \mathbb{C}^* \to G$ such that the limit $z = \lim_{t\to 0} \lambda(t)(y)$ exists in Y, we have $\operatorname{wt}(l|_z) \geq 0$. Let $|{*}|$ be the Weyl-invariant norm on $N_{\mathbb{R}}$. The above subset of l–semistable points fits into the *Kempf–Ness* (*KN*) *stratification*

$$Y = S_1 \sqcup S_2 \sqcup \cdots \sqcup S_N \sqcup Y^{l\text{-ss}}. \tag{2-4}$$

Here for each $1 \leq i \leq N$ there exists a one-parameter subgroup $\lambda_i \colon \mathbb{C}^* \to T \subset G$, an open and closed subset Z_i of $\bigl(Y \setminus \bigcup_{i'<i} S_{i'}\bigr)^{\lambda_i=0}$ (called the *center* of S_i) such that

$$S_i = G\cdot Y_i \quad \text{and} \quad Y_i := \bigl\{y \in Y^{\lambda_i\geq 0} : \lim_{t\to 0}\lambda_i(t)(y) \in Z_i\bigr\}.$$

Moreover, by setting the slope to be

$$\mu_i := -\frac{\operatorname{wt}(l|_{Z_i})}{|\lambda_i|} \in \mathbb{R},$$

we have the inequalities $\mu_1 > \mu_2 > \cdots > 0$. We have (see [Halpern-Leistner 2015, Definition 2.2]) the diagram

$$\begin{tikzcd}
{[Y_i/G^{\lambda_i\geq 0}]} \arrow[r, "\cong"] \arrow[d] & {[S_i/G]} \arrow[r, hook, "q_i"] \arrow[dl, "p_i"] & {\bigl[\bigl(Y \setminus \bigcup_{i'<i} S_{i'}\bigr)/G\bigr]} \\
{[Z_i/G^{\lambda_i=0}]} \arrow[urr, hook, "\tau_i"'] & &
\end{tikzcd} \tag{2-5}$$

Here the left vertical arrow is given by taking the $t \to 0$ limit of the action of $\lambda_i(t)$ for $t \in \mathbb{C}^*$, and τ_i and q_i are induced by the embeddings $Z_i \hookrightarrow Y$ and $S_i \hookrightarrow Y$, respectively.

Let $\eta_i \in \mathbb{Z}$ be defined by

$$\eta_i := \mathrm{wt}_{\lambda_i}(\det(N^{\vee}_{S_i/Y}|_{Z_i})). \tag{2-6}$$

In the case that Y is a G–representation, it is also written as

$$\eta_i = \langle \lambda_i, (Y^{\vee})^{\lambda_i>0} - (\mathfrak{g}^{\vee})^{\lambda_i>0}\rangle.$$

Here for a G–representation W and a one-parameter subgroup $\lambda\colon \mathbb{C}^* \to T$, we denote by $W^{\lambda>0} \in K(BT)$ the subspace of W spanned by weights which pair positively with λ. We will use the following version of the window theorem.

Theorem 2.2 [Halpern-Leistner 2015; Ballard et al. 2019] *For each i, take $m_i \in \mathbb{R}$. For $N' \leq N$, let*

$$\mathbb{W}^l_{m_\bullet}\Big(\Big[\Big(Y \setminus \bigcup_{1\leq i\leq N'} S_i\Big)/G\Big]\Big) \subset D^b\Big(\Big[\Big(Y \setminus \bigcup_{1\leq i\leq N'} S_i\Big)/G\Big]\Big) \tag{2-7}$$

be the subcategory of objects $\mathcal{P}$ satisfying the condition

$$\tau_i^*(\mathcal{P}) \in \bigoplus_{j\in[m_i, m_i+\eta_i)} D^b([Z_i/G^{\lambda_i=0}])_{\lambda_i\text{-wt}=j} \tag{2-8}$$

for all $N' < i \leq N$. Then the composition functor

$$\mathbb{W}^l_{m_\bullet}\Big(\Big[\Big(Y \setminus \bigcup_{1\leq i\leq N'} S_i\Big)/G\Big]\Big) \hookrightarrow D^b\Big(\Big[\Big(Y \setminus \bigcup_{1\leq i\leq N'} S_i\Big)/G\Big]\Big) \twoheadrightarrow D^b([Y^{l\text{-ss}}/G])$$

is an equivalence.

Let $w\colon Y \to \mathbb{A}^1$ be a G–invariant function. We will apply Theorem 2.2 for a KN stratification of $\mathrm{Crit}(w)$

$$\mathrm{Crit}(w) = S'_1 \sqcup S'_2 \sqcup \cdots \sqcup S'_N \sqcup \mathrm{Crit}(w)^{l\text{-ss}}$$

in the following way. After discarding KN strata $S_i \subset Y$ with $\mathrm{Crit}(w) \cap S_i = \varnothing$, the above stratification is obtained by restricting a KN stratification (2-4) for Y to $\mathrm{Crit}(w)$. Let $\lambda_i\colon \mathbb{C}^* \to G$ be a one-parameter subgroup for S'_i with center $Z'_i \subset S'_i$. We define $\overline{Z}_i \subset Y$ to be the union of connected components of the λ_i–fixed part of Y which contains Z'_i, and $\overline{Y}_i \subset Y$ is the set of points $y \in Y$ with $\lim_{t\to 0} \lambda_i(t)y \in \overline{Z}_i$. Similarly to (2-5), we have the diagram

$$\begin{tikzcd}
{[\overline{Y}_i/G^{\lambda_i\geq 0}]} \arrow[r, hook] \arrow[d, "\overline{p}_i"'] \arrow[rr, bend left, "\overline{q}_i"] & {[Y^{\lambda_i\geq 0}/G^{\lambda_i\geq 0}]} \arrow[r] \arrow[d] & {[Y/G]} \\
{[\overline{Z}_i/G^{\lambda_i=0}]} \arrow[r, hook] & {[Y^{\lambda_i=0}/G^{\lambda_i=0}]} \arrow[ur] &
\end{tikzcd}$$

Here the left horizontal arrows are open and closed immersions. Using the equivalence (2-1), we also have the following version of window theorem for factorization categories; see [Toda 2021, Section 2.4].

Theorem 2.3 *For each i, we take $m_i \in \mathbb{R}$. For $N' \le N$, let*

$$\mathbb{W}^l_{m_\bullet}\Big(\Big[\Big(Y \setminus \bigcup_{1\le i\le N'} S'_i\Big)/G\Big], w\Big) \subset \mathrm{MF}\Big(\Big[\Big(Y \setminus \bigcup_{1\le i\le N'} S'_i\Big)/G\Big], w\Big)$$

be the subcategory consisting of factorizations $(\mathcal{P}, d_{\mathcal{P}})$ such that

$$(\mathcal{P}, d_{\mathcal{P}})|_{[(\bar{Z}_i \setminus \bigcup_{i'<i} S'_{i'})/G^{\lambda_i=0}]} \in \bigoplus_{j\in[m_i, m_i+\bar{\eta}_i)} \mathrm{MF}\Big(\Big[\Big(\bar{Z}_i \setminus \bigcup_{i'<i} S'_{i'}\Big)/G^{\lambda_i=0}\Big], w|_{\bar{Z}_i}\Big)_{\lambda_i\text{-wt}=j} \tag{2-9}$$

for all $N' < i \le N$. Here $\bar{\eta}_i = \mathrm{wt}_{\lambda_i} \det(\mathbb{L}_{\bar{q}_i})^\vee|_{\bar{Z}_i}$. Then the composition functor

$$\mathbb{W}^l_{m_\bullet}\Big(\Big[\Big(Y \setminus \bigcup_{1\le i\le N'} S'_i\Big)/G\Big], w\Big) \hookrightarrow \mathrm{MF}\Big(\Big[\Big(Y \setminus \bigcup_{1\le i\le N'} S'_i\Big)/G\Big], w\Big) \twoheadrightarrow \mathrm{MF}([Y^{l\text{-ss}}/G], w)$$

is an equivalence.

2.4 Knörrer periodicity

Let Y be a smooth affine scheme and G be an affine algebraic group which acts on Y. Let W be a G–representation, which determines a vector bundle $\mathcal{W} \to \mathcal{Y} := [Y/G]$. Given a function $w\colon \mathcal{Y} \to \mathbb{A}^1$, we have another function on the total space of $\mathcal{W} \oplus \mathcal{W}^\vee$,

$$w + q\colon \mathcal{W} \oplus \mathcal{W}^\vee \to \mathbb{A}^1, \quad \text{where } q(x, x') = \langle x, x'\rangle.$$

We have the diagram

$$\begin{tikzcd} \mathcal{W}^\vee \arrow[r, hook, "i"] \arrow[d, "\mathrm{pr}"'] \arrow[dr, "w"] & \mathcal{W} \oplus \mathcal{W}^\vee \arrow[d, "w+q"] \\ \mathcal{Y} \arrow[r, "w"'] & \mathbb{A}^1 \end{tikzcd}$$

Here $i(x) = (0, x)$. The following is a version of Knörrer periodicity; cf [Hirano 2017a, Theorem 4.2].

Theorem 2.4 *The composition functor*

$$\Phi := i_* \,\mathrm{pr}^*\colon \mathrm{MF}(\mathcal{Y}, w) \xrightarrow{\mathrm{pr}^*} \mathrm{MF}(\mathcal{W}^\vee, w) \xrightarrow{i_*} \mathrm{MF}(\mathcal{W} \oplus \mathcal{W}^\vee, w+q) \tag{2-10}$$

is an equivalence.

Remark 2.5 Here in applying [Hirano 2017a, Theorem 4.2], we view $\mathcal{Y}$ as a closed substack $\mathcal{Y} \hookrightarrow \mathcal{W}$ cut out by the tautological section of the vector bundle $\mathcal{W} \oplus \mathcal{W} \to \mathcal{W}$ given by the second projection, and view $\mathcal{W} \oplus \mathcal{W}^\vee$ as the dual vector bundle of $\mathcal{W} \oplus \mathcal{W} \to \mathcal{W}$.

The equivalence (2-10) is given by taking the tensor product over $\mathcal{O}_{\mathcal{Y}}$ with the following factorization of q on $\mathcal{W} \oplus \mathcal{W}^\vee$

$$i_*\mathcal{O}_{\mathcal{W}^\vee} \rightleftarrows 0.$$

The above factorization is isomorphic to the Koszul factorization of q on $\mathcal{W} \oplus \mathcal{W}^\vee$, which is of the form

$$\Big(\bigwedge\nolimits^{\mathrm{even}} \mathcal{W}^\vee\Big) \otimes_{\mathcal{O}_{\mathcal{Y}}} \mathcal{O}_{\mathcal{W}\oplus\mathcal{W}^\vee} \rightleftarrows \Big(\bigwedge\nolimits^{\mathrm{odd}} \mathcal{W}^\vee\Big) \otimes_{\mathcal{O}_{\mathcal{Y}}} \mathcal{O}_{\mathcal{W}\oplus\mathcal{W}^\vee}. \tag{2-11}$$

(See [Ballard et al. 2014, Proposition 3.20].) Here each differential is given by $\lrcorner s + \wedge t$, where

$$s \colon \mathcal{W}^{\vee} \otimes \mathcal{O}_{\mathcal{W}\oplus\mathcal{W}^{\vee}} \to \mathcal{O}_{\mathcal{W}\oplus\mathcal{W}^{\vee}}$$

corresponds to the tautological section of $\mathcal{W} \oplus \mathcal{W} \to \mathcal{W}$ pulled back to $\mathcal{W} \oplus \mathcal{W}^{\vee}$, and

$$t \in \mathcal{W}^{\vee} \otimes_{\mathcal{O}_{\mathcal{Y}}} \mathcal{W} \subset \mathcal{W}^{\vee} \otimes_{\mathcal{O}_{\mathcal{Y}}} \mathcal{O}_{\mathcal{W}\oplus\mathcal{W}^{\vee}}$$

corresponds to $\mathrm{id} \in \mathrm{Hom}_{\mathcal{Y}}(\mathcal{W}, \mathcal{W})$.

Let $\lambda \colon \mathbb{C}^* \to G$ be a one-parameter subgroup. We have the diagrams

$$\begin{tikzcd}
\mathcal{Y}^{\lambda\geq 0} \arrow[r, "p_\lambda"] \arrow[d, "q_\lambda"'] \arrow[dr, "w^{\lambda\geq 0}"] & \mathcal{Y}, \arrow[d, "w"] \\
\mathcal{Y}^{\lambda=0} \arrow[r, "w^{\lambda=0}"'] & \mathbb{A}^1
\end{tikzcd}
\quad \text{and} \quad
\begin{tikzcd}
(\mathcal{W}\oplus\mathcal{W}^{\vee})^{\lambda\geq 0} \arrow[r, "p'_\lambda"] \arrow[d, "q'_\lambda"'] \arrow[dr, "(w+q)^{\lambda\geq 0}"] & \mathcal{W}\oplus\mathcal{W}^{\vee} \arrow[d, "w+q"] \\
(\mathcal{W}\oplus\mathcal{W}^{\vee})^{\lambda=0} \arrow[r, "w^{\lambda=0}+q^{\lambda=0}"'] & \mathbb{A}^1
\end{tikzcd}$$

of attracting loci. Note that we have equivalences

$$\Phi^{\lambda=0} \colon \mathrm{MF}(\mathcal{Y}^{\lambda=0}, w^{\lambda=0}) \xrightarrow{\sim} \mathrm{MF}((\mathcal{W}\oplus\mathcal{W}^{\vee})^{\lambda=0}, w^{\lambda=0} + q^{\lambda=0}),$$
$$\Phi^{\lambda\geq 0} \colon \mathrm{MF}(\mathcal{Y}^{\lambda\geq 0}, w^{\lambda\geq 0}) \xrightarrow{\sim} \mathrm{MF}((\mathcal{W}^{\lambda\geq 0} \oplus (\mathcal{W}^{\lambda\geq 0})^{\vee}, w^{\lambda\geq 0} + q^{\lambda\geq 0}),$$

by applying Theorem 2.4 for $\mathcal{W}^{\lambda=0} \to \mathcal{Y}^{\lambda=0}$ and $\mathcal{W}^{\lambda\geq 0} \to \mathcal{Y}^{\lambda\geq 0}$, respectively.

Proposition 2.6 *The following diagram commutes*:

$$\begin{tikzcd}
\mathrm{MF}(\mathcal{Y}^{\lambda=0}, w^{\lambda=0}) \arrow[r, "p_{\lambda*}q_\lambda^*"] \arrow[d, "\Phi^{\lambda=0}\circ\otimes(\det\mathcal{W}^{\lambda>0})^{\vee}{[\dim W^{\lambda>0}]}"'] & \mathrm{MF}(\mathcal{Y}, w) \arrow[d, "\Phi"] \\
\mathrm{MF}((\mathcal{W}\oplus\mathcal{W}^{\vee})^{\lambda=0}, w^{\lambda=0} + q^{\lambda=0}) \arrow[r, "p'_{\lambda*}q'^{*}_{\lambda}"] & \mathrm{MF}(\mathcal{W}\oplus\mathcal{W}^{\vee}, w+q)
\end{tikzcd}$$

Proof We have the diagram

(2-12)
$$\begin{tikzcd}
\mathcal{W}^{\lambda=0} \arrow[d] & \mathcal{W}^{\lambda\geq 0} \arrow[l] \arrow[r] \arrow[d] & \mathcal{W} \arrow[d] \\
\mathcal{Y}^{\lambda=0} & \mathcal{Y}^{\lambda\geq 0} \arrow[l, "q_\lambda"'] \arrow[r, "p_\lambda"] & \mathcal{Y}
\end{tikzcd}$$

Here all horizontal diagrams are diagrams of attracting loci, and vertical arrows are projections. From the above diagram, we construct the diagram

(2-13)
$$\begin{tikzcd}
(\mathcal{W}\oplus\mathcal{W}^{\vee})^{\lambda\geq 0} \arrow[r, "r_1"'] \arrow[d, "r_2"] \arrow[rr, bend left, "p'_\lambda"] \arrow[dd, bend right=60, "q'_\lambda"'] \arrow[dr, phantom, "\Box"] & \mathcal{W}^{\lambda\geq 0} \oplus p_\lambda^*\mathcal{W}^{\vee} \arrow[r, "f_2"'] \arrow[d, "g_2"] & \mathcal{W}\oplus\mathcal{W}^{\vee} \arrow[dd, "w+q"] \\
\mathcal{W}^{\lambda\geq 0} \oplus q_\lambda^*(\mathcal{W}^{\lambda=0})^{\vee} \arrow[r, "g_1"'] \arrow[d, "f_1"] & \mathcal{W}^{\lambda\geq 0} \oplus (\mathcal{W}^{\lambda\geq 0})^{\vee} \arrow[dr, "w^{\lambda\geq 0}+q^{\lambda\geq 0}"] & \\
\mathcal{W}^{\lambda=0} \oplus (\mathcal{W}^{\lambda=0})^{\vee} \arrow[rr, "w^{\lambda=0}+q^{\lambda=0}"] & & \mathbb{A}^1
\end{tikzcd}$$

Here (f_1, f_2) is induced by the top horizontal diagram in (2-12), (g_1, g_2) is induced by the duals of the morphisms of vector bundles

$$q_\lambda^* \mathcal{W}^{\lambda=0} \leftarrow \mathcal{W}^{\lambda\geq 0} \rightarrow p_\lambda^* \mathcal{W}$$

on $\mathcal{Y}^{\lambda\geq 0}$, and (r_1, r_2) is induced by the diagram of attracting loci $(\mathcal{W}^\vee)^{\lambda=0} \leftarrow (\mathcal{W}^\vee)^{\lambda\geq 0} \rightarrow \mathcal{W}^\vee$ for $\mathcal{W}^\vee$.

By applying Lemma 6.1 for the right square of (2-12) (and also noting that p_λ and f_2 are proper), we have the commutative diagram

$$\begin{tikzcd}
\mathrm{MF}(\mathcal{Y}^{\lambda\geq 0}, w^{\lambda\geq 0}) \arrow[r, "p_{\lambda *}"] \arrow[d, "\Phi^{\lambda\geq 0}"'] & \mathrm{MF}(\mathcal{Y}, w) \arrow[d, "\Phi"'] \\
\mathrm{MF}(\mathcal{W}^{\lambda\geq 0} \oplus (\mathcal{W}^{\lambda\geq 0})^\vee, w^{\lambda\geq 0} + q^{\lambda\geq 0}) \arrow[r, "f_{2*}g_2^*"'] & \mathrm{MF}(\mathcal{W} \oplus \mathcal{W}^\vee, w+q).
\end{tikzcd}$$

Similarly by applying Lemma 6.2 for the left square of (2-12), we have the commutative diagram

$$\begin{tikzcd}
\mathrm{MF}(\mathcal{Y}^{\lambda=0}, w^{\lambda=0}) \arrow[r, "q_\lambda^*"] \arrow[d, "\Phi^{\lambda=0}"'] & \mathrm{MF}(\mathcal{Y}^{\lambda\geq 0}, w^{\lambda\geq 0}) \arrow[d, "\Phi^{\lambda\geq 0}"'] \\
\mathrm{MF}((\mathcal{W} \oplus \mathcal{W}^\vee)^{\lambda=0}, w^{\lambda=0} + q^{\lambda=0}) \arrow[r, "g_{1!}f_1^*"'] & \mathrm{MF}(\mathcal{W}^{\lambda\geq 0} \oplus (\mathcal{W}^{\lambda\geq 0})^\vee, w^{\lambda\geq 0} + q^{\lambda\geq 0})
\end{tikzcd}$$

Note that we have

$$g_{1!}(-) = g_{1*}\big(- \otimes f_1^* \,\mathrm{pr}^* \det(\mathcal{W}^{\lambda>0})^\vee[\dim W^{\lambda>0}]\big).$$

Here $\mathrm{pr}\colon (\mathcal{W} \oplus \mathcal{W}^\vee)^{\lambda=0} \rightarrow \mathcal{Y}^{\lambda=0}$ is the projection. By the diagram (2-13) and the base change, we have the isomorphism of functors

$$f_{2*}g_2^* g_{1*} f_1^* \cong p'_{\lambda *} q_\lambda'^* \colon \mathrm{MF}(\mathcal{W}^{\lambda\geq 0} \oplus (\mathcal{W}^{\lambda\geq 0})^\vee, w^{\lambda\geq 0} + q^{\lambda\geq 0}) \rightarrow \mathrm{MF}(\mathcal{W} \oplus \mathcal{W}^\vee, w+q).$$

Therefore the proposition holds. $\square$

3 Categorified Hall products for quivers with superpotentials

In this section, we review categorified Hall products for quivers with superpotentials introduced in [Pădurariu 2019; 2023].

3.1 Moduli stacks of representations of quivers

A *quiver* consists of data $Q = (Q_0, Q_1, s, t)$, where Q_0, Q_1 are finite sets and $s, t\colon Q_1 \rightarrow Q_0$ are maps. The set Q_0 is the set of vertices, Q_1 is the set of edges, and s, t are maps which assign source and target of each edge. A *Q–representation* consists of data

$$\mathbb{V} = \{(V_i, u_e) : i \in Q_0, u_e \in \mathrm{Hom}(V_{s(e)}, V_{t(e)})\},$$

where each V_i is a finite-dimensional vector space. The dimension vector $v(\mathbb{V})$ of $\mathbb{V}$ is $(\dim V_i)_{i \in Q_0}$.

For $v=(v_i)_{i\in Q_0}\in \mathbb{Z}_{\geq 0}^{Q_0}$, let $R_Q(v)$ be the vector space

$$R_Q(v)=\bigoplus_{e\in Q_1}\mathrm{Hom}(V_{s(e)},V_{t(e)}),$$

where $\dim V_i=v_i$. The algebraic group $G(v):=\prod_{i\in Q_0}\mathrm{GL}(V_i)$ acts on $R_Q(v)$ by conjugation. The stack of Q–representations of dimension vector v is given by the quotient stack

$$\mathcal{M}_Q(v):=[R_Q(v)/G(v)].$$

We discuss King's θ–stability condition [1994] on $Q^\dagger$–representations. We take

$$\theta=(\theta_i)_{i\in Q_0}\in\mathbb{R}^{Q_0}.$$

For a dimension vector $v\in\mathbb{Z}_{\geq 0}^{Q_0}$, we set $\theta(v)=\sum_{i\in Q_0}\theta_i v_i$. For a Q–representation $\mathbb{V}$, we set $\theta(\mathbb{V}):=\theta(v(\mathbb{V}))$.

Definition 3.1 A Q–representation $\mathbb{V}$ is called θ–(semi)stable if $\theta(\mathbb{V})=0$ and for any nonzero subobject $\mathbb{V}'\subsetneq\mathbb{V}$ we have $\theta(\mathbb{V}')<(\leq)\,0$.

There is an open substack

$$\mathcal{M}_Q^{\theta\text{-ss}}(v)\subset\mathcal{M}_Q(v)$$

corresponding to θ–semistable representations. By [King 1994, Proposition 3.1], if each θ_i is an integer, the above open substack corresponds to the GIT semistable locus with respect to the character

$$\chi_\theta\colon G(v)\to\mathbb{C}^*,\quad (g_i)_{i\in Q_0}\mapsto\prod_{i\in Q_0}\det g_i^{-\theta_i}.$$

By taking the GIT quotient, it admits a good moduli space [Alper 2013]

$$\pi_M\colon\mathcal{M}_Q^{\theta\text{-ss}}(v)\to M_Q^{\theta\text{-ss}}(v) \tag{3-1}$$

such that each closed point of $M_Q^{\theta\text{-ss}}(v)$ corresponds to a θ–polystable Q–representation.

Let $(a_i,b_i)\in\mathbb{Z}_{\geq 0}^2$ be a pair of nonnegative integers for each vertex $i\in Q_0$, and take $c\in\mathbb{Z}_{\geq 0}$. We define the extended quiver $Q^\dagger$ so that its vertex set is $\{\infty\}\cup Q_0$, with edges consist of edges in Q and

$$\sharp(\infty\to i)=a_i,\quad \sharp(i\to\infty)=b_i,\quad \sharp(\infty\to\infty)=c.$$

The $\mathbb{C}^*$–rigidified moduli stack of $Q^\dagger$–representations of dimension vector $(1,d)$ is given by

$$\mathcal{M}_Q^\dagger(v):=[R_{Q^\dagger}(1,v)/G(v)],$$

where 1 is the dimension vector at ∞. Note that there is a natural morphism $\mathcal{M}_{Q^\dagger}(1,v)\to\mathcal{M}_Q^\dagger(v)$ which is a trivial $\mathbb{C}^*$–gerbe. For $\theta=(\theta_\infty,\theta_i)_{i\in Q_0}$ with $\theta(1,v)=0$, the open substack of θ–semistable representations

$$\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)\subset\mathcal{M}_Q^\dagger(v)$$

is defined in a similar way. The condition $\theta(1,v)=0$ determines θ_∞ by $\theta_\infty=-\sum_{i\in Q_0}\theta_i v_i$, so we just write $\theta=(\theta_i)_{i\in Q_0}$.

Remark 3.2 A reason for considering the extended quiver $Q^{\dagger}$ is to rigidify automorphisms of representations of quivers so that the resulting moduli spaces become schemes rather than stacks. Namely $(1, v)$ is the primitive dimension vector of $Q^{\dagger}$–representations so that $\mathcal{M}_{Q}^{\dagger,\theta\text{-ss}}(v)$ is a scheme (indeed smooth quasiprojective variety) for a generic choice of θ. Adding an extended vertex $\{\infty\}$ corresponds to giving a framing in PT stable pair theory.

3.2 Categorified Hall products

For a dimension vector $v \in \mathbb{Z}_{\geq 0}^{Q_0}$, let us take a decomposition

$$v = v^{(1)} + \cdots + v^{(l)}, \quad \text{where } v^{(j)} \in \mathbb{Z}_{\geq 0}^{Q_0}.$$

Let $V_i = \bigoplus_{j=1}^{l} V_i^{(j)}$ be a direct sum decomposition such that $\{V_i^{(j)}\}_{i \in Q_0}$ has dimension vector $v^{(j)}$. We take integers $\lambda^{(1)} > \cdots > \lambda^{(l)}$, and a one-parameter subgroup $\lambda\colon \mathbb{C}^* \to G(v)$ which acts on $V_i^{(j)}$ by weight $\lambda^{(j)}$. We have the stack of attracting loci

$$\mathcal{M}_Q(v^{\bullet}) := [R_Q(v)^{\lambda \geq 0} / G(v)^{\lambda \geq 0}].$$

The above stack is isomorphic to the stack of filtrations of Q–representations

$$0 = \mathbb{V}^{(0)} \subset \mathbb{V}^{(1)} \subset \cdots \subset \mathbb{V}^{(v)} = \mathbb{V} \tag{3-2}$$

such that each $\mathbb{V}^{(j)}/\mathbb{V}^{(j-1)}$ has dimension vector $v^{(j)}$. Moreover, we have

$$\prod_{j=1}^{l} \mathcal{M}_Q(v^{(j)}) = [R_Q(v)^{\lambda=0} / G(v)^{\lambda=0}],$$

and we have the diagram

$$\begin{tikzcd} \mathcal{M}_Q(v^{\bullet}) \arrow[r, "p_\lambda"] \arrow[d, "q_\lambda"'] & \mathcal{M}_Q(v) \\ \prod_{j=1}^{l} \mathcal{M}_Q(v^{(j)}) & \end{tikzcd} \tag{3-3}$$

Here p_λ sends a filtration (3-2) to $\mathbb{V}$, and q_λ sends a filtration (3-2) to its associated graded Q–representation. Since p_λ is proper, we have the functor (called the *categorified Hall product*)

$$p_{\lambda *} q_\lambda^* \colon \bigboxtimes_{j=1}^{l} D^b(\mathcal{M}_Q(v^{(j)})) \to D^b(\mathcal{M}_Q(v)). \tag{3-4}$$

For $\mathscr{E}^{(j)} \in D^b(\mathcal{M}_Q(v^{(j)}))$, we set

$$\mathscr{E}^{(1)} * \cdots * \mathscr{E}^{(l)} := p_{\lambda *} q_\lambda^* (\mathscr{E}^{(1)} \boxtimes \cdots \boxtimes \mathscr{E}^{(l)}).$$

The above $*$–product is associative, ie

$$(\mathscr{E}^{(1)} * \mathscr{E}^{(2)}) * \mathscr{E}^{(3)} \cong \mathscr{E}^{(1)} * (\mathscr{E}^{(2)} * \mathscr{E}^{(3)}) \cong \mathscr{E}^{(1)} * \mathscr{E}^{(2)} * \mathscr{E}^{(3)}.$$

We take $\theta = (\theta_i)_{i \in Q_0}$ such that $\theta(v^{(j)}) = 0$ for all j. Then the diagram (3-3) restricts to the diagram

$$\begin{tikzcd} \mathcal{M}_Q^{\theta\text{-ss}}(v^{\bullet}) \arrow[r, "p_\lambda"] \arrow[d, "q_\lambda"'] & \mathcal{M}_Q^{\theta\text{-ss}}(v) \\ \prod_{j=1}^{l} \mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)}) & \end{tikzcd} \tag{3-5}$$

Similarly we have the functor

$$p_{\lambda *} q_\lambda^* \colon \boxtimes_{j=1}^{l} D^b(\mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)})) \to D^b(\mathcal{M}_Q^{\theta\text{-ss}}(v)),$$

which coincides with (3-4) when $\theta = 0$.

Similarly, applying the above construction for the extended quiver $Q^{\dagger}$, for a decomposition

$$v = v^{(1)} + \cdots + v^{(l)} + v^{(\infty)}, \quad \text{where } v^{(j)} \in \mathbb{Z}^{Q_0}, \tag{3-6}$$

such that $\theta(v^{(j)}) = 0$ for $1 \le j \le l$, we have the functor

$$\boxtimes_{j=1}^{l} D^b(\mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)})) \boxtimes D^b(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v^{(\infty)})) \to D^b(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)). \tag{3-7}$$

By setting $l = 1$, it gives a left action of $\bigoplus_{\theta(v)=0} D^b(\mathcal{M}_Q^{\theta\text{-ss}}(v))$ on $\bigoplus_v D^b(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v))$.

3.3 Categorified Hall products for quivers with superpotentials

Let W be a superpotential of a quiver Q, ie $W \in \mathbb{C}[Q]/[\mathbb{C}[Q], \mathbb{C}[Q]]$, where $\mathbb{C}[Q]$ is the path algebra of Q. Then there is a function

$$w := \mathrm{Tr}(W) \colon \mathcal{M}_Q(v) \to \mathbb{A}^1 \tag{3-8}$$

whose critical locus is identified with the moduli stack of (Q, W)–representations $\mathcal{M}_{(Q,W)}(v)$, ie Q–representations satisfying the relation ∂W.

The diagram (3-5) is extended to the diagram

$$\begin{tikzcd} \mathcal{M}_Q^{\theta\text{-ss}}(v^{\bullet}) \arrow[r, "p_\lambda"] \arrow[d, "q_\lambda"'] & \mathcal{M}_Q^{\theta\text{-ss}}(v) \arrow[d, "w"] \\ \prod_{j=1}^{l} \mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)}) \arrow[r, "\sum_{j=1}^{l} w^{(j)}"] & \mathbb{A}^1 \end{tikzcd} \tag{3-9}$$

Here $w^{(j)}$ is the function (3-8) on $\mathcal{M}_Q(v^{(j)})$. Similarly to (3-4), we have the functor between triangulated categories of factorizations

$$p_{\lambda *} q_\lambda^* \colon \boxtimes_{j=1}^{l} \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)}), w^{(j)}) \to \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v), w),$$

called the *categorified Hall products* for representations of quivers with superpotentials.

The superpotential naturally defines the superpotential of the extended quiver $Q^{\dagger}$, so we have the regular function $w\colon \mathcal{M}_Q^{\dagger}(v) \to \mathbb{A}^1$ as in (3-8). Similarly to (3-7), for a decomposition (3-6) we have the left action

$$\boxtimes_{j=1}^{l} \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)}), w^{(j)}) \boxtimes \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v^{(\infty)}), w^{(\infty)}) \to \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w). \tag{3-10}$$

Note that we have the decomposition (2-2) with respect to the diagonal torus $\mathbb{C}^* \subset G(v)$

$$\mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v), w) = \bigoplus_{j\in\mathbb{Z}} \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v), w)_j.$$

We will often restrict the functor (3-10) to the fixed weight spaces

$$\boxtimes_{j=1}^{l} \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)}), w^{(j)})_{i_j} \boxtimes \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v^{(\infty)}), w^{(\infty)}) \to \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w).$$

3.4 Base change to formal fibers

Later we will take a base change of the categorified Hall product to a formal neighborhood of a point in the good moduli space (3-1). The diagram (3-9) extends to the commutative diagram

$$\begin{tikzcd} \mathcal{M}_Q^{\theta\text{-ss}}(v^{\bullet}) \arrow[r, "p_\lambda"] \arrow[d, "q_\lambda"'] & \mathcal{M}_Q^{\theta\text{-ss}}(v) \arrow[dd, "\pi_M"] \arrow[ddr, "w"] & \\ \prod_{j=1}^{l} \mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)}) \arrow[d] & & \\ \prod_{j=1}^{l} M_Q^{\theta\text{-ss}}(v^{(j)}) \arrow[r, "\oplus"] & M_Q^{\theta\text{-ss}}(v) \arrow[r] & \mathbb{A}^1 \end{tikzcd} \tag{3-11}$$

Here the bottom arrow is the morphism taking the direct sum of θ–polystable representations, which is a finite morphism (see [Meinhardt and Reineke 2019, Lemma 2.1]), and the left bottom vertical arrow is the good moduli space morphism. For a closed point $p \in M_Q^{\theta\text{-ss}}(v)$, we consider the following formal fiber

$$\widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v)_p := \mathcal{M}_Q^{\theta\text{-ss}}(v) \times_{M_Q^{\theta\text{-ss}}(v)} \widehat{M}_Q^{\theta\text{-ss}}(v)_p \to \widehat{M}_Q^{\theta\text{-ss}}(v)_p := \operatorname{Spec} \widehat{\mathcal{O}}_{M_Q^{\theta\text{-ss}}(v),p}.$$

Let $(p^{(1)}, \dots, p^{(l)}) \in \prod_{j=1}^{l} M_Q^{\theta\text{-ss}}(v^{(j)})$ be a point such that $\oplus(p^{(1)}, \dots, p^{(l)}) = p$. By taking the fiber product of the diagram (3-11) by $\widehat{M}_Q^{\theta\text{-ss}}(v)_p \to M_Q^{\theta\text{-ss}}(v)$, we obtain the diagram

$$\begin{tikzcd} \widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v^{\bullet})_p \arrow[r, "\widehat{p}_\lambda"] \arrow[d, "\widehat{q}_\lambda"'] & \widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v)_p \\ \coprod_{p^{(\bullet)} \in \oplus^{-1}(p)} \prod_{j=1}^{l} \widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v^{(j)})_{p^{(j)}} & \end{tikzcd} \tag{3-12}$$

The above diagram is a diagram of attracting loci for $\widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v)_p$; see [Toda 2021, Lemma 4.11].

By the derived base change, we have the commutative diagram

$$\begin{tikzcd} \boxtimes_{j=1}^{l} D^b(\mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)})) \arrow[r, "p_{\lambda*}q_\lambda^*"] \arrow[d] & D^b(\mathcal{M}_Q^{\theta\text{-ss}}(v)) \arrow[d] \\ \bigoplus_{p^{(\bullet)}\in\oplus^{-1}(p)} \boxtimes_{j=1}^{l} D^b(\widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v^{(j)})_{p^{(j)}}) \arrow[r, "\widehat{p}_{\lambda*}\widehat{q}_\lambda^*"] & D^b(\widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v)_p) \end{tikzcd} \tag{3-13}$$

Here the vertical arrows are pullbacks to formal fibers.

We denote by $\widehat{w}_p \colon \widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v) \to \mathbb{A}^1$ the pullback of the function (3-8) to the formal fiber. Similarly to (3-13), we have the commutative diagram

$$\begin{tikzcd} \boxtimes_{j=1}^{l} \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v^{(j)}), w) \arrow[r, "p_{\lambda*}q_\lambda^*"] \arrow[d] & \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(v)) \arrow[d] \\ \bigoplus_{p^{(\bullet)}\in\oplus^{-1}(p)} \boxtimes_{j=1}^{l} \mathrm{MF}(\widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v^{(j)})_{p^{(j)}}, \widehat{w}^{(j)}_{p^{(j)}}) \arrow[r, "\widehat{p}_{\lambda*}\widehat{q}_\lambda^*"] & \mathrm{MF}(\widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(v)_p, \widehat{w}_p) \end{tikzcd} \tag{3-14}$$

4 Derived categories of Grassmannian flips

In this section, we use categorified Hall products to refine the result of [Ballard et al. 2021, Theorem 5.4.4] on variation of derived categories under Grassmannian flips.

4.1 Grassmannian flips

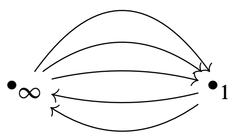
Let V be a vector space with dimension d, and let A and B be other vector spaces such that

$$a := \dim A \quad \text{and} \quad b := \dim B, \quad \text{with } a \geq b.$$

We form the quotient stack

$$\mathcal{G}_{a,b}(d) := [(\mathrm{Hom}(A, V) \oplus \mathrm{Hom}(V, B))/\mathrm{GL}(V)]. \tag{4-1}$$

Remark 4.1 The stack $\mathcal{G}_{a,b}(d)$ is the $\mathbb{C}^*$–rigidified moduli stack of representations of the quiver $Q_{a,b}$ of dimension vector $(1, d)$, where the vertex set is $\{\infty, 1\}$, the number of arrows from ∞ to 1 is a, that from 1 to ∞ is b, and there are no self-loops; see Section 3.1. For instance the quiver $Q_{3,2}$ is described by

(4-2) $Q_{3,2} =$ 

Below we fix a basis of V, and take the maximal torus $T \subset \mathrm{GL}(V)$ to be consisting of diagonal matrices. For a one-parameter subgroup $\lambda \colon \mathbb{C}^* \to T$, we use the following notation for the diagram of attracting loci as in (2-3):

$$\begin{tikzcd} \mathcal{G}_{a,b}(d)^{\lambda\geq 0} \arrow[r, "p_\lambda"] \arrow[d, "q_\lambda"'] & \mathcal{G}_{a,b}(d) \\ \mathcal{G}_{a,b}(d)^{\lambda=0} & \end{tikzcd} \tag{4-3}$$

We use the determinant character

$$\chi_0 \colon \mathrm{GL}(V) \to \mathbb{C}^*, \quad g \mapsto \det(g), \tag{4-4}$$

and often regard it as a line bundle on $\mathcal{G}_{a,b}(d)$. There exist two GIT quotients with respect to $\chi_0^{\pm 1}$ given by open substacks

$$G^{\pm}_{a,b}(d) \subset \mathcal{G}_{a,b}(d).$$

Here χ_0–semistable locus $G^+_{a,b}(d)$ consists of $(\alpha, \beta) \in \mathrm{Hom}(A, V) \oplus \mathrm{Hom}(V, B)$ such that $\alpha \colon A \to V$ is surjective, and χ_0^{-1}–semistable locus $G^-_{a,b}(d)$ consists of (α, β) such that $\beta \colon V \to B$ is injective. We have the diagram

$$\begin{tikzcd} G^+_{a,b}(d) \arrow[r, hook] \arrow[dr] & \mathcal{G}_{a,b}(d) \arrow[d] & G^-_{a,b}(d) \arrow[l, hook'] \arrow[dl] \\ & G^0_{a,b}(d) & \end{tikzcd} \tag{4-5}$$

Here the middle vertical arrow is the good moduli space for $\mathcal{G}_{a,b}(d)$.

Remark 4.2 When $a \geq d$ and $b = 0$, then $G^-_{a,0}(d) = \varnothing$ and $G^+_{a,0}(d)$ is the Grassmannian parametrizing surjections $A \twoheadrightarrow V$. If $a \geq b \geq d$, then $G^{\pm}_{a,b}(d) \to G^0_{a,b}(d)$ are birational and $G^+_{a,b}(d) \dashrightarrow G^-_{a,b}(d)$ is a flip $(a > b)$, flop $(a = b)$.

We have the KN stratifications with respect to $\chi_0^{\pm 1}$,

$$\mathcal{G}_{a,b}(d) = \mathcal{S}_0^{\pm} \sqcup \mathcal{S}_1^{\pm} \sqcup \cdots \sqcup \mathcal{S}_{d-1}^{\pm} \sqcup G^{\pm}_{a,b}(d),$$

where $\mathcal{S}_i^+$ consists of (α, β) such that the image of $\alpha \colon A \to V$ has dimension i, and $\mathcal{S}_i^-$ consists of (α, β) such that the kernel of $\beta \colon V \to B$ has dimension $d-i$. The associated one-parameter subgroups $\lambda_i^{\pm} \colon \mathbb{C}^* \to T$ are taken as

$$\lambda_i^+(t) = (\overbrace{1, \ldots, 1}^{i}, \overbrace{t^{-1}, \ldots, t^{-1}}^{d-i}), \;\; \lambda_i^-(t) = (\overbrace{t, \ldots, t}^{d-i}, \overbrace{1, \ldots, 1}^{i}). \tag{4-6}$$

(See [Halpern-Leistner 2015, Example 4.12].)

4.2 Window subcategories for Grassmannian flips

We fix a Borel subgroup $B \subset \mathrm{GL}(V)$ to be consisting of upper-triangular matrices, and set roots of B to be negative roots. Let $M = \mathbb{Z}^d$ be the character lattice for $\mathrm{GL}(V)$, and $M^+_{\mathbb{R}} \subset M_{\mathbb{R}}$ the dominant chamber. By the above choice of negative roots, we have

$$M^+_{\mathbb{R}} = \{(x_1, x_2, \ldots, x_d) \in \mathbb{R}^d : x_1 \leq x_2 \leq \cdots \leq x_d\}.$$

We set $M^+ := M^+_{\mathbb{R}} \cap M$. For $c \in \mathbb{Z}$, we set

$$\mathbb{B}_c(d) := \{(x_1, x_2, \ldots, x_d) \in M^+ : 0 \leq x_i \leq c - d\}. \tag{4-7}$$

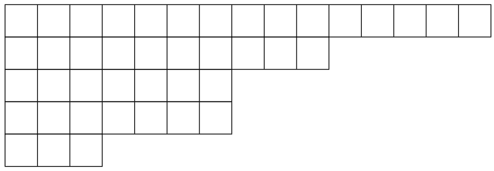


Figure 2: $(3,7,7,10,15) \in \mathbb{B}_c(d)$ for $d=5$ and $c \geq 20$.

Here $\mathbb{B}_c(d)=\varnothing$ if $c<d$. For $\chi \in \mathbb{B}_c(d)$, we assign the Young diagram whose number of boxes at the i^{th} row is x_{d-i+1}. The above assignment identifies $\mathbb{B}_c(d)$ with the set of Young diagrams with height less than or equal to d, width less than or equal to $c-d$. For example, Figure 2 illustrates the case of $(3,7,7,10,15) \in \mathbb{B}_c(d)$ for $d=5$ and $c \geq 20$.

We define the triangulated subcategory

$$\mathbb{W}_c(d) \subset D^b(\mathscr{G}_{a,b}(d)) \tag{4-8}$$

to be the smallest thick triangulated subcategory which contains $V(\chi) \otimes \mathcal{O}_{\mathscr{G}_{a,b}(d)}$ for $\chi \in \mathbb{B}_c(d)$. Here $V(\chi)$ is the irreducible $\mathrm{GL}(V)$ representation with highest weight χ, ie it is a Schur power of V associated with the Young diagram corresponding to χ. The following proposition is well-known (see [Donovan and Segal 2014, Proposition 3.6]), which gives window subcategories for Grassmannian flips. We reprove it here using Theorem 2.2.

Proposition 4.3 *The following composition functors are equivalences*:

$$\begin{aligned} &\mathbb{W}_b(d) \subset D^b(\mathscr{G}_{a,b}(d)) \twoheadrightarrow D^b(G^-_{a,b}(d)), \\ &\mathbb{W}_a(d) \subset D^b(\mathscr{G}_{a,b}(d)) \twoheadrightarrow D^b(G^+_{a,b}(d)). \end{aligned} \tag{4-9}$$

Proof We only prove the statement for $+$. Let λ_i^+ be the one-parameter subgroup in (4-6). Then η_i^+ given in (2-6) is

$$\eta_i^+ = \langle \lambda_i^+, (\mathrm{Hom}(A,V)^\vee \oplus \mathrm{Hom}(V,B)^\vee)^{\lambda_i^+>0} - \mathrm{End}(V)^{\lambda_i^+>0}\rangle = (a-i)(d-i).$$

Let $\chi'=(x'_1,\dots,x'_d)$ be a T–weight of $V(\chi)$ for $\chi \in \mathbb{B}_a(d)$. Then $0 \leq x'_j \leq a-d$ for $1 \leq j \leq d$, so

$$-\eta_i^+ < -(a-d)(d-i) \leq \langle \chi', \lambda_i^+\rangle = -\sum_{j=i+1}^{d} x'_j \leq 0.$$

Therefore by setting $m_i = -\eta_i^+ + \varepsilon$ for $0<\varepsilon \ll 1$ and $l=\chi_0$ in (2-7), we have

$$\mathbb{W}_a(d) \subset \mathbb{W}^{\chi_0}_{m_\bullet}(\mathscr{G}_{a,b}(d)) \subset D^b(\mathscr{G}_{a,b}(d)).$$

It follows that the second composition functor in (4-9) is fully faithful.

In order to show that it is essentially surjective, note that the projection to $\mathrm{Hom}(A,V)$ defines a morphism

$$G^+_{a,b}(d) \to \mathrm{Gr}(a,d), \tag{4-10}$$

where $\mathrm{Gr}(a,d)$ is the Grassmannian which parametrizes d–dimensional quotients of A. By the above morphism, $G^+_{a,b}(d)$ is the total space of a vector bundle over $\mathrm{Gr}(a,d)$. The objects $V(\chi)\otimes\mathcal{O}_{\mathcal{G}_{a,b}(d)}$ for $\chi\in\mathbb{B}_a(d)$ restricted to the zero section of (4-10) forms Kapranov's exceptional collection [1984]. Since $D^b(G^+_{a,b}(d))$ is generated by pullbacks of objects from $D^b(\mathrm{Gr}(a,d))$, the essential surjectivity of (4-9) holds. □

4.3 Resolutions of categorified Hall products

Let $d=d^{(1)}+\cdots+d^{(l)}+d^{(\infty)}$ be a decomposition of d. Note that from Section 3.1, we have the categorified Hall product

$$\boxtimes_{j=1}^{l} D^b(B\mathrm{GL}(d^{(j)}))\boxtimes D^b(\mathcal{G}_{a,b}(d^{(\infty)}))\to D^b(\mathcal{G}_{a,b}(d)).$$

In particular by setting $d^{(1)}=1$ and $d^{(\infty)}=d-1$, we have the functor

$$*\colon D^b(B\mathbb{C}^*)\boxtimes D^b(\mathcal{G}_{a,b}(d-1))\to D^b(\mathcal{G}_{a,b}(d)). \tag{4-11}$$

It is explicitly given as follows. Let $\lambda\colon\mathbb{C}^*\to T$ be given by

$$\lambda(t)=(t,1,\dots,1). \tag{4-12}$$

Then we have the decomposition $V=V^{\lambda>0}\oplus V^{\lambda=0}$, where $V^{\lambda>0}$ is one-dimensional. Then

$$\begin{aligned}\mathcal{G}_{a,b}(d)^{\lambda=0}&=[B\mathrm{GL}(V^{\lambda>0})]\times\big[\big(\mathrm{Hom}(A,V^{\lambda=0})\oplus\mathrm{Hom}(V^{\lambda=0},B)\big)/\mathrm{GL}(V^{\lambda=0})\big]\\&=B\mathbb{C}^*\times\mathcal{G}_{a,b}(d-1).\end{aligned}$$

The functor (4-11) is given by $p_{\lambda*}q_\lambda^*(-)$ in the diagram (4-3). The stack $\mathcal{G}_{a,b}(d)^{\lambda\geq 0}$ is the moduli stack of exact sequences of $Q_{a,b}$–representations

$$0\to\mathbb{V}^{\lambda>0}\to\mathbb{V}\to\mathbb{V}^{\lambda=0}\to 0 \tag{4-13}$$

such that $\mathbb{V}^{\lambda>0}$ has dimension vector $(0,1)$. We will often use the following lemmas:

Lemma 4.4 *For $\mathcal{E}_1\in D^b(B\mathbb{C}^*)$ and $\mathcal{E}_2\in D^b(\mathcal{G}_{a,b}(d-1))$, we have*

$$(\mathcal{E}_1*\mathcal{E}_2)\otimes\chi_0^j=(\mathcal{E}_1\otimes\mathcal{O}_{B\mathbb{C}^*}(j))*(\mathcal{E}_2\otimes\chi_0^j).$$

Here we have used the same symbol χ_0 for the determinant character of $\mathrm{GL}(V^{\lambda=0})$.

Proof The lemma follows using $p_\lambda^*\chi_0=\mathcal{O}_{B\mathbb{C}^*}(1)\boxtimes\chi_0$ and the definition of the functor (4-11). □

Lemma 4.5 *For $\mathcal{E}\in\mathbb{W}_c(d)$ and $j\geq 0$, we have $\mathcal{E}\otimes\chi_0^j\in\mathbb{W}_{c+j}(d)$.*

Proof The lemma follows since $V(\chi)\otimes\chi_0^j=V(\chi')$, where $\chi'=\chi+(j,j,\dots,j)$. □

The following proposition is essentially proved in [Donovan and Segal 2014; Ballard et al. 2021], which we interpret in terms of categorified Hall products.

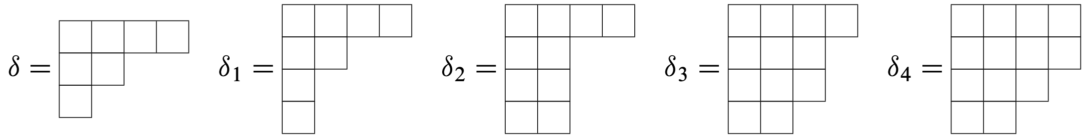


Figure 3: Algorithm for $\chi = (4, 2, 1)$, $d = 4$, $c = b = 7$.

Proposition 4.6 [Donovan and Segal 2014, Theorem A.7; Ballard et al. 2021, Proposition 5.4.6] *For $\chi \in \mathbb{B}_c(d-1)$ with $c \geq b$, let δ be the corresponding Young diagram. Then the object*

$$\mathcal{O}_{B\mathbb{C}^*} * (V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})$$

is a sheaf which fits into an exact sequence

$$\begin{aligned}(4\text{-}14)\quad 0 \to V(\chi_K) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}^{\oplus m_K} \to \cdots \to V(\chi_1) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}^{\oplus m_1}& \\ \to V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)} \to \mathcal{O}_{B\mathbb{C}^*} * (V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}) &\to 0.\end{aligned}$$

Here $\chi \in \mathbb{B}_c(d-1)$ is regarded as an element of $\mathbb{B}_{c+1}(d)$ by $(x_2, \ldots, x_d) \mapsto (0, x_2, \ldots, x_d)$, and each $\chi_i \in \mathbb{B}_{c+1}(d)$ in (4-14) corresponds to a Young diagram δ_i obtained from δ by the following algorithm:

- *The Young diagram δ_1 is obtained from δ by adding boxes to the first column until it reaches to height d.*
- *The diagram δ_i is obtained from δ_{i-1} by adding boxes to the i^{th} column until its height is one more than the height of the $(i-1)^{th}$ column of δ.*

See Figure 3. Moreover, $m_i = \dim \bigwedge^{s_i} B$ for $s_i = |\delta_i| - |\delta|$, and the sequence (4-14) terminates when we reach a positive integer K such that $s_{K+1} > b$.

Proof Let λ be the one-parameter subgroup (4-12). Then we have

$$\begin{aligned}(\mathrm{Hom}(A, V) \oplus \mathrm{Hom}(V, B))^{\lambda \geq 0} &= \mathrm{Hom}(A, V) \oplus \mathrm{Hom}(V^{\lambda=0}, B) \\ &\cong \mathrm{Hom}(V^\vee, A^\vee) \oplus \mathrm{Hom}(B^\vee, (V^{\lambda=0})^\vee).\end{aligned}$$

The parabolic subgroup $\mathrm{GL}(V)^{\lambda \geq 0}$ is the subgroup of $\mathrm{GL}(V)$ which preserves $V^{\lambda>0} \subset V$. Therefore there is an isomorphism of quotient stacks

$$\begin{aligned}&[(\mathrm{Hom}(A, V) \oplus \mathrm{Hom}(V, B))^{\lambda \geq 0} / \mathrm{GL}(V)^{\lambda \geq 0}] \\ &\qquad \xrightarrow{\cong} [\mathrm{Hom}(V^\vee, A^\vee) \oplus \mathrm{Hom}(B^\vee, (V^{\lambda=0})^\vee) \oplus \mathrm{Hom}^{\mathrm{inj}}((V^{\lambda=0})^\vee, V^\vee) / \mathrm{GL}(V) \times \mathrm{GL}(V^{\lambda=0})].\end{aligned}$$

Here $\mathrm{Hom}^{\mathrm{inj}}((V^{\lambda=0})^\vee, V^\vee) \subset \mathrm{Hom}((V^{\lambda=0})^\vee, V^\vee)$ is the subset consisting of injective homomorphisms. The above isomorphism is induced by the embedding into the direct summand $(V^{\lambda=0})^\vee \hookrightarrow V^\vee$ together with the natural inclusion $\mathrm{GL}(V)^{\lambda \geq 0} \hookrightarrow \mathrm{GL}(V) \times \mathrm{GL}(V^{\lambda=0})$. Under the above isomorphism, the morphism

$$p_\lambda \colon [(\mathrm{Hom}(A, V) \oplus \mathrm{Hom}(V, B))^{\lambda \geq 0} / \mathrm{GL}(V)^{\lambda \geq 0}] \to \mathcal{G}_{a,b}(d)$$

from the diagram (3-3) is identified with the morphism

$$[\mathrm{Hom}(V^\vee, A^\vee) \oplus \mathrm{Hom}(B^\vee, (V^{\lambda=0})^\vee) \oplus \mathrm{Hom}^{\mathrm{inj}}((V^{\lambda=0})^\vee, V^\vee)/\mathrm{GL}(V) \times \mathrm{GL}(V^{\lambda=0})] \xrightarrow{p_\lambda} [\mathrm{Hom}(V^\vee, A^\vee) \oplus \mathrm{Hom}(B^\vee, V^\vee)/\mathrm{GL}(V)]$$

induced by the composition of maps. The above morphism is nothing but the one considered in [Donovan and Segal 2014, Theorem A.7; Ballard et al. 2021, Proposition 5.4.6]. We then directly apply the computation of $p_{\lambda *}(-)$ for vector bundles given by Schur powers in [Donovan and Segal 2014, Theorem A.7; Ballard et al. 2021, Proposition 5.4.6] to obtain the resolution (4-14).

We also check that each χ_i in (4-14) is an element of $\mathbb{B}_{c+1}(d)$, ie δ_i has at most height d and width $c-d+1$. It is obvious that δ_i has at most height d. Let μ_j be the number of boxes of δ at the j^{th} column. Then from the algorithm defining δ_i, we have

$$s_i = (d-\mu_1) + (\mu_1 + 1 - \mu_2) + \cdots + (\mu_{i-1} + 1 - \mu_i) = d + i - 1 - \mu_i.$$

Since $\chi \in \mathbb{B}_c(d-1)$, we have $\mu_{c-d+2} = 0$, so $s_{c-d+2} = c+1 > b$. Therefore we have $K \leq c-d+1$. Since δ has width at most $c-d+1$, it follows that δ_i also has width at most $c-d+1$. □

Using the above proposition, we have the following lemma:

Lemma 4.7 *For $0 \leq j \leq c-1$, we have*

$$\mathbb{O}_{B\mathbb{C}^*}(j) * (\mathbb{W}_{c-1-j}(d-1) \otimes \chi_0^j) \subset \mathbb{W}_c(d). \tag{4-15}$$

Proof We have

$$\mathbb{O}_{B\mathbb{C}^*}(j) * (\mathbb{W}_{c-1-j}(d-1) \otimes \chi_0^j) = (\mathbb{O}_{B\mathbb{C}^*} * \mathbb{W}_{c-1-j}(d-1)) \otimes \chi_0^j \subset \mathbb{W}_{c-j}(d) \otimes \chi_0^j \subset \mathbb{W}_c(d).$$

Here we have used Lemma 4.4 for the first identity, Proposition 4.6 for the first inclusion and Lemma 4.5 for the last inclusion. □

4.4 Generation of window subcategories

We show that for $c \geq b$ the category $\mathbb{W}_c(d)$ is generated by its subcategory $\mathbb{W}_b(d)$ and subcategories (4-15) for $0 \leq j \leq c-b-1$. We first prove the case of $c = b+1$, which is a variant of [Ballard et al. 2021, Lemma 5.4.9].

Lemma 4.8 *The subcategory $\mathbb{W}_{b+1}(d) \subset D^b(\mathscr{G}_{a,b}(d))$ is generated by $\mathbb{W}_b(d)$ and $\mathbb{O}_{B\mathbb{C}^*} * \mathbb{W}_b(d-1)$.*

Proof For $\chi \in \mathbb{B}_{b+1}(d)$, it is enough to show that $V(\chi) \otimes \mathbb{O}_{\mathscr{G}_{a,b}(d)}$ is generated by

$$\mathbb{W}_b(d) \quad \text{and} \quad \mathbb{O}_{B\mathbb{C}^*} * \mathbb{W}_b(d-1).$$

Let δ be the Young diagram corresponding to χ, and we denote by μ_j the number of boxes of δ at the j^{th} column. We may assume that the width of δ is exactly $b-d+1$, ie $\mu_j \geq 1$ for $1 \leq j \leq b-d+1$ and $\mu_{b-d+2} = 0$.

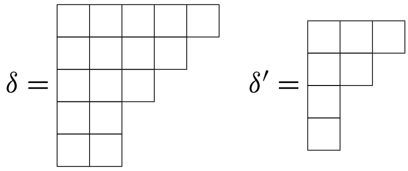


Figure 4: The diagrams δ and δ' for $\chi = (2,2,3,4,5) \in \mathbb{B}_{10}(5)$.

Suppose that the height of δ is exactly d, ie $\mu_1 = d$. We define another Young diagram δ' whose number of boxes at the j^{th} column is $\mu_{j+1} - 1$. Then the height of δ' is at most $d-1$, and the width of δ' is at most $b-d$; see Figure 4. Let $\chi' \in \mathbb{B}_{b-1}(d-1)$ be the character corresponding to δ'. As $\mathbb{B}_{b-1}(d-1) \subset \mathbb{B}_b(d-1)$, we apply Proposition 4.6 to obtain a resolution

$$(4\text{-}16) \quad 0 \to V(\chi'_K) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}^{\oplus m_K} \to \cdots \to V(\chi'_1) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}^{\oplus m_1} \\ \to V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)} \to \mathcal{O}_{B\mathbb{C}^*} * (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}) \to 0.$$

for $\chi'_i \in \mathbb{B}_{b+1}(d)$. Note that we have

$$|\delta| - |\delta'| = (d - \mu_2 + 1) + (\mu_2 - \mu_3 + 1) + \cdots + (\mu_{b-d} - \mu_{b-d+1} + 1) + \mu_{b-d+1} = b.$$

From the construction of δ', the Young diagram δ is reconstructed from δ' by the algorithm given in Proposition 4.6 at the $(b-d+1)^{\text{th}}$ step. Therefore, from the above identity, it follows that there are exactly $b-d+1$ terms of the resolution (4-16), ie $K = b-d+1$, and also $m_K = 1$, $\chi'_K = \chi$. Moreover, since the width of δ' is as most $b-d$, we also have $\chi'_i \in \mathbb{B}_b(d)$ for $0 \le i < b-d+1$. Therefore $V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}$ is generated by objects in $\mathbb{W}_b(d)$ and $\mathcal{O}_{B\mathbb{C}^*} * (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}) \in \mathcal{O}_{B\mathbb{C}^*} * \mathbb{W}_b(d-1)$.

Suppose that the height of δ is less than d. Then we have $\chi \in \mathbb{B}_b(d-1)$. By applying Proposition 4.6, we see that $V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}$ is generated by $\mathcal{O}_{B\mathbb{C}^*} * (V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}) \in \mathcal{O}_{B\mathbb{C}^*} * \mathbb{W}_b(d-1)$ and $V(\chi_i) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}$ for $\chi_i \in \mathbb{B}_{b+1}(d)$. By the algorithm in Proposition 4.6, the Young diagram corresponding to χ_i has a full column, ie the height of χ_i is exactly d. Therefore by the above argument, each $V(\chi_i) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}$ is generated by $\mathbb{W}_b(d)$ and $\mathcal{O}_{B\mathbb{C}^*} * \mathbb{W}_b(d-1)$. $\square$

We then show the generation for $\mathbb{W}_c(d)$:

Lemma 4.9 *For $c \ge b$, the subcategory $\mathbb{W}_c(d) \subset D^b(\mathcal{G}_{a,b}(d))$ is generated by $\mathbb{W}_b(d)$ and*

$$\mathcal{O}_{B\mathbb{C}^*}(j) * (\mathbb{W}_{c-1-j}(d-1) \otimes \chi_0^j) \quad \text{for } 0 \le j \le c-b-1.$$

Proof The case of $c = b+1$ is proved in Lemma 4.8. We prove the lemma for $c > b+1$ by induction on c. For $\chi \in \mathbb{B}_c(d)$, suppose that the corresponding Young diagram δ has a full column. Let δ'' be the Young diagram obtained by removing the first column, and χ'' the corresponding character. Then $\chi'' \in \mathbb{B}_{c-1}(d)$, so by the induction hypothesis $V(\chi'') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}$ is generated by $\mathbb{W}_b(d)$ and $\mathcal{O}_{B\mathbb{C}^*}(j) * (\mathbb{W}_{c-2-j}(d-1) \otimes \chi_0^j)$ for $0 \le j \le c-b-2$. By taking the tensor product with χ_0 and setting $j' = j+1$, we see that

$V(\chi)\otimes \mathcal{O}_{\mathscr{G}_{a,b}(d)}$ is generated by $\mathbb{W}_b(d)\otimes\chi_0$ and $\mathcal{O}_{B\mathbb{C}^*}(j')*(\mathbb{W}_{c-1-j'}(d-1)\otimes\chi_0^{j'})$ for $1\le j'\le c-b-1$. Since $\mathbb{W}_b(d)\otimes\chi_0\subset\mathbb{W}_{b+1}(d)$, and the latter is generated by $\mathbb{W}_b(d)$ and $\mathcal{O}_{B\mathbb{C}^*}*\mathbb{W}_b(d-1)\subset\mathcal{O}_{B\mathbb{C}^*}*\mathbb{W}_{c-1}(d-1)$ by Lemma 4.8, we have the desired generation for $V(\chi)\otimes\mathcal{O}_{\mathscr{G}_{a,b}(d)}$ when δ has a full column.

If δ does not have a full column, then $\chi\in\mathbb{B}_{c-1}(d-1)$. By applying Proposition 4.6, we see that $V(\chi)\otimes\mathcal{O}_{\mathscr{G}_{a,b}(d)}$ is generated by $\mathcal{O}_{B\mathbb{C}^*}*(V(\chi)\otimes\mathcal{O}_{\mathscr{G}_{a,b}(d-1)})$ and $V(\chi_i)\otimes\mathcal{O}_{\mathscr{G}_{a,b}(d)}$ for $\chi_i\in\mathbb{B}_c(d)$. Since each Young diagram corresponding to χ_i has a full column, the desired generation of $V(\chi)\otimes\mathcal{O}_{\mathscr{G}_{a,b}(d)}$ is reduced to the case of the existence of a full column, which is proved above. □

Remark 4.10 The results of Lemmas 4.8 and 4.9 are essentially proved using [Ballard et al. 2021, Lemma 5.4.9] combined with the definition of $O_{d,s}$ in [Ballard et al. 2021, Definition 5.4.2]. Also several cohomology vanishing calculations, which will be given in Section 4.5, are also given in loc. cit. We have re-proved them in order to make them compatible with our notation, and to state them in terms of categorical Hall products.

The above generation result is stated in terms of iterated Hall products as follows:

Proposition 4.11 *For $c\ge b$, the subcategory $\mathbb{W}_c(d)\subset D^b(\mathscr{G}_{a,b}(d))$ is generated by the subcategories*

$$\mathscr{C}_{j_\bullet}:=\mathcal{O}_{B\mathbb{C}^*}(j_1)*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_l)*(\mathbb{W}_b(d-l)\otimes\chi_0^{j_l})\subset D^b(\mathscr{G}_{a,b}(d)) \tag{4-17}$$

for $0\le l\le d$ and $0\le j_1\le\cdots\le j_l\le c-b-l$. Here when $l=0$, the above subcategory is set to be $\mathbb{W}_b(d)$.

Proof We first show that (4-17) are subcategories of $\mathbb{W}_c(d)$ by the induction on c. By Lemma 4.4, the subcategory (4-17) is written as

$$\mathcal{O}_{B\mathbb{C}^*}(j_1)*\big((\mathcal{O}_{B\mathbb{C}^*}(j_2-j_1)*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_l-j_1)*\big(\mathbb{W}_b((d-1)-(l-1))\otimes\chi_0^{j_l-j_1}\big)\big)\otimes\chi_0^{j_1}\big).$$

Since $j_l-j_1\le(c-1-j_1)-b-(l-1)$, by the induction hypothesis we have

$$\mathcal{O}_{B\mathbb{C}^*}(j_2-j_1)*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_l-j_1)*\big(\mathbb{W}_b((d-1)-(l-1))\otimes\chi_0^{j_l-j_1}\big)\in\mathbb{W}_{c-1-j_1}(d-1).$$

Therefore (4-17) is a subcategory of $\mathbb{W}_c(d)$ by Lemma 4.7.

We next show that $\mathbb{W}_c(d)$ is generated by subcategories (4-17) by the induction on c. By Lemma 4.9, the subcategory $\mathbb{W}_c(d)$ is generated by $\mathbb{W}_b(d)$ and $\mathcal{O}_{B\mathbb{C}^*}(j)*(\mathbb{W}_{c-1-j}(d-1)\otimes\chi_0^{j})$ for $0\le j\le c-b-1$. By the induction hypothesis and Lemma 4.4, $\mathcal{O}_{B\mathbb{C}^*}(j)*(\mathbb{W}_{c-1-j}(d-1)\otimes\chi_0^{j})$ is generated by

$$\begin{aligned}\mathcal{O}_{B\mathbb{C}^*}(j)*\big((\mathcal{O}_{B\mathbb{C}^*}(j_1)*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_{l'})*(\mathbb{W}_b(d-1-l')\otimes\chi_0^{j_{l'}}))\otimes\chi_0^{j}\big)\\ =\mathcal{O}_{B\mathbb{C}^*}(j)*\mathcal{O}_{B\mathbb{C}^*}(j+j_1)*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j+j_{l'})*(\mathbb{W}_b(d-1-l')\otimes\chi_0^{j+j_{l'}})\end{aligned}$$

for $0\le l'\le d-1$ and $0\le j_1\le\cdots\le j_{l'}\le(c-1-j)-b-l'$. Since $j+j_{l'}\le c-b-(l'+1)$, the above subcategory is of the form (4-17) for $l=l'+1$. Therefore we obtain the desired generation. □

Remark 4.12 Let $F_j(-) := (\mathcal{O}_{B\mathbb{C}^*} * (-)) \otimes \chi_0^j = \mathcal{O}_{B\mathbb{C}^*}(j) * ((-) \otimes \chi_0^j)$. Then the repeated use of Lemma 4.4 implies that

$$\mathcal{C}_{j_\bullet} = F_{j_1} \circ F_{j_2 - j_1} \circ \cdots \circ F_{j_l - j_{l-1}}(\mathbb{W}_b(d-l)).$$

Similarly, for an intermediate step, we have

$$\mathcal{O}_{B\mathbb{C}^*}(j_i) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_l) * (\mathbb{W}_b(d-l) \otimes \chi_0^{j_l}) = F_{j_i} \circ F_{j_{i+1}-j_i} \circ \cdots \circ F_{j_l - j_{l-1}}(\mathbb{W}_b(d-l)).$$

By the repeated use of Lemma 4.7, the above category is a subcategory of $\mathbb{W}_{b+l-i+1+j_l}(d-i+1)$.

4.5 Semiorthogonal decompositions under Grassmannian flips

We show that the subcategories in Proposition 4.11 form a semiorthogonal decomposition. We prepare some lemmas.

Lemma 4.13 *For any $\chi \in \mathbb{B}_b(d)$ and $\chi' \in \mathbb{B}_c(d-1)$ for some $c \geq 0$, we have the vanishing*

$$\mathrm{Hom}_{\mathcal{G}_{a,b}(d)}\big(\mathcal{O}_{B\mathbb{C}^*}(j) * (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}), V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}\big) = 0 \quad \textit{for } j \geq 0. \tag{4-18}$$

Proof Let $\lambda\colon \mathbb{C}^* \to T$ be the one-parameter subgroup given by (4-12). Using the notation of the diagram (4-3), the left-hand side of (4-18) is

$$\begin{aligned}\mathrm{Hom}\big(&p_{\lambda *} q_\lambda^*\big(\mathcal{O}_{B\mathbb{C}^*}(j) \boxtimes (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big), V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}\big) \\ &= \mathrm{Hom}\big(q_\lambda^*\big(\mathcal{O}_{B\mathbb{C}^*}(j) \boxtimes (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big), p_\lambda^!\big(V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}\big)\big).\end{aligned} \tag{4-19}$$

We have the formula

$$p_\lambda^!(-) = (-) \otimes (\det V^{\lambda>0})^{d-b-1} \otimes (\det V^{\lambda=0})^{-1}[d-b-1]$$

(cf [Donovan and Segal 2014, Section A.1; Ballard et al. 2021, (5.8)]). Since $\chi \in \mathbb{B}_b(d)$ and it is a highest weight of $V(\chi)$, any T–weight $\chi'' = (x_1'', \dots, x_d'')$ of $V(\chi)$ satisfies $x_i'' \leq b-d$. Therefore any T–weight of $V(\chi) \otimes (\det V^{\lambda>0})^{d-b-1} \otimes (\det V^{\lambda=0})^{-1}$ pairs negatively with λ. On the other hand, a pairing of λ with any T–weight of the $\mathrm{GL}(V)^{\lambda=0}$–representation $(\det V^{\lambda>0})^j \boxtimes V(\chi')$ is $j \geq 0$. Therefore we have the vanishing of (4-19) by Lemma 2.1. □

Lemma 4.14 *For $\chi, \chi' \in \mathbb{B}_c(d-1)$ for some $c \geq 0$, we have the vanishing*

$$\mathrm{Hom}_{\mathcal{G}_{a,b}(d)}(\mathcal{O}_{B\mathbb{C}^*}(j) * (V(\chi) \otimes \chi_0^j \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}), \mathcal{O}_{B\mathbb{C}^*}(j') * (V(\chi) \otimes \chi_0^{j'} \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})) = 0 \tag{4-20}$$

for $j > j'$.

Proof By Lemma 4.4, we may assume that $j' = 0$. We use the notation in the proof of Lemma 4.13. Using Lemma 4.4 and the adjunction, the left-hand side of (4-20) is

$$\begin{aligned}&\mathrm{Hom}\big(\big(p_{\lambda *} q_\lambda^*(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}))\big) \otimes \chi_0^j, p_{\lambda *} q_\lambda^*\big(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big)\big) \\ &\quad \cong \mathrm{Hom}\big(p_\lambda^*\big(\big(p_{\lambda *} q_\lambda^*(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}))\big) \otimes \chi_0^j\big), q_\lambda^*\big(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big)\big).\end{aligned}$$

By Proposition 4.6, the object

$$p_{\lambda *}q_{\lambda}^{*}(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})) \in D^b(\mathcal{G}_{a,b}(d))$$

is resolved by vector bundles of the form $V(\chi'') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}$, where χ'' is either $\chi'' = \chi$, or $\chi'' \in \mathbb{B}_{c+1}(d)$ whose corresponding Young diagram has a full column. In the latter case, any T–weight of $V(\chi'')$ pairs positively with λ. Therefore in both cases, any T–weight of $V(\chi'') \otimes \chi_0^j$ for $j > 0$ pairs positively with λ. On the other hand any λ–weight of $\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})$ is zero, so the desired vanishing (4-20) follows from Lemma 2.1. □

Lemma 4.15 *In the situation of Lemma 4.14, for $j \in \mathbb{Z}$ we have the isomorphism*

$$\begin{aligned} &\mathrm{Hom}_{\mathcal{G}_{a,b}(d-1)}(V(\chi)\otimes\chi_0^j\otimes\mathcal{O}_{\mathcal{G}_{a,b}(d-1)}, V(\chi')\otimes\chi_0^j\otimes\mathcal{O}_{\mathcal{G}_{a,b}(d-1)}) \\ &\xrightarrow{\cong} \mathrm{Hom}_{\mathcal{G}_{a,b}(d)}\big(\mathcal{O}_{B\mathbb{C}^*}(j)*(V(\chi)\otimes\chi_0^j\otimes\mathcal{O}_{\mathcal{G}_{a,b}(d-1)}), \mathcal{O}_{B\mathbb{C}^*}(j)*(V(\chi')\otimes\chi_0^j\otimes\mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big). \end{aligned} \tag{4-21}$$

Proof By Lemma 4.4, we may assume that $j = 0$. Let χ'' be a weight which appeared in the proof of Lemma 4.14. Note that we observed that any T–weight of $V(\chi'')$ pairs positively with λ except $\chi'' = \chi$. Therefore by Lemma 2.1(i), the right-hand side of (4-21) is isomorphic to

$$\mathrm{Hom}\big(p_{\lambda}^{*}(V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d)}), q_{\lambda}^{*}\big(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big)\big). \tag{4-22}$$

Since $\mathcal{G}_{a,b}(d)^{\lambda\geq 0}$ parametrizes exact sequences (4-13), the object $p_{\lambda}^{*}(V(\chi)\otimes\mathcal{O}_{\mathcal{G}_{a,b}(d)})$ admits a filtration whose associated graded is of the form $q_{\lambda}^{*}(\mathcal{O}_{B\mathbb{C}^*}(j)\boxtimes(V(\chi''')\otimes\mathcal{O}_{\mathcal{G}_{a,b}(d-1)}))$ for $j \geq 0$ and $\chi''' \in \mathbb{B}_c(d-1)$, and $j = 0$ if and only if $\chi''' = \chi$. Therefore by Lemma 2.1(i)–(ii), the above (4-22) is isomorphic to

$$\begin{aligned} &\mathrm{Hom}\big(q_{\lambda}^{*}\big(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big), q_{\lambda}^{*}\big(\mathcal{O}_{B\mathbb{C}^*} \boxtimes (V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)})\big)\big) \\ &\qquad \cong \mathrm{Hom}_{\mathcal{G}_{a,b}(d-1)}(V(\chi) \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}, V(\chi') \otimes \mathcal{O}_{\mathcal{G}_{a,b}(d-1)}). \end{aligned} \tag{4-23}$$

□

In order to state the order of semiorthogonal decompositions, we take a lexicographical order on $\mathbb{Z}^d$, ie for $m_\bullet = (m_1, \dots, m_d) \in \mathbb{Z}^d$ and $m'_\bullet = (m'_1, \dots, m'_d) \in \mathbb{Z}^d$, we write $m_\bullet \succ m'_\bullet$ if $m_i = m'_i$ for $1 \leq i \leq k$ for some $k \geq 0$ and $m_{k+1} > m'_{k+1}$.

Definition 4.16 For $j_\bullet = (j_1, j_2, \dots, j_l)$ and $j'_\bullet = (j'_1, j'_2 \dots, j'_{l'})$ with $l, l' \leq d$, we define $j_\bullet \succ j'_\bullet$ if we have $\widetilde{j}_\bullet \succ \widetilde{j}'_\bullet$, where $\widetilde{j}_\bullet$ is defined by

$$\widetilde{j}_\bullet = (j_1, j_2, \dots, j_l, -1, \dots, -1) \in \mathbb{Z}^d. \tag{4-24}$$

The next proposition shows the semiorthogonality of subcategories (4-17) with respect to the above order.

Proposition 4.17 *For*

$$\begin{aligned} j_\bullet &= (j_1, j_2, \dots, j_l) \quad \text{with } 0 \leq l \leq d \text{ and } 0 \leq j_1 \leq j_2 \leq \dots \leq j_l, \\ j'_\bullet &= (j'_1, j'_2, \dots, j'_{l'}) \quad \text{with } 0 \leq l' \leq d \text{ and } 0 \leq j'_1 \leq j'_2 \leq \dots \leq j'_{l'}, \end{aligned}$$

suppose that $j_\bullet \succ j'_\bullet$. Then we have $\mathrm{Hom}(\mathcal{C}_{j_\bullet}, \mathcal{C}_{j'_\bullet}) = 0$. Here $\mathcal{C}_\bullet$ is defined in (4-17).

Proof Let us take $P \in \mathbb{W}_b(d-l)$ and $P' \in \mathbb{W}_b(d-l')$. We need to show the vanishing of

$$\mathrm{Hom}\big(\mathcal{O}_{B\mathbb{C}^*}(j_1) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_l) * (P \otimes \chi_0^{j_l}), \mathcal{O}_{B\mathbb{C}^*}(j_1) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_{l'}) * (P' \otimes \chi_0^{j'_l})\big). \tag{4-25}$$

We note that, by Remark 4.12, for each $i \le l, l'$ the objects

$$\mathcal{O}_{B\mathbb{C}^*}(j_{i+1}) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_l) * (P \otimes \chi_0^{j_l}) \quad \text{and} \quad \mathcal{O}_{B\mathbb{C}^*}(j_{i+1}) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_{l'}) * (P' \otimes \chi_0^{j'_l}) \tag{4-26}$$

are objects in $\mathbb{W}_{c'}(d-i)$ for some $c' \ge 0$.

From $j_\bullet \succ j'_\bullet$, we have two cases:

(i) $l > l'$ and $j_i = j'_i$ for $1 \le i \le l'$.

(ii) There is $1 \le m < l, l'$ such that $j_i = j'_i$ for $1 \le i \le m$ and $j_{m+1} > j'_{m+1}$.

In the first case, we have

$$\begin{aligned}(4\text{-}25) &= \mathrm{Hom}\big(\mathcal{O}_{B\mathbb{C}^*}(j_1) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_{l'}) * (\mathcal{O}_{B\mathbb{C}^*}(j_{l'+1}) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_l) * (P \otimes \chi_0^{j_l})), \\ &\qquad\qquad \mathcal{O}_{B\mathbb{C}^*}(j_1) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_{l'}) * (P' \otimes \chi_0^{j_{l'}})\big) \\ &\cong \mathrm{Hom}(\mathcal{O}_{B\mathbb{C}^*}(j_{l'+1}) * \cdots * \mathcal{O}_{B\mathbb{C}^*}(j_l) * (P \otimes \chi_0^{j_l}), P' \otimes \chi_0^{j_{l'}}) \cong 0.\end{aligned}$$

Here the first isomorphism follows from the repeated use of Lemma 4.15, noting that (4-26) are objects in $\mathbb{W}_{c'}(d-i)$; and the second isomorphism follows from Lemma 4.13. In the second case, a similar argument as above shows that

$$\begin{aligned}(4\text{-}25) &= \mathrm{Hom}\big(\mathcal{O}_{B\mathbb{C}^*}(j_1)*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_m)*(\mathcal{O}_{B\mathbb{C}^*}(j_{m+1})*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_l)*(P\otimes\chi_0^{j_l})), \\ &\qquad \mathcal{O}_{B\mathbb{C}^*}(j_1)*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_m)*(\mathcal{O}_{B\mathbb{C}^*}(j'_{m+1})*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j'_{l'})*(P'\otimes\chi_0^{j'_{l'}}))\big) \\ &\cong \mathrm{Hom}\big(\mathcal{O}_{B\mathbb{C}^*}(j_{m+1})*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j_l)*(P\otimes\chi_0^{j_l}), \mathcal{O}_{B\mathbb{C}^*}(j'_{m+1})*\cdots*\mathcal{O}_{B\mathbb{C}^*}(j'_{l'})*(P'\otimes\chi_0^{j'_{l'}})\big) \cong 0.\end{aligned}$$

Here the first isomorphism follows from the repeated use of Lemma 4.15, and the second isomorphism follows from Lemma 4.14. □

The following is the main result in this section, which gives a refinement of a semiorthogonal decomposition in [Ballard et al. 2021, Theorem 5.4.4]:

Theorem 4.18 *For $c \ge b$, there exists a semiorthogonal decomposition*

$$\mathbb{W}_c(d) = \langle \mathcal{C}_{j_\bullet} : 0 \le j \le d, \ j_\bullet = (0 \le j_1 \le \cdots \le j_l \le c-b-l) \rangle,$$

where $\mathrm{Hom}(\mathcal{C}_{j_\bullet}, \mathcal{C}_{j'_\bullet}) = 0$ for $j_\bullet \succ j'_\bullet$, and for each $j_\bullet$ we have an equivalence $\mathbb{W}_b(d-l) \xrightarrow{\sim} \mathcal{C}_{j_\bullet}$.

Proof The generation of $\mathbb{W}_c(d)$ by $\mathcal{C}_{j_\bullet}$ is proved in Proposition 4.11, and the semiorthogonality is proved in Proposition 4.17. The equivalence $\mathbb{W}_b(d-l) \xrightarrow{\sim} \mathcal{C}_{j_\bullet}$ follows from repeated use of Lemma 4.15. □

By applying the above theorem to $c = a$ and using equation (4-9), we obtain the following corollary, which relates derived categories under Grassmannian flips.

Corollary 4.19 *There exists a semiorthogonal decomposition*

$$D^b(G^+_{a,b}(d)) = \langle D^b(G^-_{a,b}(d-l))_{j_1,\dots,j_l} : 0 \le l \le d, 0 \le j_1 \le \cdots \le j_l \le a-b-l \rangle.$$

Here, $D^b(G^-_{a,b}(d-l))_{j_1,\dots,j_l}$ is a copy of $D^b(G^-_{a,b}(d-l))$.

Remark 4.20 When $b=0$, from Remark 4.2 the semiorthogonal decomposition in Corollary 4.19 is

$$D^b(G^+_{a,0}(d)) = \langle D^b(\operatorname{Spec}\mathbb{C})_{j_1,\dots,j_d} : 0 \le j_1 \le \cdots \le j_d \le a-d \rangle.$$

Each factor $D^b(\operatorname{Spec}\mathbb{C})_{j_1,\dots,j_d}$ is generated by a vector bundle, which forms Kapranov's exceptional collection [Kapranov 1984] of the Grassmannian $G^+_{a,0}(d)$.

Remark 4.21 When $d=1$, the birational map $G^+_{a,b}(1) \dashrightarrow G^-_{a,b}(1)$ is a standard toric flip. In this case, the semiorthogonal decomposition in Corollary 4.19 is

$$D^b(G^+_{a,b}(1)) = \langle D^b(G^-_{a,b}(1)), D^b(\mathrm{pt})_{(0)}, \dots, D^b(\mathrm{pt})_{(a-b-1)} \rangle.$$

The above semiorthogonal decomposition is a (mutation of) a well-known semiorthogonal decomposition for a standard flip; see [Kawamata 2018, Example 8.8(2)].

Remark 4.22 For a fixed (a,b,l), the set of sequences of integers $(j_1,\dots,j_l)$ satisfying

$$0 \le j_1 \le \cdots \le j_l \le a-b-l$$

consists of $\binom{a-b}{l}$ elements. Therefore Corollary 4.19 implies (1-9). The same applies to Corollary 5.18 and Corollary 5.24 below, so that they imply (1-5) and (1-2), respectively.

4.6 Applications to categories of factorizations

We will use the following variant of Corollary 4.19. Let Z be a smooth scheme with a closed point $z \in Z$. Let us take the formal completion of $G^0_{a,b}(d) \times Z$, where $G^0_{a,b}(d)$ is the good moduli space for $\mathcal{G}_{a,b}(d)$,

$$\widehat{G}^0_{a,b}(d)_Z := \operatorname{Spec} \widehat{\mathcal{O}}_{G^0_{a,b}(d)\times Z,(0,z)}.$$

We also take a regular function w on it,

$$w \colon \widehat{G}^0_{a,b}(d)_Z \to \mathbb{A}^1, \quad w(0,z)=0.$$

By taking the product of the diagram (4-5) with Z and pulling it back via $\widehat{G}^0_{a,b}(d)_Z \to G^0_{a,b}(d) \times Z$, we obtain the diagram

$$(4\text{-}27)\qquad \begin{tikzcd} \widehat{G}^+_{a,b}(d)_Z \arrow[r, hook] \arrow[dr] \arrow[ddr, bend right, "w"'] & \widehat{\mathcal{G}}_{a,b}(d)_Z \arrow[d] & \widehat{G}^-_{a,b}(d)_Z \arrow[l, hook'] \arrow[dl] \arrow[ddl, bend left, "w"] \\ & \widehat{G}^0_{a,b}(d)_Z \arrow[d, "w"] & \\ & \mathbb{A}^1 & \end{tikzcd}$$

Similarly to (4-11), we have the categorified Hall product for formal fibers (see Section 3.4)

$$*\colon \mathrm{MF}(B\mathbb{C}^*,0)\boxtimes \mathrm{MF}(\widehat{\mathscr{G}}_{a,b}(d-1)_Z,w)\to \mathrm{MF}(\widehat{\mathscr{G}}_{a,b}(d)_Z,w).$$

The subcategory

$$\widehat{\mathbb{W}}_c(d)\subset \mathrm{MF}(\widehat{\mathscr{G}}_{a,b}(d)_Z,w)$$

is also defined, similarly to (4-8), to be the smallest thick triangulated subcategory which contains factorizations with entries $V(\chi)\otimes\mathscr{O}$ for $\chi\in\mathbb{B}_c(d)$. Note that we have the decomposition (2-2)

$$\mathrm{MF}(B\mathbb{C}^*,0)=\bigoplus_{j\in\mathbb{Z}}\mathrm{MF}(\operatorname{Spec}\mathbb{C},0)_j$$

such that $\mathrm{MF}(\operatorname{Spec}\mathbb{C},0)_j$ is equivalent to $\mathrm{MF}(\operatorname{Spec}\mathbb{C},0)$. We then define

$$\widehat{\mathscr{C}}_{j_\bullet}:=\mathrm{MF}(\operatorname{Spec}\mathbb{C},0)_{j_1}*\cdots*\mathrm{MF}(\operatorname{Spec}\mathbb{C},0)_{j_l}*(\widehat{\mathbb{W}}_b(d-l)\otimes\chi_0^{j_l})\subset\mathrm{MF}(\widehat{\mathscr{G}}_{a,b}(d)_Z,w)) \tag{4-28}$$

for $0\le l\le d$ and $0\le j_1\le\cdots\le j_l\le c-b-l$. We have the following variant of Theorem 4.18.

Corollary 4.23 *For $c\ge b$, there exists a semiorthogonal decomposition*

$$\widehat{\mathbb{W}}_c(d)=\langle\widehat{\mathscr{C}}_{j_\bullet}:0\le l\le d,\ j_\bullet=(0\le j_1\le\cdots\le j_l\le c-b-l)\rangle,$$

where $\mathrm{Hom}(\widehat{\mathscr{C}}_{j_\bullet},\widehat{\mathscr{C}}_{j'_\bullet})=0$ for $j_\bullet\succ j'_\bullet$, and for each $j_\bullet$ we have an equivalence $\widehat{\mathbb{W}}_b(d-l)\xrightarrow{\sim}\widehat{\mathscr{C}}_{j_\bullet}$.

Proof The argument of the proof of Theorem 4.18 implies an analogous semiorthogonal decomposition for $D^b(\widehat{\mathscr{G}}_{a,b}(d)_Z)$. Then it is well-known that the above semiorthogonal decomposition induces the one for categories of factorizations; cf [Halpern-Leistner and Pomerleano 2020, Lemmas 1.17 and 1.18; Orlov 2006, Proposition 1.10; Pădurariu 2019, Proposition 2.7; 2023, Proposition 2.1]. □

5 Categorical Donaldson–Thomas theory for the resolved conifold

In this section, we use the result in the previous section to prove Theorem 1.2.

5.1 Geometry and algebras for the resolved conifold

Let X be the resolved conifold

$$X:=\mathrm{Tot}_{\mathbb{P}^1}(\mathscr{O}_{\mathbb{P}}(-1)^{\oplus 2}).$$

Here we recall some well-known geometry and algebras for the resolved conifold; see [Van den Bergh 2004; Nagao and Nakajima 2011] for details. There is a birational contraction

$$f\colon X\to Y:=\{xy+zw=0\}\subset\mathbb{C}^4,$$

which contracts the zero section $C=\mathbb{P}^1\subset X$ to the conifold singularity $0\in Y$. Let $\mathscr{E}:=\mathscr{O}_X\oplus\mathscr{O}_X(1)$, and $A:=\mathrm{End}(\mathscr{E})$. Then there is an equivalence by Van den Bergh [2004],

$$\Phi:=\mathbf{R}\mathrm{Hom}(\mathscr{E},-)\colon D^b(X)\xrightarrow{\sim}D^b(\operatorname{mod}A). \tag{5-1}$$

Here $\operatorname{mod} A$ is the abelian category finitely generated right A–modules. The noncommutative algebra A is isomorphic to the path algebra associated with a quiver with superpotential (Q, W) given by

$$Q = \begin{tikzcd} \bullet_0 \arrow[r, bend left=60, "a_1"] \arrow[r, bend left=20, "a_2"'] & \bullet_1 \arrow[l, bend left=20, "b_1"'] \arrow[l, bend left=60, "b_2"] \end{tikzcd} \qquad \text{and} \qquad W = a_1 b_1 a_2 b_2 - a_1 b_2 a_2 b_1.$$

The equivalence (5-1) restricts to the equivalences of abelian subcategories

$$\Phi \colon \operatorname{Per}(X/Y) \xrightarrow{\sim} \operatorname{mod} A \quad \text{and} \quad \Phi \colon \operatorname{Per}_c(X/Y) \xrightarrow{\sim} \operatorname{mod}_{\mathrm{fd}}(A).$$

Here $\operatorname{Per}(X/Y)$ is the abelian category of Bridgeland's perverse coherent sheaves [2002], given by

$$\operatorname{Per}(X/Y) = \\ \{E \in D^b(X) : \mathcal{H}^i(E) = 0 \text{ for } i \neq -1, 0,\ R^1 f_* \mathcal{H}^0(E) = f_* \mathcal{H}^{-1}(E) = 0, \operatorname{Hom}(\mathcal{H}^0(E), \mathcal{O}_C(-1)) = 0\}.$$

The subcategory $\operatorname{Per}_c(X/Y) \subset \operatorname{Per}(X/Y)$ consists of compactly supported objects, and $\operatorname{mod}_{\mathrm{fd}}(A) \subset \operatorname{mod}(A)$ consists of finite-dimensional A–modules. The simple (Q, W)–representations corresponding to the vertex $\{0, 1\}$ are given by

$$\{\mathcal{O}_C, \mathcal{O}_C(-1)[1]\} \subset \operatorname{Per}_c(X/Y).$$

An object $F \in \operatorname{Per}_c(X/Y)$ is supported on C or a zero-dimensional subscheme in X. For $F \in \operatorname{Per}_c(X/Y)$, we set

$$\operatorname{cl}(F) := (\beta, n) \in \mathbb{Z}^{\oplus 2}, \quad \text{with } [F] = \beta[C],\ \chi(F) = n,$$

where $[F]$ is the fundamental one-cycle of F. Under the equivalence Φ, an object $F \in \operatorname{Per}_c(X/Y)$ with $\operatorname{cl}(F) = (\beta, n)$ corresponds to a (Q, W)–representation with dimension vector $(n, n - \beta)$.

Following [Nagao and Nakajima 2011, Section 1], a *perverse coherent system* is defined to be a pair

$$(F, s), \quad \text{with } F \in \operatorname{Per}_c(X/Y),\ s \colon \mathcal{O}_X \to F. \tag{5-2}$$

Let $(Q^\dagger, W)$ be a quiver with superpotential, given by

$$Q^\dagger = \begin{tikzcd} \bullet_\infty \arrow[d] & \\ \bullet_0 \arrow[r, bend left=60, "a_1"] \arrow[r, bend left=20, "a_2"'] & \bullet_1 \arrow[l, bend left=20, "b_1"'] \arrow[l, bend left=60, "b_2"] \end{tikzcd} \qquad \text{and} \qquad W = a_1 b_1 a_2 b_2 - a_1 b_2 a_2 b_1.$$

Note that $Q^\dagger$ is an extended quiver obtained from Q as in Section 3.1. By the equivalence (5-1), giving a perverse coherent system with $\operatorname{cl}(F) = (\beta, n)$ is equivalent to giving a representation of $(Q^\dagger, W)$ with dimension vector $(v_\infty, v_0, v_1) = (1, n, n - \beta)$.

5.2 Categorical DT invariants for the resolved conifold

For a dimension vector $v=(v_0,v_1)$ of Q, let V_0 and V_1 be vector spaces with dimensions v_0 and v_1, respectively. The $\mathbb{C}^*$–rigidified moduli stack of $Q^\dagger$–representations of dimension vector $(1,v)$ in Section 3.1 is explicitly written as

$$\mathcal{M}_Q^\dagger(v)=[R_{Q^\dagger}(v)/G(v)]=[V_0\oplus \mathrm{Hom}(V_0,V_1)^{\oplus 2}\oplus \mathrm{Hom}(V_1,V_0)^{\oplus 2}/\mathrm{GL}(V_0)\times \mathrm{GL}(V_1)].$$

Let w be the function

$$w=\mathrm{Tr}(W)\colon \mathcal{M}_Q^\dagger(v)\to \mathbb{A}^1,\quad w(v,A_1,A_2,B_1,B_2)=\mathrm{Tr}(A_1B_1A_2B_2-A_1B_2A_2B_1). \tag{5-3}$$

Then its critical locus

$$\mathcal{M}_{(Q,W)}^\dagger(v):=\mathrm{Crit}(w)\subset w^{-1}(0)\subset \mathcal{M}_Q^\dagger(v) \tag{5-4}$$

is the $\mathbb{C}^*$–rigidified moduli stack of $(Q^\dagger,W)$–representations of dimension vector $(1,v)$. Here the first inclusion follows from the fact that w is a homogeneous function on $R_{Q^\dagger}(v)$ of degree four. By the equivalence (5-1), $\mathcal{M}_{(Q,W)}^\dagger(v)$ is isomorphic to the moduli stack of perverse coherent systems (5-2) satisfying $\mathrm{cl}(F)=(v_0-v_1,v_0)$.

For $\theta=(\theta_0,\theta_1)\in\mathbb{R}^2$, we denote by

$$\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)=[R_{Q^\dagger}^{\theta\text{-ss}}(v)/G(v)]\subset \mathcal{M}_Q^\dagger(v)$$

the open substack of θ–semistable $Q^\dagger$–representations. We also have the open substack

$$\mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-ss}}(v):=\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)\cap \mathcal{M}_{(Q,W)}^\dagger(v)\subset \mathcal{M}_{(Q,W)}^\dagger(v)$$

corresponding to θ–semistable $(Q^\dagger,W)$–representations. If $\theta_i\in\mathbb{Z}$, as mentioned in Section 3.1 these open substacks are GIT semistable locus with respect to the character

$$\chi_\theta\colon G(v)=\mathrm{GL}(V_0)\times \mathrm{GL}(V_1)\to\mathbb{C}^*,\quad (g_0,g_1)\mapsto \det(g_0)^{-\theta_0}\det(g_1)^{-\theta_1}. \tag{5-5}$$

We have the good moduli spaces by taking GIT quotients

$$\pi_Q^\dagger\colon \mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)\to M_Q^{\dagger,\theta\text{-ss}}(v)\quad\text{and}\quad \pi_{(Q,W)}^\dagger\colon \mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-ss}}(v)\to M_{(Q,W)}^{\dagger,\theta\text{-ss}}(v). \tag{5-6}$$

We will consider the triangulated category

$$\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v),w)$$

and call it the *categorical DT invariant* for the conifold quiver $(Q^\dagger,W)$. The above triangulated category (or more precisely its dg-enhancement) recovers the numerical DT invariant considered in [Nagao and Nakajima 2011]:

Lemma 5.1 *For a generic $\theta \in \mathbb{R}^2$, there is an equality*

$$e_{\mathbb{C}(\!(u)\!)}(\mathrm{HP}_*(\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v), w)) = (-1)^{v_1}\, \mathrm{DT}^{\theta}(v).$$

Here $\mathrm{HP}_*(-)$ is the periodic cyclic homology which is a $\mathbb{Z}/2$–graded $\mathbb{C}(\!(u)\!)$–vector space [Keller 1999], $e_{\mathbb{C}(\!(u)\!)}(-)$ is the Euler characteristic of $\mathbb{Z}/2$–graded $\mathbb{C}(\!(u)\!)$–vector space, and $\mathrm{DT}^{\theta}(v) \in \mathbb{Z}$ is the numerical DT invariant counting $(Q^{\dagger}, w)$–representations with dimension vector $(1, v)$.

Proof Since θ is generic and the dimension vector $(1, v)$ of $Q^{\dagger}$ is primitive, the stack $M = \mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v)$ consists of only θ–stable objects and it is a smooth quasiprojective scheme. By [Efimov 2018, Theorem 5.4], there is an isomorphism of $\mathbb{Z}/2$–graded vector spaces over $\mathbb{C}(\!(u)\!)$,

$$\mathrm{HP}_*(\mathrm{MF}(M, w)) \cong H^*(M, \phi_w(\mathbb{Q}_M)) \otimes_{\mathbb{Q}} \mathbb{C}(\!(u)\!) \cong H^{*+\dim M}(M, \phi_w(\mathrm{IC}_M)) \otimes_{\mathbb{Q}} \mathbb{C}(\!(u)\!).$$

Here $\phi_w(-)$ is the vanishing cycle functor and u has degree two, and $\mathrm{IC}_M = \mathbb{Q}_M[\dim M]$. We take the Euler characteristics of both sides as $\mathbb{Z}/2$–graded vector spaces over $\mathbb{C}(\!(u)\!)$. Since we have

$$e(H^*(M, \phi_w(\mathrm{IC}_M))) = \int_M \chi_B \, de =: \mathrm{DT}^{\theta}(v),$$

where χ_B is the Behrend function [2009] on M, it is enough to show that $(-1)^{v_1} = (-1)^{\dim M}$. Let E be a $Q^{\dagger}$–representation with dimension vector $(1, v_0, v_1)$. Then we have

$$\dim M = 1 + \dim \mathrm{Ext}^1_{Q^{\dagger}}(E, E) - \dim \mathrm{Hom}_{Q^{\dagger}}(E, E) = v_0 - v_0^2 - v_1^2 + 4v_0v_1.$$

Here we have used Lemma 5.3 below for the second identity. Therefore $(-1)^{v_1} = (-1)^{\dim M}$ holds. □

Remark 5.2 If θ lies in a DT chamber in Figure 1, the invariant $\mathrm{DT}^{\theta}(v)$ reduces to the DT invariant counting ideal sheaves of compactly supported closed subschemes $Z \hookrightarrow X$ satisfying $\mathrm{cl}(\mathcal{O}_Z) = (v_0 - v_1, v_0)$, considered in [Maulik et al. 2006].

The following lemma follows immediately from the Euler pairing computations of quiver representations [Brion 2012, Corollary 1.4.3].

Lemma 5.3 *For $Q^{\dagger}$–representations E, E' with dimension vector (v_{∞}, v_0, v_1), $(v'_{\infty}, v'_0, v'_1)$, we have*

$$\dim \mathrm{Hom}_{Q^{\dagger}}(E, E') - \dim \mathrm{Ext}^1_{Q^{\dagger}}(E, E') = v_{\infty}v'_{\infty} - v_{\infty}v'_0 + v_0v'_0 - 2v_0v'_1 - 2v_1v'_0 + v_1v'_1.$$

We have the unstable locus

$$\mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-us}}(v) := \mathcal{M}_{(Q,W)}^{\dagger}(v) \setminus \mathcal{M}_{(Q,W)}^{\dagger,\theta\text{–ss}}(v).$$

Then we have the open immersion

$$\mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v) \subset \mathcal{M}_Q^{\dagger}(v) \setminus \mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-us}}(v).$$

The next lemma shows that the categorical DT invariant can be also defined on a bigger ambient space.

Lemma 5.4 *The restriction functor*

$$\mathrm{MF}(\mathcal{M}_Q^{\dagger}(v) \setminus \mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-us}}(v), w) \xrightarrow{\sim} \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w)$$

is an equivalence.

Proof The lemma follows since the category of factorizations only depends on an open neighborhood of the critical locus; see the equivalence (2-1). □

5.3 Wall-chamber structure

There is a wall-chamber structure for the θ–stability as in Figure 1 in the introduction, taken from [Nagao and Nakajima 2011, Figure 1].

In Figure 1, if θ lies in the first quadrant then $\mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-ss}}(v) = \varnothing$ unless $v = 0$, so it is called an *empty chamber*. In this case, the categorical DT invariants are given in the following lemma.

Lemma 5.5 *Let $\theta_{\mathrm{en}} \in \mathbb{R}^2$ lie in an empty chamber. Then*

$$\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_{\mathrm{en}}\text{-ss}}(v), w) = \begin{cases} \mathrm{MF}(\operatorname{Spec}\mathbb{C}, 0) & \textit{if } v = 0, \\ 0 & \textit{if } v \neq 0. \end{cases}$$

Proof If $v \neq 0$, then $\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_{\mathrm{en}}\text{-ss}}(v), w) = 0$ by the equivalence (2-1), since the critical locus of w is empty. If $v = 0$, then $\mathcal{M}_Q^{\dagger,\theta_{\mathrm{en}}\text{-ss}}(v) = \operatorname{Spec}\mathbb{C}$ and $w = 0$. □

We focus on the walls in the second quadrant, classified by $m \in \mathbb{Z}_{\geq 1}$:

$$W_m := \mathbb{R}_{>0} \cdot (1-m, m) \subset \mathbb{R}^2.$$

If θ lies between W_m and W_{m+1}, then the moduli stack $\mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-ss}}(v)$ is constant, consisting of θ–stable objects. So $\mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-ss}}(v)$ is a quasiprojective scheme, and the good moduli space morphism $\pi_{(Q,W)}^{\dagger}$ in (5-6) is an isomorphism. If θ is also sufficiently close to the wall W_m, then $\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)$ also consists of θ–stable objects and the morphism $\pi_Q^{\dagger}$ in (5-6) is an isomorphism. The categorical DT invariant is also constant when θ deforms inside a chamber:

Lemma 5.6 *The triangulated category $\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w)$ is constant (up to equivalence) when θ deforms inside a chamber in Figure 1.*

Proof Suppose that θ lies in a chamber in Figure 1. Although θ does not lie in a wall for $(Q^{\dagger}, W)$–representations, it may lie on a wall for $Q^{\dagger}$–representations. However, the destabilizing locus in $\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)$ is disjoint from $\mathrm{Crit}(w) = \mathcal{M}_{(Q,W)}^{\dagger,\theta\text{-ss}}(v)$, so by (2-1) the triangulated categories $\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w)$ are equivalent under wall-crossing inside a chamber of Figure 1. □

For $\theta \in W_m$, there is a unique (up to isomorphism) θ–stable (Q, W)–representation S_m — that is, $(Q^\dagger, W)$–representation whose dimension vector at ∞ is zero — given by

$$S_m := \left(\begin{tikzcd} \mathbb{C}^m \arrow[r, bend left=50, "0"] \arrow[r, shift left, "0"] & \mathbb{C}^{m-1} \arrow[l, shift left, "B_1^0"] \arrow[l, bend left=50, "B_2^0"] \end{tikzcd} \right), \quad B_1^0(f_i) = e_i, \quad B_2^0(f_i) = e_{i+1}. \tag{5-7}$$

See [Nagao and Nakajima 2011, Theorem 3.5]. Here $\{e_1, \dots, e_m\}$, $\{f_1, \dots, f_{m-1}\}$ are bases of $\mathbb{C}^m$ and $\mathbb{C}^{m-1}$, respectively. Note that S_m has dimension vector $s_m = (m, m-1)$ so that $\theta(S_m) = 0$ when $\theta \in W_m$. Under the equivalence Φ in (5-1), we have the relation

$$\Phi(\mathcal{O}_C(m-1)) = S_m. \tag{5-8}$$

See [Nagao and Nakajima 2011, Remark 3.6]. Since $s_m = (m, m-1)$ is primitive, the moduli stack $\mathcal{M}_Q^{\theta\text{–ss}}(s_m)$ consists of θ–stable Q–representations, and the good moduli space morphism

$$\mathcal{M}_Q^{\theta\text{–ss}}(s_m) \to M_Q^{\theta\text{–ss}}(s_m) \tag{5-9}$$

is a $\mathbb{C}^*$–gerbe. There is a function defined similarly to (5-3),

$$w = \operatorname{Tr}(W) \colon \mathcal{M}_Q^{\theta\text{–ss}}(s_m) \to \mathbb{A}^1,$$

whose critical locus $\mathcal{M}_{(Q,W)}^{\theta\text{–ss}}(s_m)$ is the moduli stack of θ–stable (Q, W)–representation. Note that $\mathcal{M}_{(Q,W)}^{\theta\text{–ss}}(s_m)$ consists of one point, corresponding to the unique θ–stable (Q, W)–representation S_m.

Lemma 5.7 *For any $j \in \mathbb{Z}$, there is an equivalence*

$$\operatorname{MF}(\mathcal{M}_Q^{\theta\text{–ss}}(s_m), w)_j \simeq \operatorname{MF}(\operatorname{Spec} \mathbb{C}, 0).$$

Proof Let $V_0 = \mathbb{C}^m$, $V_1 = \mathbb{C}^{m-1}$ and $B_1^0, B_2^0 \colon V_1 \to V_0$ be maps as in (5-7). Note that we have

$$\mathcal{M}_Q^{\theta\text{–ss}}(s_m) = [(\operatorname{Hom}(V_0, V_1)^{\oplus 2} \oplus \operatorname{Hom}(V_1, V_0)^{\oplus 2})^{\theta\text{–ss}} / \operatorname{GL}(V_0) \times \operatorname{GL}(V_1)].$$

It admits a projection

$$\mathcal{M}_Q^{\theta\text{–ss}}(s_m) \to [\operatorname{Hom}(V_1, V_0)^{\oplus 2} / \operatorname{GL}(V_0) \times \operatorname{GL}(V_1)]. \tag{5-10}$$

The target of the above morphism is identified with the moduli stack of representations of the Kronecker quiver Q_K (ie two vertices $\{0, 1\}$ with two arrows from 1 to 0). We have the Beilinson equivalence

$$\mathbf{R}\operatorname{Hom}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(1), -) \colon D^b(\mathbb{P}^1) \xrightarrow{\sim} D^b(\operatorname{Rep}(Q_K)).$$

Under the above equivalence, $\mathcal{O}_{\mathbb{P}^1}(m-1)$ corresponds to (B_1^0, B_2^0), which is a θ–stable Q_K–representation. Since $\mathcal{O}_{\mathbb{P}^1}(m-1)$ is rigid in $\mathbb{P}^1$ with automorphism $\mathbb{C}^*$, there is a $\operatorname{GL}(V_0) \times \operatorname{GL}(V_1)$–invariant open neighborhood

$$(B_1^0, B_2^0) \in \mathcal{U} \subset (\operatorname{Hom}(V_1, V_0)^{\oplus 2})^{\theta\text{–ss}}$$

such that $[\mathcal{U}/\mathrm{GL}(V_0)\times\mathrm{GL}(V_1)]$ is isomorphic to $B\mathbb{C}^*$, where $\mathbb{C}^*$ is the diagonal torus in $\mathrm{GL}(V_0)\times\mathrm{GL}(V_1)$, ie $t\mapsto(t\cdot\mathrm{id}_{V_0},t\cdot\mathrm{id}_{V_1})$. By pulling it back by the projection (5-10), we see that there is an open immersion

$$\mathrm{Hom}(V_0,V_1)^{\oplus 2}\times B\mathbb{C}^*\subset\mathcal{M}_Q^{\theta\text{–ss}}(s_m),\quad (A_1,A_2)\mapsto(A_1,A_2,B_1^0,B_2^0), \tag{5-11}$$

such that the image of $0\in\mathrm{Hom}(V_0,V_1)^{\oplus 2}$ is $\{S_m\}=\mathrm{Crit}(w)$. By the equivalence (2-1), the restriction functor gives an equivalence

$$\mathrm{MF}(\mathcal{M}_Q^{\theta\text{–ss}}(s_m),w)\xrightarrow{\sim}\mathrm{MF}(\mathrm{Hom}(V_0,V_1)^{\oplus 2}\times B\mathbb{C}^*,w). \tag{5-12}$$

The function w restricted (5-11) is a quadratic function by the definition of w, which must be nondegenerate as its critical locus is one point. Since $\mathrm{Hom}(V_0,V_1)^{\oplus 2}$ is even-dimensional, for a suitable choice of basis of $\mathrm{Hom}(V_0.V_1)^{\oplus 2}$ the function w is written as $y_1z_1+\cdots+y_nz_n$, where n is the dimension of $\mathrm{Hom}(V_0,V_1)$. Therefore the right-hand side of (5-12) is equivalent to $\mathrm{MF}(B\mathbb{C}^*,0)$ by the Knörrer periodicity in Theorem 2.4. □

5.4 Descriptions of formal fibers

By the above classification of θ–stable (Q,W)–representations, a θ–polystable $(Q^\dagger,W)$–representation of dimension vector $(1,v_0,v_1)$ at the wall $\theta\in W_m$ is of the form

$$R=R_\infty\oplus(V\otimes S_m), \tag{5-13}$$

where V is a finite-dimensional vector space and R_∞ is a θ–stable $(Q^\dagger,W)$–representation. By setting $d:=\dim V$, the dimension vector of R_∞ is $(1,v_0-dm,v_1-d(m-1))$. By regarding R as a θ–polystable $Q^\dagger$–representation, it determines a point $p\in M_Q^{\dagger,\theta\text{–ss}}(v)$.

Remark 5.8 The vector space V in (5-13) will play the same role of the vector space V in Section 4. Below we fix a basis of V and use the same convention of the dominant chamber in Section 4.2.

Below, we fix a θ–polystable object (5-13), and $p\in M_Q^{\dagger,\theta\text{–ss}}(v)$ is the corresponding point as above. We will give a description of the formal fiber of the good moduli space morphism $\pi_Q^\dagger:\mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v)\to M_Q^{\dagger,\theta\text{–ss}}(v)$ at p. We set

$$G_p:=\mathrm{Aut}(R)=\mathrm{GL}(V).$$

It acts on $\mathrm{Ext}^1_{Q^\dagger}(R,R)$ by the conjugation, and we have the good moduli space morphism

$$[\mathrm{Ext}^1_{Q^\dagger}(R,R)/G_p]\to\mathrm{Ext}^1_{Q^\dagger}(R,R)/\!\!/G_p. \tag{5-14}$$

Let $q\in R_{Q^\dagger}(v)$ be a point corresponding to the polystable object (5-13). Note that $\mathrm{Ext}^1_{Q^\dagger}(R,R)$ is the tangent space of the stack $\mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v)$ at q. By Luna's étale slice theorem, there exists a G_p–invariant locally closed subset $q \in W_p \subset R_{Q^\dagger}(v)$ and a commutative diagram

$$\begin{tikzcd} \big([\mathrm{Ext}^1_{Q^\dagger}(R,R)/G_p],0\big) \arrow[d] \arrow[dr, phantom, "\square"] & ([W_p/G_p],q) \arrow[l] \arrow[r] \arrow[d] \arrow[dr, phantom, "\square"] & (\mathcal{M}^{\dagger,\theta\text{-ss}}_Q(v),q) \arrow[d] \\ (\mathrm{Ext}^1_{Q^\dagger}(R,R)/\!\!/G_p,0) & (W_p/\!\!/G_p,q) \arrow[l] \arrow[r] & (M^{\dagger,\theta\text{-ss}}_Q(v),p) \end{tikzcd} \tag{5-15}$$

such that each horizontal arrows are étale.

We have the following decomposition of $\mathrm{Ext}^1_{Q^\dagger}(R,R)$ as G_p–representations:

$$\begin{aligned} \mathrm{Ext}^1_{Q^\dagger}(R,R) &= \mathrm{Ext}^1_{Q^\dagger}(R_\infty,R_\infty) \oplus (V\otimes \mathrm{Ext}^1_{Q^\dagger}(R_\infty,S_m)) \\ &\qquad \oplus (V^\vee \otimes \mathrm{Ext}_{Q^\dagger}(S_m,R_\infty)) \oplus (\mathrm{End}(V)\otimes \mathrm{Ext}^1_Q(S_m,S_m)) \\ &= (\mathrm{Ext}^1_{Q^\dagger}(R_\infty,R_\infty)\oplus \mathrm{Ext}^1_Q(S_m,S_m)) \oplus (V\otimes \mathrm{Ext}^1_{Q^\dagger}(R_\infty,S_m)) \\ &\qquad \oplus (V^\vee\otimes \mathrm{Ext}^1_{Q^\dagger}(S_m,R_\infty)) \oplus (\mathrm{End}_0(V)\otimes \mathrm{Ext}^1_Q(S_m,S_m)). \end{aligned} \tag{5-16}$$

Here $\mathrm{End}_0(V)$ is the kernel of the trace map $\mathrm{Tr}\colon \mathrm{End}(V)\to\mathbb{C}$, which is an irreducible G_p–representation. The last identity gives a direct-sum decomposition of $\mathrm{Ext}^1_{Q^\dagger}(R,R)$ into its irreducible G_p–representations whose irreducible factors are $\mathbb{C}$ (the trivial representation), V, $V^\vee$ and $\mathrm{End}_0(V)$. The number of summands is calculated as follows:

Lemma 5.9 *We have the identities*

$$\begin{aligned} a_{v,m,d} &:= \mathrm{ext}^1_{Q^\dagger}(R_\infty,S_m) = C_{v,m}+m+d(-2m^2+2m+1), \\ b_{v,m,d} &:= \mathrm{ext}^1_{Q^\dagger}(S_m,R_\infty) = C_{v,m}+d(-2m^2+2m+1), \\ C_{v,m} &:= (m-2)v_0+(m+1)v_1, \\ c_m &:= \mathrm{ext}^1_{Q^\dagger}(S_m,S_m) = 2m^2-2m. \end{aligned} \tag{5-17}$$

Proof The lemma easily follows from Lemma 5.3, noting that

$$\mathrm{Hom}(R_\infty,S_m)=\mathrm{Hom}(S_m,R_\infty)=0 \quad\text{and}\quad \mathrm{Hom}(T_m,S_m)=\mathbb{C}.$$

For example, since the dimension vectors of R_∞ and S_m are $(1,v_0-md,v_1-(m-1)d)$ and $(0,m,m-1)$, respectively, we have

$$\begin{aligned} -a_{v,m,d} &= \hom(R_\infty,S_m)-\mathrm{ext}^1(R_\infty,S_m) \\ &= -m+(v_0-md)m-2(v_0-md)(m-1)-2(v_1-(m-1)d)m+(m-1)(v_1-(m-1)d) \\ &= -(m-2)v_0-(m+1)v_1-m-d(-2m^2+2m+1). \qquad \square \end{aligned}$$

The left vertical arrow in (5-15) is also identified with a moduli stack of some quiver representations and its good moduli space. We define Q_p to be the Ext quiver for $\{S_m\}$ and $Q_p^\dagger$ to be the Ext quiver for $\{R_\infty, S_m\}$. Namely Q_p is the quiver with one vertex $\{1\}$ and the number of loops at 1 is c_m. The quiver $Q_p^\dagger$ consists of two vertices $\{\infty, 1\}$, the number of arrows from ∞ to 1 is $a_{v,m,d}$, from 1 to ∞ is $b_{v,m,d}$, and the number of loops at ∞ (resp. 1) is $\mathrm{ext}^1_{Q^\dagger}(R_\infty, R_\infty)$ (resp. c_m). From (5-16), we have the identification

$$\begin{tikzcd} \mathcal{M}^\dagger_{Q_p}(d) \arrow[r, equal] \arrow[d] & {[\mathrm{Ext}^1_{Q^\dagger}(R,R)/G_p]} \arrow[d] \\ M^\dagger_{Q_p}(d) \arrow[r, equal] & \mathrm{Ext}^1_{Q^\dagger}(R,R)/\!\!/G_p \end{tikzcd} \tag{5-18}$$

By combining the diagrams (5-15), (5-18) and taking the formal fibers, we have a commutative diagram

$$\begin{tikzcd} \mathcal{M}^\dagger_{Q_p}(d) \arrow[d] \arrow[dr, phantom, "\Box"] & \widehat{\mathcal{M}}^\dagger_{Q_p}(d) \arrow[l] \arrow[r, equal] \arrow[d] \arrow[rr, bend left, "\eta_p"] & \Big[\widehat{\mathrm{Ext}}^1_{Q^\dagger}(R,R)/G_p\Big] \arrow[r, "\cong"] \arrow[d] & \widehat{\mathcal{M}}^{\dagger,\theta\text{-ss}}_{Q}(v)_p \arrow[r] \arrow[d] \arrow[dr, phantom, "\Box"] & \mathcal{M}^{\dagger,\theta\text{-ss}}_{Q}(v) \arrow[d] \\ M^\dagger_{Q_p}(d) & \widehat{M}^\dagger_{Q_p}(d) \arrow[l] \arrow[r, equal] & \widehat{\mathrm{Ext}}^1_{Q^\dagger}(R,R)/\!\!/G_p \arrow[r, "\cong"] & \widehat{M}^{\dagger,\theta\text{-ss}}_{Q}(v)_p \arrow[r] & M^{\dagger,\theta\text{-ss}}_{Q}(v) \end{tikzcd} \tag{5-19}$$

Here each vertical arrow is a good moduli space morphism, the vertical arrow second from the right (resp. left) is the formal fiber of the right (resp. left) one at p (resp. origin), the middle vertical arrow is the formal fiber of the morphism (5-14) at the origin. The square second from the right is obtained by the formal completions of good moduli spaces in the diagram (5-15), where the horizontal arrows are isomorphisms since the horizontal arrows in the diagram (5-15) are étale.

We then compare the semistable loci under the isomorphism η_p in the diagram (5-19). We take $\theta = (\theta_0, \theta_1) \in W_m$ and $\theta_\pm$ of the form

$$\theta_\pm = (\theta_0 \mp \varepsilon, \theta_1 \pm \varepsilon), \quad \text{with } \varepsilon > 0. \tag{5-20}$$

We take (θ_0, θ_1) and ε to be integers, and $\theta_\pm$ to lie on chambers adjacent to W_m which are sufficiently close to W_m, eg take $\varepsilon = 1$ and $(\theta_0, \theta_1) = N \cdot (1-m, m)$ for a sufficiently large integer N. We have the open substacks

$$\mathcal{M}^{\dagger,\theta_\pm\text{-ss}}_{Q}(v) \subset \mathcal{M}^{\dagger,\theta\text{-ss}}_{Q}(v) \quad \text{and} \quad \widehat{\mathcal{M}}^{\dagger,\theta_\pm\text{-ss}}_{Q}(v)_p \subset \widehat{\mathcal{M}}^{\dagger,\theta\text{-ss}}_{Q}(v)_p$$

corresponding to $\theta_\pm$–semistable representations.

On the other hand, as in (4-4) we set $\chi_0 \colon \mathrm{GL}(V) \to \mathbb{C}^*$ to be the determinant character $g \mapsto \det(g)$. We have the open substacks

$$\mathcal{M}^{\dagger,\chi_0^{\pm 1}\text{-ss}}_{Q_p}(d) \subset \mathcal{M}^\dagger_{Q_p}(d) \quad \text{and} \quad \widehat{\mathcal{M}}^{\dagger,\chi_0^{\pm 1}\text{-ss}}_{Q_p}(d) \subset \widehat{\mathcal{M}}^\dagger_{Q_p}(d)$$

corresponding to $\chi_0^{\pm 1}$–semistable $Q_p^\dagger$–representations. We have the following lemma.

Lemma 5.10 *The isomorphism η_p in (5-19) restricts to the isomorphisms*

$$\eta_p \colon \widehat{\mathcal{M}}^{\dagger,\chi_0^{\pm 1}\text{-ss}}_{Q_p}(d) \xrightarrow{\cong} \widehat{\mathcal{M}}^{\dagger,\theta_\pm\text{-ss}}_{Q}(v)_p. \tag{5-21}$$

Proof Let us consider the composition

$$G_p = \mathrm{GL}(V) \hookrightarrow \mathrm{GL}(V_0) \times \mathrm{GL}(V_1) \xrightarrow{\chi_{\theta_\pm}} \mathbb{C}^*. \tag{5-22}$$

We see that the above composition is given by $g \mapsto \det(g)^{\pm\varepsilon}$, where $\chi_{\theta_\pm}$ is the character (5-5) applied to $\theta_\pm$. Indeed, we have

$$V_0 = (V \otimes \mathbb{C}^m) \oplus \mathbb{C}^{v_0 - dm} \quad \text{and} \quad V_1 = (V \otimes \mathbb{C}^{m-1}) \oplus \mathbb{C}^{v_1 - d(m-1)}.$$

The embedding $\mathrm{GL}(V) \hookrightarrow \mathrm{GL}(V_0) \times \mathrm{GL}(V_1)$ is given by

$$g \mapsto ((g \otimes 1_{\mathbb{C}^m}) \oplus 1_{\mathbb{C}^{v_0 - dm}}, (g \otimes 1_{\mathbb{C}^{m-1}}) \oplus 1_{\mathbb{C}^{v_1 - d(m-1)}}).$$

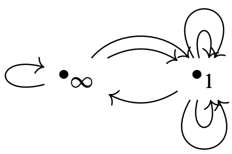
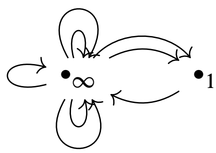
By composing it with $\chi_{\theta_\pm}$, we see that the composition (5-22) is given by $g \mapsto \det(g)^{\pm\varepsilon}$. Therefore under the isomorphism η_p in (5-19), the line bundle on $\widehat{\mathcal{M}}^{\dagger,\theta\text{-ss}}_{Q}(v)$ determined by $\chi_{\theta_\pm}$ corresponds to that on $\widehat{\mathcal{M}}^{\dagger}_{Q_p}(d)$ determined by $\chi_0^{\pm\varepsilon}$. Therefore the lemma holds. □

5.5 Reduced Ext quiver

We define the *reduced Ext quiver* $Q_p^{\mathrm{red},\dagger}$ to be the quiver obtained from $Q_p^{\dagger}$ by removing all the loops at the vertex $\{1\}$, and adding c_m loops at the vertex $\{\infty\}$, where c_m is as given in (5-17). It contains the full subquiver

$$Q_p^{\mathrm{red}} \subset Q_p^{\mathrm{red},\dagger} \tag{5-23}$$

consisting of the vertex $\{1\}$ and no loops. See the diagrams

$Q_p^{\dagger} =$  and $Q_p^{\mathrm{red},\dagger} =$ 

Let $\mathcal{M}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$ be the $\mathbb{C}^*$–rigidified moduli stack of $Q_p^{\mathrm{red},\dagger}$–representations with dimension vector $(1,d)$. It is described as

$$\begin{aligned} \mathcal{M}^{\dagger}_{Q_p^{\mathrm{red}}}(d) &= \left[\left(\mathbb{C}^{\mathrm{ext}^1_{Q^\dagger}(R_\infty,R_\infty)+c_m} \oplus V^{\oplus a_{v,m,d}} \oplus (V^\vee)^{\oplus b_{v,m,d}}\right)/\mathrm{GL}(V)\right], \\ &= \mathbb{C}^{\mathrm{ext}^1_{Q^\dagger}(R_\infty,R_\infty)+c_m} \times \mathcal{G}_{a_{v,m,d},b_{v,m,d}}(d), \end{aligned} \tag{5-24}$$

ie it is obtained from (5-16) by removing the last factor $\mathrm{End}_0(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m)$, and taking the quotient by $\mathrm{GL}(V)$. Here $\mathcal{G}_{a,b}(d)$ is the quotient stack (4-1) studied in Section 4. We also denote by $\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$ the formal fiber for the good moduli space morphism

$$\mathcal{M}^{\dagger}_{Q_p^{\mathrm{red}}}(d) \to M^{\dagger}_{Q_p^{\mathrm{red}}}(d)$$

at the origin.

By restricting the function (5-3) to the formal fiber of the good moduli space morphism (5-6) and pulling it back by the isomorphism η_p in (5-19), we have the function

$$w_p \colon \widehat{\mathcal{M}}^{\dagger}_{Q_p}(d) = [\widehat{\mathrm{Ext}}^1_{Q^{\dagger}}(R,R)/G_p] \to \mathbb{A}^1.$$

We see that the above function is a sum of a function from $\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$ and some nondegenerate quadratic form. Let us take a (noncanonical) isomorphism of $\mathbb{C}$–vector spaces

$$\mathrm{Ext}^1_Q(S_m, S_m) \cong H \oplus H^{\vee}, \tag{5-25}$$

where the dimension of H is $m^2 - m$. Since there is also an isomorphism $\mathrm{End}_0(V) \cong \mathrm{End}_0(V)^{\vee}$ of G_p–representations, we have an isomorphism of G_p–representations

$$\mathrm{End}_0(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m) \cong W \oplus W^{\vee},$$

where $W = \mathrm{End}_0(V) \otimes H$. In particular, we have the nondegenerate symmetric quadratic form

$$q = \langle -,- \rangle \colon \mathrm{End}_0(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m) \to \mathbb{A}^1 \tag{5-26}$$

defined to be the natural pairing on W and $W^{\vee}$. Note that (5-25) is a summand of $\mathrm{Ext}^1_{Q^{\dagger}}(R,R)$ by the decomposition (5-16). We will use the following proposition, whose proof will be given in Section 6.4.

Proposition 5.11 *By replacing the isomorphisms in (5-19) and (5-25) if necessary, the function w_p is written as*

$$w_p = w_p^{\mathrm{red}} + q. \tag{5-27}$$

Here w_p^{red} is nonzero and does not contain variables from $\mathrm{End}_0(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m)$ under the decomposition (5-16).

The $\mathrm{GL}(V)$–representation W determines the vector bundle $\mathcal{W}$ on $\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$. By Proposition 5.11, we have the commutative diagram

$$\begin{tikzcd}
\mathcal{W}^{\vee} \arrow[r, hook, "i"] \arrow[d, "\mathrm{pr}"'] & \mathcal{W} \oplus \mathcal{W}^{\vee} \arrow[dr, "w_p^{\mathrm{red}}+q"'] & \widehat{\mathcal{M}}^{\dagger}_{Q_p}(d) \arrow[l, "\iota"'] \arrow[r, "\eta_p", "\cong"'] \arrow[d, "w_p"] & \widehat{\mathcal{M}}^{\dagger,\theta\text{-ss}}_{Q}(v)_p \arrow[dl, "w_p"] \\
\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d) \arrow[rr, "w_p^{\mathrm{red}}"'] & & \mathbb{A}^1 &
\end{tikzcd} \tag{5-28}$$

Here pr is the projection, i is given by $i(x) = (0, x)$, ι is the natural morphism by the formal completion (see Lemma 6.4) and η_p is the isomorphism in (5-19).

Proposition 5.12 *There is an equivalence*

$$\Phi_p := \iota^* i_* \mathrm{pr}^* \colon \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}) \xrightarrow{\sim} \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p}(d), w_p). \tag{5-29}$$

Proof The composition functor

$$i_* \operatorname{pr}^* \colon \operatorname{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}) \xrightarrow{\operatorname{pr}^*} \operatorname{MF}(\mathcal{W}^{\vee}, w_p^{\mathrm{red}}) \xrightarrow{i_*} \operatorname{MF}(\mathcal{W} \oplus \mathcal{W}^{\vee}, w_p^{\mathrm{red}} + q) \tag{5-30}$$

is an equivalence by Theorem 2.4. By Lemma 6.4, the functor

$$\iota^* \colon \operatorname{MF}(\mathcal{W} \oplus \mathcal{W}^{\vee}, w_p^{\mathrm{red}} + q) \to \operatorname{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p}(d), w_p) \tag{5-31}$$

is fully faithful with dense image. By Lemma 6.3 and the equivalence (5-30), the left-hand side of (5-31) is idempotent complete, so the functor (5-31) is an equivalence. Therefore we obtain the proposition. □

5.6 Window subcategories

In this subsection, we define several window subcategories for moduli stacks of representations of quivers and their formal fibers discussed in the previous subsection. The notation is summarized in Table 1.

Global window subcategory $\mathbb{W}^{\theta_\pm}_{\mathrm{glob}}(v)$ We take $\theta \in W_m$ and $\theta_\pm$ as in (5-20) which are sufficiently close to the wall W_m. Then the KN stratification of $\mathcal{M}^{\dagger}_Q(v)$ for $\chi_{\theta_\pm}$ is finer than those for χ_θ. So we have KN stratifications for $\mathcal{M}^{\dagger,\theta\text{-ss}}_Q(v)$ with respect to $\chi_{\theta_\pm}$,

$$\mathcal{M}^{\dagger,\theta\text{-ss}}_Q(v) = \mathcal{S}^{\pm}_1 \sqcup \cdots \sqcup \mathcal{S}^{\pm}_{N^\pm} \sqcup \mathcal{M}^{\dagger,\theta_\pm\text{-ss}}_Q(v), \tag{5-32}$$

with associated one-parameter subgroups $\lambda_i^\pm \colon \mathbb{C}^* \to \operatorname{GL}(V_0) \times \operatorname{GL}(V_1)$ and the associated number $\eta_i^\pm \in \mathbb{Z}$ as in (2-6). By Theorem 2.3 (and also noting Lemma 5.4), for each choice of real numbers $m_\bullet^\pm = \{(m_i^\pm)\}_{1 \le i \le N^\pm}$ we have the subcategories

$$\mathbb{W}^{\theta_\pm}_{m_\bullet^\pm}(v) \subset \operatorname{MF}(\mathcal{M}^{\dagger,\theta\text{-ss}}_Q(v), w) \tag{5-33}$$

such that the compositions

$$\mathbb{W}^{\theta_\pm}_{m_\bullet^\pm}(v) \hookrightarrow \operatorname{MF}(\mathcal{M}^{\dagger,\theta\text{-ss}}_Q(v), w) \twoheadrightarrow \operatorname{MF}(\mathcal{M}^{\dagger,\theta_\pm\text{-ss}}_Q(v), w) \tag{5-34}$$

are equivalences. The subcategory (5-33) consists of objects whose $\lambda_i^\pm$–weights at each center of $\mathcal{S}_i^\pm$ are contained in $[m_i^\pm, m_i^\pm + \eta_i^\pm)$.

We define the character

$$\chi_0 \colon \operatorname{GL}(V_0) \times \operatorname{GL}(V_1) \to \mathbb{C}^*, \quad (g_0, g_1) \mapsto \det(g_0) \cdot \det(g_1)^{-1},$$

ie $\chi_0 = \chi_{(-1,1)}$ in equation (5-5).

	moduli stack	formal fiber	windows
conifold quiver $Q^\dagger$	$\mathcal{M}^{\dagger,\theta\text{-ss}}_Q(v)$	$\widehat{\mathcal{M}}^{\dagger,\theta\text{-ss}}_Q(v)_p$	$\mathbb{W}^{\theta_\pm}_{\mathrm{glob}}(v)$, $\mathbb{W}^{\theta_\pm}_{\mathrm{loc}}(v)_p$
Ext quiver $Q_p^\dagger$	$\mathcal{M}^{\dagger}_{Q_p}(d)$	$\widehat{\mathcal{M}}^{\dagger}_{Q_p}(d)$	$\mathbb{W}^{\pm}(d)_p$
reduced Ext quiver $Q_p^{\mathrm{red},\dagger}$	$\mathcal{M}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$	$\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$	$\mathbb{W}_c(d)_p$

Table 1: Notation of moduli spaces and windows.

As we discussed in (5-22), the composition

$$G_p = \mathrm{GL}(V) \hookrightarrow \mathrm{GL}(V_0) \times \mathrm{GL}(V_1) \xrightarrow{\chi_0} \mathbb{C}^*$$

coincides with the determinant character $\chi_0 \colon \mathrm{GL}(V) \to \mathbb{C}^*$. For $m_\bullet^\pm$ we use the special choices

$$\begin{aligned} m_i^+ &= -\tfrac{1}{2}\eta_i^+ + \left(\tfrac{1}{2}C_{v,m} + \tfrac{1}{2}m\right)\langle \lambda_i^+, \chi_0\rangle, \\ m_i^- &= -\tfrac{1}{2}\eta_i^- + \tfrac{1}{2}C_{v,m}\langle \lambda_i^-, \chi_0\rangle. \end{aligned} \tag{5-35}$$

Here $C_{v,m}$ is given in (5-17). We then define

$$\mathbb{W}_{\mathrm{glob}}^{\theta\pm}(v) \subset \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w) \tag{5-36}$$

to be the window subcategories (5-33) for the choices of $m_\bullet^\pm$ as (5-35).

Local window subcategories $\mathbb{W}_{\mathrm{loc}}^{\theta\pm}(v)_p$ Let us take a θ–polystable object R as in (5-13), and the corresponding closed point $p \in M_Q^{\dagger,\theta\text{-ss}}(v)$. Then we have the diagram of formal fibers (5-19). By restricting the KN stratification (5-32) to the formal fiber, we obtain the KN stratification of $\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p$

$$\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p = \widehat{\mathcal{S}}_{1,p}^\pm \sqcup \cdots \sqcup \widehat{\mathcal{S}}_{N^\pm,p}^\pm \sqcup \widehat{\mathcal{M}}_Q^{\dagger,\theta_\pm\text{-ss}}(v)_p. \tag{5-37}$$

We define local window subcategories

$$\mathbb{W}_{\mathrm{loc}}^{\theta\pm}(v)_p \subset \mathrm{MF}(\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p, w_p)$$

similarly to (5-36) as in Theorem 2.3, with respect to the KN stratifications (5-37) and the choices of $m_\bullet^\pm$ in (5-35). The following lemma follows immediately from the definition of window subcategories:

Lemma 5.13 *An object $\mathcal{E} \in \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v), w)$ is an object in $\mathbb{W}_{\mathrm{glob}}^{\theta\pm}(v)$ if and only if for any closed point $p \in M_Q^{\dagger,\theta\text{-ss}}(v)$ represented by an object of the form (5-13) we have $\mathcal{E}|_{\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p} \in \mathbb{W}_{\mathrm{loc}}^{\theta\pm}(v)_p$.*

Proof Since the defining conditions of window subcategories $\mathbb{W}_{\mathrm{glob}}^{\theta\pm}(v)$ are local on the good moduli space, $\mathcal{E}$ is an object in $\mathbb{W}_{\mathrm{glob}}^{\theta\pm}(v)$ if and only if $\mathcal{E}|_{\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p} \in \mathbb{W}_{\mathrm{loc}}^{\theta\pm}(v)_p$ for any $p \in M_Q^{\dagger,\theta\text{-ss}}(v)$. If p is not represented by an object of the form (5-13), then the formal fiber $\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p$ does not intersect with the critical locus of w, so $\mathrm{MF}(\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p, w_p) = 0$. □

Window subcategories $\mathbb{W}^\pm(d)_p$ for the Ext quiver By pulling the KN stratification (5-37) back to $\widehat{\mathcal{M}}_{Q_p}^\dagger(d)$ by the isomorphism η_p in (5-19), we have the KN stratification of $\widehat{\mathcal{M}}_{Q_p}^\dagger(d)$ with respect to $\chi_0^{\pm 1}$

$$\widehat{\mathcal{M}}_{Q_p}^\dagger(d) = \widetilde{\mathcal{S}}_{1,p}^\pm \sqcup \cdots \sqcup \widetilde{\mathcal{S}}_{N^\pm,p}^\pm \sqcup \widehat{\mathcal{M}}_{Q_p}^{\dagger,\chi_0^{\pm 1}}(d). \tag{5-38}$$

We define window subcategories

$$\mathbb{W}^\pm(d)_p \subset \mathrm{MF}(\widehat{\mathcal{M}}_{Q_p}^\dagger(d), w_p)$$

as in Theorem 2.3, with respect to the KN stratifications (5-38) and the choices of $m_\bullet^\pm$ in (5-35). By the isomorphism η_p in (5-19), we have the equivalence

$$\eta_p^* \colon \mathbb{W}_{\mathrm{loc}}^{\theta\pm}(v)_p \xrightarrow{\sim} \mathbb{W}^\pm(d)_p. \tag{5-39}$$

Window subcategories $\mathbb{W}_c(d)_p$ for the reduced Ext quiver For $c \in \mathbb{Z}_{\geq 0}$, we also define

$$\mathbb{W}_c(d)_p \subset \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}})$$

to be the thick closure of matrix factorizations whose entries are of the form $V(\chi) \otimes \mathcal{O}$ for $\chi \in \mathbb{B}_c(d)$, where $\mathbb{B}_c(d)$ is as defined in (4-7). By the description (5-24) of $\mathcal{M}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$ in terms of the stack $\mathcal{G}_{a_{v,m,d}, b_{v,m,d}}(d)$, the argument of Proposition 4.3 (also see the argument of Corollary 4.23) implies that the following composition functors are equivalences:

$$\begin{aligned}
&\mathbb{W}_{a_{v,m,d}}(d)_p \hookrightarrow \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}) \twoheadrightarrow \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger, \chi_0\text{-ss}}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}), \\
&\mathbb{W}_{b_{v,m,d}}(d)_p \hookrightarrow \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}) \twoheadrightarrow \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger, \chi_0^{-1}\text{-ss}}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}).
\end{aligned} \tag{5-40}$$

5.7 Comparison of window subcategories

We compare the window subcategories in the previous subsection under the Knörrer periodicity:

Proposition 5.14 *The equivalence (5-29) restricts to the equivalences*

$$\begin{aligned}
\Phi_p &: \mathbb{W}_{a_{v,m,d}}(d)_p \otimes \chi_0^{d(m^2-m)} \xrightarrow{\sim} \mathbb{W}^+(d)_p, \\
\Phi_p &: \mathbb{W}_{b_{v,m,d}}(d)_p \otimes \chi_0^{d(m^2-m)} \xrightarrow{\sim} \mathbb{W}^-(d)_p.
\end{aligned}$$

Proof We only give a proof for the $+$ part. Let $\mathcal{W} \to \widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$ be the vector bundle as in the diagram (5-28). The KN stratifications (5-32) are pullbacks of the KN stratifications

$$\mathcal{W} \oplus \mathcal{W}^{\vee} = \overline{\mathcal{S}}_1^{\pm} \sqcup \cdots \sqcup \overline{\mathcal{S}}^{\pm}_{N^{\pm}} \sqcup (\mathcal{W} \oplus \mathcal{W}^{\vee})^{\chi_0^{\pm 1}\text{-ss}} \tag{5-41}$$

of $\mathcal{W} \oplus \mathcal{W}^{\vee}$ with respect to $\chi_0^{\pm 1}$ by the morphism ι in (5-28). We denote by

$$\overline{\mathbb{W}}^{\pm}(d)_p \subset \mathrm{MF}(\mathcal{W} \oplus \mathcal{W}^{\vee}, w_p^{\mathrm{red}} + q)$$

the window subcategories with respect to the above stratifications (5-41) and $m_{\bullet}^{\pm} \in \mathbb{R}$ given by (5-35). By the definition of the above window subcategories, the equivalence (5-31) restricts to the equivalence

$$\iota^* : \overline{\mathbb{W}}^{\pm}(d)_p \xrightarrow{\sim} \mathbb{W}^{\pm}(d)_p.$$

Therefore it is enough to show that the equivalence (5-30) restricts to the equivalence

$$i_* \mathrm{pr}^* : \mathbb{W}_{a_{v,m,d}}(d)_p \otimes \chi_0^{d(m^2-m)} \xrightarrow{\sim} \overline{\mathbb{W}}^{+}(d)_p.$$

We have the commutative diagram

$$\begin{tikzcd}
\mathbb{W}_{a_{v,m,d}}(d)_p \otimes \chi_0^{d(m^2-m)} \arrow[r, hook] \arrow[d, dotted] & \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}) \arrow[r] \arrow[d, "i_* \mathrm{pr}^*"', "\sim"] & \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger, \chi_0\text{-ss}}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}) \arrow[d, dotted, "\sim"] \\
\overline{\mathbb{W}}^{+}(d)_p \arrow[r, hook] & \mathrm{MF}(\mathcal{W} \oplus \mathcal{W}^{\vee}, w_p^{\mathrm{red}} + q) \arrow[r] & \mathrm{MF}((\mathcal{W} \oplus \mathcal{W}^{\vee})^{\chi_0\text{-ss}}, w_p^{\mathrm{red}} + q)
\end{tikzcd} \tag{5-42}$$

The composition of top arrows is an equivalence by the equivalence in (5-40), and that of bottom arrows is also an equivalence by Theorem 2.3. We see that the middle vertical arrow descends to an equivalence of the right vertical dotted arrow. Note that we have the isomorphism

$$\mathrm{Crit}(w_p^{\mathrm{red}}) \cap \widehat{\mathcal{M}}^{\dagger,\chi_0\text{-ss}}_{Q_p^{\mathrm{red}}} \xhookrightarrow{\cong} \mathrm{Crit}(w_p^{\mathrm{red}}+q) \cap (\mathcal{W} \oplus \mathcal{W}^\vee)^{\chi_0\text{-ss}}$$

induced by the zero section $\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d) \hookrightarrow \mathcal{W} \oplus \mathcal{W}^\vee$. In particular, we have the inclusion

$$\mathrm{Crit}(w_p^{\mathrm{red}}+q) \cap (\mathcal{W} \oplus \mathcal{W}^\vee)^{\chi_0\text{-ss}} \subset (\mathcal{W} \oplus \mathcal{W}^\vee) \times_{\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)} \widehat{\mathcal{M}}^{\dagger,\chi_0\text{-ss}}_{Q_p^{\mathrm{red}}}(d). \tag{5-43}$$

The desired equivalence is given by the composition

$$\begin{aligned}\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger,\chi_0\text{-ss}}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}) &\xrightarrow{\sim} \mathrm{MF}\big((\mathcal{W} \oplus \mathcal{W}^\vee) \times_{\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)} \widehat{\mathcal{M}}^{\dagger,\chi_0\text{-ss}}_{Q_p^{\mathrm{red}}}(d), w_p^{\mathrm{red}}+q\big) \\ &\xrightarrow{\sim} \mathrm{MF}((\mathcal{W} \oplus \mathcal{W}^\vee)^{\chi_0\text{-ss}}, w_p^{\mathrm{red}}+q).\end{aligned}$$

Here the first equivalence is Knörrer periodicity in Theorem 2.4, and the second equivalence follows from (5-43) and the equivalence (2-1).

Therefore it is enough to show that the middle vertical arrow in (5-42) restricts to the left dotted arrow, ie for $\mathcal{P} \in \mathbb{W}_{a_{v,m,d}}(d)_p \otimes \chi_0^{d(m^2-m)}$, we show that the object $i_* \,\mathrm{pr}^*(\mathcal{P})$ lies in $\overline{\mathbb{W}}^+(d)_p$. Note that the critical locus of $w_p^{\mathrm{red}}+q$ lies in the zero section $\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d) \subset \mathcal{W} \oplus \mathcal{W}^\vee$. From Theorem 2.3, it is enough to show that $i_* \,\mathrm{pr}^*(\mathcal{P})$ satisfies the condition (2-9) for one-parameter subgroups which appear in the KN stratification of $\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$. From the description (5-24) of $\mathcal{M}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$, its KN stratifications with respect to $\chi_0^{\pm 1}$ are KN stratifications of $\mathcal{G}_{a_{v,m,d},b_{v,m,d}}(d)$ discussed in Section 4.1, up to a product with a trivial factor. Therefore they are of the form

$$\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d) = \mathcal{S}_0^{\pm} \sqcup \cdots \sqcup \mathcal{S}_{d-1}^{\pm} \sqcup \widehat{\mathcal{M}}^{\dagger,\chi_0^{\pm 1}\text{-ss}}_{Q_p^{\mathrm{red}}}(d)$$

such that each associated one-parameter subgroup $\lambda_i^{\pm}: \mathbb{C}^* \to G_p = \mathrm{GL}(V)$ is given by (4-6), ie λ_i^+ is

$$\lambda_i^+(t) = (\overbrace{1,\dots,1}^{i}, \overbrace{t^{-1},\dots,t^{-1}}^{d-i}). \tag{5-44}$$

Therefore in order to show that the object $i_* \,\mathrm{pr}^*(\mathcal{P})$ lies in $\overline{\mathbb{W}}^+(d)_p$, it is enough to check the weight conditions (2-9) for the above λ_i^+.

Since the object $i_* \,\mathrm{pr}^*(\mathcal{P})$ is given by taking the tensor product with the Koszul factorization (2-11), it is isomorphic to a direct summand of a matrix factorization whose entries are of the form

$$V(\chi) \otimes \textstyle\bigwedge^k W \otimes \chi_0^{d(m^2-m)} \otimes \mathcal{O}, \quad \text{where } \chi \in \mathbb{B}_{a_{v,m,d}}(d) \text{ and } 0 \le k \le \dim W.$$

For each one-parameter subgroup $\lambda: \mathbb{C}^* \to G_p$, we set

$$\gamma_\lambda := \langle \lambda, W^{\lambda>0} \rangle = -\langle \lambda, W^{\lambda<0} \rangle,$$

where the second identity holds as $W = \mathrm{End}_0(V) \otimes \mathbb{C}^{m^2-m}$ is a self-dual G_p–representation. Then we have the following inclusions of the set of λ_i^+–weights of $V(\chi) \otimes \bigwedge^k W \otimes \chi_0^{d(m^2-m)}$:

$$\begin{aligned}
&\mathrm{wt}_{\lambda_i^+}\big(V(\chi) \otimes \textstyle\bigwedge^k W \otimes \chi_0^{d(m^2-m)}\big) \\
&\quad\subset \bigcup_{\chi' \in \mathrm{wt}(V(\chi))} \Big[-\sum_{j=i+1}^{d} x'_j - (d-i)\cdot d(m^2-m) - \gamma_{\lambda_i^+}, -\sum_{j=i+1}^{d} x'_j - (d-i)\cdot d(m^2-m) + \gamma_{\lambda_i^+}\Big] \\
&\quad\subset [(d-i)(-a_{v,m,d} + d - dm^2 + dm) - \gamma_{\lambda_i^+}, (d-i)(-dm^2 + dm) + \gamma_{\lambda_i^+}].
\end{aligned}$$

Here $\mathrm{wt}(V(\chi))$ is the set of T–weights of $V(\chi)$ for the maximal torus $T \subset G_p$, and we have written $\chi' \in \mathrm{wt}(V(\chi))$ as $\chi' = (x'_1, \dots, x'_d)$ satisfying $0 \le x'_j \le a_{v,m,d} - d$.

We show that the above set of weights is contained in $[m_i^+, m_i^+ + \eta_i^+)$. From the decomposition (5-16), the $\eta_i^+ \in \mathbb{Z}$ which appear in (5-35) for the one-parameter subgroup (5-44) are calculated as in the proof of Proposition 4.3:

$$\begin{aligned}
\eta_i^+ &= \langle \lambda_i^+, (\mathrm{Ext}^1_{Q^\dagger}(R,R)^\vee)^{\lambda_i^+>0} - (\mathfrak{g}_p^\vee)^{\lambda_i^+>0}\rangle \\
&= \langle \lambda_i^+, ((V^\vee)^{\oplus a_{v,m,d}} \oplus V^{\oplus b_{b,m,d}} \oplus W \oplus W^\vee)^{\lambda_i^+>0} - \mathrm{End}(V)^{\lambda_i^+>0}\rangle \\
&= (a_{v,m,d} - i)(d-i) + 2\gamma_{\lambda_i^+}.
\end{aligned}$$

Here $\mathfrak{g}_p = \mathrm{End}(V)$ is the Lie algebra of $G_p = \mathrm{GL}(V)$. Therefore we have

$$\begin{aligned}
&[m_i^+, m_i^+ + \eta_i^+) \\
&\quad= \big[-\tfrac{1}{2}\eta_i^+ + \big(\tfrac{1}{2}C_{v,m} + \tfrac{1}{2}m\big)\langle \lambda_i^+, \chi_0\rangle, \tfrac{1}{2}\eta_i^+ + \big(\tfrac{1}{2}C_{v,m} + \tfrac{1}{2}m\big)\langle \lambda_i^+, \chi_0\rangle\big) \\
&\quad= \big[(d-i)\big(-a_{v,m,d} + \tfrac{1}{2}i + \tfrac{1}{2}d - dm^2 + dm\big) - \gamma_{\lambda_i^+}, (d-i)\big(-dm^2 + dm + \tfrac{1}{2}d - \tfrac{1}{2}i\big) + \gamma_{\lambda_i^+}\big).
\end{aligned}$$

Since $0 \le i \le d-1$, we conclude the inclusion

$$\mathrm{wt}_{\lambda_i^+}\big(V(\chi) \otimes \textstyle\bigwedge^k W \otimes \chi_0^{d(m^2-m)}\big) \subset [m_i^+, m_i^+ + \eta_i^+).$$

Therefore the weight condition (2-9) for $i_* \mathrm{pr}^* \mathcal{P}$ with respect to λ_i^+ is satisfied. $\square$

Let $s_m = (m, m-1)$ be the dimension vector of the stable Q–representation S_m, defined in (5-7). Let $q_m \in M_Q^{\theta\text{-ss}}(s_m)$ be the corresponding closed point. We consider the formal fiber of the good moduli space morphism (5-9) at q_m

$$\widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(s_m) \to \widehat{M}_Q^{\theta\text{-ss}}(s_m).$$

Similarly to (5-15), the étale slice theorem implies an isomorphism

$$\widehat{\mathcal{M}}_{Q_p}(1) = [\widehat{\mathrm{Ext}}^1_Q(S_m, S_m)/\mathrm{Aut}(S_m)] \xrightarrow{\cong} \widehat{\mathcal{M}}_Q^{\theta\text{-ss}}(s_m). \tag{5-45}$$

Here $\mathrm{Aut}(S_m) = \mathbb{C}^*$ acts on $\mathrm{Ext}^1_Q(S_m, S_m)$ trivially. We will also use the following lemma, which compares window subcategories for quivers without framings.

Lemma 5.15 *For any $j \in \mathbb{Z}$, we have equivalences*

$$\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1), w_p^{\mathrm{red}})_j \xrightarrow{\sim} \mathrm{MF}(\widehat{\mathcal{M}}_{Q_p}(1), w_p)_j \xrightarrow{\sim} \mathrm{MF}(\widehat{\mathcal{M}}_Q^{\theta\text{–ss}}(s_m), w)_j, \tag{5-46}$$

and all of them are equivalent to $\mathrm{MF}(\operatorname{Spec}\mathbb{C}, 0)$. *Here the first equivalence is given by the Knörrer periodicity in Theorem 2.4.*

Proof By the definition of Q_p^{red} in (5-23), we have $(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1), w_p^{\mathrm{red}}) = (B\mathbb{C}^*, 0)$. On the other hand, the isomorphisms (5-25), (5-45) and an argument of Proposition 5.11 imply an isomorphism

$$(\widehat{\mathcal{M}}_{Q_p}(1), w_p) \cong \bigl([\widehat{(H \oplus H^\vee)}/\mathbb{C}^*], q\bigr), \tag{5-47}$$

where $\mathbb{C}^*$ acts on $H = \mathbb{C}^{m^2-m}$ trivially and q is the natural paring on H and its dual. By the Knörrer periodicity in Theorem 2.4, we have an equivalence

$$\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1), w_p^{\mathrm{red}})_j = \mathrm{MF}(\operatorname{Spec}\mathbb{C}, 0) \xrightarrow{\sim} \mathrm{MF}(H \oplus H^\vee, q).$$

The natural functor by the formal completion

$$\mathrm{MF}(H \oplus H^\vee, q) \to \mathrm{MF}(\widehat{H \oplus H^\vee}, q)$$

is an equivalence; see [Brown 2016, Remark 2.18]. Therefore we obtain the desired equivalences (5-46). □

5.8 Comparison of Hall products

As in the previous subsections, we take a stability condition on the wall $\theta \in W_m$ for $m \geq 1$. As in Section 3.3, we have the categorified Hall product

$$\begin{aligned}\mathrm{MF}(\mathcal{M}_Q^{\theta\text{–ss}}(s_m), w)_{j_1} \boxtimes \cdots \boxtimes \mathrm{MF}(\mathcal{M}_Q^{\theta\text{–ss}}(s_m), w)_{j_l} \boxtimes \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v - l s_m), w)& \\ \to \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta\text{–ss}}(v), w).&\end{aligned} \tag{5-48}$$

Here $s_m = (m, m-1)$ is the dimension vector of S_m. We take a θ–polystable representation $Q^\dagger$–representation R of the form (5-13), ie $R = R_\infty \oplus (V \otimes S_m)$ with $\dim V = d$, and the corresponding closed point $p \in M_Q^{\dagger,\theta\text{–ss}}(v)$. By taking the base change of the above categorified Hall product to the formal completion at p (see Section 3.4), we obtain the functor

$$\begin{aligned}\mathrm{MF}(\widehat{\mathcal{M}}_Q^{\theta\text{–ss}}(s_m), w)_{j_1} \boxtimes \cdots \boxtimes \mathrm{MF}(\widehat{\mathcal{M}}_Q^{\theta\text{–ss}}(s_m), w)_{j_l} \boxtimes \mathrm{MF}(\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{–ss}}(v - l s_m)_{p_l}, w)& \\ \to \mathrm{MF}(\widehat{\mathcal{M}}_Q^{\dagger,\theta\text{–ss}}(v)_p, w).&\end{aligned} \tag{5-49}$$

Here $p_l \in M_Q^{\dagger,\theta\text{–ss}}(v - l s_m)$ corresponds to the θ–polystable representation $R_\infty \oplus (V' \otimes S_m)$ with $\dim V' = d - l$. We note that by the isomorphism η_p in (5-19) and the isomorphism (5-45), the above functor is identified with the functor

$$\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p}(1), w_p)_{j_1} \boxtimes \cdots \boxtimes \mathrm{MF}(\widehat{\mathcal{M}}_{Q_p}(1), w_p)_{j_l} \boxtimes \mathrm{MF}(\widehat{\mathcal{M}}^\dagger_{Q_p}(d-l), w_p) \to \mathrm{MF}(\widehat{\mathcal{M}}^\dagger_{Q_p}(d), w_p) \tag{5-50}$$

obtained by the categorified Hall products for $Q_p^\dagger$–representations and the completions at the origins.

A similar construction also gives the categorified Hall product for $Q_p^{\mathrm{red},\dagger}$–representations

$$\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)_{j_1}\boxtimes\cdots\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)_{j_l}\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d-l),w_p^{\mathrm{red}})\to\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d),w_p^{\mathrm{red}}). \tag{5-51}$$

We compare the above categorified Hall products under the Knörrer periodicity:

Proposition 5.16 *The following diagram commutes*:

$$\begin{tikzcd} \boxtimes_{i=1}^{l}\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)_{j_i}\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d-l),w_p^{\mathrm{red}}) \arrow[r] \arrow[d] & \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d),w_p^{\mathrm{red}}) \arrow[d] \\ \boxtimes_{i=1}^{l}\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p}(1),w_p)_{j_i+(2i-d-1)(m^2-m)}\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p}(d-l),w_p) \arrow[r] & \mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p}(d),w_p) \end{tikzcd} \tag{5-52}$$

Here the horizontal arrows are given by categorized Hall products (5-50) and (5-51), the right vertical arrow is given in Proposition 5.12, and the left vertical arrow is a composition of the functors in Proposition 5.12 and Lemma 5.15 with the equivalence

$$\begin{aligned}\boxtimes_{i=1}^{l}\otimes\mathcal{O}_{B\mathbb{C}^*}((2i-d-1)(m^2-m))\boxtimes\otimes\chi_0^{l(m^2-m)}\big[\big(dl-\tfrac{1}{2}l-\tfrac{1}{2}l^2\big)(m^2-m)\big]&:\\ \boxtimes_{i=1}^{l}\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)_{j_i}\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d-l),w_p^{\mathrm{red}})&\\ \xrightarrow{\sim}\boxtimes_{i=1}^{l}\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)_{j_i+(2i-d-1)(m^2-m)}\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d-l),w_p^{\mathrm{red}})&.\end{aligned} \tag{5-53}$$

Proof We take $\lambda\colon\mathbb{C}^*\to G_p=\mathrm{GL}(V)$ by

$$\lambda(t)=(t^l,t^{l-1},\dots,t,\overbrace{1,\dots,1}^{d-l}).$$

Then the top arrow in the diagram (5-52) is obtained from the diagram of attracting loci for $\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$ with respect to the above λ. In the diagram (5-28), the vector bundle $\mathcal{W}\to\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d)$ is induced by the $\mathrm{GL}(V)$–representation $W=\mathrm{End}_0(V)\otimes H$ for $H=\mathbb{C}^{m^2-m}$ by its definition. By Proposition 2.6, the categorified Hall products in (5-52) commute with Knörrer periodicity equivalences up to equivalence:

$$\begin{aligned}\otimes\det(\mathrm{End}_0(V)^{\lambda>0}\otimes H)^{\vee}[\dim(\mathrm{End}_0(V)^{\lambda>0}\otimes H)]&:\\ \boxtimes_{i=1}^{l}\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d-l),w_p^{\mathrm{red}})&\\ \xrightarrow{\sim}\boxtimes_{i=1}^{l}\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)\boxtimes\mathrm{MF}(\widehat{\mathcal{M}}^{\dagger}_{Q_p^{\mathrm{red}}}(d-l),w_p^{\mathrm{red}})&.\end{aligned} \tag{5-54}$$

It is enough to show that the equivalence (5-54) restricts to the equivalence (5-53). Let $V=\bigoplus_{i=0}^{l}V_i$ be the decomposition into λ–weight parts, ie V_i has λ–weight i so that $\dim V_i=1$ for $1\le i\le l$ and $\dim V_0=d-l$. We have

$$\mathrm{End}_0(V,V)^{\lambda>0}=\Big(\bigoplus_{0\le i<j\le l}V_i^{\vee}\otimes V_j\Big).$$

We compute that

$$\det(\mathrm{End}_0(V,V)^{\lambda>0})^{\vee}=\bigotimes_{0\le i<j\le l}\det(V_i\otimes V_j^{\vee})=\bigotimes_{1\le j\le l}\det(V_0\otimes V_j^{\vee})\otimes\bigotimes_{1\le i<j\le l}\det(V_i\otimes V_j^{\vee})$$
$$=(\det V_0)^l\otimes\bigotimes_{i=1}^{l}(\det V_i)^{2l-2i+1-d}.$$

We note that $\otimes\det V_i=\otimes\mathcal{O}_{B\mathbb{C}^*}(1)$ on the factor $\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)_{j_{l-i+1}}$. We also have

$$\dim\mathrm{End}_0(V,V)^{\lambda>0}=dl-\tfrac{1}{2}l-\tfrac{1}{2}l^2.$$

Therefore the equivalence (5-54) restricts to the equivalence (5-53). □

5.9 Semiorthogonal decomposition of global window subcategories

The following is the main result in this section:

Theorem 5.17 *For $l\ge 0$ and $0\le j_1\le\cdots\le j_l\le m-l$, the categorified Hall product (5-48) restricts to the fully faithful functor*

$$\Upsilon_{j_\bullet}\colon\boxtimes_{i=1}^{l}\mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(s_m),w)_{j_i+(2i-1)(m^2-m)}\boxtimes\left(\mathbb{W}_{\mathrm{glob}}^{\theta_-}(v-ls_m)\otimes\chi_0^{j_l+2l(m^2-m)}\right)\to\mathbb{W}_{\mathrm{glob}}^{\theta_+}(v)\tag{5-55}$$

such that, by setting $\mathcal{C}_{j_\bullet}$ to be the essential image of the above functor $\Upsilon_{j_\bullet}$, we have the semiorthogonal decomposition

$$\mathbb{W}_{\mathrm{glob}}^{\theta_+}(v)=\langle\mathcal{C}_{j_\bullet}:l\ge 0,0\le j_1\le\cdots\le j_l\le m-l\rangle,\tag{5-56}$$

where $\mathrm{Hom}(\mathcal{C}_{j_\bullet},\mathcal{C}_{j'_\bullet})=0$ *for* $j_\bullet\succ j'_\bullet$ (see Definition 4.16).

Proof We take a θ–polystable representation R of the form (5-13), ie $R=R_\infty\oplus(V\otimes S_m)$ with $\dim V=d$, the corresponding closed point $p\in M_Q^{\dagger,\theta\text{-ss}}(v)$, and consider the quivers $Q_p^{\dagger}$, $Q_p^{\dagger,\mathrm{red}}$ as in the previous subsections. Note that if we remove the loops at the vertex $\{\infty\}$ from $Q_p^{\dagger}$, then we obtain the quiver $Q_{a,b}$ for $a=a_{v,m,d}$ and $b=b_{v,m,d}$ considered in Remark 4.1. By applying Corollary 4.23 for the above $Q_{a,b}$, and then taking the tensor product with $\chi_0^{d(m^2-m)}$, we obtain the semiorthogonal decomposition

$$\mathbb{W}_{a_{v,m,d}}(d)_p\otimes\chi_0^{d(m^2-m)}$$
$$=\left\langle\boxtimes_{i=1}^{l}\mathrm{MF}(\widehat{\mathcal{M}}_{Q_p^{\mathrm{red}}}(1),0)_{j_i+d(m^2-m)}\boxtimes\left(\mathbb{W}_{b_{v,m,d}}(d-l)_{p_l}\otimes\chi_0^{(d-l)(m^2-m)}\otimes\chi_0^{j_l+l(m^2-m)}\right)\right\rangle.$$

Here $l\ge 0$, $p_l\in M_Q^{\dagger,\theta\text{-ss}}(v-ls_m)$ corresponds to $R_\infty\oplus(V'\otimes S_m)$ with $\dim V'=d-l$, and

$$0\le j_1\le\cdots\le j_l\le a_{v,m,d}-b_{v,m,d}-l=m-l.\tag{5-57}$$

Applying Proposition 5.14, Lemma 5.15 and Proposition 5.16, we get the semiorthogonal decomposition

$$\mathbb{W}^{+}(d)_p = \left\langle \boxtimes_{i=1}^{l} \mathrm{MF}(\widehat{\mathcal{M}}_{Q_p}(1), w_p)_{j_i+(2i-1)(m^2-m)} \boxtimes \left(\mathbb{W}^{-}(d-l)_{p_l} \otimes \chi_0^{j_l+2l(m^2-m)}\right)\right\rangle.$$

By the identification of categorified Hall products (5-49) with (5-50) together with the equivalence (5-39), we obtain the semiorthogonal decomposition

$$\mathbb{W}^{\theta+}_{\mathrm{loc}}(v)_p = \left\langle \boxtimes_{i=1}^{l} \mathrm{MF}(\widehat{\mathcal{M}}_{Q}^{\theta\text{-ss}}(s_m), w)_{j_i+(2i-1)(m^2-m)} \boxtimes \left(\mathbb{W}^{\theta-}_{\mathrm{loc}}(v-l s_m)_{p_l} \otimes \chi_0^{j_l+2l(m^2-m)}\right)\right\rangle. \tag{5-58}$$

A key observation is that in the above semiorthogonal decomposition there is no term involving $d = \dim V$ (which depends on a choice of θ–polystable object (5-13)), so we can globalize it. Indeed we have globally defined functors (5-55) and, noting Lemma 5.13, in order to show that they are fully faithful and forms a semiorthogonal decomposition it is enough to check these properties formally locally at each closed point of $M_Q^{\dagger,\theta\text{-ss}}(v)$ corresponding to a θ–polystable $(Q^{\dagger}, W)$–representation; see the arguments in [Toda 2021, Proposition 6.9, Theorem 6.11], for example.

Here we give some more details for how to derive the global semiorthogonal decomposition (5-56) from the formal local one (5-58). We first note that the categorified Hall product (5-48) restricts to the functor (5-55). This follows from the fact that the categorified Hall products commute with base change to the formal completion of good moduli spaces (see the diagram (3-14)), the fact (which follows from (5-58)) that formally locally over $M_Q^{\dagger,\theta\text{-ss}}(v)$ the categorified Hall product restricts to the functor

$$\boxtimes_{i=1}^{l} \mathrm{MF}(\widehat{\mathcal{M}}_{Q}^{\theta\text{-ss}}(s_m), w)_{j_i+(2i-1)(m^2-m)} \boxtimes \left(\mathbb{W}^{\theta-}_{\mathrm{loc}}(v-l s_m)_{p_l} \otimes \chi_0^{j_l+2l(m^2-m)}\right) \to \mathbb{W}^{\theta+}_{\mathrm{loc}}(v)_p,$$

and noting Lemma 5.13.

By Lemma 6.6 below, the functor $\Upsilon_{j_\bullet}$ admits a right adjoint $\Upsilon_{j_\bullet}^{R}$. Now in order to show that $\Upsilon_{j_\bullet}$ is fully faithful, it is enough to show that the adjunction morphism

$$(-) \to \Upsilon_{j_\bullet}^{R} \circ \Upsilon_{j_\bullet}(-)$$

is an isomorphism. Equivalently, it is enough to show that the cone of the above morphism is zero. By Lemma 6.5, this is a property formally locally over $M_Q^{\dagger,\theta\text{-ss}}(v)$. So from the semiorthogonal decomposition (5-58) we conclude that $\Upsilon_{j_\bullet}$ is fully faithful. A similar argument also shows that $\mathscr{C}_{j_\bullet}$ for $j_\bullet$ given in (5-57) are semiorthogonal.

In order to show that $\mathscr{C}_{j_\bullet}$ for $j_\bullet$ given in (5-57) generate $\mathbb{W}^{\theta+}_{\mathrm{glob}}(v)$, let us take $\mathscr{E} \in \mathbb{W}^{\theta+}_{\mathrm{glob}}(v)$ and $j_\bullet$ so that $j_\bullet$ is maximal in the order of Definition 4.16. We have the distinguished triangle

$$\Upsilon_{j_\bullet}\Upsilon_{j_\bullet}^{R}(\mathscr{E}) \to \mathscr{E} \to \mathscr{E}', \quad \text{where } \mathscr{E}' \in \mathscr{C}_{j_\bullet}^{\perp}.$$

By applying the above construction for $\mathscr{E}'$ and the second maximal $j_\bullet$ and repeating, we obtain the distinguished triangle

$$\mathscr{E}_1 \to \mathscr{E} \to \mathscr{E}_2, \quad \text{where } \mathscr{E}_1 \in \langle \mathscr{C}_{j_\bullet} \rangle \text{ and } \mathscr{E}_2 \in \langle \mathscr{C}_{j_\bullet} \rangle^{\perp}.$$

Here $\langle \mathscr{C}_{j_\bullet} \rangle$ is the right-hand side of (5-56). From the semiorthogonal decomposition (5-58), we have $\mathscr{E}_2|_{\widehat{\mathscr{M}}_Q^{\dagger,\theta\text{-ss}}(v)_p} = 0$ for any closed point $p \in M_Q^{\dagger,\theta\text{-ss}}(v)$, therefore $\mathscr{E}_2 = 0$ by Lemma 6.5. Therefore $\mathscr{E} \in \langle \mathscr{C}_{j_\bullet} \rangle$, and we have the desired semiorthogonal decomposition (5-56). □

The following corollary, which is an immediate consequence from Theorem 5.17, categorifies wall-crossing formula of the associated DT invariants in [Nagao and Nakajima 2011].

Corollary 5.18 *There exists a semiorthogonal decomposition of the form*

$$\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta+}(v), w) = \langle \mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta-}(v - l s_m), w)_{j_\bullet} : l \geq 0, 0 \leq j_1 \leq \cdots \leq j_l \leq m - l \rangle.$$

Here $\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta-}(v - l s_m), w)_{j_\bullet}$ *is a copy of* $\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta-}(v - l s_m), w)$.

Proof By the equivalences (5-34), the left-hand side of (5-56) is equivalent to $\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta+}(v), w)$. On the other hand, the subcategory $\mathscr{C}_{j_\bullet}$ in (5-56) is equivalent to $\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta-}(v - l s_m), w)$ by the equivalences (5-34) together with Lemma 5.7. □

Remark 5.19 The semiorthogonal decomposition in Corollary 5.18 recovers the numerical wall-crossing formula (1-4). Indeed the periodic cyclic homologies are additive with respect to semiorthogonal decompositions [Tabuada 2005, Theorem 6.3, Section 6.1], so we have

$$\mathrm{HP}_*(\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta+\text{-ss}}(v), w)) = \bigoplus_{l \geq 0} \mathrm{HP}_*(\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta-\text{-ss}}(v - l s_m), w))^{\oplus \binom{m}{l}}.$$

By taking the Euler characteristics and using Lemma 5.1, we obtain the formula (1-4).

By applying Corollary 5.18 from the empty chamber in Figure 1 to the wall-crossing at W_m, and noting Lemma 5.6, we obtain the following.

Corollary 5.20 *For* $\theta \in W_m$, *there exists a semiorthogonal decomposition*

$$\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta+}(v), w) = \langle \mathscr{C}_{j_\bullet^{(*)}} \rangle.$$

Here each $\mathscr{C}_{j_\bullet^{(*)}}$ *is equivalent to* $\mathrm{MF}(\operatorname{Spec}\mathbb{C}, 0)$ *and* $j_\bullet^{(*)}$ *is a collection of nonpositive integers of the form*

$$j_\bullet^{(*)} = \{(0 \leq j_1^{(i)} \leq \cdots \leq j_{l_i}^{(i)} \leq i - l_i)\}_{1 \leq i \leq m}$$

for some integers $l_i \geq 0$ *satisfying*

$$(v_0, v_1) = \sum_{i=1}^{m} l_i \cdot (i, i-1).$$

We have $\mathrm{Hom}(\mathscr{C}_{j_\bullet^{(*)}}, \mathscr{C}_{j_\bullet'^{(*)}}) = 0$ *if* $j_\bullet^{(i)} = j_\bullet'^{(i)}$ *for* $k < i \leq m$ *for some* k *and* $j_\bullet^{(k)} \succ j_\bullet'^{(k)}$.

Proof Let $\theta_{\mathrm{en}} \in \mathbb{R}^2$ lie in the empty chamber in Figure 1. By Lemma 5.6, a successive application of Corollary 5.18 gives the semiorthogonal decomposition

$$\mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta+}(v), w) = \langle \mathrm{MF}(\mathscr{M}_Q^{\dagger,\theta_{\mathrm{en}}}(v - l_m s_m - l_{m-1} s_{m-1} - \cdots - l_1 s_1), w)_{j_\bullet^{(m)}, j_\bullet^{(m-1)}, \ldots, j_\bullet^{(1)}} \rangle.$$

Here $l_i \geq 0$ are integers and $0 \leq j_1^{(1)} \leq \cdots \leq j_{l_i}^{(i)} \leq i - l_i$ for $1 \leq i \leq m$. By applying Lemma 5.5, we obtain the corollary. $\square$

Remark 5.21 The arguments of Theorem 5.17 and Corollary 5.18 work for other walls except walls at $\{\theta_0 + \theta_1 = 0\}$. For example, let us consider, in Figure 1, the wall

$$W'_m := \mathbb{R}_{>0}(-m-1, m), \quad \text{where } m \in \mathbb{Z}_{\geq 0}.$$

Then for $\theta \in W'_m$, there is a unique θ–stable (Q, W)–representation S'_m of dimension vector $s'_m = (m, m+1)$, which corresponds to $\mathcal{O}_C(-m-1)[1]$ under the equivalence Φ in (5-1); see [Nagao and Nakajima 2011, Remark 3.6]. The arguments of Theorem 5.17 and Corollary 5.18 work verbatim by replacing S_m and s_m with S'_m and s'_m, so that we have the semiorthogonal decomposition

$$\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_+}(v), w) = \langle \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_-}(v - l s'_m), w)_{j_\bullet} : l \geq 0, 0 \leq j_1 \leq \cdots \leq j_l \leq m - l \rangle.$$

Remark 5.22 On the other hand, the above arguments do not work at walls in $\{\theta_0 + \theta_1 = 0\}$. For example at the DT/PT wall $\theta \in \mathbb{R}_{>0}(-1, 1)$, there exist an infinite number of θ–stable (Q, W)–representations corresponding to closed points in X, and the associated Ext quivers are more complicated. At the DT/PT wall, we expect the categorical wall-crossing formula

$$\mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_+}(v), w) = \left\langle \bigboxtimes_{i=1}^{k} \mathbb{S}(d_i)_{v_i} \boxtimes \mathrm{MF}\left(\mathcal{M}_Q^{\dagger,\theta_-}\left(v - \sum_{i=1}^{k} d_i \cdot s_\infty\right), w\right)\right\rangle.$$

Here $s_\infty = (1, 1)$, $(d_i, v_i) \in \mathbb{Z}_{>0} \times \mathbb{Z}$ satisfy $-1 < v_1/d_1 < \cdots < v_k/d_k \leq 1$ and $\mathbb{S}(v)_d$ is the subcategory

$$\mathbb{S}(d)_v \subset \mathrm{MF}(\mathcal{M}_Q^{\theta\text{-ss}}(d s_\infty), w)_v$$

defined similarly to the quasi-BPS category in [Pădurariu and Toda 2022]. Some details may be pursued in a future work.

5.10 Semiorthogonal decompositions of categorical stable pair theory

By definition a *PT stable pair* [Pandharipande and Thomas 2009] on X is a pair (F, s) where F is a pure one-dimensional coherent sheaf on X and $s : \mathcal{O}_X \to F$ is surjective in dimension one. For $(\beta, n) \in \mathbb{Z}^2$, we denote by

$$P_n(X, \beta)$$

the moduli space of PT stable pair moduli space (F, s) on X satisfying $[F] = \beta[C]$ and $\chi(F) = n$, where $[F]$ is the fundamental one-cycle of F. Since any such a sheaf F is supported on C, the moduli space $P_n(X, \beta)$ is a projective scheme.

Nakao and Nakajima [2011, Proposition 2.11] proved that the equivalence (5-1) induces the isomorphism

$$\Phi_* : P_n(X, \beta) \xrightarrow{\cong} \mathcal{M}_{(Q,W)}^{\dagger,\theta_{\mathrm{PT}}}(n, n - \beta),$$

where $\theta_{\mathrm{PT}} := (-1+\varepsilon, 1+\varepsilon)$ for $0<\varepsilon \ll 1$. The right-hand side is the critical locus of the function $w\colon \mathcal{M}_Q^{\dagger,\theta_{\mathrm{PT}}}(n, n-\beta) \to \mathbb{A}^1$ defined by (5-3). Based on the above isomorphism, the categorical PT invariant is defined as follows.

Definition 5.23 We define the categorical PT invariant for the resolved conifold X to be

$$\mathcal{DT}(P_n(X,\beta)) := \mathrm{MF}(\mathcal{M}_Q^{\dagger,\theta_{\mathrm{PT}}}(n, n-\beta), w).$$

Similarly to Lemma 5.1, the categorical PT invariant recovers the numerical PT invariant by

$$P_{n,\beta} = (-1)^{n+\beta} e_{\mathbb{C}((u))}\big(\mathrm{HP}_*\big(\mathcal{DT}(P_n(X,\beta))\big)\big). \tag{5-59}$$

By applying Corollary 5.20 for $m \gg 0$, we obtain the following.

Corollary 5.24 *For any $(\beta,n)\in\mathbb{Z}^2$, there exists a semiorthogonal decomposition*

$$\mathcal{DT}(P_n(X,\beta)) = \langle \mathcal{C}_{j_\bullet^{(*)}} \rangle.$$

Here each $\mathcal{C}_{j_\bullet^{(*)}}$ is equivalent to $\mathrm{MF}(\mathrm{Spec}\,\mathbb{C},0)$ and $j_\bullet^{(*)}$ is a collection of nonpositive integers of the form

$$j_\bullet^{(*)} = \{(0 \le j_1^{(i)} \le \cdots \le j_{l_i}^{(i)} \le i - l_i)\}_{i\ge 1}$$

for some integers $l_i \ge 0$ satisfying

$$(\beta,n) = \sum_{i\ge 1} l_i\cdot(1,i).$$

We have $\mathrm{Hom}(\mathcal{C}_{j_\bullet^{(*)}}, \mathcal{C}_{j_\bullet'^{(*)}}) = 0$ if $j_\bullet^{(i)} = j_\bullet'^{(i)}$ for $i>k$ for some k and $j_\bullet^{(k)} \succ j_\bullet'^{(k)}$.

Remark 5.25 Similarly to Remark 5.19, the semiorthogonal decomposition in Corollary 5.24 implies

$$\mathrm{HP}_*(\mathcal{DT}(P_n(X,\beta))) = \mathrm{HP}_*(\mathrm{MF}(\mathrm{Spec}\,\mathbb{C},0))^{\oplus a_{n,\beta}},$$

where $a_{n,\beta}$ is given by (1-3). Taking the Euler characteristics of both sides, we obtain $P_{n,\beta} = (-1)^{n+\beta}a_{n,\beta}$, which recovers the formula (1-1).

6 Some technical lemmas

In this section, we give proofs of some postponed technical lemmas.

6.1 Functoriality of Knörrer periodicity

Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be stacks of the form $\mathcal{Y}_i = [Y_i/G_i]$, where Y_i is a smooth affine scheme and G_i is a reductive algebraic group which acts on Y_i. Let $\mathcal{W}_i \to \mathcal{Y}_i$ be vector bundles. Then by Theorem 2.4, we have equivalences

$$\Phi_i\colon \mathrm{MF}(\mathcal{Y}_i, w_i) \xrightarrow{\sim} \mathrm{MF}(\mathcal{W}_i \oplus \mathcal{W}_i^\vee, w_i + q_i), \tag{6-1}$$

where q_i is a natural quadratic form on $\mathcal{W}_i \oplus \mathcal{W}_i^\vee$, ie $q_i(x, x') = \langle x, x' \rangle$. On the other hand, the categories of quasicoherent factorizations $\mathrm{MF}_{\mathrm{qcoh}}(\mathcal{Y}_i, w_i)$ are compactly generated by $\mathrm{MF}(\mathcal{Y}_i, w_i)$ (see [Ballard et al. 2014, Proposition 3.15]), so it is equivalent to the ind-completion of $\mathrm{MF}(\mathcal{Y}_i, w_i)$. Therefore by taking ind-completions of both sides in (6-1), the above equivalences extend to equivalences

$$\Phi_i \colon \mathrm{MF}_{\mathrm{qcoh}}(\mathcal{Y}_i, w_i) \xrightarrow{\sim} \mathrm{MF}_{\mathrm{qcoh}}(\mathcal{W}_i \oplus \mathcal{W}_i^\vee, w_i + q_i).$$

Suppose that we have a commutative diagram

$$\begin{tikzcd} \mathcal{W}_1 \arrow[r, "g"] \arrow[d] & \mathcal{W}_2 \arrow[d] \\ \mathcal{Y}_1 \arrow[r, "f"] & \mathcal{Y}_2 \end{tikzcd}$$

where f is a morphism of stacks, and the top arrow is induced by a morphism of vector bundles $g \colon \mathcal{W}_1 \to f^*\mathcal{W}_2$. We have the induced diagram

$$\mathcal{W}_1 \oplus \mathcal{W}_1^\vee \xleftarrow{h_1} \mathcal{W}_1 \oplus f^*\mathcal{W}_2^\vee \xrightarrow{h_2} \mathcal{W}_2 \oplus \mathcal{W}_2^\vee, \tag{6-2}$$

where $h_1 = (\mathrm{id}_{\mathcal{W}_1}, g^\vee)$ and $h_2 = (g, f)$. The following lemma is a variant of [Toda 2019, Lemma 2.4.4].

Lemma 6.1 *The following diagram commutes*:

$$\begin{tikzcd} \mathrm{MF}_{\mathrm{qcoh}}(\mathcal{Y}_1, w_1) \arrow[r, "f_*"] \arrow[d, "\Phi_1"', "\sim"] & \mathrm{MF}_{\mathrm{qcoh}}(\mathcal{Y}_2, w_2) \arrow[d, "\sim"', "\Phi_2"] \\ \mathrm{MF}_{\mathrm{qcoh}}(\mathcal{W}_1 \oplus \mathcal{W}_1^\vee, w_1 + q_1) \arrow[r, "h_{2*}h_1^*"] & \mathrm{MF}_{\mathrm{qcoh}}(\mathcal{W}_2 \oplus \mathcal{W}_2^\vee, w_2 + q_2) \end{tikzcd} \tag{6-3}$$

Proof We have the commutative diagram

$$\begin{tikzcd} & f^*\mathcal{W}_2^\vee \arrow[r, hook, "h_4"] \arrow[rr, bend left, "h_6"] \arrow[dl, "h_5"'] \arrow[d, "h_3"'] \arrow[dr, phantom, "\square"] & \mathcal{W}_1 \oplus f^*\mathcal{W}_2^\vee \arrow[r, "h_2"] \arrow[d, "h_1"'] & \mathcal{W}_2 \oplus \mathcal{W}_2^\vee \\ \mathcal{Y}_1 & \mathcal{W}_1^\vee \arrow[l, "\mathrm{pr}_1"'] \arrow[r, hook, "i_1"] & \mathcal{W}_1 \oplus \mathcal{W}_1^\vee & \end{tikzcd}$$

Here pr_1 is the projection and $i_1(x) = (0, x)$. By the above diagram together with derived base change, we have

$$h_{2*}h_1^*\Phi_1(-) \cong h_{2*}h_1^* i_{1*}\,\mathrm{pr}_1^*(-) \cong h_{2*}h_{4*}h_3^*\,\mathrm{pr}_1^*(-) \cong h_{6*}h_5^*(-).$$

On the other hand, we have the commutative diagram

$$\begin{tikzcd} f^*\mathcal{W}_2^\vee \arrow[r, "h_7"] \arrow[rr, bend left, "h_6"] \arrow[d, "h_5"'] \arrow[dr, phantom, "\square"] & \mathcal{W}_2^\vee \arrow[r, hook, "i_2"] \arrow[d, "\mathrm{pr}_2"'] & \mathcal{W}_2 \oplus \mathcal{W}_2^\vee \\ \mathcal{Y}_1 \arrow[r, "f"] & \mathcal{Y}_2 & \end{tikzcd}$$

Here pr_2 is the projection and $i_2(x)=(0,x)$. Similarly we have

$$\Phi_2 f_*(-)\cong i_{2*}\,\mathrm{pr}_2^* f_* \cong i_{2*}h_{7*}h_5^* \cong h_{6*}h_5^*(-).$$

Therefore the diagram (6-3) commutes. □

We also have the following lemma, which is a variant of [Toda 2019, Lemma 2.4.7].

Lemma 6.2 *Suppose that $g\colon \mathcal{W}_1\to f^*\mathcal{W}_2$ is a surjective morphism of vector bundles on $\mathcal{Y}_1$. Then we have the commutative diagram*

$$\begin{tikzcd}
\mathrm{MF}(\mathcal{Y}_2,w_2) \arrow[r, "f^*"] \arrow[d, "\Phi_2"', "\sim"] & \mathrm{MF}(\mathcal{Y}_1,w_1) \arrow[d, "\sim"', "\Phi_1"] \\
\mathrm{MF}(\mathcal{W}_2\oplus\mathcal{W}_2^\vee, w_2+q_2) \arrow[r, "h_{1!}h_2^*"] & \mathrm{MF}(\mathcal{W}_1\oplus\mathcal{W}_1^\vee, w_1+q_1)
\end{tikzcd}$$

Proof The assumption that $g\colon \mathcal{W}_1\to f^*\mathcal{W}_2$ is surjective implies that the morphism h_1 in (6-2) is a closed immersion, hence $h_{1!}$ gives a left adjoint of h_1^*. The lemma now follows by taking left adjoints of horizontal arrows in (6-3) and restrict to coherent factorizations. □

6.2 The categories of factorizations on formal fibers

Let G be a reductive algebraic group and Y be a finite-dimensional G–representation. We denote by $\widehat{Y}$ the formal fiber of the quotient morphism $Y\to Y/\!\!/G$ at the origin; see Section 1.6 for the definition of formal fiber. Then

$$[\widehat{Y}/G]\to \widehat{Y}/\!\!/G=\operatorname{Spec}\widehat{\mathcal{O}}_{Y/\!\!/G,0}$$

is a good moduli space for $[\widehat{Y}/G]$, and is isomorphic to the formal fiber of the morphism $[Y/G]\to Y/\!\!/G$ at 0. We take an element $w\in\Gamma(\mathcal{O}_{[\widehat{Y}/G]})=\widehat{\mathcal{O}}_{Y/\!\!/G,0}$ with $w(0)=0$. We have the following lemma:

Lemma 6.3 *For $w\neq 0$, the triangulated category $\mathrm{MF}([\widehat{Y}/G],w)$ is idempotent complete.*

Proof Let $\widehat{Z}\subset\widehat{Y}$ be the closed subscheme defined by the zero locus of w. We have the following version of Orlov equivalence [2009] relating the categories of factorizations and those of singularities (see [Polishchuk and Vaintrob 2011, Theorem 3.14])

$$\mathrm{MF}([\widehat{Y}/G],w)\xrightarrow{\sim} D^b([\widehat{Z}/G])/\mathrm{Perf}([\widehat{Z}/G]).$$

Let $\boldsymbol{m}_0\subset\mathcal{O}_{\widehat{Z}}$ be the maximal ideal which defines $0\in\widehat{Z}$, and denote by $\widehat{\mathcal{O}}_{\widehat{Z}}$ the formal completion of $\mathcal{O}_{\widehat{Z}}$ at $\boldsymbol{m}_0$. Let $Z^{(n)}:=\operatorname{Spec}\mathcal{O}_{\widehat{Z}}/\boldsymbol{m}_0^n$ and $\overline{Z}:=\operatorname{Spec}\widehat{\mathcal{O}}_{\widehat{Z}}$. By the coherent completeness for the stacks $[\widehat{Z}/G]$ and $[\overline{Z}/G]$ (see [Alper et al. 2019, Theorem 1.6]), we have the equivalences

$$\mathrm{Coh}([\widehat{Z}/G])\xrightarrow{\sim}\varprojlim_n \mathrm{Coh}([Z^{(n)}/G])\xleftarrow{\sim}\mathrm{Coh}([\overline{Z}/G]).$$

In particular, we have an equivalence

$$D^b([\widehat{Z}/G])\xrightarrow{\sim} D^b([\overline{Z}/G]),$$

which restricts to the equivalence for subcategories of perfect objects. Therefore we obtain the equivalence

$$\mathrm{MF}([\widehat{Y}/G], w) \xrightarrow{\sim} D^b([\overline{Z}/G])/\mathrm{Perf}([\overline{Z}/G]).$$

Since $\widehat{\mathcal{O}}_{\widehat{Z}}$ is a complete local ring, the singularity category $D^b(\overline{Z})/\mathrm{Perf}(\overline{Z})$ is well-known to be idempotent complete; for example, see [Dyckerhoff 2011, Lemma 5.6; Kalck and Yang 2018, Lemma 5.5]. The argument can be easily extended to the G–equivariant setting. Indeed, following the proof of [Kalck and Yang 2018, Lemma 5.5], it is enough to show that for a G–equivariant maximal Cohen–Macaulay $\widehat{\mathcal{O}}_{\widehat{Z}}$–module M and an idempotent $e \in \underline{\mathrm{End}}^G(M)$, it is lifted to a G–invariant idempotent in $\mathrm{End}(M)$. Here $\underline{\mathrm{End}}^G(M)$ is the set of morphisms in the G–equivariant stable category of maximal Cohen–Macaulay modules over $\widehat{\mathcal{O}}_{\widehat{Z}}$. For an idempotent $e \in \underline{\mathrm{End}}^G(M)$, we lift it to $a \in \mathrm{End}(M)$, which we can assume to be G–invariant as G is reductive. Then as in the proof of [Curtis and Reiner 1981, Theorem 6.7], the limit $\widetilde{e} := \lim f_j(a)$ exists, and is an idempotent in $\mathrm{End}(M)$ which lifts e. Here $f_j(x)$ is given by

$$f_j(x) = \sum_{i=0}^{n} \binom{2n}{i} x^{2n-i}(1-x)^i.$$

By construction $\widetilde{e}$ is G–invariant, so we obtain the desired lifting property of the idempotents. □

Let W be another finite-dimensional G–representation and $q\colon W \to \mathbb{A}^1$ be a G–invariant nondegenerate quadratic form. We take $w \in \widehat{\mathcal{O}}_{Y/\!\!/G,0}$ with $w(0) = 0$. We have the following lemma:

Lemma 6.4 *There is a natural morphism of stacks*

$$\iota\colon [\widehat{(Y \oplus W)}/G] \to [(\widehat{Y} \times W)/G] \tag{6-4}$$

such that the induced functor

$$\iota^*\colon \mathrm{MF}\big([(\widehat{Y} \times W)/G], w+q\big) \to \mathrm{MF}\big([\widehat{(Y \oplus W)}/G], w+q\big) \tag{6-5}$$

is fully faithful with dense image.

Proof Let π_Y and $\pi_{Y\oplus W}$ be the quotient morphisms

$$\pi_Y\colon Y \to Y/\!\!/G, \quad \pi_{Y\oplus W}\colon Y \oplus W \to (Y \oplus W)/\!\!/G.$$

Then we have $\pi_{Y\oplus W}^{-1}(0,0) \subset \pi_Y^{-1}(0) \times W$, therefore we have the induced natural morphism (6-4) by the definition of formal fibers.

Note that we have $\mathrm{Crit}(w+q) = \mathrm{Crit}(w) \times \{0\}$, so the morphism (6-4) induces the isomorphism of critical loci of $w+q$ on $\widehat{Y} \times W$ and $\widehat{Y \oplus W}$, and also their formal neighborhoods. Therefore the functor (6-5) is fully faithful with dense image by [Orlov 2011, Theorem 2.10] (in loc. cit. it is stated without G–action, but the same argument applies verbatim to the G–equivariant setting). □

Suppose that Y is quasiprojective variety with an action of a reductive algebraic group G such that the good moduli space $\pi\colon [Y/G] \to Y/\!\!/G$ exists. For each closed point $y \in Y/\!\!/G$, we denote by $[\widehat{Y}_y/G]$ the formal fiber of π at y. For a regular function $w\colon [Y/G] \to \mathbb{A}^1$, we denote by $\widehat{w}_y$ its restriction to $[\widehat{Y}_y/G]$, and $\widehat{\pi}_y\colon [\widehat{Y}_y/G] \to \widehat{Y}_y /\!\!/ G$ its good moduli space. We have the following lemma:

Lemma 6.5 *For $\mathscr{E} \in \mathrm{MF}([Y/G], w)$, suppose that $\mathscr{E}|_{[\widehat{Y}_y/G]} \in \mathrm{MF}([\widehat{Y}_y/G], \widehat{w}_y)$ is isomorphic to zero for any closed point $y \in Y /\!\!/ G$. Then we have $\mathscr{E} \cong 0$.*

Proof The inner homomorphism $\mathscr{H}om^\bullet(\mathscr{E}, \mathscr{E})$ is an object in $\mathrm{MF}([Y/G], 0)$, which is equivalent to the $\mathbb{Z}/2$–periodic derived category of coherent sheaves on $[Y/G]$. By the derived base change, we have

$$\pi_* \mathscr{H}om^\bullet(\mathscr{E}, \mathscr{E}) \otimes_{\mathscr{O}_{Y/\!\!/ G}} \widehat{\mathscr{O}}_{Y/\!\!/ G, y} \cong \widehat{\pi}_{y*} \mathscr{H}om^\bullet(\mathscr{E}|_{[\widehat{Y}_y/G]}, \mathscr{E}|_{[\widehat{Y}_y/G]}) \cong 0$$

in the $\mathbb{Z}/2$–periodic derived category of quasicoherent sheaves on $\widehat{Y}_y /\!\!/ G$. The object $\pi_* \mathscr{H}om^\bullet(\mathscr{E}, \mathscr{E})$ is an object in the $\mathbb{Z}/2$–periodic derived category of quasicoherent sheaves on $Y /\!\!/ G$ whose formal completions at any $y \in Y /\!\!/ G$ is zero, so it is isomorphic to zero. Then we have $\mathrm{Hom}^\bullet(\mathscr{E}, \mathscr{E}) = \mathbf{R}\Gamma(\mathscr{H}om^\bullet(\mathscr{E}, \mathscr{E})) = 0$, so $\mathscr{E} \cong 0$. ◻

6.3 Right adjoint functor

Lemma 6.6 *The functor $\Upsilon_{j_\bullet}$ in (5-55) admits a right adjoint $\Upsilon_{j_\bullet}^R$.*

Proof We consider the diagram

(6-6)
$$\begin{tikzcd}
 & \mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v^\bullet) \arrow[d, "q"'] \arrow[dr, "p"] & & \\
\begin{array}{c}\mathcal{M}_Q^{\theta\text{-ss}}(s_m)^{\times l} \\ \times \mathcal{M}_Q^{\dagger,\theta_-\text{-ss}}(v - l s_m)\end{array} \arrow[r, hook] \arrow[d] & \begin{array}{c}\mathcal{M}_Q^{\theta\text{-ss}}(s_m)^{\times l} \\ \times \mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v - l s_m)\end{array} \arrow[d] & \mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v) \arrow[d] & \mathcal{M}_Q^{\dagger,\theta_+\text{-ss}}(v) \arrow[l, hook'] \arrow[d] \\
\begin{array}{c}M_Q^{\theta\text{-ss}}(s_m)^{\times l} \\ \times M_Q^{\dagger,\theta_-\text{-ss}}(v - l s_m)\end{array} \arrow[r] & \begin{array}{c}M_Q^{\theta\text{-ss}}(s_m)^{\times l} \\ \times M_Q^{\dagger,\theta\text{-ss}}(v - l s_m)\end{array} \arrow[r, "\oplus"'] & M_Q^{\dagger,\theta\text{-ss}}(v) \arrow[d, "w"'] & M_Q^{\dagger,\theta_+\text{-ss}}(v) \arrow[l] \\
 & & \mathbb{A}^1 &
\end{tikzcd}$$

Similarly to (5-36), let

$$\widetilde{\mathbb{W}}_{\mathrm{glob}}^{\theta_\pm}(v) \subset D^b(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v))$$

be the window subcategory (2-7) for the choice $m_\bullet^\pm$ in (5-35). We consider the composition functor

(6-7)
$$\begin{aligned}
D^b(M_Q^{\theta\text{-ss}}(s_m))^{\boxtimes l} \boxtimes D^b(M_Q^{\dagger,\theta\text{-ss}}(v - l s_m)) \\
\xrightarrow{\sim} \mathop{\boxtimes}_{i=1}^{l} D^b(\mathcal{M}_Q^{\theta\text{-ss}}(s_m))_{j_i + (2i-1)(m^2-m)} \boxtimes \widetilde{\mathbb{W}}_{\mathrm{glob}}^{\theta_-}(v - l s_m) \\
\to D^b(\mathcal{M}_Q^{\dagger,\theta\text{-ss}}(v)) \to D^b(\mathcal{M}_Q^{\dagger,\theta_+\text{-ss}}(v)).
\end{aligned}$$

Here the first equivalence is due to window theorem in Theorem 2.2 together with the fact that (5-9) is a $\mathbb{C}^*$–gerbe, the second arrow is the categorified Hall product (ie $p_* q^*$ in the diagram (6-6)), and the last arrow is the restriction to the semistable locus. The first arrow is of Fourier–Mukai type by Lemma 6.7 below, and the second and the third arrows are also of Fourier–Mukai type by their constructions. Therefore the above composition functor is of Fourier–Mukai type. So we have the kernel object

$$\mathcal{P} \in D^b\big((M_Q^{\theta\text{-ss}}(s_m)^{\times l} \times M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m)) \times M_Q^{\dagger,\theta_+\text{-ss}}(v)\big).$$

Moreover the kernel objects of the second and the third arrows in (6-7) are pushforwards from the fiber products over $M_Q^{\dagger,\theta\text{-ss}}(v)$ by their constructions. By Lemma 6.7 below, the kernel object of the first arrow in (6-7) is a pushforward from the fiber product over $\mathbb{A}^1$ and supported on the fiber product over $M_Q^{\dagger,\theta\text{-ss}}(v)$. Therefore the object $\mathcal{P}$ is a pushforward of an object

$$\mathcal{P}_w \in D^b\big((M_Q^{\theta\text{-ss}}(s_m)^{\times l} \times M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m)) \times_{\mathbb{A}^1} M_Q^{\dagger,\theta_+\text{-ss}}(v)\big) \tag{6-8}$$

supported on the fiber product over $M_Q^{\dagger,\theta\text{-ss}}(v)$. Since $M_Q^{\dagger,\theta_+\text{-ss}}(v)$ and $M_Q^{\theta\text{-ss}}(s_m)^{\times l} \times M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m)$ are proper over $M_Q^{\dagger,\theta\text{-ss}}(v)$, the functor (6-7) admits a right adjoint given by the Fourier–Mukai kernel $\mathcal{P}^R$ defined by

$$\mathcal{P}^R := \mathcal{P}^\vee \boxtimes \omega_{M_Q^{\theta\text{-ss}}(s_m)^{\times l} \times M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m)}[\dim M_Q^{\theta\text{-ss}}(s_m)^{\times l} \times M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m)].$$

By Theorem 2.3, the functor $\Upsilon_{j_\bullet}$ in (5-55) is regarded as a functor

$$\Upsilon_{j_\bullet} \colon \mathrm{MF}(M_Q^{\theta\text{-ss}}(s_m), w)^{\boxtimes l} \boxtimes \mathrm{MF}(M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m), w) \to \mathrm{MF}(M_Q^{\dagger,\theta_+\text{-ss}}(v), w). \tag{6-9}$$

The above functor is a Fourier–Mukai functor with kernel given by $\Xi(\mathcal{P}_w)$, where Ξ is the natural functor (see [Hirano 2017b, Theorem 5.5])

$$\begin{aligned}\Xi \colon D^b\big((M_Q^{\theta\text{-ss}}(s_m)^{\times l} \times M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m)\big) \times_{\mathbb{A}^1} M_Q^{\dagger,\theta_+\text{-ss}}(v)\big) \\ \to \mathrm{MF}\big(\big(M_Q^{\theta\text{-ss}}(s_m)^{\times l} \times M_Q^{\dagger,\theta_-\text{-ss}}(v-l s_m)\big) \times M_Q^{\dagger,\theta_+\text{-ss}}(v), w \boxplus (-w)\big)\end{aligned}$$

By the Grothendieck Riemann–Roch theorem, the object $\mathcal{P}^R$ is the pushforward of an object $\mathcal{P}_w^R$ in the right-hand side of (6-8). Then the right adjoint of (6-9) is obtained by the Fourier–Mukai kernel $\Xi(\mathcal{P}_w^R)$. □

Lemma 6.7 *In the setting of Theorem 2.2, let $\mathcal{Y} = [Y/G]$ and $\mathcal{Y}^{\mathrm{ss}} = [Y^{l\text{-ss}}/G]$, and assume that $\mathcal{Y}^{\mathrm{ss}}$ is a projective scheme over $Y/\!\!/G$. Then the splitting of $D^b(\mathcal{Y}) \twoheadrightarrow D^b(\mathcal{Y}^{\mathrm{ss}})$ in Theorem 2.2 (applied to $N'=0$) is of Fourier–Mukai type, with kernel object $\mathcal{P} \in D^b(\mathcal{Y} \times \mathcal{Y}^{\mathrm{ss}})$ supported on $\mathcal{Y} \times_{Y/\!\!/G} \mathcal{Y}^{\mathrm{ss}}$. Moreover for any nonconstant $w \colon Y/\!\!/G \to \mathbb{A}^1$, we have $\mathcal{P} = i_* \mathcal{P}_w$ for some $\mathcal{P}_w \in D^b(\mathcal{Y} \times_{\mathbb{A}^1} \mathcal{Y}^{\mathrm{ss}})$. Here $\mathcal{Y} \times_{\mathbb{A}^1} \mathcal{Y}^{\mathrm{ss}}$ is given by the diagram*

$$\begin{tikzcd} \mathcal{Y} \times_{\mathbb{A}^1} \mathcal{Y}^{\mathrm{ss}} \arrow[r, "i"] \arrow[d] & \mathcal{Y} \times \mathcal{Y}^{\mathrm{ss}} \arrow[d, "w \boxplus (-w)"] \\ 0 \arrow[r] & \mathbb{A}^1 \end{tikzcd}$$

Proof The KN stratification of $\mathcal{Y}$ pulls back to the one on $\mathcal{Y} \times \mathcal{Y}^{\mathrm{ss}}$ via the first projection, thus by a choice of $m_{\bullet}$ in Theorem 2.2 we have the splitting Ψ of $D^b(\mathcal{Y} \times \mathcal{Y}^{\mathrm{ss}}) \twoheadrightarrow D^b(\mathcal{Y}^{\mathrm{ss}} \times \mathcal{Y}^{\mathrm{ss}})$. From its construction, Ψ is linear over $\mathrm{Perf}(Y /\!\!/ G \times Y /\!\!/ G)$. Therefore for any nonconstant w, by [Halpern-Leistner 2015, Proposition 5.5] there is a splitting Φ_w of $D^b(\mathcal{Y} \times_{\mathbb{A}^1} \mathcal{Y}^{\mathrm{ss}}) \twoheadrightarrow D^b(\mathcal{Y}^{\mathrm{ss}} \times_{\mathbb{A}^1} \mathcal{Y}^{\mathrm{ss}})$ such that the following diagram commutes:

$$\begin{tikzcd} D^b(\mathcal{Y}^{\mathrm{ss}} \times_{\mathbb{A}^1} \mathcal{Y}^{\mathrm{ss}}) \arrow[r, "\Phi_w"] \arrow[d, "i_*"'] & D^b(\mathcal{Y} \times_{\mathbb{A}^1} \mathcal{Y}^{\mathrm{ss}}) \arrow[d, "i_*"'] \\ D^b(\mathcal{Y}^{\mathrm{ss}} \times \mathcal{Y}^{\mathrm{ss}}) \arrow[r, "\Phi"] & D^b(\mathcal{Y} \times \mathcal{Y}^{\mathrm{ss}}) \end{tikzcd}$$

Since $\mathcal{Y}^{\mathrm{ss}}$ is a quasiprojective scheme, we have $\mathcal{O}_\Delta \in D^b(\mathcal{Y}^{\mathrm{ss}} \times \mathcal{Y}^{\mathrm{ss}})$. We set $\mathcal{P} = \Phi(\mathcal{O}_\Delta)$ and $\mathcal{P}_w = \Phi_w(\mathcal{O}_\Delta)$. Then $\mathcal{P} = i_* \mathcal{P}_w$. Since this holds for any w, the object $\mathcal{P}$ is supported on $\mathcal{Y} \times_{Y /\!\!/ G} \mathcal{Y}^{\mathrm{ss}}$. Then the object $\mathcal{P}$ induces the Fourier–Mukai functor $D^b(\mathcal{Y}^{\mathrm{ss}}) \to D^b(\mathcal{Y})$ which gives the splitting in Theorem 2.2 by the argument in [Halpern-Leistner 2015, Section 2.3]. □

6.4 Proof of Proposition 5.11

Proof The assertion is trivial if $\dim V \leq 1$. Below we assume that $\dim V \geq 2$. Note that $\mathrm{ord}_0(w_p) \geq 2$, where $\mathrm{ord}_0(w_p)$ is the vanishing order of w_p at 0. This is because $w_p(0) = 0$ by the first inclusion in (5-4) together with the fact that $0 \in \mathrm{Crit}(w_p) \neq \varnothing$.

Let us consider the Hessian of w_p,

$$\mathrm{Hess}(w_p)\colon \mathrm{Ext}^1_{Q^\dagger}(R, R) \otimes \mathcal{O}_{\widehat{\mathcal{M}}^\dagger_{Q_p}(d)} \to \mathrm{Ext}^1_{Q^\dagger}(R, R)^\vee \otimes \mathcal{O}_{\widehat{\mathcal{M}}^\dagger_{Q_p}(d)}.$$

The kernel of the above morphism at the origin is $\mathrm{Ext}^1_{(Q^\dagger, W)}(R, R)$. By the relation (5-8), we have

$$\mathrm{Ext}^1_{(Q,W)}(S_m, S_m) = \mathrm{Ext}^1_X(\mathcal{O}_C(m-1), \mathcal{O}_C(m-1)) = 0.$$

It follows that

$$\mathrm{Ker}(\mathrm{Hess}(w_p)|_0) \cap (\mathrm{End}(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m)) = 0. \tag{6-10}$$

By Lemma 6.8 below, by replacing the isomorphism η_p in (5-19) if necessary, there exist linear subspaces

$$W_1 \subset \mathrm{Ext}^1_{Q^\dagger}(R_\infty, R_\infty), \quad W_2 \subset \mathrm{Ext}^1_{Q^\dagger}(R_\infty, S_m), \quad W_3 \subset \mathrm{Ext}^1_{Q^\dagger}(S_m, R_\infty)$$

such that $w_p = w_1 + w_2$, where w_1 does not contain variables from $\mathrm{End}(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m)$ with $\deg(w_1) \geq 3$, and w_2 is a nondegenerate G–invariant quadratic form on

$$\begin{aligned} &W_1 \oplus (W_2 \otimes V) \oplus (W_3 \otimes V^\vee) \oplus (\mathrm{End}(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m)) \\ &\qquad = (W_1 \oplus \mathrm{Ext}_Q(S_m, S_m)) \oplus (W_2 \otimes V) \oplus (W_3 \otimes V^\vee) \oplus (\mathrm{End}_0(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m)). \end{aligned}$$

As we assumed that $\dim V \geq 2$, the $\mathrm{GL}(V)$–representation $\mathrm{End}_0(V)$ is a nontrivial irreducible $\mathrm{GL}(V)$–representation, and it is not isomorphic to V nor $V^\vee$. Therefore $w_2 = w_3 + q$, where w_3 does not contain variables from $\mathrm{End}_0(V) \otimes \mathrm{Ext}^1_Q(S_m, S_m)$ and q is a nondegenerate $\mathrm{GL}(V)$–invariant quadratic form on $\mathrm{End}_0(V)\otimes \mathrm{Ext}^1_Q(S_m,S_m)$. Moreover, w_3 is nonzero, since otherwise it contradicts with (6-10) and $\mathrm{End}_0(V)\subsetneq \mathrm{End}(V)$. By replacing the isomorphism (5-25) if necessary, we can also assume that q coincides with (5-26). Therefore we obtain a desired form (5-27). $\square$

We have used the following lemma, whose proof is a variant of [Joyce 2015, Proposition 2.24]:

Lemma 6.8 *Let G be a reductive algebraic group and V be a finite-dimensional G–representation. Let $w\colon \widehat{V}\to \mathbb{A}^1$ be a G–invariant formal function such that $\mathrm{ord}_0(w)\geq 2$. Let V_1 be the kernel of the Hessian at the origin*

$$V_1=\mathrm{Ker}\big(\mathrm{Hess}(w)|_0\colon V\to V^{\vee}\big).$$

Then there exists a direct sum decomposition $V=V_1\oplus V_2$ of G–representations and a G–equivariant isomorphism $\phi\colon \widehat{V}\xrightarrow{\cong}\widehat{V}$ such that $\phi^*w=w_1+w_2$, where $w_1\in \mathcal{O}_{\widehat{V}_1}$ is G–invariant with $\mathrm{ord}_0(w_1)\geq 3$, and $w_2\in \mathrm{Sym}^2(V_2^{\vee})$ is a G–invariant nondegenerate quadratic form on V_2.

Proof As w is G–invariant, the Hessian of w at the origin $\mathrm{Hess}(w)|_0\colon V\to V^{\vee}$ is G–equivariant. As G is reductive, there is a splitting $V=V_1\oplus V_2$ as G–representations and the Hessian at the origin is written as

$$\mathrm{Hess}(w)|_0=\begin{pmatrix}0&0\\0&q\end{pmatrix}\colon V_1\oplus V_2\to V_1^{\vee}\oplus V_2^{\vee},$$

where q is a G–equivariant isomorphism $q\colon V_2\xrightarrow{\cong} V_2^{\vee}$ with $q^{\vee}=q$. We identify q as an element $q\in \mathrm{Sym}^2(V_2^{\vee})^G$, which is a G–invariant nondegenerate quadratic form q on V_2. For $(y_1,y_2)\in V_1\oplus V_2$, we can write $w(y_1,y_2)$ as

$$w(y_1,y_2)=w^{\geq 3}(y_1,y_2)+q(y_2), \tag{6-11}$$

where $w^{\geq 3}(y_1,y_2)$ consists of terms with degrees bigger than or equal to three. We set $d_i=\dim V_i$ and fix bases of V_1 and V_2, writing their elements as $y_1=\{y_1^{(i)}\}_{1\leq i\leq d_1}$ and $y_2=\{y_2^{(i)}\}_{1\leq i\leq d_2}$, respectively. Here we take an orthonormal basis for V_2, so q is written as

$$q(y_2)=\frac{1}{2}\sum_{i=1}^{d_1}(y_2^{(i)})^2.$$

Then the closed subscheme

$$\left\{\frac{\partial w}{\partial y_2^{(i)}}=0:1\leq i\leq d_2\right\}=\left\{y_2^{(i)}+\frac{\partial w^{\geq 3}}{\partial y_2^{(i)}}=0:1\leq i\leq d_2\right\}\subset \widehat{V}$$

is smooth of codimension d_2. By the variable change

$$y_2^{(i)}\mapsto \frac{\partial w}{\partial y_2^{(i)}}=y_2^{(i)}+\frac{\partial w^{\geq 3}}{\partial y_2^{(i)}}, \tag{6-12}$$

we may assume that $\mathrm{Crit}(w)$ is contained in $\{y_2=0\}\subset \widehat{V}$. The variable change (6-12) can be described without coordinates as follows. Let dw be the morphism given by the derivation of w,

$$dw\colon V\otimes \mathcal{O}_{\widehat{V}}\to \mathcal{O}_{\widehat{V}}. \tag{6-13}$$

We have the morphisms

$$\phi\colon V^\vee = V_1^\vee \oplus V_2^\vee \xrightarrow{(\mathrm{id},q^{-1})} V_1^\vee \oplus V_2 \xrightarrow{(\mathrm{id},dw|_{V_2})} \widehat{\mathcal{O}}_V.$$

The above composition induces the isomorphism $\mathcal{O}_{\widehat{V}} \xrightarrow{\cong} \mathcal{O}_{\widehat{V}}$, which is identified with the variable change (6-12). The above construction is G–equivariant, so the variable change (6-12) is G–equivariant.

The condition that $\mathrm{Crit}(w) \subset \{y_2 = 0\}$ implies that each $y_2^{(i)}$ is written as

$$y_2^{(i)} = \sum_{j=1}^{d_1} a_{ij} \frac{\partial w}{\partial y_1^{(j)}} + \sum_{j=1}^{d_2} b_{ij} \frac{\partial w}{\partial y_2^{(j)}}$$

for some $a_{ij}, b_{ij} \in \mathcal{O}_{\widehat{V}}$. By writing $b_{ij} = b_{ij}(0) + b_{ij}^{\geq 1}$ and comparing the degree-one terms for y_2, we see that $b_{ij}(0) = \delta_{ij}$. Therefore we obtain the relation

$$-\frac{\partial w^{\geq 3}}{\partial y_2^{(i)}} = \sum_{j=1}^{d_1} a_{ij} \frac{\partial w^{\geq 3}}{\partial y_1^{(i)}} + \sum_{j=1}^{d_2} b_{ij}^{\geq 1} \biggl(y_2^{(j)} + \frac{\partial w^{\geq 3}}{\partial y_2^{(j)}} \biggr).$$

The Nakayama lemma implies the inclusion of ideals

$$\biggl(\frac{\partial w^{\geq 3}}{\partial y_2^{(i)}} : 1 \leq i \leq d_2 \biggr) \subset \biggl(\frac{\partial w^{\geq 3}}{\partial y_1^{(j)}}, y_2^{(i)} : 1 \leq j \leq d_1,\ 1 \leq i \leq d_2 \biggr) \tag{6-14}$$

in $\widehat{\mathcal{O}}_{\widehat{V}}$, the formal completion at the maximal ideal of $\mathcal{O}_{\widehat{V}}$. Since these are G–invariant ideals, by the coherent completeness of $[\widehat{V}/G]$ the inclusion (6-14) also holds in $\mathcal{O}_{\widehat{V}}$; see the proof of Lemma 6.3. In particular, there is a relation of the form

$$\frac{\partial w}{\partial y_2^{(i)}}\bigg|_{y_2=0} = \sum_{i,j} c_{ij} \frac{\partial w}{\partial y_1^{(j)}}\bigg|_{y_2=0} \tag{6-15}$$

for some $c_{ij} \in \mathcal{O}_{\widehat{V}_1}$. We apply the variable change

$$\tilde{y}_1^{(i)} = y_1^{(i)} + \sum_j c_{ij} y_2^{(i)} \quad \text{and} \quad \tilde{y}_2^{(i)} = y_2^{(i)}. \tag{6-16}$$

Then we have

$$\frac{\partial w}{\partial \tilde{y}_2^{(i)}}\bigg|_{\tilde{y}_2=0} = \biggl(\sum_j \frac{\partial y_1^{(j)}}{\partial \tilde{y}_2^{(i)}} \frac{\partial w}{\partial y_1^{(j)}} + \sum_j \frac{\partial y_2^{(j)}}{\partial \tilde{y}_2^{(i)}} \frac{\partial w}{\partial y_2^{(j)}} \biggr)\bigg|_{\tilde{y}_2=0} = -\sum_j c_{ij} \frac{\partial w}{\partial y_1^{(j)}}\bigg|_{\tilde{y}_2=0} + \frac{\partial w}{\partial y_2^{(i)}}\bigg|_{\tilde{y}_2=0} = 0.$$

It follows that we can assume that $(\partial w/\partial y_2^{(i)})|_{y_2=0} = 0$.

We see that the variable change (6-16) can be taken to be G–equivariant. For the morphism (6-13), we can write $dw \otimes \mathcal{O}_{\widehat{V}_1}$ as

$$dw \otimes \mathcal{O}_{\widehat{V}_1} = \alpha^{(1)} \oplus \alpha^{(2)} \colon (V_1 \otimes \mathcal{O}_{\widehat{V}_1}) \oplus (V_2 \otimes \mathcal{O}_{\widehat{V}_1}) \to \mathcal{O}_{\widehat{V}_1}.$$

Then the ideals of $\mathcal{O}_{\widehat{V}_1}$

$$I_1 = \left(\frac{\partial w}{\partial y_1^{(i)}} \bigg|_{y_2=0} \right) \quad \text{and} \quad I_2 = \left(\frac{\partial w}{\partial y_2^{(i)}} \bigg|_{y_2=0} \right)$$

are generated by the images of $\alpha^{(1)}$ and $\alpha^{(2)}$, respectively, so in particular they are G–invariant. By the relation (6-15) we have $I_2 \subset I_1$. We have the G–equivariant diagram

$$\begin{tikzcd} V_2 \otimes \mathcal{O}_{\widehat{V}_1} \arrow[r, "\alpha^{(2)}"] \arrow[d, dotted, "\phi"'] & I_2 \arrow[d, hook] \\ V_1 \otimes \mathcal{O}_{\widehat{V}_1} \arrow[r, "\alpha^{(1)}"] & I_1 \end{tikzcd}$$

where each horizontal arrow is a surjection. As G is reductive, from the above diagram there is a G–equivariant dotted arrow ϕ which makes the above diagram commutative. A choice of ϕ corresponds to a choice of c_{ij} in (6-15). Then we have the G–equivariant morphism

$$V^{\vee} = V_1^{\vee} \oplus V_2^{\vee} \xrightarrow{(\mathrm{id}+\phi^{\vee},\mathrm{id})} \mathcal{O}_{\widehat{V}}.$$

The above morphism induces the G–equivariant isomorphism $\widehat{\mathcal{O}}_V \xrightarrow{\cong} \widehat{\mathcal{O}}_V$, which corresponds to the variable change (6-16). In particular, we can choose c_{ij} so that (6-16) is G–equivariant.

Finally, we set

$$g(y_1, y_2) := w(y_1, y_2) - w(y_1, 0).$$

Then from the above arguments we have $g(y_1, 0) = 0$ and $(\partial g/\partial y_2^{(i)})|_{y_2=0} = 0$. It follows that $g(y_1, y_2)$ is written as

$$g(y_1, y_2) = \sum_{i,j} y_2^{(i)} y_2^{(j)} Q_{ij}(y_1, y_2)$$

for some $Q_{ij} \in \mathcal{O}_{\widehat{V}}$. As the quadratic term of $g(y_1, y_2)$ coincides with q by (6-11), we have $Q_{ij}(0) = \frac{1}{2}\delta_{ij}$. It follows that the critical locus of $g(y_1, y_2)$ is $\{y_2 = 0\} \subset \widehat{V}$, so the G–equivariant Morse lemma (see [Arnold et al. 1985, Section 17.3]) applied for g implies that by a G–equivariant variable change of the form $\tilde{y}_1^{(i)} = y_1^{(i)}$ and $\tilde{y}_2^{(i)} = \sum_{i,j} \alpha^{(ij)}(y_1, y_2) y_2^{(j)}$ we can make $g(\tilde{y}_1, \tilde{y}_2) = q(\tilde{y}_2)$. As $\mathrm{ord}_0(w(y_1, 0)) \geq 3$ from (6-11), the lemma is proved. □

References

[Alper 2013] **J Alper**, *Good moduli spaces for Artin stacks*, Ann. Inst. Fourier (Grenoble) 63 (2013) 2349–2402 MR Zbl

[Alper et al. 2019] **J Alper**, **J Hall**, **D Rydh**, *The étale local structure of algebraic stacks*, preprint (2019) arXiv 1912.06162

[Arnold et al. 1985] **V I Arnold**, **S M Guseĭn-Zade**, **A N Varchenko**, *Singularities of differentiable maps, I: The classification of critical points, caustics and wave fronts*, Monogr. Math. 82, Birkhäuser, Boston, MA (1985) MR Zbl

[Ballard et al. 2014] **M Ballard**, **D Favero**, **L Katzarkov**, *A category of kernels for equivariant factorizations and its implications for Hodge theory*, Publ. Math. Inst. Hautes Études Sci. 120 (2014) 1–111 MR Zbl

[Ballard et al. 2019] **M Ballard**, **D Favero**, **L Katzarkov**, *Variation of geometric invariant theory quotients and derived categories*, J. Reine Angew. Math. 746 (2019) 235–303 MR Zbl

[Ballard et al. 2021] **M R Ballard**, **N K Chidambaram**, **D Favero**, **P K McFaddin**, **R R Vandermolen**, *Kernels for Grassmann flops*, J. Math. Pures Appl. 147 (2021) 29–59 MR Zbl

[Behrend 2009] **K Behrend**, *Donaldson–Thomas type invariants via microlocal geometry*, Ann. of Math. 170 (2009) 1307–1338 MR Zbl

[Behrend and Bryan 2007] **K Behrend**, **J Bryan**, *Super-rigid Donaldson–Thomas invariants*, Math. Res. Lett. 14 (2007) 559–571 MR Zbl

[Bondal and Orlov 1995] **A Bondal**, **D Orlov**, *Semiorthogonal decomposition for algebraic varieties*, preprint (1995) arXiv alg-geom/9506012

[Brav et al. 2019] **C Brav**, **V Bussi**, **D Joyce**, *A Darboux theorem for derived schemes with shifted symplectic structure*, J. Amer. Math. Soc. 32 (2019) 399–443 MR Zbl

[Bridgeland 2002] **T Bridgeland**, *Flops and derived categories*, Invent. Math. 147 (2002) 613–632 MR Zbl

[Bridgeland 2011] **T Bridgeland**, *Hall algebras and curve-counting invariants*, J. Amer. Math. Soc. 24 (2011) 969–998 MR Zbl

[Brion 2012] **M Brion**, *Representations of quivers*, from "Geometric methods in representation theory, I" (M Brion, editor), Sémin. Congr. 24, Soc. Math. France, Paris (2012) 103–144 MR Zbl

[Brown 2016] **M K Brown**, *Knörrer periodicity and Bott periodicity*, Doc. Math. 21 (2016) 1459–1501 MR Zbl

[Calabrese 2016] **J Calabrese**, *Donaldson–Thomas invariants and flops*, J. Reine Angew. Math. 716 (2016) 103–145 MR Zbl

[Cao and Toda 2023] **Y Cao**, **Y Toda**, *Counting perverse coherent systems on Calabi–Yau 4–folds*, Math. Ann. 385 (2023) 1379–1429 MR Zbl

[Curtis and Reiner 1981] **C W Curtis**, **I Reiner**, *Methods of representation theory*, *I*: *With applications to finite groups and orders*, Wiley, New York (1981) MR Zbl

[Davison 2017] **B Davison**, *The critical CoHA of a quiver with potential*, Q. J. Math. 68 (2017) 635–703 MR Zbl

[Donovan and Segal 2014] **W Donovan**, **E Segal**, *Window shifts, flop equivalences and Grassmannian twists*, Compos. Math. 150 (2014) 942–978 MR Zbl

[Dyckerhoff 2011] **T Dyckerhoff**, *Compact generators in categories of matrix factorizations*, Duke Math. J. 159 (2011) 223–274 MR Zbl

[Efimov 2018] **A I Efimov**, *Cyclic homology of categories of matrix factorizations*, Int. Math. Res. Not. 2018 (2018) 3834–3869 MR Zbl

[Efimov and Positselski 2015] **A I Efimov**, **L Positselski**, *Coherent analogues of matrix factorizations and relative singularity categories*, Algebra Number Theory 9 (2015) 1159–1292 MR Zbl

[Halpern-Leistner 2015] **D Halpern-Leistner**, *The derived category of a GIT quotient*, J. Amer. Math. Soc. 28 (2015) 871–912 MR Zbl

[Halpern-Leistner and Pomerleano 2020] **D Halpern-Leistner**, **D Pomerleano**, *Equivariant Hodge theory and noncommutative geometry*, Geom. Topol. 24 (2020) 2361–2433 MR Zbl

[Halpern-Leistner and Sam 2020] **D Halpern-Leistner**, **S V Sam**, *Combinatorial constructions of derived equivalences*, J. Amer. Math. Soc. 33 (2020) 735–773 MR Zbl

[Hirano 2017a] **Y Hirano**, *Derived Knörrer periodicity and Orlov's theorem for gauged Landau–Ginzburg models*, Compos. Math. 153 (2017) 973–1007 MR Zbl

[Hirano 2017b] **Y Hirano**, *Equivalences of derived factorization categories of gauged Landau–Ginzburg models*, Adv. Math. 306 (2017) 200–278 MR Zbl

[Jiang 2021] **Q Jiang**, *Derived categories of Quot schemes of locally free quotients*, preprint (2021) arXiv 2107.09193

[Joyce 2015] **D Joyce**, *A classical model for derived critical loci*, J. Differential Geom. 101 (2015) 289–367 MR Zbl

[Joyce and Song 2012] **D Joyce**, **Y Song**, *A theory of generalized Donaldson–Thomas invariants*, Mem. Amer. Math. Soc. 1020, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[Kalck and Yang 2018] **M Kalck**, **D Yang**, *Relative singularity categories*, *II*: *DG models*, preprint (2018) arXiv 1803.08192

[Kapranov 1984] **M M Kapranov**, *Derived category of coherent sheaves on Grassmann manifolds*, Izv. Akad. Nauk SSSR Ser. Mat. 48 (1984) 192–202 MR Zbl In Russian; translated in Math. USSR-Izv. 24 (1985) 183–192

[Kawamata 2002] **Y Kawamata**, *D–equivalence and K–equivalence*, J. Differential Geom. 61 (2002) 147–171 MR Zbl

[Kawamata 2018] **Y Kawamata**, *Birational geometry and derived categories*, from "Surveys in differential geometry" (H-D Cao, J Li, R M Schoen, S-T Yau, editors), Surv. Differ. Geom. 22, International, Somerville, MA (2018) 291–317 MR Zbl

[Keller 1999] **B Keller**, *On the cyclic homology of exact categories*, J. Pure Appl. Algebra 136 (1999) 1–56 MR Zbl

[King 1994] **A D King**, *Moduli of representations of finite-dimensional algebras*, Q. J. Math. 45 (1994) 515–530 MR Zbl

[Kinjo ≥ 2024] **T Kinjo**, *Cohomological DT/PT correspondence for the resolved conifold*, in preparation

[Kontsevich and Soibelman 2008] **M Kontsevich**, **Y Soibelman**, *Stability structures*, *motivic Donaldson–Thomas invariants and cluster transformations*, preprint (2008) arXiv 0811.2435

[Kontsevich and Soibelman 2011] **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

[Koseki 2021] **N Koseki**, *Birational geometry of moduli spaces of perverse coherent sheaves on blow-ups*, Math. Z. 299 (2021) 2379–2404 MR Zbl

[Maulik et al. 2006] **D Maulik**, **N Nekrasov**, **A Okounkov**, **R Pandharipande**, *Gromov–Witten theory and Donaldson–Thomas theory, I*, Compos. Math. 142 (2006) 1263–1285 MR Zbl

[Meinhardt and Reineke 2019] **S Meinhardt**, **M Reineke**, *Donaldson–Thomas invariants versus intersection cohomology of quiver moduli*, J. Reine Angew. Math. 754 (2019) 143–178 MR Zbl

[Morrison et al. 2012] **A Morrison**, **S Mozgovoy**, **K Nagao**, **B Szendrői**, *Motivic Donaldson–Thomas invariants of the conifold and the refined topological vertex*, Adv. Math. 230 (2012) 2065–2093 MR Zbl

[Nagao and Nakajima 2011] **K Nagao**, **H Nakajima**, *Counting invariant of perverse coherent sheaves and its wall-crossing*, Int. Math. Res. Not. 2011 (2011) 3885–3938 MR Zbl

[Nakajima and Yoshioka 2011] **H Nakajima**, **K Yoshioka**, *Perverse coherent sheaves on blow-up, I: A quiver description*, from "Exploring new structures and natural constructions in mathematical physics" (K Hasegawa, T Hayashi, S Hosono, Y Yamada, editors), Adv. Stud. Pure Math. 61, Math. Soc. Japan, Tokyo (2011) 349–386 MR Zbl

[Orlov 2006] **D O Orlov**, *Triangulated categories of singularities, and equivalences between Landau–Ginzburg models*, Mat. Sb. 197 (2006) 117–132 MR Zbl In Russian; translated in Sb. Math. 197 (2006) 1827–1840

[Orlov 2009] **D Orlov**, *Derived categories of coherent sheaves and triangulated categories of singularities*, from "Algebra, arithmetic, and geometry, II" (Y Tschinkel, Y Zarhin, editors), Progr. Math. 270, Birkhäuser, Boston, MA (2009) 503–531 MR Zbl

[Orlov 2011] **D Orlov**, *Formal completions and idempotent completions of triangulated categories of singularities*, Adv. Math. 226 (2011) 206–217 MR Zbl

[Orlov 2012] **D Orlov**, *Matrix factorizations for nonaffine LG-models*, Math. Ann. 353 (2012) 95–108 MR Zbl

[Pandharipande and Thomas 2009] **R Pandharipande**, **R P Thomas**, *Curve counting via stable pairs in the derived category*, Invent. Math. 178 (2009) 407–447 MR Zbl

[Pantev et al. 2013] **T Pantev**, **B Toën**, **M Vaquié**, **G Vezzosi**, *Shifted symplectic structures*, Publ. Math. Inst. Hautes Études Sci. 117 (2013) 271–328 MR Zbl

[Polishchuk and Vaintrob 2011] **A Polishchuk**, **A Vaintrob**, *Matrix factorizations and singularity categories for stacks*, Ann. Inst. Fourier (Grenoble) 61 (2011) 2609–2642 MR Zbl

[Porta and Sala 2023] **M Porta**, **F Sala**, *Two-dimensional categorified Hall algebras*, J. Eur. Math. Soc. 25 (2023) 1113–1205 MR Zbl

[Pădurariu 2019] **T Pădurariu**, *K–theoretic Hall algebras for quivers with potential*, preprint (2019) arXiv 1911.05526

[Pădurariu 2023] **T Pădurariu**, *Categorical and K–theoretic Hall algebras for quivers with potential*, J. Inst. Math. Jussieu 22 (2023) 2717–2747 MR Zbl

[Pădurariu 2024] **T Pădurariu**, *Generators for K–theoretic Hall algebras of quivers with potential*, Selecta Math. 30 (2024) art. id. 4 MR Zbl

[Pădurariu and Toda 2022] **T Pădurariu**, **Y Toda**, *Categorical and K–theoretic Donaldson–Thomas theory of $\mathbb{C}^3$, I* (2022) arXiv 2207.01899 To appear in Duke Math. J.

[Špenko and Van den Bergh 2017] **Š Špenko**, **M Van den Bergh**, *Non-commutative resolutions of quotient singularities for reductive groups*, Invent. Math. 210 (2017) 3–67 MR Zbl

[Stoppa and Thomas 2011] **J Stoppa**, **R P Thomas**, *Hilbert schemes and stable pairs: GIT and derived category wall crossings*, Bull. Soc. Math. France 139 (2011) 297–339 MR Zbl

[Szendrői 2008] **B Szendrői**, *Non-commutative Donaldson–Thomas invariants and the conifold*, Geom. Topol. 12 (2008) 1171–1202 MR Zbl

[Tabuada 2005] **G Tabuada**, *Une structure de catégorie de modèles de Quillen sur la catégorie des dg-catégories*, C. R. Math. Acad. Sci. Paris 340 (2005) 15–19 MR Zbl

[Toda 2010] **Y Toda**, *Curve counting theories via stable objects, I: DT/PT correspondence*, J. Amer. Math. Soc. 23 (2010) 1119–1157 MR Zbl

[Toda 2013] **Y Toda**, *Curve counting theories via stable objects, II: DT/ncDT flop formula*, J. Reine Angew. Math. 675 (2013) 1–51 MR Zbl

[Toda 2019] **Y Toda**, *Categorical Donaldson–Thomas theory for local surfaces*, preprint (2019) arXiv 1907.09076

[Toda 2021] **Y Toda**, *Semiorthogonal decompositions for categorical Donaldson–Thomas theory via Θ–stratifications*, preprint (2021) arXiv 2106.05496

[Toda 2022] **Y Toda**, *Birational geometry for d-critical loci and wall-crossing in Calabi–Yau 3–folds*, Algebr. Geom. 9 (2022) 513–573 MR Zbl

[Van den Bergh 2004] **M Van den Bergh**, *Three-dimensional flops and noncommutative rings*, Duke Math. J. 122 (2004) 423–455 MR Zbl

Kavli Institute for the Physics and Mathematics of the Universe (WPI), University of Tokyo
Kashiwa, Japan

yukinobu.toda@ipmu.jp

Proposed: Jim Bryan
Seconded: Mark Gross, Richard P Thomas

Received: 20 January 2022
Revised: 14 August 2022

Geometry & Topology 28:3 (2024) 1409–1436
DOI: 10.2140/gt.2024.28.1409
Published: 10 May 2024

Nonnegative Ricci curvature, metric cones and virtual abelianness

JIAYIN PAN

Let M be an open n–manifold with nonnegative Ricci curvature. We prove that if its escape rate is not $\frac{1}{2}$ and its Riemannian universal cover is conic at infinity (that is, every asymptotic cone (Y, y) of the universal cover is a metric cone with vertex y), then $\pi_1(M)$ contains an abelian subgroup of finite index. If in addition the universal cover has Euclidean volume growth of constant at least L, we can further bound the index by a constant $C(n, L)$.

53C20, 53C23; 53C21, 57S30

1 Introduction

We study the virtual abelianness/nilpotency of fundamental groups of open (complete and noncompact) manifolds with $\mathrm{Ric} \geq 0$. According to the work of Kapovitch and Wilking [8], these fundamental groups always have nilpotent subgroups with index at most $C(n)$; also see Gromov [7] and Milnor [9]. In general, these fundamental groups may not contain any abelian subgroups with finite index, because Wei [16] has constructed examples with torsion-free nilpotent fundamental groups. This is different from manifolds with $\sec \geq 0$, whose fundamental groups are always virtually abelian; see Cheeger and Gromoll [4].

A question raised from here is, for an open manifold M with $\mathrm{Ric} \geq 0$, under what conditions is $\pi_1(M)$ virtually abelian? To answer this question, one naturally looks for indications from the geometry of nonnegative sectional curvature. Ideally, if the manifold M fulfills some geometric conditions modeled on nonnegative sectional curvature, even in a much weaker form, then $\pi_1(M)$ may turn out to be virtually abelian. In other words, when $\pi_1(M)$ is not virtually abelian, some aspects of M should be drastically different from the geometry of nonnegative sectional curvature.

We have explored this direction in [12; 13], from the viewpoint of escape rate. Recall that each element γ in $\pi_1(M, p)$ can be represented by a geodesic loop at p, denoted by c_γ, with the minimal length in its homotopy class. If M has $\sec \geq 0$, then all these representing loops must stay in a bounded ball; however, this property in general does not hold for nonnegative Ricci curvature. The escape rate measures how fast these loops escape from bounded balls:

$$E(M, p) := \limsup_{|\gamma| \to \infty} \frac{d_H(p, c_\gamma)}{|\gamma|},$$

where $|\gamma|$ is the length of c_γ and d_H is the Hausdorff distance. As the main result of [13], if M satisfies $\mathrm{Ric} \geq 0$ and $E(M,p) \leq \epsilon(n)$, then $\pi_1(M)$ is virtually abelian. We also mention that from the definition, the escape rate always takes values between 0 and $\frac{1}{2}$. To the author's best knowledge, all known examples of open manifolds with $\mathrm{Ric} \geq 0$ have escape rate strictly less than $\frac{1}{2}$.

In this paper, we study how virtual abelianness/nilpotency is related to conic asymptotic geometry. Recall that an asymptotic cone of M is the pointed Gromov–Hausdorff limit of a sequence

$$(r_i^{-1}M, p) \xrightarrow{\mathrm{GH}} (Y, y),$$

where $r_i \to \infty$. We say that M is *conic at infinity* if any asymptotic cone (Y, y) of M is a metric cone with vertex y. We do not assume the asymptotic cone to be unique in this definition. If the manifold has $\sec \geq 0$, then its asymptotic cone (Y, y) is unique as a metric cone with vertex y. If the manifold has $\mathrm{Ric} \geq 0$ and Euclidean volume growth, then it is conic at infinity; see Cheeger and Colding [1].

We state the main result of this paper.

Theorem 1.1 *Let (M,p) be an open n–manifold with* $\mathrm{Ric} \geq 0$ *and $E(M,p) \neq \frac{1}{2}$.*

(1) *If its Riemannian universal cover is conic at infinity, then $\pi_1(M)$ is virtually abelian.*

(2) *If its Riemannian universal cover has Euclidean volume growth of constant at least L, then $\pi_1(M)$ has an abelian subgroup of index at most $C(n,L)$, a constant only depending on n and L.*

The contrapositive of Theorem 1.1(1) shows that the nilpotency of $\pi_1(M)$ leads to asymptotic geometry of the universal cover that is very different from the one with nonnegative sectional curvature; also see Conjecture 1.3.

Before proceeding further, we make some comments about the conditions in Theorem 1.1.

We emphasize that in Theorem 1.1(1), the conic at infinity condition is imposed on the Riemannian universal cover of M, not M itself. In fact, Wei's example [16] has the half-line $([0,\infty), 0)$ as the unique asymptotic cone of M. Therefore, in general, virtual abelianness does not hold when M is conic at infinity.

Regarding the condition $E(M,p) \neq \frac{1}{2}$ in Theorem 1.1, it is unclear to the author whether it can be dropped. On the one hand, at present we do not know any examples with $\mathrm{Ric} \geq 0$ and $E(M,p) = \frac{1}{2}$. On the other hand, we are unable to show $E(M,p) \neq \frac{1}{2}$ even when $\pi_1(M) = \mathbb{Z}$.

Question 1.2 Let (M,p) be an open n–manifold with $\mathrm{Ric} \geq 0$ and a finitely generated fundamental group. Is it true that $E(M,p) < \frac{1}{2}$?

The converse of Question 1.2 is known to be true: if $E(M,p) < \frac{1}{2}$, then $\pi_1(M)$ is finitely generated; see Sormani [15, Lemma 5]. We believe that answering Question 1.2 will lead to a better understanding of the Milnor conjecture [9]: the fundamental group of any open n–manifold with $\mathrm{Ric} \geq 0$ is finitely generated.

We compare Theorem 1.1 with previous results. In [11; 10] we have shown that if the universal cover is conic at infinity and satisfies certain stability conditions (for example, the universal cover has a unique asymptotic cone), then $\pi_1(M)$ is finitely generated and virtually abelian; in fact, these manifolds have zero escape rate [12, Corollary 4.7]. In contrast, here we do not assume any additional stability conditions in Theorem 1.1(1), and the escape rate may not be equal or close to 0 in general. In Theorem 1.1(2), if L is sufficiently close to 1, then the universal cover fulfills the stability condition in [10] and thus $E(M,p)=0$. Theorem 1.1(2) also confirms [10, Conjecture 0.2] on the condition $E(M,p)\neq\frac{1}{2}$.

We briefly state our approach to proving Theorem 1.1 and give some indications as to why $\pi_1(M)$ cannot be the discrete Heisenberg 3–group $H^3(\mathbb{Z})$. Let γ be a generator of the center of $H^3(\mathbb{Z})$. Then γ^b can be expressed as a word in terms of two elements α and β outside the center; moreover, this word has word length comparable to $b^{1/2}$ as $b\to\infty$. This expression provides an upper bound on the length growth of γ:

$$|\gamma^b|\le C\cdot b^{1/2}$$

for all b large. If one could find a lower bound violating the above upper bound, then it would end in a desired contradiction. To prove a lower bound, we study the equivariant asymptotic cones of $(\widetilde{M},\langle\gamma\rangle)$:

$$(r_i^{-1}\widetilde{M},\tilde{p},\langle\gamma\rangle)\xrightarrow{\mathrm{GH}}(Y,y,H),$$

where $r_i\to\infty$, $\widetilde{M}$ is the universal cover of M, and $\langle\gamma\rangle$ is the group generated by γ. The limit (Y,y,H) may depend on the sequence r_i, so we shall study all equivariant asymptotic cones. One key intermediate step is to show that the asymptotic orbit Hy is always homeomorphic to $\mathbb{R}$ (Proposition 4.1). This actually requires some understanding of equivariant asymptotic cones of $(\widetilde{M},\pi_1(M,p))$ beforehand. Therefore, we shall first show that for any equivariant asymptotic cone (Y,y,G) of $(\widetilde{M},\pi_1(M,p))$, the asymptotic orbit Gy is homeomorphic to $\mathbb{R}^k$ (Proposition 3.6), by applying a critical rescaling argument and the metric cone structure. After knowing the orbit Hy is homeomorphic to $\mathbb{R}$, we can further deduce some uniform controls on Hy by using the metric cone structure; see Lemmas 4.8 and 4.9. These estimates lead to an almost linear growth estimate of γ: for any $s\in(0,1)$, we have

$$|\gamma^b|\ge C\cdot b^{1-s}$$

for all b large (Theorem 5.3). The index bound in Theorem 1.1(2) follows from the above almost linear growth estimate and the universal bounds in [8; 10]. We point out that this almost linear growth estimate and its proof are quite different from the case of small escape rate [12; 13], where a stronger almost translation estimate holds and the understanding of Hy is not required; see Remark for details.

Motivated by [14] and the work in this paper, we propose the conjecture below.

Conjecture 1.3 *Let (M,p) be an open n–manifold with* $\mathrm{Ric}\ge 0$ *and $E(M,p)\neq\frac{1}{2}$. Suppose that $\pi_1(M)$ contains a torsion-free nilpotent subgroup of nilpotency length l. Then there exist an asymptotic cone (Y,y) of the universal cover and a closed $\mathbb{R}$–subgroup of* $\mathrm{Isom}(Y)$ *such that the orbit $\mathbb{R}y$ is homeomorphic to $\mathbb{R}$ but has Hausdorff dimension at least l.*

The first examples of asymptotic cones that admit isometric $\mathbb{R}$–orbits with Hausdorff dimension strictly larger than 1 were discovered in Pan and Wei [14]. It was suspected in [14, Remark 1.7] that this feature of extra Hausdorff dimension might be related to the nilpotency of fundamental groups. Conjecture 1.3 is a formal description of the question first raised in [14, Remark 1.7].

We organize the paper as follows. We start with some preliminaries in Section 2; this includes some results of nilpotent isometric actions on metric cones and the escape rate. In Section 3, assuming that M satisfies the conditions in Theorem 1.1(1) and $\pi_1(M)$ is nilpotent, we study the equivariant asymptotic geometry of $(\widetilde{M}, \pi_1(M,p))$. Then in Section 4, we further study the asymptotic orbits coming from $\langle\gamma\rangle$–action, where $\gamma \in \pi_1(M,p)$ has infinite order. The properties of these asymptotic orbits lead to the almost linear growth estimate and virtual abelianness in Section 5.

Acknowledgements The author was supported by a Fields Postdoctoral Fellowship from the Fields Institute during the preparation of this work. The author is grateful to Vitali Kapovitch and Guofang Wei for many helpful conversations.

2 Preliminaries

2.1 Equivariant asymptotic geometry and metric cones

Let M be an open n–manifold with $\mathrm{Ric} \geq 0$. Recall that for any sequence $r_i \to \infty$, after passing to a subsequence if necessary, the corresponding blowdown sequence converges in the pointed Gromov–Hausdorff topology:

$$(r_i^{-1}M, p) \xrightarrow{\mathrm{GH}} (X, x).$$

We call the limit space (X, x) an *asymptotic cone* of M. In general, the limit X may not be unique and may not be a metric cone; see examples in [2].

Recall that an open n–manifold M is said to have *Euclidean volume growth of constant L* if

$$\lim_{R\to\infty} \frac{\mathrm{vol}(B_R(p))}{\mathrm{vol}(B_R^n(0))} = L > 0,$$

where $p \in M$ and $B_R^n(0)$ is an R–ball in Euclidean space $\mathbb{R}^n$. By Bishop–Gromov volume comparison, the above limit always exists and is no greater than 1.

Theorem 2.1 [1] *Let M be an open n–manifold with* $\mathrm{Ric} \geq 0$. *If M has Euclidean volume growth, then M is conic at infinity.*

For the purpose of understanding fundamental groups, we study the asymptotic geometry of the Riemannian universal cover $\widetilde{M}$ with the isometric Γ–action, where $\Gamma = \pi_1(M,p)$. Let $r_i \to \infty$ be a sequence. We can pass to a subsequence and consider the pointed equivariant Gromov–Hausdorff convergence [6]

$$(r_i^{-1}\widetilde{M}, \tilde{p}, \Gamma) \xrightarrow{\mathrm{GH}} (Y, y, G),$$

where G is a closed subgroup of the isometry group of Y. It follows from the work of Colding and Naber [5] that G is always a Lie group. We call the above limit space (Y, y, G) an *equivariant asymptotic cone* of $(\widetilde{M}, \Gamma)$.

As a matter of fact, if M has nonnegative sectional curvature, then the equivariant asymptotic cone of $(\widetilde{M}, \Gamma)$ is unique as $(C(Z), z, G)$, where $C(Z)$ is a metric cone with vertex z; moreover, the orbit Gz is a Euclidean factor of $C(Z)$. In fact, by the splitting theorem [3] we write $\widetilde{M}$ as a metric product $\mathbb{R}^k \times W$, where W does not contain any line. The isometry group of $\widetilde{M}$ also splits as $\mathrm{Isom}(\mathbb{R}^k) \times \mathrm{Isom}(W)$; we denote by q_1 and q_2 the natural projections from $\mathrm{Isom}(\widetilde{M})$ to $\mathrm{Isom}(\mathbb{R}^k)$ and $\mathrm{Isom}(W)$, respectively. By setting the basepoint p in a soul of M [4], all minimal representing loops of $\Gamma = \pi_1(M, p)$ are contained in the soul. Then one can show that the set $\{q_2(\gamma \cdot w) \mid \gamma \in \Gamma\}$ is bounded in W, where $\tilde{p} = (0, w)$; see [12, Proof of Proposition A.1]. Thus the limit orbit Gz can be viewed as the one in the asymptotic cone of $(\mathbb{R}^k, 0, q_1(\Gamma))$, which is a Euclidean subspace in $\mathbb{R}^k$. (Also see [11, Theorem 1.4 and Remark 5.12].)

Let $\Omega(\widetilde{M}, \Gamma)$ be the set of all equivariant asymptotic cones of $(\widetilde{M}, \Gamma)$.

Proposition 2.2 *The set $\Omega(\widetilde{M}, \Gamma)$ is compact and connected in the pointed equivariant Gromov–Hausdorff topology.*

See [11, Proposition 2.1] for a proof. The compactness allows us to obtain new spaces in $\Omega(\widetilde{M}, \Gamma)$ as the Gromov–Hausdorff limit of any sequence $\{(Y_j, y_j, G_j)\} \subseteq \Omega(\widetilde{M}, \Gamma)$. Also, note that if $(Y, y, G) \in \Omega(\widetilde{M}, \Gamma)$, then its scaling (sY, y, G) is also an equivariant asymptotic cone for any $s > 0$. Therefore, the Gromov–Hausdorff limit of any sequence $(s_j Y_j, y_j, G_j)$ in $\Omega(\widetilde{M}, \Gamma)$, where $s_j > 0$, is an equivariant asymptotic cone as well. The connectedness of $\Omega(\widetilde{M}, \Gamma)$ will be used implicitly in Section 3.

For an asymptotic cone that is a metric cone, by Cheeger and Colding's splitting theorem [1], any line in the space must split off isometrically. Therefore, we can write such a metric cone (Y, y) as $(\mathbb{R}^k \times C(Z), (0, z))$, where $C(Z)$ is a metric cone without lines and z is the unique vertex of $C(Z)$. The isometry group of Y also splits as

$$\mathrm{Isom}(Y) = \mathrm{Isom}(\mathbb{R}^k) \times \mathrm{Isom}(C(Z)).$$

After setting a point in $\mathbb{R}^k$ as the origin of $\mathbb{R}^k$, we can express any isometry $g \in \mathrm{Isom}(Y)$ as

$$g = (A, v, \alpha),$$

where $(A, v) \in \mathrm{Isom}(\mathbb{R}^k) = O(k) \ltimes \mathbb{R}^k$ and $\alpha \in \mathrm{Isom}(C(Z))$. The multiplication in $\mathrm{Isom}(Y)$ is given by

$$(A_1, v_1, \alpha_1) \cdot (A_2, v_2, \alpha_2) = (A_1 A_2, A_1 v_2 + v_1, \alpha_1 \alpha_2).$$

Also note that since z is the unique vertex of $C(Z)$, any isometry of $C(Z)$ must fix the vertex z. Consequently, for any element $g \in \mathrm{Isom}(Y)$, the orbit point gy must be contained in the Euclidean factor $\mathbb{R}^k \times \{z\}$.

Next, we go through some basic results about nilpotent isometric actions on metric cones. Recall that a group N is called *nilpotent* if its lower central series terminates at the trivial identity subgroup, that is,

$$N = C_0(N) \triangleright C_1(N) \triangleright \cdots \triangleright C_l(N) = \{e\},$$

where the subgroup $C_{j+1}(N) = [C_j(N), N]$ is inductively defined. The smallest integer l such that $C_l(N) = \{e\}$ is called the *nilpotency length* of N.

Lemma 2.3 *Let G be a nilpotent subgroup of* $\mathrm{Isom}(\mathbb{R}^k)$*. Then two elements (A, v) and (B, w) in G commute if and only if A and B commute.*

See [10, Lemma 2.4] for a proof.

Lemma 2.4 *Let $(Y, y) \in \mathcal{M}(n, 0)$ be a metric cone with vertex y. Let G be a closed nilpotent subgroup of the isometry group of Y. Then the center of G has finite index in G; in particular, the identity component subgroup G_0 must be central in G.*

Proof Write $Y = \mathbb{R}^k \times C(Z)$, where $C(Z)$ does not contain lines, and consider the group homomorphism

$$\psi : \mathrm{Isom}(Y) \to O(k) \times \mathrm{Isom}(C(Z)), \quad (A, v, \alpha) \mapsto (A, \alpha).$$

Note that $\overline{\psi(G)}$, the closure of $\psi(G)$, is a compact nilpotent Lie group. It follows from a standard result of group theory that the identity component of $\overline{\psi(G)}$, denoted by K, is central and of finite index in $\overline{\psi(G)}$; see, for example, [10, Lemma 5.7]. Now we consider the subgroup H of G defined by

$$H = \psi^{-1}(K) \cap G,$$

which has finite index in G. It follows from Lemma 2.3 that H is central in G. □

Let (M, p) be an open manifold with the assumptions in Theorem 1.1(1) and a nilpotent fundamental group Γ. Lemma 2.4 implies that G is always virtually abelian for any $(Y, y, G) \in \Omega(\widetilde{M}, \Gamma)$. One may wonder whether the virtual abelianness of all asymptotic limit groups of Γ indicates that Γ itself should be virtually abelian as well. However, it is possible that Γ is a torsion-free nilpotent nonabelian group, while all asymptotic limit groups of Γ are abelian; see the appendix for the example. In other words, the nilpotency length of Γ may not be well-preserved in the asymptotic limits.

2.2 Escape rate

The notion of escape rate was introduced in [12] to study the structure of fundamental groups. It measures where the minimal representing geodesic loops of $\pi_1(M, p)$ are positioned in M. We assign two natural quantities to any loop c based at $p \in M$: its length and its size. Here, size means the smallest radius R such that c is contained in the closed ball $\overline{B}_R(p)$, or equivalently, the Hausdorff distance between the loop c and the basepoint p. For each element $\gamma \in \pi_1(M, p)$, we choose a representing geodesic loop of γ at p, denoted by c_γ, such that c_γ has the minimal length in its homotopy class; if there are multiple choices of c_γ, we choose the one with the smallest size. We write

$$|\gamma| := d(\gamma\tilde{p}, \tilde{p}) = \text{length}(c_\gamma)$$

for convenience. The escape rate of (M, p) is defined as

$$E(M, p) = \limsup_{|\gamma|\to\infty} \frac{\text{size}(c_\gamma)}{\text{length}(c_\gamma)}.$$

As a convention, if $\pi_1(M)$ is a finite group, then we set $E(M, p) = 0$.

In [15], Sormani proved that if $\pi_1(M)$ is not finitely generated, then there is a consequence of elements $\gamma_i \in \pi_1(M, p)$ with representing geodesic loops c_i that are minimal up to halfway. In other words, if $E(M, p) \neq \frac{1}{2}$, then $\pi_1(M)$ is finitely generated.

Lemma 2.5 *Let (M, p) be an open manifold with $\mathrm{Ric} \geq 0$, and let $F\colon (\widehat{M}, \hat{p}) \to (M, p)$ be a finite cover. Then $E(\widehat{M}, \hat{p}) \leq E(M, p)$.*

Proof First note that we can naturally identify $\pi_1(\widehat{M}, \hat{p})$ as a subgroup of $\pi_1(M, p)$, namely we have $F_\star(\pi_1(\widehat{M}, \hat{p})) \subseteq \pi_1(M, p)$ via the injection $F_\star\colon \pi_1(\widehat{M}, \hat{p}) \to \pi_1(M, p)$. Let $\gamma \in \pi_1(\widehat{M}, \hat{p})$ and let σ be a minimal representing geodesic loop of γ at $\hat{p}$. Then $F(\sigma)$ is a minimal representing geodesic loop of $F_\star(\gamma) \in \pi_1(M, p)$. We have

$$d_H(p, F(\sigma)) = \inf\{R > 0 \mid B_R(p) \supseteq F(\sigma)\} = \inf\{R > 0 \mid B_R(F^{-1}(p)) \supseteq \sigma\}.$$

Because F is a finite cover, $F^{-1}(p)$ consists of finitely many points. Let $D > 0$ be the diameter of $F^{-1}(p)$. Then

$$B_{R+D}(\hat{p}) \supseteq B_R(F^{-1}(p))$$

for all $R > 0$. It follows that

$$\inf\{R > 0 \mid B_R(F^{-1}(p)) \supseteq \sigma\} \geq \inf\{R > 0 \mid B_R(\hat{p}) \supseteq \sigma\} - D = d_H(\hat{p}, \sigma) - D.$$

For a sequence of elements $\gamma_i \in \pi_1(\widehat{M}, \hat{p})$ and their corresponding minimal representing geodesic loops σ_i such that

$$\lim_{i\to\infty} \frac{d_H(\hat{p}, \sigma_i)}{\text{length}(\sigma_i)} = E(\widehat{M}, \hat{p}),$$

we have

$$\frac{d_H(p, F(\sigma_i))}{\text{length}(F(\sigma_i))} \geq \frac{d_H(\hat{p}, \sigma_i)}{\text{length}(\sigma_i)} \to E(\widehat{M}, \hat{p}).$$

This shows that $E(M, p) \geq E(\widehat{M}, \hat{p})$. □

With Lemma 2.5, we can assume that $\pi_1(M, p)$ is nilpotent without loss of generality when proving Theorem 1.1. In fact, because $E(M, p) \neq \frac{1}{2}$, $\pi_1(M)$ is finitely generated. By [9; 7], $\pi_1(M)$ has a nilpotent subgroup N of finite index; moreover, according to [8], we can assume that the index of N is bounded by some constant $C(n)$. Let $\widehat{M} = \widetilde{M}/N$ be an intermediate cover of M and let $\hat{p} \in \widehat{M}$ be a lift of p. Lemma 2.5 assures that $E(\widehat{M}, \hat{p}) \neq \frac{1}{2}$. In order to prove Theorem 1.1, it suffices to show that N is virtually abelian and further bound the index when $\widetilde{M}$ has Euclidean volume growth.

3 Asymptotic orbits of nilpotent group actions

In this section, we always assume that (M, p) is an open n–manifold with $\mathrm{Ric} \geq 0$ and $E(M, p) \neq \frac{1}{2}$. Due to Lemma 2.5, we will also assume that $\pi_1(M, p)$ is an infinite nilpotent group, denoted by N.

The goal of this section is to study the properties of asymptotic equivariant cones of $(\widetilde{M}, N)$ when $\widetilde{M}$ is conic at infinity. In particular, we will show that there is an integer k such that for any $(Y, y, G) \in \Omega(\widetilde{M}, N)$, the orbit Gy must be homeomorphic to $\mathbb{R}^k$; see Proposition 3.6.

Lemma 3.1 *Let $(Y, y, G) \in \Omega(\widetilde{M}, N)$. For any point $gy \in Gy$ that is not y, there is a minimal geodesic σ from y to gy and an orbit point $g'y \in Gy$ such that*

$$d(m, g'y) < \tfrac{1}{2} \cdot d(y, gy),$$

where m is the midpoint of σ.

Proof Let $E = E(M, p) < \frac{1}{2}$. Let $r_i \to \infty$ such that

$$(r_i^{-1}\widetilde{M}, p, N) \xrightarrow{\mathrm{GH}} (Y, y, G),$$

and let $\gamma_i \in N$ such that $\gamma_i \xrightarrow{\mathrm{GH}} g \in G$ with respect to the above convergence. Let c_i be a sequence of minimal geodesic loops based at p representing γ_i. By the definition of $E(M, p)$, we have

$$\limsup_{i\to\infty} \frac{d_H(p, c_i)}{\mathrm{length}(c_i)} \leq E.$$

For each i, we lift c_i to $\tilde{c}_i$ as a minimal geodesic from $\tilde{p}$ to $\gamma_i \tilde{p}$. Let

$$R_i = d_H(p, c_i) \quad \text{and} \quad d_i = \mathrm{length}(c_i) = \mathrm{length}(\tilde{c}_i).$$

R_i is also the smallest radius such that $\overline{B_{R_i}(N\tilde{p})}$ covers $\tilde{c}_i$. Passing to a subsequence, we obtain

$$(r_i^{-1}\widetilde{M}, p, N, \tilde{c}_i) \xrightarrow{\mathrm{GH}} (Y, y, G, \sigma), \quad r_i^{-1} d_i \to d(y, gy) \quad \text{and} \quad r_i^{-1} R_i \to R.$$

The above σ is a limit minimal geodesic from y to gy. Moreover, σ is contained in $\overline{B_R(Gy)}$; in particular, let m be the midpoint of σ, then $d(m, g'y) \leq R$ for some $g' \in G$. Thus

$$d(m, g'y) \leq R = \lim_{i\to\infty} r_i^{-1} R_i = \lim_{i\to\infty} \frac{r_i^{-1} R_i}{r_i^{-1} d_i} \cdot r_i^{-1} d_i \leq E \cdot d(y, gy). \qquad \square$$

Lemma 3.1 states that for some minimal geodesic σ from y to gy, its midpoint is closer to Gy than the endpoints of σ. Next, we show that this property implies the connectedness of Gy.

Proposition 3.2 *The orbit Gy is connected for all $(Y, y, G) \in \Omega(\widetilde{M}, N)$.*

Proof We argue by contradiction. Suppose that Gy is not connected. Let $\mathscr{C}_0$ be the connected component of Gy containing y. Note that $\mathscr{C}_0 = G_0 y$, where G_0 is the identity component subgroup of G. Because

G is a Lie group and Gy is not connected, there is a different component $\mathscr{C}_1$ of Gy such that

$$d(\mathscr{C}_0,\mathscr{C}_1)=d(y,\mathscr{C}_1)=\min_{gy\in Gy-\mathscr{C}_0} d(y,gy)>0.$$

Let $gy\in\mathscr{C}_1$ be such that $d(y,gy)=d(y,\mathscr{C}_1)$.

Claim $$d(y,\mathscr{C}_1)=d(\mathscr{C}_0,\mathscr{C}_1).$$

In fact, suppose that $z_0\in\mathscr{C}_0$ and $z_1\in\mathscr{C}_1$ are such that $d(z_0,z_1)<d(y,\mathscr{C}_1)$. We can write $z_0=g_0y$ and $z_1=g_1y$, where $g_0\in G_0$ and $g_1\in G-G_0$. Thus

$$d(y,\mathscr{C}_1)>d(g_0y,g_1y)=d(y,g_0^{-1}g_1y).$$

It follows from the choice of $\mathscr{C}_1$ that $g_0^{-1}g_1y\in\mathscr{C}_0=G_0y$. This leads to

$$z_1=g_1y\in G_0y=\mathscr{C}_0,$$

a contradiction.

By Lemma 3.1, there is a minimal geodesic σ from y to gy and a point $g'y\in Gy$ such that

$$d(m,g'y)<\tfrac{1}{2}\cdot d(y,gy),$$

where m is the midpoint of σ. Then

$$\begin{aligned} d(y,g'y)&\le d(y,m)+d(m,g'y)<d(y,gy)=d(\mathscr{C}_0,\mathscr{C}_1),\\ d(gy,g'y)&\le d(gy,m)+d(m,g'y)<d(gy,y)=d(\mathscr{C}_1,\mathscr{C}_0).\end{aligned}$$

By our choice of $\mathscr{C}_1$ as the closest component to $\mathscr{C}_0$, $\mathscr{C}_0$ is also the closest component to $\mathscr{C}_1$. The first inequality above implies $g'y\in\mathscr{C}_0$, while the second one implies $g'y\in\mathscr{C}_1$, a contradiction. □

Starting from Lemma 3.3 below, we will assume that the universal cover $\widetilde{M}$ is conic at infinity for the rest of this section.

Lemma 3.3 *Let $(Y,y,G)\in\Omega(\widetilde{M},N)$. Then $g_1g_2y=g_2g_1y$ for all $g_1,g_2\in G$.*

Proof Let G_0 be the identity component subgroup of G. Because Gy is connected by Proposition 3.2, we have $Gy=G_0y$. In other words, for any $g\in G$, we can write $gy=hy$ for some $h\in G_0$. Then any $g\in G$ can be written as the product of an element in G_0 and an element in the isotropy subgroup at y; namely, $g=h\cdot(h^{-1}g)$, where $h\in G_0$ and $h^{-1}g$ fixes y.

Let $g_1,g_2\in G$. We write

$$g_1=h_1\cdot\alpha_1 \quad\text{and}\quad g_2=h_2\cdot\alpha_2,$$

where $h_1,h_2\in G_0$ and α_1 and α_2 belong to the isotropy subgroup at y. Because N is nilpotent, G must be nilpotent as well. According to Lemma 2.4, G_0 is central in G. Thus

$$g_1g_2y=h_1\alpha_1h_2\alpha_2y=h_1h_2y=h_2h_1y=h_2\alpha_2h_1\alpha_1y=g_2g_1y. \qquad\square$$

Definition 3.4 Let (Y, y, G) be a space, where G is a closed nilpotent Lie subgroup of $\mathrm{Isom}(Y)$. Let T be a maximal torus of G_0. Let $k \in \mathbb{N}$ and $d \in [0, \infty)$. We say that (Y, y, G) is of *type* (k, d) if the orbit Gy is connected and

$$\dim G - \dim T = k \quad \text{and} \quad \mathrm{diam}(Ty) = d.$$

Lemma 3.5 *Let $C(Z)$ be a metric cone with a vertex z. Let G be a closed nilpotent subgroup of* $\mathrm{Isom}(C(Z))$. *Suppose that $(C(Z), z, G)$ is of type (k, d). For a sequence $r_i \to \infty$, we consider the corresponding blowdown sequence*

$$(r_i^{-1} C(Z), z, G) \xrightarrow{\mathrm{GH}} (C(Z), z, G').$$

Then the orbit $G'z$ is a k–dimensional Euclidean factor in $C(Z)$.

Proof Because the orbit Gz is connected and contained in a Euclidean factor of $C(Z)$, it suffices to prove the statement when $C(Z)$ is a Euclidean space $\mathbb{R}^l$ and G is a connected Lie group. By Lemma 2.4, G is abelian. We set z as the origin 0 of $\mathbb{R}^l$, then we can write any element of G in the form of $(A, v) \in \mathrm{SO}(l) \ltimes \mathbb{R}^l$. Let

$$\psi\colon \mathrm{Isom}(\mathbb{R}^l) \to \mathrm{SO}(l), \quad (A, v) \mapsto A,$$

be the natural projection. Because $\psi(G)$ is abelian, we can decompose $\mathbb{R}^k$ into an orthogonal direct sum $E + E^{\perp}$, where E is the maximal subspace such that $A|_E = \mathrm{id}|_E$ for all $A \in \psi(G)$. Note that by the above construction, any translation in E and any $g \in G$ must commute.

Let $(A, v) \in G$. We write $v = v_1 + v_2$, where $v_1 \in E$ and $v_2 \in E^{\perp}$. Let $\delta \in \mathrm{Isom}(\mathbb{R}^l)$ be the translation by $-v_1$. We claim that $\delta g = (A, v_2)$ must have a fixed point. The proof of this claim is by linear algebra. We argue by contradiction. Because $\psi(G)$ is an abelian subgroup of $\mathrm{SO}(l)$, we can further decompose $E^{\perp}$ into an orthogonal direct sum of subspaces with dimension at most 2,

$$E^{\perp} = E^1 + E^2 + \cdots + E^m,$$

so that each E^i is $\psi(G)$–invariant, where $i = 1, \ldots, m$. We write $v_2 = \sum_{i=1}^m v^i$, where $v^i \in E^i$. By the hypothesis that (A, v_2) does not have fixed points, there exists $j \in \{1, \ldots, m\}$ such that $A|_{E^j} = \mathrm{id}|_{E^j}$ and $v^j \neq 0$. From the maximality of E in its definition, we can find some element $(B, w) \in G$ such that $Bv^j \neq v^j$. We write

$$w = w_1 + w_2 = w_1 + \left(\sum_{i=1}^m w^i \right),$$

where $w_1 \in E$, $w_2 \in E^{\perp}$ and $w^i \in E^i$. As (A, v) commutes with (B, w), by direct calculation we have

$$(B - I)v^i = (A - I)w^i \quad \text{for all } i = 1, \ldots, m.$$

Taking $i = j$, we derive that

$$0 \neq (B - I)v^j = (A - I)w^j = 0,$$

a contradiction. We have verified the claim.

By the claim, any element $g \in G$ can be written as a product $g = \delta\alpha = \alpha\delta$, where δ is a translation in E and α has a fixed point; moreover, this expression is unique. This enables us to define a group homomorphism

$$F: G \to E, \quad g = \delta\alpha \mapsto \delta.$$

Noting that $(\mathbb{R}^l, z, G)$ is of type (k, d), we can write $G = H \times T$, where H is a closed subgroup isomorphic to $\mathbb{R}^k$ and T is a torus subgroup. We remark that the choice of H is not unique in general. It is clear that $F(h) \neq 0$ for all nontrivial elements $h \in H$; otherwise, $\langle h \rangle$ would be contained in a compact subgroup of $\mathrm{Isom}(\mathbb{R}^l)$, which is not true. Therefore, $F(H)$ consists of all translations in a k–dimensional subspace in V_1. After blowing down

$$(r_i^{-1}\mathbb{R}^l, z, G) \xrightarrow{\mathrm{GH}} (\mathbb{R}^l, z, G'),$$

it is clear that the limit orbit $G'z$ is formed exactly by translations in $F(H)$. In particular, $G'z$ is a k–dimensional Euclidean subspace of $\mathbb{R}^l$. □

Let $(Y, y, G) \in \Omega(\widetilde{M}, N)$. By the proof of [12, Lemma 3.1], the finite generation of N implies that the orbit Gy is always noncompact. In other words, letting (k, d) be the type of (Y, y, G), we always have $k \geq 1$.

Proposition 3.6 *Let (M, p) be an open n–manifold with the assumptions in Theorem 1.1(1). Suppose that the fundamental group N is an infinite nilpotent group. Then there is an integer k such that all $(Y, y, G) \in \Omega(\widetilde{M}, N)$ are of type $(k, 0)$.*

The proof of Proposition 3.6 is by contradiction and a critical rescaling argument, which implicitly uses the connectedness of $\Omega(\widetilde{M}, N)$ (Proposition 2.2). This kind of argument is also used in [11; 10; 12; 13], in different contexts, to prove certain uniform properties among all equivariant asymptotic cones. This method requires an equivariant Gromov–Hausdorff distance gap between certain spaces, which we establish below.

Lemma 3.7 *Given any integer $n \geq 2$, there is a constant $\delta(n) > 0$ such that the following holds.*

Let $(C(Z_j), z_j) \in \mathcal{M}(n, 0)$ be a metric cone with vertex z_j and let G_j be a closed nilpotent subgroup of $\mathrm{Isom}(C(Z_j))$, where $j = 1, 2$. Suppose that

1. *the orbit $G_1 z_1$ is a k_1–dimensional Euclidean factor of $C(Z_1)$, and*
2. *the orbit $G_2 z_2$ is connected and is of type (k_2, d_2), where $k_2 > k_1$.*

Then

$$d_{GH}((C(Z_1), z_1, G_1), (C(Z_2), z_2, G_2)) \geq \delta(n).$$

Proof We set $\delta(n) = 1/(100n^2)$. Suppose that

$$d_{GH}((C(Z_1), z_1, G_1), (C(Z_2), z_2, G_2)) < \delta(n).$$

Let $e_1, \dots, e_{k_1} \in G_1$ be such that their orbit points $\{e_1 z_1, \dots, e_{k_1} z_1\}$ form an orthogonal basis of $G_1 z_1 \simeq \mathbb{R}^{k_1}$ and $d(e_j z_1, z_1) = 1/n$ for all $j = 1, \dots, k_1$. Letting L be the subgroup generated by

$\{e_1,\dots,e_{k_1}\}$, then Lz_1 is 1–dense in G_1z_1. Let $e'_1,\dots,e'_{k_1}\in G_2$ such that each e'_j is $\delta(n)$–close to e_j, where $j=1,\dots,k_1$. Let L' be the subgroup generated by $\{e'_1,\dots,e'_{k_1}\}$. Though elements in L' may not be commutative, by Lemma 3.3, the orbit $L'z_2$ can be identified as

$$L'z_2=\left\{\prod_{j=1}^{k_1}(e'_j)^{l_j}z_2 \;\middle|\; l_j\in\mathbb{Z}\right\}\subseteq G_2z_2.$$

By the inequality $k_2>k_1$ in the second condition, there exists an element $g'\in G_2$ such that

$$d(g'z_2,z_2)=d(g'z_2,L'z_2)\in(7,8).$$

Take a $g\in G_1$ that is $\delta(n)$–close to g'. Because Lz_1 is 1–dense in G_1z_1, there exists some element $h=\prod_{j=1}^{k_1}e_j^{l_j}\in L$ such that $d(hz_1,gz_1)\le 1$. By the triangle inequality, we have

$$d(hz_1,z_1)\le d(hz_1,gz_1)+d(gz_1,z_1)\le 1+d(gz_1,g'z_2)+d(g'z_2,z_2)+d(z_2,z_1)\le 10.$$

According to [12, Lemma 4.10], G_1 acts as translations on G_1z_1. Recall that each $e_j\in G_1$ has displacement $1/n$ at z_1, thus each $l_j\le 10n$. Together with the choice of $\delta(n)$, we see that $h'=\prod_{j=1}^{k_1}(e'_j)^{l_j}\in G_2$ is $\frac{1}{10}$–close to $h\in G_1$. Thus

$$d(g'z_2,L'z_2)\le d(g'z_2,h'z_2)\le d(gz_1,hz_1)+d(gz_1,g'z_2)+d(hz_1,h'z_2)\le 2.$$

This is a contradiction to $d(g'z_2,L'z_2)>7$, and thus

$$d_{GH}\big((C(Z_1),z_1,G_1),(C(Z_2),z_2,G_2)\big)\ge\delta(n). \qquad\square$$

Now we use Lemma 3.7 and a critical rescaling argument to prove Proposition 3.6.

Proof of Proposition 3.6 We argue by contradiction.

Claim 1 *Suppose that the statement is not true. Then there exist spaces (Y_1,y_1,G_1) and (Y_2,y_2,G_2) in $\Omega(\widetilde{M},N)$ such that for $j=1,2$, the orbit G_jy_j is a Euclidean factor of dimension k_j, with $k_1>k_2$.*

In fact, if the statement of Proposition 3.6 fails, then either there exists a space $(W,w,H)\in\Omega(\widetilde{M},N)$ of type (k,d) with $d>0$, or for $j=1,2$ there exist $(W_j,w_j,H_j)\in\Omega(\widetilde{M},N)$ of type $(k_j,0)$, with $k_1>k_2$. For the first case above, we consider the blowup and blowdown limits of (W,w,H):

$$(lW,w,H)\xrightarrow{\mathrm{GH}}(W,w,H_1),\qquad (l^{-1}W,w,H)\xrightarrow{\mathrm{GH}}(W,w,H_2)$$

where $l\to+\infty$. Because $d>0$, it is clear that the orbit H_1w is a Euclidean factor with dimension strictly larger than k. By Lemma 3.5, the orbit H_2w is a k–dimensional Euclidean factor. Then (W,w,H_1) and (W,w,H_2) are the desired spaces in Claim 1. For the second case, for $j=1,2$ the blowup limit of (W_j,w_j,H_j) clearly satisfies the requirements; alternatively, one can also use the blowdown limits of (W_j,w_j,H_j) and then apply Lemma 3.5. This proves Claim 1.

For $j=1,2$ let $(Y_j,y_j,G_j)\in\Omega(\widetilde{M},N)$ be as described in Claim 1. Let $r_i,s_i\to\infty$ be such that

$$(r_i^{-1}\widetilde{M},\tilde{p},N)\xrightarrow{\mathrm{GH}}(Y_1,y_1,G_1)\quad\text{and}\quad(s_i^{-1}\widetilde{M},\tilde{p},N)\xrightarrow{\mathrm{GH}}(Y_2,y_2,G_2).$$

By passing to a suitable subsequence of r_i or s_i, we can assume that $t_i := r_i/s_i \to \infty$. We put

$$(M_i, q_i, N_i) = (r_i^{-1}\widetilde{M}, \tilde{p}, N).$$

Then

$$(M_i, q_i, N_i) \xrightarrow{\text{GH}} (Y_1, y_1, G_1) \quad \text{and} \quad (t_i M_i, q_i, N_i) \xrightarrow{\text{GH}} (Y_2, y_2, G_2).$$

Let $\delta(n)$ be the constant in Lemma 3.7. For each i, we define a set of scales L_i by

$$L_i = \bigl\{l \in [1, t_i] \mid d_{GH}((l M_i, q_i, N_i), (W, w, H)) \le \tfrac{1}{10}\delta(n), \text{where } (W, w, H) \in \Omega(\widetilde{M}, N) \text{ has the orbit } Hw \text{ as a Euclidean factor with dimension } < k_1\bigr\}.$$

Recall that in (Y_2, y_2, G_2), the orbit $G_2 y$ is a k_2–dimensional Euclidean factor with $k_2 < k_1$, thus $t_i \in L_i$ for all i large; in particular, L_i is nonempty. We choose $l_i \in L_i$ with $\inf L_i \le l_i \le \inf L_i + 1$ as a sequence of critical scales.

Claim 2 $$l_i \to \infty.$$

Suppose that l_i subconverges to a number $l_\infty < +\infty$. Then

$$(l_i M_i, q_i, N_i) \xrightarrow{\text{GH}} (l_\infty Y_1, y_1, G_1).$$

Recall that the orbit $G_1 y_1$ in (Y_1, y_1, G_1) is a k_1–dimensional Euclidean factor; thus after scaling by l_∞, the orbit $G_1 y_1$ in $(l_\infty Y_1, y_1, G_1)$ is also a k_1–dimensional Euclidean factor. On the other hand, since $l_i \in L_i$, each $(l_i M_i, q_i, N_i)$ is $\frac{1}{10}\delta(n)$–close to some $(W_i, w_i, H_i) \in \Omega(\widetilde{M}, N)$ whose orbit $H_i w_i$ is a Euclidean factor of dimension $< k_1$. It follows that

$$d_{GH}((l_\infty Y_1, y_1, G_1), (W_i, w_i, H_i)) \le \tfrac{1}{2}\delta(n)$$

for all i large; a contradiction to Lemma 3.7. This proves Claim 2.

Next, after passing to a convergent subsequence, we consider the rescaling limit

$$(l_i M_i, q_i, N_i) \xrightarrow{\text{GH}} (Y', y', G') \in \Omega(\widetilde{M}, N).$$

Let (k', d') be the type of (Y', y', G'). It has the following two possibilities.

Case 1 $(k' \ge k_1)$ Recall that each $(l_i M_i, q_i, N_i)$ satisfies

$$d_{GH}((l_i M_i, q_i, N_i), (W_i, w_i, H_i)) \le \tfrac{1}{10}\delta(n)$$

for some $(W_i, w_i, H_i) \in \Omega(\widetilde{M}, N)$ whose orbit $H_i w_i$ is a Euclidean factor of dimension $< k_1$. Since $(l_i M_i, q_i, N_i)$ converges to (Y', y', G'), we have

$$d_{GH}((Y', y', G'), (W_i, w_i, H_i)) \le \tfrac{1}{2}\delta(n)$$

for all i large, where $G' y'$ is of type (k', d') with $k' \ge k_1$ and $H_i w_i$ is a Euclidean factor of dimension $< k_1$. This contradicts Lemma 3.7. Thus Case 1 cannot happen.

Case 2 $(k' < k_1)$ We consider the blowdown limit of (Y', y', G'):

$$(j^{-1}Y', y', G') \xrightarrow{\mathrm{GH}} (Y', y', H'),$$

where $j \to \infty$. By Lemma 3.5, the orbit $H'y'$ is a Euclidean factor of dimension k'. Let $J \in \mathbb{N}$ be large such that

$$d_{GH}((J^{-1}Y', y', G'), (Y', y', H')) \le \tfrac{1}{100}\delta(n).$$

Note that

$$(J^{-1}l_i M_i, q_i, N_i) \xrightarrow{\mathrm{GH}} (J^{-1}Y', y', G');$$

thus,

$$d_{GH}((J^{-1}l_i M_i, q_i, N_i), (Y', y', H')) \le \tfrac{1}{10}\delta(n)$$

for all i large. Because $l_i \to \infty$ and $H'y'$ is a Euclidean factor of dimension $< k_1$, we conclude that $J^{-1}l_i \in L_i$ for all i large. However, this contradicts our choice of l_i as $\inf L_i \le l_i \le \inf L_i + 1$. Thus Case 2 cannot happen.

With all possibilities of (Y', y', G') being ruled out, we reach the desired contradiction and thus complete the proof of statement. $\square$

As a direct consequence of Proposition 3.6, in any $(Y, y, G) \in \Omega(\widetilde{M}, N)$, any compact subgroup of G must fix the basepoint y. This implies the lemmas below.

Lemma 3.8 *Let $(Y, y, G) \in \Omega(\widetilde{M}, N)$ and let $h_1, h_2 \in G$. If $h_1^m y = h_2^m y$ for some integer $m \ge 2$, then $h_1 y = h_2 y$.*

Proof We first prove that if $h^m y = y$ for some integer $m \ge 2$, then $hy = y$. In fact, let $H \subseteq G$ be the closure of the subgroup generated by h. Because $h^m y = y$, the orbit Hy consists of at most $m-1$ points; in particular, the orbit Hy is closed and bounded. Thus H is a compact subgroup of G. Because (Y, y, G) is of type $(k, 0)$ by Proposition 3.6, H must fix y. Thus $hy = y$.

Now, let $h_1, h_2 \in G$ be such that $h_1^m y = h_2^m y \ne y$ for some integer $m \ge 2$. By Lemma 3.3,

$$y = h_1^{-m} h_2^m y = (h_1^{-1} h_2)^m y.$$

It follows from the previous paragraph that $h_1 y = h_2 y$. $\square$

Lemma 3.9 *Let $(Y, y, G) \in \Omega(\widetilde{M}, N)$. Let H be a closed $\mathbb{R}$–subgroup of N and let $\beta \in G$ be such that βy is outside of Hy. Then $d(\beta^m y, Hy)$ is unbounded as $m \to \infty$.*

Proof We argue by contradiction. Suppose that there is a number $C > 0$ such that $d(\beta^m y, Hy) \le C$ for all $m \in \mathbb{Z}$. By the connectedness of Gy, we can assume that $\beta \in G_0$ without loss of generality. Because H is central in G by Lemma 2.4, we can consider the quotient of (Y, y, G) by the H–action, denoted by $(Y/H, \bar{y}, G/H)$. Let $\bar{\beta} \in G/H$ be the quotient of β. By hypothesis, we have

$$d(\bar{\beta}\bar{y}, \bar{y}) > 0 \quad \text{and} \quad d(\bar{\beta}^m \bar{y}, \bar{y}) \le C$$

for all $m \in \mathbb{Z}$. Let $K \subseteq G/H$ be the closure of the subgroup generated $\overline{\beta}$. Then K is a compact subgroup in the identity component of G/H with

$$0 < \operatorname{diam}(K\overline{y}) \leq C.$$

On the other hand, because G_0 is abelian and (Y, y, G) is of type $(k, 0)$, we can write $G_0 = \mathbb{R}^k \times T$, where T is a torus group fixing y. After taking the quotient by the $\mathbb{R}$–subgroup H, any compact subgroup in G_0/H must fix $\overline{y}$; a contradiction. □

4 Asymptotic orbits of $\mathbb{Z}$–actions

Throughout this section, we always assume that an open manifold M satisfies the assumptions in Theorem 1.1(1) and has an infinite nilpotent fundamental group N. We fix an element $\gamma \in N$ with infinite order. We will study the equivariant asymptotic cones of $(\widetilde{M}, \langle\gamma\rangle)$. Our first goal of this section is to prove the result below.

Proposition 4.1 *Any space $(Y, y, H) \in \Omega(\widetilde{M}, \langle\gamma\rangle)$ must be of type $(1, 0)$. Consequently, the orbit Hy is connected and homeomorphic to $\mathbb{R}$.*

We remark that the group H could be strictly larger than $\mathbb{R}$, because H may have a nontrivial isotropy subgroup at y. Also, recall that H is a closed subgroup of $\operatorname{Isom}(Y)$, thus the orbit Hy is embedded in Y, that is, the subspace topology of Hy matches with the quotient topology from H/K, where K is the isotropy subgroup of H at y.

Here is the rough idea to prove Proposition 4.1: suppose that (Y, y, H) is not of type $(1, 0)$. Then we shall find a space $(Y', y', G') \in \Omega(\widetilde{M}, N)$ violating Proposition 3.6.

We need some preparations first.

Definition 4.2 Let G be a group. We say a subset S of G is *symmetric* if S satisfies the following conditions:

(1) $\mathrm{id} \in S$.

(2) If $g \in S$, then $g^{-1} \in S$.

Definition 4.3 Let (X_i, x_i, G_i) be a pointed equivariant Gromov–Hausdorff convergent sequence with limit (Y, y, H). Recall that this means there is a sequence of triples of ϵ_i–approximation maps (f_i, φ_i, ψ_i):

$$f_i \colon B_{1/\epsilon_i}(x_i) \to Y, \quad \varphi \colon G_i(1/\epsilon_i) \to H(1/\epsilon_i), \quad \psi \colon H(1/\epsilon_i) \to G_i(1/\epsilon_i),$$

with the properties described in [6, Definition 3.3]). For each i, let S_i be a closed symmetric subset of G_i. We write $\overline{\varphi_i(S_i)}$ for the closure of $\varphi_i(S_i)$ in H. We say that the sequence S_i Gromov–Hausdorff converges to a limit closed symmetric subset $S \subseteq H$, denoted by

$$(X_i, x_i, S_i) \xrightarrow{\mathrm{GH}} (Y, y, S),$$

if S is the limit of $\overline{\varphi_i(S_i)}$ with respect to the topology on the set of all closed subsets of H induced by the compact–open topology. Equivalently, the closed symmetric subset $S \subseteq H$ satisfies:

(1) For any $h \in S$, there is a sequence of isometries $g_i \in S_i$ converging to h.

(2) Any convergent sequence of isometries $g_i \in S_i$ has the limit h in S.

It follows directly from the proof of [6, Proposition 3.6] that we have the precompactness result below.

Proposition 4.4 *Let (X_i, x_i, G_i) be a pointed equivariant Gromov–Hausdorff convergent sequence with limit (Y, y, H). For each i, let S_i be a closed symmetric subset of G_i. Then passing to a subsequence, we have the convergence*

$$(X_i, x_i, S_i) \xrightarrow{\mathrm{GH}} (Y, y, S)$$

for some limit closed symmetric subset S of H.

Definition 4.5 Let $(Y, y, G) \in \Omega(\widetilde{M}, N)$ and let $gy \in Gy - \{y\}$. Because the orbit Gy is connected, we can assume $g \in G_0$. Let exp be the exponential map from the Lie algebra of G_0 to the Lie group G_0; note that exp is surjective because G_0 is abelian. Then $g = \exp(v)$ for some v in the Lie algebra. We define the following subsets of Gy:

$$P(g)y = \{\exp(tv)y \mid t \in [-1, 1]\} \quad \text{and} \quad \mathbb{R}(g)y = \{\exp(tv)y \mid t \in \mathbb{R}\}.$$

Lemma 4.6 *In Definition 4.5, the sets $P(g)y$, and thus $\mathbb{R}(g)y$, are uniquely determined by the orbit point gy.*

Proof We first show that the set $P(g)y$ is independent of the choice of v in Definition 4.5. Suppose that

$$g = \exp(v) = \exp(w),$$

where v, w are elements in the Lie algebra of G_0. By Lemma 3.8, we have

$$\exp\Big(\frac{1}{b}v\Big)y = \exp\Big(\frac{1}{b}w\Big)y$$

for any integer $b \in \mathbb{Z}_+$. Then for any integer $a \in \mathbb{Z}_+$, it follows from Lemma 3.3 that

$$\exp\Big(\frac{a}{b}v\Big)y = \exp\Big(\frac{a-1}{b}v\Big)\exp\Big(\frac{1}{b}v\Big)y = \exp\Big(\frac{1}{b}w\Big)\exp\Big(\frac{a-1}{b}v\Big)y = \cdots = \exp\Big(\frac{a}{b}w\Big)y.$$

In other words, we have shown that

$$\exp(tv)y = \exp(tw)y$$

holds for all $t \in \mathbb{Q}$. Because $P(g)y$ is the closure of the set

$$\{\exp(tv)y \mid t \in [-1, 1] \cap \mathbb{Q}\},$$

we conclude that $P(g)y$ is independent of the choice of v.

Next, we show that $P(g)y$ only depends on the orbit point gy, but not the choice of $g \in G_0$. Suppose that $h \in G_0$ is such that $gy = hy$. Let v and w be vectors in the Lie algebra of G_0 such that

$$\exp(v) = g \quad \text{and} \quad \exp(w) = h.$$

Following a similar argument to that in the first paragraph of the proof and applying Lemmas 3.3 and 3.8, one can clearly verify the result. □

Lemma 4.7 *Let $(Y, y, G) \in \Omega(\widetilde{M}, N)$ and let S be a closed symmetric subset of G. Suppose that the set Sy satisfies the following properties*:

1. *Sy is closed under multiplication*; *that is*, *if $g_1, g_2 \in S$*, *then $g_1 g_2 y \in Sy$.*
2. *Sy is bounded.*

Then $Sy = \{y\}$.

Proof Let H be the closure of the subgroup generated by S. The first assumption implies that $Hy = Sy$. Because $Hy = Sy$ is bounded, we conclude that H must be a compact subgroup of G. Since $(Y, y, G) \in \Omega(\widetilde{M}, N)$ is of type $(k, 0)$ by Proposition 3.6, H fixes y. In other words, we have $Sy = Hy = \{y\}$. □

We are in a position to prove Proposition 4.1.

Proof of Proposition 4.1 Let $r_i \to \infty$ be a sequence. We consider the convergence

$$(r_i^{-1}\widetilde{M}, \tilde{p}, N, \langle\gamma\rangle) \xrightarrow{\text{GH}} (Y, y, G, H).$$

We shall show that (Y, y, H) is of type $(1, 0)$.

For each i, let

$$l_i = \min\{l \in \mathbb{Z}_+ \mid d(\gamma^l \tilde{p}, \tilde{p}) \ge r_i\}$$

and let

$$S_\gamma(l_i) = \{\mathrm{id}, \gamma^{\pm 1}, \ldots, \gamma^{\pm l_i}\}$$

be a sequence of symmetric subsets of $\langle\gamma\rangle$. By the triangle inequality, we have

$$r_i \le |\gamma^{l_i}| \le r_i + |\gamma|.$$

Passing to a subsequence, we obtain convergence

$$(r_i^{-1}\widetilde{M}, \tilde{p}, \gamma^{l_i}, S_\gamma(l_i)) \xrightarrow{\text{GH}} (Y, y, g, A),$$

where A is a closed symmetric subset of H and $g \in A$ with $d(gy, y) = 1$.

Claim 1 *The set Ay contains $P(g)y$.*

Let $b \in \mathbb{Z}_+$. By the choice of l_i,

$$r_i^{-1} d(\gamma^{\lfloor l_i/b \rfloor}\tilde{p}, \tilde{p}) \le 1,$$

where $\lfloor \cdot \rfloor$ means the floor function. Thus the sequence $\gamma^{\lfloor l_i/b \rfloor}$ subconverges to some limit $\alpha \in A$. Because

$$l_i \le b \cdot \lfloor l_i/b \rfloor < l_i + b,$$

passing to a subsequence if necessary, we can assume that $b \cdot \lfloor l_i/b \rfloor = l_i + b_0$ for some $b_0 = 0, \dots, b$ and all i. For this subsequence, we have

$$(r_i^{-1}\widetilde{M}, \tilde{p}, \gamma^{\lfloor l_i/b \rfloor}, \gamma^{b_0}, \gamma^{b\cdot\lfloor l_i/b \rfloor}) \xrightarrow{\mathrm{GH}} (Y, y, \alpha, g_0, g \cdot g_0),$$

where $g_0 \in A$ fixes y; moreover, $g \cdot g_0 = \alpha^b$. Thus α satisfies

$$\alpha^b y = g \cdot g_0 y = gy.$$

It follows from Lemma 3.8 that

$$\alpha y = \exp\Big(\frac{1}{b} v\Big) y,$$

where $\exp(v) = g$. By construction, the limit symmetric subset A contains the set $\{\mathrm{id}, \alpha^{\pm 1}, \dots, \alpha^{\pm b}\}$. Therefore, Ay contains the orbit points

$$\Big\{y, \exp\Big(\pm\frac{1}{b} vy\Big), \exp\Big(\pm\frac{2}{b} vy\Big), \dots, \exp(\pm v)y\Big\}.$$

Because b is an arbitrary positive integer and Ay is closed, we conclude that Ay contains $P(g)y$.

By Claim 1, the limit orbit Hy must contain $\mathbb{R}(g)y$. To this end, we argue by contradiction to prove Proposition 4.1. Suppose that (Y, y, H) is not of type $(1,0)$, then there exists an element $\beta \in H$ such that $\beta y \notin \mathbb{R}(g)y$. Because (Y, y, G) is of type $(k, 0)$, by Lemma 3.9 we can choose an element as a power of β, denoted by h, such that $d(hy, \mathbb{R}(g)y) \ge 2$. Let $m_i \to \infty$ be such that

$$(r_i^{-1}\widetilde{M}, \tilde{p}, \gamma^{m_i}) \xrightarrow{\mathrm{GH}} (Y, y, h).$$

Because $d(hy, y) \ge 2$, it is clear that $m_i > l_i$ by our choice of l_i.

Claim 2 $$m_i / l_i \to \infty.$$

Suppose that $m_i/l_i \to C \in [1, \infty)$ for a subsequence. We write

$$m_i = \lfloor C \rfloor \cdot l_i + o_i,$$

where $0 \le o_i \le l_i$. Note that

$$(r_i^{-1}\widetilde{M}, \tilde{p}, \gamma^{\lfloor C \rfloor \cdot l_i}, \gamma^{o_i}) \xrightarrow{\mathrm{GH}} (Y, y, g^{\lfloor C \rfloor}, \delta)$$

with $g^{\lfloor C \rfloor} y \in \mathbb{R}(g)(y)$ and $\delta y \in Ay$. Since $d(\delta y, y) \le 1$, we have

$$d(hy, \mathbb{R}(g)y) = d(g^{\lfloor C \rfloor}\delta y, \mathbb{R}(g)y) = d(\delta y, \mathbb{R}(g)y) \le 1.$$

We result in a contradiction to $d(hy, \mathbb{R}(g)y) \ge 2$. This proves Claim 2.

For each i, let

$$d_i := \max\{d(\gamma^k \tilde{p}, \tilde{p}) \mid k = l_i, l_i + 1, \dots, m_i\} \to \infty.$$

It is clear that $d_i \ge r_i$.

Claim 3 $$d_i / r_i \to \infty.$$

Suppose the contrary, that is, $d_i/r_i \to C \in [1, \infty)$. Let

$$S_\gamma(m_i) = \{\mathrm{id}, \gamma^{\pm 1}, \dots, \gamma^{\pm m_i}\}.$$

Then we obtain convergence

$$(d_i^{-1}\widetilde{M},\tilde{p},S_\gamma(l_i),S_\gamma(m_i)) \xrightarrow{\mathrm{GH}} (C^{-1}Y,y,A,B).$$

Recall that Ay contains $P(g)y$ by Claim 1. Together with Claim 2, that $m_i/l_i\to\infty$, we see that By must contain $\mathbb{R}(g)y$, which is unbounded. On the other hand, by the choice of d_i, By should be contained in $\bar{B}_1(y)$; a contradiction. This proves Claim 3.

Next, we consider the convergence

$$(d_i^{-1}\widetilde{M},\tilde{p},N,\langle\gamma\rangle,\gamma^{m_i},S_\gamma(m_i)) \xrightarrow{\mathrm{GH}} (Y',y',G',H',h',B').$$

Due to the choice of d_i, it is clear that

$$d_H(B'y',y')=1.$$

Also, it follows from Claim 3 that $h'y'=y'$.

Claim 4 *The set $B'y'$ is closed under multiplication; that is, if $\beta_1,\beta_2\in B'$, then $\beta_1\beta_2y'\in B'y'$.*

In fact, let $b_{i,1},b_{i,2}\in[-m_i,m_i]$ be two sequences of integers such that

$$(d_i^{-1}\widetilde{M},\tilde{p},\gamma^{b_{i,1}},\gamma^{b_{i,2}}) \xrightarrow{\mathrm{GH}} (Y',y',\beta_1,\beta_2).$$

If $b_{i,1}+b_{i,2}\in[-m_i,m_i]$, then $\beta_1\beta_2\in B'$ and the claim holds trivially. If not, we can write

$$b_{i,1}+b_{i,2}=\pm m_i+o_i,$$

where $o_i\in[-m_i,m_i]$. Let $\beta_0\in B'$ be the limit of γ^{o_i} after passing to a convergent subsequence. Then Claim 4 follows because

$$\beta_1\beta_2y'=\lim_{i\to\infty}\gamma^{o_i}\cdot\gamma^{\pm m_i}\tilde{p}=\beta_0(h')^{\pm 1}y'=\beta_0y'\in B'y'.$$

Lastly, we apply Lemma 4.7 to B' and conclude that $B'y'=y'$. We end in a contradiction to $d_H(B'y',y')=1$. This contradiction completes the proof of Proposition 4.1. □

Let $z\in Hy$ be an orbit point. Because Hy is connected, we can write $z=hy$ for some $h\in H_0$. Let v in the Lie algebra of H_0 be such that $\exp(v)=h$. For convenience, in the rest of the paper, we will denote the orbit point $\exp(tv)y$ by $(th)y$, where $t\in\mathbb{R}$. By the proof of Lemma 4.6, this point $(th)y$ is independent of the choice of v and h. Also, with Proposition 4.1, we have $Hy=\mathbb{R}(h)y$.

For the rest of this section, we prove some uniform controls on the path $P(h)y$ that will be used later in Section 5.

Lemma 4.8 *There exists a constant $C_1=C_1(\widetilde{M},\gamma)$ such that the following holds. For any $(Y,y,H)\in\Omega(\widetilde{M},\langle\gamma\rangle)$ and any $h\in H_0$ with $d(hy,y)\neq 0$, we have*

$$d((th)y,y)\le C_1\cdot d(hy,y)\quad \textit{for all } t\in[0,1].$$

Proof Without loss of generality, we assume that $d(hy,y)=1$ by scaling (Y,y,H). We argue by contradiction to prove the lemma. Suppose that we have a sequence of spaces $(Y_j,y_j,H_j)\in\Omega(\widetilde{M},\langle\gamma\rangle)$

and $h_j \in H_j$ with $d(h_j y_j, y_j) = 1$, but

$$R_j := \max_{t\in[0,1]} d((th_j)y_j, y_j) \to \infty.$$

Scaling the sequence by R_j^{-1} and passing to a convergent subsequence, we obtain

$$(R_j^{-1} Y_j, y_j, H_j) \xrightarrow{\mathrm{GH}} (Y', y', H') \in \Omega(\widetilde{M}, \langle\gamma\rangle).$$

The hypothesis implies $h_j y_j \xrightarrow{\mathrm{GH}} y'$ with respect to the above convergence. We consider the closed symmetric subset $S_j = \{th_j \mid t \in [0,1]\}$ of H_j and let $S' \subset H'$ be its limit symmetric subset, that is,

$$(R_j^{-1} Y_j, y_j, S_j) \xrightarrow{\mathrm{GH}} (Y', y', S').$$

We claim that the set $S'y'$ is closed under multiplication; the proof is similar to Claim 4 in the proof of Proposition 4.1. In fact, for any $\beta_1, \beta_2 \in S'$, we have $t_{j,1}, t_{j,2} \in [-1,1]$ such that

$$(R_j^{-1} Y_j, y_j, t_{j,1}h_j, t_{j,2}h_j) \xrightarrow{\mathrm{GH}} (Y', y', \beta_1, \beta_2).$$

If $t_{j,1} + t_{j,2} \in [-1,1]$, then it is clear that $\beta_1\beta_2 \in S'$. If not, we write

$$t_{j,1} + t_{j,2} = \pm 1 + o_j,$$

where $o_j \in [-1,1]$. The sequence $o_j h_j \in S_j$ subconverges to a limit $\beta_0 \in S'$. Then

$$\beta_1\beta_2 y' = \lim_{j\to\infty} (o_j h_j)\cdot(\pm h_j) y_j = \beta_0 y' \in S'y'.$$

Since the set $S'y'$ is closed under multiplication and is contained in $\overline{B}_1(y')$, by Lemma 4.7, we obtain $S'y' = y'$. On the other hand, by the construction of S_j and R_j, $S'y'$ must have a point at distance 1 from y'; a contradiction. □

Lemma 4.9 *Given $s, \epsilon \in (0,1)$, there exists a constant $L_0(\widetilde{M}, \gamma, s, \epsilon)$ such that for any $(Y, y, H) \in \Omega(\widetilde{M}, \langle\gamma\rangle)$ and any $h \in H_0$ with $d(hy, y) = 1$, there exists an integer $2 \le L \le L_0$ with*

$$L^{1-s} \cdot d\Big(\Big(\frac{1}{L}h\Big)y, y\Big) \le \epsilon.$$

Proof We argue by contradiction. Suppose that for each integer $L_j = j$, there are $(Y_j, y_j, H_j) \in \Omega(\widetilde{M}, \langle\gamma\rangle)$ and $h_j \in H_j$ such that $d(h_j y_j, y_j) = 1$ and

$$L^{1-s} \cdot d\Big(\Big(\frac{1}{L}h_j\Big)y_j, y_j\Big) > \epsilon$$

for all $2 \le L \le L_j$. After passing to a subsequence, we consider the convergence

$$(Y_j, y_j, H_j, h_j) \xrightarrow{\mathrm{GH}} (Y', y', H', h').$$

Claim *For any integer $L \ge 2$, we have*

$$\Big(Y_j, \Big(\frac{1}{L}h_j\Big)y_j\Big) \xrightarrow{\mathrm{GH}} \Big(Y', \Big(\frac{1}{L}h'\Big)y'\Big).$$

In fact, due to Lemma 4.8, there is a constant C_1 such that

$$d\Big(\Big(\frac{1}{L}h_j\Big)y_j, y_j\Big) \le C_1$$

for any integer $L \ge 2$. Thus after passing to a subsequence, we can assume that $(1/L)h_j$ converges to some limit isometry $\beta \in H'$ as $j \to \infty$. Note that

$$\beta^L y' = \lim_{j\to\infty}\Big(\frac{1}{L}h_j\Big)^L y_j = \lim_{j\to\infty} h_j y_j = h'y'.$$

Applying Lemma 3.8, we see that $\beta y' = ((1/L)h')y'$ and the claim follows.

The above claim and the hypothesis together imply that for any integer $L \ge 2$,

$$L^{1-s}\cdot d\Big(\Big(\frac{1}{L}h'\Big)y', y'\Big) = \lim_{j\to\infty} L^{1-s}\cdot d\Big(\Big(\frac{1}{L}h_j\Big)y_j, y_j\Big) \ge \epsilon.$$

Thus

$$L\cdot d\Big(\Big(\frac{1}{L}h'\Big)y', y'\Big) \ge L^s\epsilon \to \infty$$

as $L \to \infty$. This shows that in (Y', y', H'), the path $P(h')y'$ from y' to $h'y'$ has infinite length, which cannot be true since $P(h')y'$ comes from an $\mathbb{R}$–orbit of some isometric actions embedded in a Euclidean factor $\mathbb{R}^k$. □

5 Almost linear growth and virtual abelianness

We prove the almost linear growth estimate (Theorem 5.3) and Theorem 1.1 in this section.

In Lemmas 5.1 and 5.2 below, we always assume that the manifold M satisfies the assumptions in Theorem 1.1(1) and the fundamental group is an infinite nilpotent group N. We fix γ as an element of infinite order in N. The purpose of Lemma 5.2 is to transfer Lemma 4.9, as an estimate in the asymptotic limits, to an estimate on $\widetilde{M}$ at large scale.

Lemma 5.1 *Let $b_i \to \infty$ be a sequence of positive integers and let $r_i = d(\gamma^{b_i}\tilde{p}, \tilde{p})$. We consider the convergence*

$$(r_i^{-1}\widetilde{M}, \tilde{p}, \langle\gamma\rangle, \gamma^{b_i}) \xrightarrow{\mathrm{GH}} (Y, y, H, h).$$

Then for any integer $L \in \mathbb{Z}_+$, we have

$$(r_i^{-1}\widetilde{M}, \gamma^{\lceil b_i/L\rceil}\tilde{p}) \xrightarrow{\mathrm{GH}} \Big(Y, \Big(\frac{1}{L}h\Big)y\Big),$$

where $\lceil\cdot\rceil$ means the ceiling function.

The above statement also holds if one replaces the power $\lceil b_i/L\rceil$ by $\lfloor b_i/L\rfloor$. We use the ceiling function in Lemma 5.1 for later applications.

Proof of Lemma 5.1 The statement of Lemma 5.1 is to some extent similar to the claim in the proof of Lemma 4.9, but at the moment we don't have an estimate similar to Lemma 4.8 on the sequence. So, we need to first derive a similar estimate on the distance.

Claim *There is a number C such that*

$$r_i^{-1} d(\gamma^{\lceil b_i/L \rceil} \tilde{p}, \tilde{p}) \le C$$

for all i.

This claim assures that $\gamma^{\lceil b_i/L \rceil} \tilde{p}$ subconverges to some limit point in Y. Suppose the contrary, that is,

$$r_i^{-1} d(\gamma^{\lceil b_i/L \rceil} \tilde{p}, \tilde{p}) \to \infty.$$

For each i, we put

$$R_i := \max_{m=1,\dots,b_i} d(\gamma^m \tilde{p}, \tilde{p}).$$

If follows from the hypothesis that $r_i^{-1} R_i \to \infty$. We consider an asymptotic cone from the sequence R_i,

$$(R_i^{-1} \widetilde{M}, \tilde{p}, \langle \gamma \rangle, \gamma^{b_i}, S_\gamma(b_i)) \xrightarrow{\mathrm{GH}} (Y', y', h', B),$$

where

$$S_\gamma(b_i) = \{\mathrm{id}, \gamma^{\pm 1}, \dots, \gamma^{\pm b_i}\}$$

and $d(h'y', y') = 0$. By Lemma 4.7 and the same argument as Claim 4 in the proof of Proposition 4.1, we see that By' is closed under multiplication and thus $By' = y'$. On the other hand, by the construction $S_\gamma(b_i)$ and R_i, By' should contain a point with distance 1 to y'. This contradiction verifies the claim.

For convenience, below we write $m_i = \lceil b_i/L \rceil$. With the claim, we can pass to a subsequence such that

$$(r_i^{-1} \widetilde{M}, \tilde{p}, \gamma^{m_i}) \xrightarrow{\mathrm{GH}} (Y, y, \alpha),$$

where $\alpha \in H$. Since

$$Lm_i - L \le b_i \le Lm_i$$

for each i, we can pass to a subsequence such that $b_i = Lm_i - K$, where K is some integer between 0 and L. Thus

$$\gamma^{b_i} = \gamma^{Lm_i} \cdot \gamma^{-K} \xrightarrow{\mathrm{GH}} \alpha^L \cdot \beta$$

for some $\beta \in H$ with $\beta y = y$. Recall that $h \in H$ is the limit of γ^{b_i}. It follows that $\alpha^L \beta = h$ and

$$\left(\frac{1}{L} h\right)^L y = hy = \alpha^L \beta y = \alpha^L y.$$

Applying Lemma 3.8, we conclude that $((1/L)h)y = \alpha y$, that is, $\gamma^{m_i} \tilde{p} \xrightarrow{\mathrm{GH}} ((1/L)h)y$. $\square$

Lemma 5.2 *Give $s \in (0,1)$, there are constants $L_0 = L_0(\widetilde{M}, \gamma, s)$ and $R_0 = R_0(\widetilde{M}, \gamma, s)$ such that for all $b \in \mathbb{Z}_+$ with $|\gamma^b| \ge R_0$, there is some integer $2 \le L \le L_0$ with*

$$|\gamma^b| \ge L^{1-s} \cdot |\gamma^{\lceil b/L \rceil}|,$$

where $\lceil \cdot \rceil$ means the ceiling function.

Proof Let $L_0 = L_0\big(\widetilde{M}, \gamma, s, \frac{1}{2}\big)$, the constant in Lemma 4.9. We argue by contradiction to prove the statement. Suppose that there is a sequence $b_i \to \infty$ such that

$$|\gamma^{b_i}| \le L^{1-s} \cdot |\gamma^{\lceil b_i/L \rceil}|$$

for all $L = 2, \dots, L_0$. Let $r_i = |\gamma^{b_i}| \to \infty$. We consider

$$(r_i^{-1}\widetilde{M}, \widetilde{p}, \langle\gamma\rangle, \gamma^{b_i}) \xrightarrow{\mathrm{GH}} (Y, y, H, h),$$

where $h \in H$ satisfies $d(hy, y) = 1$. For each integer $L \geq 2$, by Lemma 5.1, $\gamma^{\lceil b_i/L\rceil}\widetilde{p} \xrightarrow{\mathrm{GH}} ((1/L)h)y$. Together with the hypothesis, we deduce

$$d\Big(\Big(\frac{1}{L}h\Big)y, y\Big) = \lim_{i\to\infty} \frac{d(\gamma^{\lceil b_i/L\rceil}\widetilde{p}, \widetilde{p})}{d(\gamma^{b_i}\widetilde{p}, \widetilde{p})} \geq \Big(\frac{1}{L}\Big)^{1-s}$$

for all $L \in \{2, \dots, L'\}$. On the other hand, by the choice $L_0 = L_0\big(\widetilde{M}, \gamma, s, \frac{1}{2}\big)$ and Lemma 4.9, we have

$$d\Big(\Big(\frac{1}{L}h\Big)y, y\Big) \leq \frac{1}{2}\cdot\Big(\frac{1}{L}\Big)^{1-s}$$

for some $L \in \{2, \dots, L_0\}$, a contradiction. □

We are ready to prove the almost linear growth estimate.

Theorem 5.3 *Let M be an open n–manifold with the assumptions in Theorem 1.1(1). Suppose that its fundamental group is an infinite nilpotent group, denoted by N. Let $\gamma \in N$ be an element of infinite order. Given any $s \in (0,1)$, there are positive constants $C_0 = C_0(\widetilde{M}, \gamma, s)$ and $P_0 = P_0(\widetilde{M}, \gamma, s)$ such that*

$$|\gamma^b| \geq C_0 \cdot b^{1-s}$$

holds for all integers $b > P_0$.

Proof Let P_0 be a large constant such that $|\gamma^b| \geq R_0(\widetilde{M}, \gamma, s)$ for all $b \geq P_0$, where $R_0(\widetilde{M}, \gamma, s)$ is the corresponding constant in Lemma 5.2.

Let $b > P_0$. By Lemma 5.2, we have

$$|\gamma^b| \geq L_1^{1-s} \cdot |\gamma^{\lceil b/L_1\rceil}|$$

for some integer $2 \leq L_1 \leq L_0$, where $L_0 = L_0(\widetilde{M}, \gamma, s)$ is the constant in Lemma 5.2. If $\lceil b/L_1\rceil < P_0$, then we stop right here. If not, we can apply Lemma 5.2 again to find some integer $2 \leq L_2 \leq L_0$ such that

$$|\gamma^b| \geq L_1^{1-s} \cdot |\gamma^{\lceil b/L_1\rceil}| \geq (L_1L_2)^{1-\epsilon} \cdot |\gamma^{\lceil\lceil b/L_1\rceil/L_2\rceil}|.$$

Repeating this process, we eventually derive

$$|\gamma^b| \geq \Big(\prod_{j=1}^k L_j\Big)^{1-s} \cdot |\gamma^{\lceil\cdots\lceil b/L_1\rceil/L_2\cdots/L_k\rceil}| \geq \Big(\prod_{j=1}^k L_j\Big)^{1-s} \cdot r_0,$$

where $\big\lceil\cdots\lceil b/L_1\rceil/L_2\cdots/L_k\big\rceil < P_0$ and $r_0 = \min_{m\in\mathbb{Z}_+} |\gamma^m| > 0$. Noting that

$$b\Big/\Big(\prod_{j=1}^k L_j\Big) \leq \big\lceil\cdots\lceil b/L_1\rceil/L_2\cdots/L_k\big\rceil < P_0,$$

we result in

$$|\gamma^b| \geq \Big(\frac{b}{P_0}\Big)^{1-s} \cdot r_0 = C_0 \cdot b^{1-s},$$

where $C_0 = r_0/(P_0^{1-s})$. □

Remark We compare the almost linear growth estimate and its proof with the methods in the small escape rate case [13].

When the escape rate is very small, for any $(Y, y, G) \in \Omega(\widetilde{M}, N)$, the orbit Gy is Gromov–Hausdorff close to a Euclidean space; see [13, Theorem 0.1]. This almost Euclidean orbit implies that an almost translation estimate:

$$|\gamma^{2b}| \geq 1.9 \cdot |\gamma^b|$$

holds for all b large (see [13, Lemma 4.7]), which is stronger than the almost linear growth estimate here. Also, [13] does not require a description of $\Omega(\widetilde{M}, \langle\gamma\rangle)$; knowing Gy as almost Euclidean orbit is sufficient for its proof.

To derive virtual abelianness from the almost linear growth in Theorem 5.3, we require the following standard result from group theory:

Lemma 5.4 *Let Γ be a group generated by at most m many elements. Suppose that the commutator subgroup $[\Gamma, \Gamma]$ is finite and has at most k elements. Then the center $Z(\Gamma)$ has index at most $C(k, m)$ in Γ.*

Proof We include a proof here for readers' convenience. Let $\{\gamma_1, \dots, \gamma_l\}$ be a set of generators of Γ, where $l \leq m$. Let $Z(\gamma_j)$ be the subgroup consisting of all elements in Γ that commute with γ_j. By assumptions, there are at most k elements in Γ conjugating to γ_j because

$$g\gamma_j g^{-1} = [g, \gamma_j] \cdot \gamma_j.$$

Thus $[\Gamma : Z(\gamma_j)] \leq k$. Noting that

$$Z(\Gamma) = \bigcap_{j=1}^{l} Z(\gamma_j),$$

we conclude

$$[\Gamma : Z(\Gamma)] \leq k^l \leq k^m. \qquad \square$$

Lemma 5.5 *Let (M, p) be an open n–manifold with $\mathrm{Ric} \geq 0$ and $E(M, p) \neq \frac{1}{2}$. Suppose that*

(1) *its Riemannian universal cover is conic at infinity,*

(2) *$N = \pi_1(M, p)$ is nilpotent.*

Then the commutator subgroup $[N, N]$ is finite.

Proof The proof is similar to [10, Lemma 4.7]. The difference is that here we use the almost linear growth estimate in Theorem 5.3 instead of the almost translation estimate in [10, Lemma 4.5]. We include the proof for completeness.

Let

$$N = C_0(N) \triangleright C_1(N) \triangleright \cdots \triangleright C_l(N) = \{e\}$$

be the lower central series of N. We prove the following statement by a reverse induction in k: if $C_{k+1}(N)$ is finite, then $C_k(N)$ is also finite. Because N is nilpotent, it suffices to show that any element of the form $[\alpha, \beta]$ has finite order, where $\alpha \in N$ and $\beta \in C_{k-1}(N)$.

We argue by contradiction and suppose that for some $\alpha \in N$ and $\beta \in C_{k-1}(N)$, $[\alpha, \beta]$ has infinite order. By the triangle inequality,

$$|[\alpha^b, \beta^b]| \le 2b(|\alpha| + |\beta|)$$

for all $b \in \mathbb{Z}_+$. On the other hand, we can apply Theorem 5.3 to obtain a lower bound for large b as follows. We can write

$$[\alpha^b, \beta^b] = [\alpha, \beta]^{b^2} \cdot h,$$

where $h \in C_{k+1}(N)$; see [10, Lemma 4.4]. By the inductive assumption that $C_{k+1}(N)$ is finite, there is $D > 0$ such that $|h| \le D$ for all $h \in C_{k+1}(N)$. Let $s = \frac{1}{4}$ and let $P_0 = P_0(\widetilde{M}, [\alpha, \beta], s)$ be the constant in Theorem 5.3. The triangle inequality and Theorem 5.3 lead to

$$|[\alpha, \beta]^{b^2} \cdot h| \ge |[\alpha, \beta]^{b^2}| - |h| \ge C \cdot (b^2)^{1-s} - D$$

for all $b^2 > P_0$, where C_0 is independent of b. Therefore, we derive that

$$C_0 \cdot b^{2-2s} - D \le 2b(|\alpha| + |\beta|)$$

holds for all b large. Recall that we have chosen $s = \frac{1}{4}$. Then the above inequality clearly results in a contradiction when b is sufficiently large. □

Proof of Theorem 1.1(1) By [9; 7], we can choose a normal nilpotent subgroup N of $\pi_1(M, p)$ with finite index. Let $\widehat{M} = \widetilde{M}/N$ be a covering space of M and let $\hat{p} \in \widehat{M}$ be a lift of $p \in M$. By Lemma 2.5, $E(\widehat{M}, \hat{p}) \neq \frac{1}{2}$. Applying Lemma 5.5 to $(\widehat{M}, \hat{p})$, we conclude that $[N, N]$ is finite. Thus the center $Z(N)$ has finite index in N by Lemma 5.4. Now the result immediately follows since $Z(N)$ has finite index in $\pi_1(M, p)$. □

To prove the universal index bound in Theorem 1.1(2), we use the results below from [8; 10].

Theorem 5.6 [8] *Given $n \in \mathbb{N}$, there are constants $C_1(n)$ and $C_2(n)$ such that the following holds.*

Let M be an open n–manifold of $\mathrm{Ric} \ge 0$ and a finitely generated $\pi_1(M)$. Then:

(1) *$\pi_1(M)$ can be generated by at most $C_1(n)$ many elements.*
(2) *$\pi_1(M)$ contains a normal nilpotent subgroup of index at most $C_2(n)$ and nilpotency length at most n.*

Theorem 5.7 [10] *Given $n \in \mathbb{N}$ and $L \in (0, 1]$, there exists a constant $C(n, L)$ such that the following holds.*

Let M be an open n–manifold of $\mathrm{Ric} \ge 0$. Suppose that:

(1) *$\widetilde{M}$ has Euclidean volume growth of constant at least L.*

(2) *$\Gamma = \pi_1(M, p)$ is finitely generated and nilpotent with nilpotency length $\le n$.*

(3) *$\#[\Gamma, \Gamma]$ is finite.*

Then $\#[\Gamma, \Gamma] \le C(n, L)$.

Proof of Theorem 1.1(2) According to Theorem 5.6(2), we can choose be a normal nilpotent subgroup N of $\pi_1(M, p)$ of index at most $C_1(n)$ and nilpotency length at most n. When $\pi_1(M)$ is finite, surely $[N, N]$ is also finite; when $\pi_1(M)$ is infinite and $E(M, p) \neq \frac{1}{2}$, we apply Lemmas 2.5 and 5.5 to obtain that $[N, N]$ is finite as well. It follows from Theorem 5.7 that the order of $[N, N]$ is bounded by some constant $C_2(n, L)$. Also, Theorem 5.6(1) gives a bound $C_3(n)$ on the number of generators of N. Thus by Lemma 5.4, we deduce

$$[N : Z(N)] \le C_4(C_2(n, L), C_3(n)) = C_5(n, L).$$

Therefore,

$$[\pi_1(M, p) : Z(N)] = [\pi_1(M, p) : N] \cdot [N : Z(N)] \le C_1(n) C_5(n, L). \qquad \square$$

Appendix A nilpotent group with abelian asymptotic limits

In this appendix, we slightly modify Wei's example [16] to construct an open manifold M with $\mathrm{Ric} > 0$ and verify that M satisfies following properties:

(1) $\pi_1(M)$ is the discrete Heisenberg 3–group; and.

(2) For any $(Y, y, G) \in \Omega(\widetilde{M}, \pi_1(M, p))$, the limit group G is abelian.

This example demonstrates that the nilpotency length of Γ may not be preserved in the asymptotic limits.

Let $\widetilde{N}$ be the simply connected 3–dimensional Heisenberg group and let Γ be the discrete Heisenberg 3–group; that is,

$$\widetilde{N} = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \,\middle|\, a, b, c \in \mathbb{R} \right\} \quad \text{and} \quad \Gamma = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \,\middle|\, a, b, c \in \mathbb{Z} \right\} \subseteq \widetilde{N}.$$

The Lie algebra of $\widetilde{N}$ has a basis

$$X_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with $[X_1, X_2] = X_3$ as the only nontrivial Lie bracket. Given $\alpha > 0$ and $\beta \ge 1$, we assign a family of norms $\|\cdot\|_r$, where $r \in [0, \infty)$ is the parameter, on this Lie algebra by

$$\|X_1\|_r = \|X_2\|_r = (1 + r^2)^{-\alpha} \quad \text{and} \quad \|X_3\|_r = (1 + r^2)^{-\beta/2 - 2\alpha}.$$

The family of norms $\|\cdot\|_r$ uniquely determines a family of left-invariant Riemannian metrics $\widetilde{g}_r$ on $\widetilde{N}$, and $\widetilde{g}_r$ satisfies an almost nonnegative Ricci curvature bound

$$\mathrm{Ric}(\widetilde{g}_r) \geq -C(1+r^2)^{-\beta},$$

where C is a positive constant. Let $N_r = (N, g_r)$ be the quotient Riemannian manifold $(\widetilde{N}, \widetilde{g}_r)/\Gamma$.

Next, we construct an open Riemannian manifold (M, g) as a warped product

$$M = [0,\infty) \times_f S^p \times N_r, \quad g = dr^2 + f(r)^2 ds_p^2 + g_r,$$

where (S^p, ds_p^2) is the standard p–dimensional sphere and

$$f(r) = r(1+r^2)^{-1/4}.$$

Following the calculation in [16], one can verify that (M, g) has positive Ricci curvature when p is sufficiently large (depending on α and β).

Let $p \in M$ at $r = 0$. We explain that for $\beta > 1$, the above constructed open manifold (M, p) satisfies the required condition (2). Let

$$\gamma_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \gamma_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

be elements in $\pi_1(M, p) = \Gamma$. Following the method in [14, Lemma 1.1], one can verify that the length estimates

$$|\gamma_1^l| = |\gamma_2^l| \sim l^{1/(1+2\alpha)} \quad \text{and} \quad \big|[\gamma_1^l, \gamma_2^l]\big| = |\gamma_3^{(l^2)}| \sim (l^2)^{1/(1+\beta+4\alpha)}$$

hold for all l large. When $\beta > 1$, $\big|[\gamma_1^l, \gamma_2^l]\big|$ is much shorter than $|\gamma_1^l|$ and $|\gamma_2^l|$ as $l \to \infty$. Below, we fix a $\beta > 1$. Let $r_i \to \infty$ be a sequence and consider an equivariant asymptotic cone

$$(r_i^{-1}\widetilde{M}, \tilde{p}, N, \langle\gamma_3\rangle) \xrightarrow{\mathrm{GH}} (Y, y, G, H).$$

By construction, it is clear that H is a closed $\mathbb{R}$–subgroup of G. Let $l_i \to \infty$ be a sequence of integers such that

$$(r_i^{-1}\widetilde{M}, \tilde{p}, \gamma_1^{l_i}, \gamma_2^{l_i}) \xrightarrow{\mathrm{GH}} (Y, y, g_1, g_2),$$

where $g_1, g_2 \in G$ satisfy

$$d(g_1 y, y) = d(g_2 y, y) = 1.$$

It follows from the length estimates that $[g_1, g_2]y = y$. Note that $[g_1, g_2]$ is also the limit of $\gamma_3^{(l_i^2)}$; thus, $[g_1, g_2] \in H$. Because H is a closed $\mathbb{R}$–subgroup of G, we see that $[g_1, g_2] = \mathrm{id}$; in other words, G is abelian.

As a side note, we mention that by a similar argument in [14], one can check that the orbit Hy has Hausdorff dimension $1 + \beta + 4\alpha \geq 2$. This supports Conjecture 1.3.

References

[1] **J Cheeger**, **T H Colding**, *Lower bounds on Ricci curvature and the almost rigidity of warped products*, Ann. of Math. 144 (1996) 189–237 MR Zbl

[2] **J Cheeger**, **T H Colding**, *On the structure of spaces with Ricci curvature bounded below, I*, J. Differential Geom. 46 (1997) 406–480 MR Zbl

[3] **J Cheeger**, **D Gromoll**, *The splitting theorem for manifolds of nonnegative Ricci curvature*, J. Differential Geom. 6 (1971) 119–128 MR Zbl

[4] **J Cheeger**, **D Gromoll**, *On the structure of complete manifolds of nonnegative curvature*, Ann. of Math. 96 (1972) 413–443 MR Zbl

[5] **T H Colding**, **A Naber**, *Sharp Hölder continuity of tangent cones for spaces with a lower Ricci curvature bound and applications*, Ann. of Math. 176 (2012) 1173–1229 MR Zbl

[6] **K Fukaya**, **T Yamaguchi**, *The fundamental groups of almost non-negatively curved manifolds*, Ann. of Math. 136 (1992) 253–333 MR Zbl

[7] **M Gromov**, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. 53 (1981) 53–73 MR Zbl

[8] **V Kapovitch**, **B Wilking**, *Structure of fundamental groups of manifolds with Ricci curvature bounded below*, preprint (2011) arXiv 1105.5955

[9] **J Milnor**, *A note on curvature and fundamental group*, J. Differential Geom. 2 (1968) 1–7 MR Zbl

[10] **J Pan**, *Nonnegative Ricci curvature, almost stability at infinity, and structure of fundamental groups*, preprint (2018) arXiv 1809.10220

[11] **J Pan**, *Nonnegative Ricci curvature, stability at infinity and finite generation of fundamental groups*, Geom. Topol. 23 (2019) 3203–3231 MR Zbl

[12] **J Pan**, *On the escape rate of geodesic loops in an open manifold with nonnegative Ricci curvature*, Geom. Topol. 25 (2021) 1059–1085 MR Zbl

[13] **J Pan**, *Nonnegative Ricci curvature and escape rate gap*, J. Reine Angew. Math. 782 (2022) 175–196 MR Zbl

[14] **J Pan**, **G Wei**, *Examples of Ricci limit spaces with non-integer Hausdorff dimension*, Geom. Funct. Anal. 32 (2022) 676–685 MR Zbl

[15] **C Sormani**, *Nonnegative Ricci curvature, small linear diameter growth and finite generation of fundamental groups*, J. Differential Geom. 54 (2000) 547–559 MR Zbl

[16] **G Wei**, *Examples of complete manifolds of positive Ricci curvature with nilpotent isometry groups*, Bull. Amer. Math. Soc. 19 (1988) 311–313 MR Zbl

Fields Institute for Research in Mathematical Sciences
Toronto, ON, Canada

Current address: *Mathematics Department, University of California, Santa Cruz*
Santa Cruz, CA, United States

jpan53@ucsc.edu

Proposed: Tobias H Colding
Seconded: David Fisher, Urs Lang

Received: 3 February 2022
Revised: 7 July 2022

Geometry & Topology **28**:3 (2024) 1437–1499
DOI: 10.2140/gt.2024.28.1437
Published: 10 May 2024

The homology of the Temperley–Lieb algebras

Rachael Boyd
Richard Hepworth

We study the homology and cohomology of the Temperley–Lieb algebra $\mathrm{TL}_n(a)$, interpreted as appropriate Tor and Ext groups. Our main result applies under the common assumption that $a = v + v^{-1}$ for some unit v in the ground ring, and states that the homology and cohomology vanish up to and including degree $n-2$. To achieve this we simultaneously prove homological stability and compute the stable homology. We show that our vanishing range is sharp when n is even.

Our methods are inspired by the tools and techniques of homological stability for families of groups. We construct and exploit a chain complex of "planar injective words" that is analogous to the complex of injective words used to prove stability for the symmetric groups. However, in this algebraic setting we encounter a novel difficulty: $\mathrm{TL}_n(a)$ is not flat over $\mathrm{TL}_m(a)$ for $m < n$, so that Shapiro's lemma is unavailable. We resolve this difficulty by constructing what we call "inductive resolutions" of the relevant modules.

Vanishing results for the homology and cohomology of Temperley–Lieb algebras can also be obtained from the existence of the Jones–Wenzl projector. Our own vanishing results are in general far stronger than these, but in a restricted case we are able to obtain additional vanishing results via the existence of the Jones–Wenzl projector.

We believe that these results, together with the second author's work on Iwahori–Hecke algebras, are the first time the techniques of homological stability have been applied to algebras that are not group algebras.

20J06, 16E40; 20F36

1 Introduction

In this work we study the homology and cohomology of the *Temperley–Lieb algebras*. In particular, we simultaneously prove that the algebras satisfy homological stability, and that their stable homology vanishes.

A sequence of groups and inclusions $G_0 \to G_1 \to G_2 \to \cdots$ is said to satisfy *homological stability* if for each degree d the induced sequence of homology groups

$$H_d(G_0) \to H_d(G_1) \to H_d(G_2) \to \cdots$$

eventually consists of isomorphisms. Homological stability can also be formulated for sequences of spaces. There are many important examples of groups and spaces for which homological stability is known to hold, such as symmetric groups [Nakaoka 1960], general linear groups [Charney 1980; Maazen 1979; van der Kallen 1980], mapping class groups of surfaces [Harer 1985; Randal-Williams 2016] and 3–manifolds [Hatcher and Wahl 2010], automorphism groups of free groups [Hatcher and Vogtmann 1998; 2004], diffeomorphism groups of high-dimensional manifolds [Galatius and Randal-Williams 2018], configuration spaces [Church 2012; Randal-Williams 2013], Coxeter groups [Hepworth 2016], Artin monoids [Boyd 2020], and many more. In almost all cases, homological stability is one of the strongest things we know about the homology of these families. It is often coupled with computations of the stable homology $\lim_{n\to\infty} H_*(G_n)$, which is equal to the homology of the G_n in the *stable range* of degrees, ie those degrees for which stability holds.

The homology and cohomology of a group G can be expressed in the language of homological algebra as

$$H_*(G) = \mathrm{Tor}_*^{RG}(\mathbb{1}, \mathbb{1}) \quad \text{and} \quad H^*(G) = \mathrm{Ext}_{RG}^*(\mathbb{1}, \mathbb{1}),$$

where R is the coefficient ring for homology and cohomology, RG is the group algebra of G and $\mathbb{1}$ is its trivial module. Thus the homology and cohomology of a group depend only on the group algebra RG and its trivial module $\mathbb{1}$. It is therefore natural to consider the homology and cohomology of an arbitrary algebra equipped with a "trivial" module. Moreover, one may ask whether homological stability occurs in this wider context.

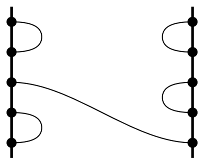
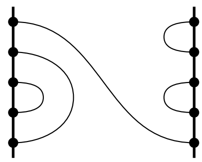
Hepworth [2022] proved homological stability for *Iwahori–Hecke* algebras of type A. These are deformations of the group rings of the symmetric groups that are important in representation theory, knot theory and combinatorics. There is a fairly standard suite of techniques used to prove homological stability, albeit with immense local variation, and the proof strategy of [Hepworth 2022] followed all the steps familiar from the setting of groups. As is typical, the hardest step was to prove that the homology of a certain (chain) complex vanishes in a large range of degrees.

In the present paper we will prove homological stability for the *Temperley–Lieb algebras*, and we will prove that the stable homology vanishes. However amongst the familiar steps in our proof lies a novel obstacle and — to counter it — a novel construction. At a certain point the usual techniques fail because Shapiro's lemma cannot be applied, as we will explain below. This is a new difficulty that never occurs in the setting of groups, but we are able to resolve it for the algebras at hand, and in fact our solution facilitates the unusually strong results that we are able to obtain. It is not surprising that the Iwahori–Hecke case is more straightforward than the Temperley–Lieb case: Iwahori–Hecke algebras are deformations of group rings, whereas the Temperley–Lieb algebras are significantly different.

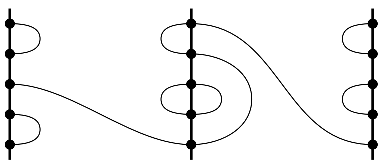
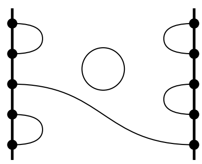
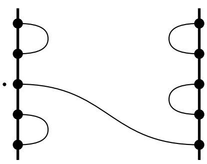
To the best of our knowledge, the present paper and [Hepworth 2022] are the first time the techniques of homological stability have been applied to algebras that are not group algebras, and together they serve as proof-of-concept for the export of homological stability techniques to the setting of algebras. The moral of [Hepworth 2022] is that the "usual" techniques of homological stability suffice, so long as the algebras involved satisfy a certain flatness condition. The moral of the present paper is that failure of the flatness condition can in some cases be overcome, using new ingredients and techniques, and can even lead to stronger results than in the flat scenario. Since the completion of this paper, we have extended our techniques to study the homology of the Brauer algebras in joint work with Patzt [Boyd et al. 2021].

1.1 Temperley–Lieb algebras

Let $n \geqslant 0$, let R be a commutative ring, and let $a \in R$. The *Temperley–Lieb algebra* $\mathrm{TL}_n(a)$ is the R–algebra with basis (by which we will always mean R–module basis) given by the planar diagrams on n strands, taken up to isotopy, and with multiplication given by pasting diagrams and replacing closed loops with factors of a. The last sentence was intentionally brief, but we hope that its meaning becomes clearer with an illustration of two elements $x, y \in \mathrm{TL}_5(a)$

$x =$  and $y =$ 

and their product

$x \cdot y =$  $=$  $= a \cdot$ 

The Temperley–Lieb algebras [1971] arose in theoretical physics in the 1970s. They were later rediscovered by Jones [1983] in his work on von Neumann algebras, and used in the first definition of the Jones polynomial [1985]. Kauffman [1987; 1990] gave the above diagrammatic interpretation of the algebras.

The Temperley–Lieb algebra $\mathrm{TL}_n(a)$ is perhaps best studied in the case where $a = v + v^{-1}$, for $v \in R$ a unit. In this case, it is a quotient of the Iwahori–Hecke algebra of type A_{n-1} with parameter $q = v^2$ (so it is closely related to the symmetric group) and it receives a homomorphism from the group algebra of the braid group on n strands. It can also be described as the endomorphism algebra of $V_q^{\otimes n}$, where V_q is a certain 2–dimensional representation of the quantum group $U_q(\mathfrak{sl}_2)$. We recommend [Ridout and Saint-Aubin 2014; Kassel and Turaev 2008] for further reading on $\mathrm{TL}_n(a)$, and [Westbury 1995; Graham and Lehrer 1996] for details on their representation theory.

1.2 Homology of Temperley–Lieb algebras

The Temperley–Lieb algebra $\mathrm{TL}_n(a)$ has a *trivial module* $\mathbb{1}$ consisting of a copy of R on which all diagrams other than the identity diagram act as multiplication by 0. It therefore has homology and cohomology groups $\mathrm{Tor}_*^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$.

Our first result is a vanishing theorem in the case that the parameter $a \in R$ is invertible.

Theorem A *Let R be a commutative ring, and a a unit in R. Then $\mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^d_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$ both vanish for $d > 0$.*

The next result holds regardless of whether or not a is invertible, and uses the common assumption that $a = v + v^{-1}$, with $v \in R^\times$. However, we see shortly that this assumption can be removed.

Theorem B *Let R be a commutative ring, let $v \in R$ be a unit, let $a = v + v^{-1}$, and let $n \geqslant 0$. Then*

$$\mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) = 0 \quad \textit{and} \quad \mathrm{Ext}^d_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) = 0$$

for $1 \leqslant d \leqslant n-2$ if n is even, and for $1 \leqslant d \leqslant n-1$ if n is odd.

Thus the map

$$\mathrm{Tor}_d^{\mathrm{TL}_{n-1}(a)}(\mathbb{1},\mathbb{1}) \to \mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$$

is an isomorphism for $d \leqslant n-3$, so that we have homological stability, and $\lim_{n\to\infty} \mathrm{Tor}_*^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) = 0$ in positive degrees, so the stable homology is trivial. The latter is reminiscent of Quillen's result [1972] on the vanishing stable homology of general linear groups of finite fields in defining characteristic, and of Szymik and Wahl's result [2019] on the acyclicity of the Thompson groups. Theorems A and B might lead us to expect that the homology and cohomology of the $\mathrm{TL}_n(a)$ are largely trivial, but in fact the results are as strong as possible, at least for n even:

Theorem C *In the setting of Theorem B above, suppose further that n is even and that $a = v + v^{-1}$ is not a unit. Then $\mathrm{Tor}_{n-1}^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \neq 0$.*

Thus Theorem A does not extend to the case of a not invertible, and the stable range in Theorem B is sharp. In fact we can say more: when n is even, $\mathrm{Tor}_{n-1}^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \cong R/bR$, where b is a multiple of a (unfortunately our methods do not allow us to say anything more concrete about b).

Remark One can compute $\mathrm{Tor}_1^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$ directly using the method of [Weibel 1994, Exercise 3.1.3]: it is R/aR for $n=2$, and vanishes otherwise. We also compute the homology and cohomology of $\mathrm{TL}_2(a)$ by an explicit resolution: $\mathrm{Tor}_*^{\mathrm{TL}_2(a)}(\mathbb{1},\mathbb{1})$ is R/aR in odd degrees, and the kernel R_a of $r\mapsto ar$ in positive even degrees, so that if a is not invertible then $\mathrm{Tor}_*^{\mathrm{TL}_2(a)}(\mathbb{1},\mathbb{1})$ is nontrivial in infinitely many degrees.

Randal-Williams [2021] showed that in fact you can remove our assumption that $a=v+v^{-1}$ for a unit $v\in R$, by applying Theorem C for an associated ring S. This yields the following strengthening of Theorem B.

Corollary [Randal-Williams 2021, Theorem B′] *Let R be a commutative ring, a be any element in R, and $n\geqslant 0$. Then*

$$\mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})=0 \quad \textit{and} \quad \mathrm{Ext}^d_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})=0$$

for $1\leqslant d\leqslant n-2$ if n is even, and for $1\leqslant d\leqslant n-1$ if n is odd.

Proof The full proof can be found in [Randal-Williams 2021], and uses the base change spectral sequence [Weibel 1994, Section 5.6]. This is applied to the faithfully flat ring homomorphism $R\to S$ where $S=R[v]/(v^2-a\cdot v+1)$, which by construction has a unit v and element a such that $a=v+v^{-1}$. The results in Theorem B for the ring S can now be transferred to analogous results for the ring R. □

1.3 Jones–Wenzl projectors

The *Jones–Wenzl projector* or *Jones–Wenzl idempotent* JW_n, if it exists, is the element of $\mathrm{TL}_n(a)$ uniquely characterised by the following two properties:

- $\mathrm{JW}_n\in 1+I_n$, and
- $\mathrm{JW}_n\cdot I_n=0=I_n\cdot \mathrm{JW}_n$,

where I_n is the two-sided ideal in $\mathrm{TL}_n(a)$ spanned by all diagrams other than the identity diagram. The Jones–Wenzl projector was first introduced by Jones [1983], was further studied by Wenzl [1987], and has since become important in representation theory, knot theory and the study of 3–manifolds.

The Jones–Wenzl projector exists if and only if the trivial module $\mathbb{1}$ is projective. Moreover, when the ground ring R is a field, there is a simple and explicit criterion for the existence of JW_n, given in terms of the parameter a. Thus, when this criterion holds, the vanishing of $\mathrm{Tor}_*^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$ in positive degrees follows immediately.

Our own Theorems A and B are in general far stronger than the vanishing results obtained from the existence of JW_n, as they do not require R to be a field, and the constraints are weaker. Indeed, in the case of n even, Theorems A and C are the final word on vanishing, since they imply that the homology and cohomology of $\mathrm{TL}_n(a)$ vanish in all positive degrees if and only if a is invertible. However, in the case of n odd and R a field, there are some situations where our theorems do not incorporate all vanishing results given by the existence of JW_n. These cases are encapsulated in the following.

Theorem D *Let $n = 2k+1$, and let R be a field whose characteristic does not divide $\binom{k}{t}$ for any $1 \leq t \leq k$. Let v be a unit in R and assume that $a = v + v^{-1} = 0$. Then $\mathrm{Tor}_*^{\mathrm{TL}_n(0)}(\mathbb{1}, \mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(0)}(\mathbb{1}, \mathbb{1})$ vanish in positive degrees.*

As with Theorem B, the assumption that $a = v + v^{-1}$ for v a unit can be removed in this result.

Combining Theorem D with Theorem A yields rather comprehensive vanishing results when R is a field with appropriate characteristic. For example, it now follows that when R is any field, the homology and cohomology of $\mathrm{TL}_3(v + v^{-1})$ vanish regardless of the choice of v. Similarly, the homology and cohomology of $\mathrm{TL}_5(v + v^{-1})$ will vanish over any field and for any value of v, except possibly in characteristic 2 when $v + v^{-1} = 0$. Since the first appearance of our paper, Sroka [2022] has used related techniques to show that when n is odd, the Tor groups vanish in all positive degrees, for any choice of R.

The next few sections of this introduction will discuss the proofs of our main results in some detail.

1.4 Planar injective words

Several proofs of homological stability for the symmetric group [Maazen 1979; Kerz 2005; Randal-Williams 2013] make use of the *complex of injective words*. This is a highly connected complex with an action of the symmetric group $\mathfrak{S}_n$. Our main tool for proving Theorems B and C is the *complex of planar injective words* $W(n)$, a Temperley–Lieb analogue of the complex of injective words that we introduce and study here for the first time. It is a chain complex of $\mathrm{TL}_n(a)$–modules, and in degree i it is given by the tensor product module $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{n-i-1}(a)} \mathbb{1}$. This is analogous to the complex of injective words, whose i–simplices form a single $\mathfrak{S}_n$–orbit with typical stabiliser $\mathfrak{S}_{n-i-1}$, which is an alternative way of saying that the i^{th} chain group is isomorphic to $R\mathfrak{S}_n \otimes_{R\mathfrak{S}_{n-i-1}} \mathbb{1}$. We show the following high-acyclicity result. In order to construct appropriate differentials for $W(n)$ we exploit a homomorphism from the group algebra of the braid group on n strands, which is not necessarily apparent from the definition of $\mathrm{TL}_n(a)$. This is where the restriction of a to $a = v + v^{-1}$ is necessary.

Theorem E *$H_d(W(n))$ vanishes in degrees $d \leqslant n-2$.*

The complex $W(n)$ has rich combinatorial properties, analogous to those of the complex of injective words, that we explore in the companion paper [Boyd and Hepworth 2021]. In particular, Theorem E tells us that the homology of $W(n)$ is concentrated in the top degree $H_{n-1}(W(n))$, and in [Boyd and Hepworth 2021] we show that when R is Noetherian the rank of this top homology group is the n^{th} *Fine number* F_n [Deutsch and Shapiro 2001], an analogue of the number of derangements on n letters. Furthermore we show that the differentials of $W(n)$ encode the *Jacobsthal numbers* [Sloane 2000]. Finally in the semisimple case we show that $H_{n-1}(W(n))$ has descriptions firstly categorifying an alternating sum for the Fine numbers, and secondly in terms of standard Young tableaux. We call the $\mathrm{TL}_n(a)$–module $H_{n-1}(W(n))$ the *Fineberg module*, and we denote it by $\mathscr{F}_n(a)$. We know little about $\mathscr{F}_n(a)$ in general, though in the cases $n = 2, 3, 4$ we give examples describing it in terms of the cell modules of $\mathrm{TL}_n(a)$.

The proof of Theorem E is perhaps the most difficult technical result in this paper. It is obtained by filtering $W(n)$ and showing that the filtration quotients are (suspensions of truncations of) copies of $W(n-1)$, and then proceeding by induction.

1.5 Spectral sequences and Shapiro's lemma

Let us now outline how we use the complex of planar injective words $W(n)$ to prove Theorems B and C. Following standard approaches to homological stability for groups, we consider a spectral sequence obtained from the complex $W(n)$. The E^1–page of our spectral sequence consists of the groups $\mathrm{Tor}_j^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{n-i-1}(a)} \mathbb{1})$. Furthermore, thanks to Theorem E, the spectral sequence converges to $\mathrm{Tor}_{*-n+1}^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathscr{F}_n(a))$, where $\mathscr{F}_n(a) = H_{n-1}(W(n))$ is the Fineberg module. Our experience from homological stability tells us to apply *Shapiro's lemma*, or in this context a *change-of-rings isomorphism*, to identify

$$\mathrm{Tor}_*^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{n-i-1}(a)} \mathbb{1}) \quad \text{with} \quad \mathrm{Tor}_*^{\mathrm{TL}_{n-i-1}(a)}(\mathbb{1}, \mathbb{1}).$$

This identification applied to the columns of our spectral sequence would allow us to implement an inductive hypothesis. However, such a change-of-rings isomorphism would only be valid if $\mathrm{TL}_n(a)$ were flat as a $\mathrm{TL}_{n-i-1}(a)$–module, and this is not the case. This failure of Shapiro's lemma is a potentially serious obstacle to proceeding further. However, we are able to identify the columns of our spectral sequence by independent means, as follows:

Theorem F *Let R be a commutative ring and let $a \in R$. Let $0 \leqslant m < n$. Then*

$$\mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}) \quad \textit{and} \quad \mathrm{Ext}^d_{\mathrm{TL}_n(a)}(\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}, \mathbb{1})$$

both vanish for $d > 0$.

In conjunction with a computation of the $d = 0$ case, this gives us the vanishing results of Theorem B. Moreover, in the case of n even we are able to analyse the rest of the spectral sequence (there is a single differential and a single extension problem) in sufficient detail to prove the sharpness result of Theorem C. This involves a careful study of the Fineberg module $\mathscr{F}_n(a)$. In general, our method identifies $\mathrm{Tor}_*^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathbb{1})$ with $\mathrm{Tor}_{*-n}^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathscr{F}_n(a))$, except in degrees $* = n-1, n$ when n is even.

1.6 Inductive resolutions

It remains for us to discuss the proofs of Theorems A and F. These results are proved by a novel method that exploits the structure of the Temperley–Lieb algebras, and in particular they lie outwith the standard toolkit of homological stability. Moreover, it is Theorem F which allows us to overcome the failure of Shapiro's lemma.

The two theorems are very similar: Theorem A is an instance of the more general statement that $\mathrm{Tor}_*^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1})$ vanishes in positive degrees for $m \leqslant n$ and a invertible, while Theorem F states that the same groups vanish for $m < n$ and a arbitrary. These are both proved by strong induction on m. The initial cases $m = 0, 1$ are immediate because then $\mathrm{TL}_m(a) = R$ so $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ is free.

The induction step is proved by constructing and exploiting a resolution of $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ whose terms have the form $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{m-1}(a)} \mathbb{1}$ and $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{m-2}(a)} \mathbb{1}$, and then applying the inductive hypothesis. We call these resolutions *inductive resolutions* since they resolve the next module in terms of those already considered.

Our technique of inductive resolutions is generalised in [Boyd et al. 2021], where we show that the homology of the Brauer algebras is isomorphic to the homology of the symmetric groups in a stable range when the parameter δ is not invertible, and in every degree when δ is invertible. This provides concrete evidence that the new techniques developed in this paper can be adapted to other algebras to obtain results of similar strength.

1.7 Discussion: homological stability for algebras

As stated earlier, we regard the present paper, together with the results of [Hepworth 2022] on Iwahori–Hecke algebras, as proof-of-concept for the export of the techniques of homological stability to the setting of algebras. And, since the first appearance of this paper, these techniques have been extended to the setting of Brauer algebras in [Boyd et al. 2021]. We hope that the present paper, together with [Hepworth 2022; Boyd et al. 2021], will be a springboard for further research in this direction.

One of the main motivations for studying the homology of groups, is that homology is a useful "measurement" of the group. Put another way, homology is a powerful invariant, where the power comes from the fact that it is both informative, and (relatively) computable. The Tor and Ext groups of algebras are likewise strong invariants, and it is our hope that homology and cohomology of algebras can be utilised as a tool to answer questions in the fields where the algebras arise. For example, modern representation theory is rich in conjectures, and home to surprising isomorphisms between apparently very different algebras [Brundan and Kleshchev 2009; Bowman et al. 2023]. Understanding the similarities and differences between naturally arising algebras is precisely the kind of question that could be investigated via Tor and Ext groups.

We will now discuss some questions arising from our work. Readers with experience in homological stability will be able to think of many new questions in this direction, so we will simply list some that are most prominent in our minds.

The Temperley–Lieb algebra can be regarded as an algebra of one-dimensional cobordisms embedded in two dimensions, and the Brauer algebra can similarly be viewed as an algebra of one-dimensional cobordisms embedded in infinite dimensions.

Question Are there analogues of the Temperley–Lieb algebra consisting of d–dimensional cobordisms embedded in n dimensions? Does homological stability hold for these algebras? And can the stability be understood in an essentially geometric way?

And more generally:

Question For which natural families of algebras does homological stability hold?

Candidate algebras, closely related to the existing cases, are: Iwahori–Hecke and Temperley–Lieb algebras of types B and D; the periodic and dilute Temperley–Lieb algebras; and the blob, partition and Birman–Murakami–Wenzl algebras. We invite the reader to think of possibilities from further afield.

There have recently been advances in building general frameworks for homological stability proofs. Randal-Williams and Wahl [2017] introduce a categorical framework that encapsulates, improves and extends several of the standard techniques used in homological stability proofs for groups. Galatius, Kupers and Randal-Williams [Galatius et al. 2018] introduce a framework that applies to E_k–algebras in simplicial modules. It exploits the notion of cellular E_k–algebras, and incorporates methods for proving *higher stability* results. This invites us to pose the following questions.

Question Does the general homological stability machinery of Randal-Williams and Wahl [2017] generalise to an R–linear version, giving a general framework to prove that a family of R–algebras $A_0 \to A_1 \to A_2 \to \cdots$ satisfies homological stability?

In this question, the most interesting issue is what form the resulting complexes will take. One might expect that for a family of algebras the relevant complexes will be constructed from tensor products, as with our complex $W(n)$. However, it may happen, as in this paper, that flatness issues arise, in which case it seems unlikely that complexes built from the honest tensor products will be sufficient.

Question Can the homological stability machinery of [Galatius et al. 2018] be applied in the setting of algebras?

It seems extremely likely that homology of Temperley–Lieb algebras will indeed fit into the framework of [Galatius et al. 2018], by using appropriate simplicial models for the $\mathrm{Tor}_*^{\mathrm{TL}_n(a)}(\mathbb{1}, \mathbb{1})$, or more precisely for the chain complexes underlying these Tor groups. Again, the difficulty will lie in identifying and computing the associated *splitting complexes*, especially when flatness issues arise.

1.8 Outline

In Section 2 we recall the definition of the Temperley–Lieb algebra, the Jones basis, the relationship with Iwahori–Hecke algebras, and we establish results on the induced modules $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ that will be important in the rest of the paper. Section 3 establishes our inductive resolutions and proves Theorems A and F. Section 4 introduces the complex of planar injective words $W(n)$ and the Fineberg module $\mathscr{F}_n(a)$. Sections 5 and 6 then use $W(n)$, in particular its high acyclicity (Theorem E), to prove Theorems B and C. Section 7 investigates our results in the case of $\mathrm{TL}_2(a)$, computing the homology directly and also in terms of the Fineberg module $\mathscr{F}_2(a)$. Section 8 proves Theorem E. Section 9 investigates the vanishing results given by the Jones–Wenzl projectors and proves Theorem D.

Acknowledgements The authors would like to thank the Max Planck Institute for Mathematics in Bonn for its support and hospitality. We are grateful to Robert Spencer and Paul Wedrich for interesting and helpful discussions about the representation theory of Temperley–Lieb algebras, and to Oscar Randal-Williams for his extension of Theorem B. We would also like to thank the referees for very insightful comments, which have much improved this paper.

2 Temperley–Lieb algebras

In this section we will cover the basic facts about the Temperley–Lieb algebra that we will need for the rest of the paper. There is some overlap between the material recalled here and in [Boyd and Hepworth 2021]. In particular, we cover the definitions by generators and relations and by diagrams; we discuss the Jones basis for $\mathrm{TL}_n(a)$; we look at the induced modules $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ that will be an essential ingredient in all that follows; and we discuss the homomorphism from the Iwahori–Hecke algebra of type A_{n-1} into $\mathrm{TL}_n(a)$. Historical references on Temperley–Lieb algebras were given in the introduction. General references for readers new to the $\mathrm{TL}_n(a)$ are Section 5.7 of Kassel and Turaev's book [2008] on the braid groups, and especially Sections 1 and 2 of Ridout and Saint-Aubin's survey [2014] on the representation theory of the $\mathrm{TL}_n(a)$.

Definition 2.1 (the Temperley–Lieb algebra $\mathrm{TL}_n(a)$) Let R be a commutative ring and let $a \in R$. Let n be a nonnegative integer. The *Temperley–Lieb algebra* $\mathrm{TL}_n(a)$ is defined to be the R–algebra with generators $U_1, \dots, U_{n-1}$ and the relations

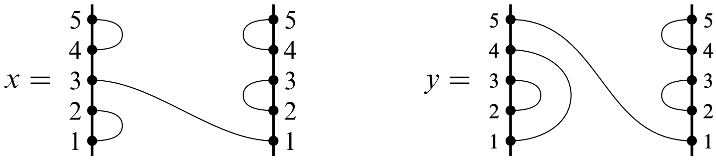
(1) $U_i U_j = U_j U_i$ for $j \neq i \pm 1$,

(2) $U_i U_j U_i = U_i$ for $j = i \pm 1$, and

(3) $U_i^2 = a U_i$ for all i.

Thus elements of the Temperley–Lieb algebra are formal sums of monomials in the U_i, with coefficients in the ground ring R, modulo the relations above. We often write $\mathrm{TL}_n(a)$ as TL_n. We note here that $\mathrm{TL}_0 = \mathrm{TL}_1 = R$.

There is an alternative definition of TL_n in terms of diagrams. In this description, an element of TL_n is an R–linear combination of *planar diagrams* (or one-dimensional cobordisms). Each planar diagram consists of two vertical lines in the plane, decorated with n dots labelled $1, \dots, n$ from bottom to top, together with a collection of n arcs joining the dots in pairs. The arcs must lie between the vertical lines, they must be disjoint, and the diagrams are taken up to isotopy. For example, here are two planar diagrams in the case $n = 5$:



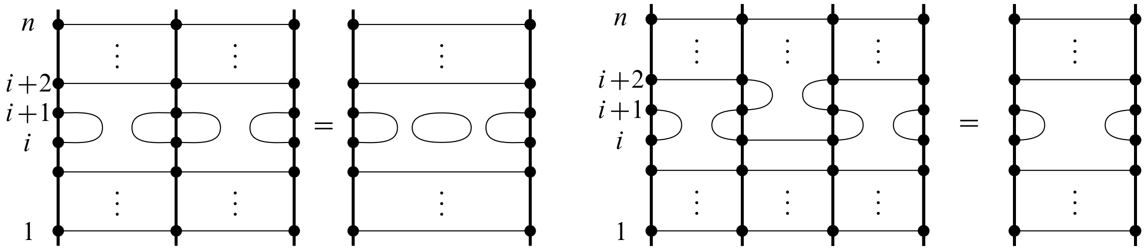
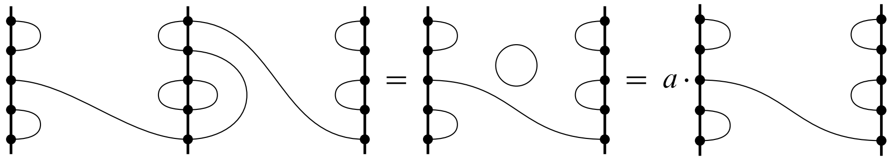


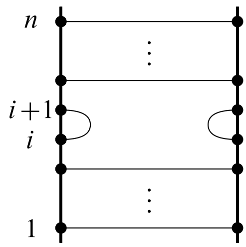
Figure 1: Diagrammatic relations in TL_n: $U_i^2 = aU_i$ (left) and $U_iU_{i+1}U_i = U_i$ (right).

We will often omit the labels on the dots. Multiplication of diagrams is given by placing them side-by-side and joining the ends. Any closed loops created by this process are then erased and replaced with a factor of a. For example, the product xy of the elements x and y above is:



(We have subscribed to the heresy of [Ridout and Saint-Aubin 2014] by drawing planar diagrams that go from left to right rather than top to bottom.)

One can pass from the generators-and-relations definition of TL_n in Definition 2.1 to the diagrammatic description of the previous paragraph as follows. For $1 \leqslant i \leqslant n-1$, to each U_i we associate the planar diagram shown below:



We refer to an arc joining adjacent dots as a *cup*. The relations for the Temperley–Lieb algebras are satisfied, and two of them are illustrated in Figure 1. The fact that this determines an isomorphism between the algebra defined by generators and relations, and the one defined by diagrams, is proved in [Ridout and Saint-Aubin 2014, Theorem 2.4; Kassel and Turaev 2008, Theorem 5.34; Kauffman 2005, Section 6].

In the rest of the paper we will refer to the diagrammatic point of view on the Temperley–Lieb algebra, but we will not rely on it for any proofs.

2.1 The Jones basis

From the diagrammatic point of view the Temperley–Lieb algebra TL_n has an evident R–basis given by the (isotopy classes of) planar diagrams. This is called the *diagram basis*. We now recall the analogue of the diagram basis given in terms of the U_i, which is called the *Jones basis* for TL_n, and we prove some additional facts about it that we will require later. See [Kassel and Turaev 2008, Section 5.7; Ridout and Saint-Aubin 2014, Section 2; Kauffman 2005, Section 6], but note that conventions vary, and see Remark 2.5 below in particular.

Definition 2.2 (Jones normal form) The *Jones normal form* for elements of $\mathrm{TL}_n(a)$ is defined as follows. Let

$$n > a_k > a_{k-1} > \cdots > a_1 > 0 \quad \text{and} \quad n > b_k > b_{k-1} > \cdots > b_1 > 0$$

be integers such that $b_i \geqslant a_i$ for all i. Let $\underline{a} = (a_k, \dots a_1)$ and $\underline{b} = (b_k, \dots b_1)$. Then set

$$x_{\underline{a},\underline{b}} = (U_{a_k} \dots U_{b_k}) \cdot (U_{a_{k-1}} \dots U_{b_{k-1}}) \cdots (U_{a_1} \dots U_{b_1}),$$

where the subscripts of the generators increase in each tuple $U_{a_i} \dots U_{b_i}$. A word written in the form $x_{\underline{a},\underline{b}}$ is said to be written in *Jones normal form* for $\mathrm{TL}_n(a)$.

Example 2.3 In TL_5 the words

$$\begin{aligned}
U_1U_2U_3U_4 &= (U_1U_2U_3U_4) = x_{(1),(4)},\\
U_4U_3U_2U_1 &= (U_4)\cdot(U_3)\cdot(U_2)\cdot(U_1) = x_{(4,3,2,1),(4,3,2,1)},\\
U_3U_4U_1U_2 &= (U_3U_4)\cdot(U_1U_2) = x_{(3,1),(4,2)},\\
U_2U_3U_1U_2 &= (U_2U_3)\cdot(U_1U_2) = x_{(2,1),(3,2)}
\end{aligned}$$

are in Jones normal form. The word $U_2U_1U_4U_2U_3$ is not, but it can be rewritten using the defining relations to give

$$U_2U_1U_4U_2U_3 = U_4U_2U_1U_2U_3 = U_4U_2U_3 = (U_4)(U_2U_3) = x_{(4,2),(4,3)}.$$

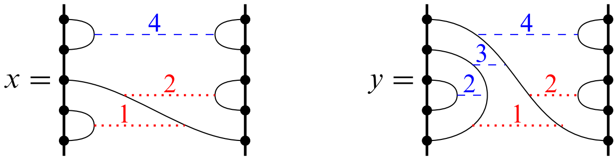
Denote the subset of TL_n consisting of all $x_{\underline{a},\underline{b}}$ with $\underline{a} = (a_1, \dots, a_k)$ and $\underline{b} = (b_1, \dots, b_k)$ by $\mathrm{TL}_{n,k}$. Then the set

$$\mathrm{TL}_{n,0} \sqcup \mathrm{TL}_{n,1} \sqcup \cdots \sqcup \mathrm{TL}_{n,n-1}$$

is a basis (recall that by basis we always mean R–module basis) of TL_n, called the *Jones basis*. For a proof of this fact see [Kassel and Turaev 2008, Corollary 5.32; Ridout and Saint-Aubin 2014, pages 967–969; Kauffman 2005, Section 6], though we again warn the reader that conventions vary.

There is an algorithm for taking a diagram and writing it as an element of the Jones basis; see [Kauffman 2005, Section 6]. We summarise the algorithm here. Let the i^{th} *row* of the diagram be the horizontal strip whose left and right ends lie between the dots i and $i+1$ on each vertical line. Take a planar diagram, and ensure that it is drawn in minimal form: all arcs connecting the same side of the diagram to itself are drawn as semicircles, and all arcs from left to right are drawn without any cups, ie transverse to all vertical lines, and such that each arc of the diagram intersects each row transversely and at most once.

Proceed along each row of the diagram, connecting the consecutive arcs encountered with a dotted horizontal line labelled by the row in question. This is done in an alternating fashion: the first arc encountered is connected to the second by a dotted line, then the third is connected to the fourth, and so on. If we start with the elements x and y used earlier in this section, this gives us the following:



A *sequence* in such a decorated diagram is taken by travelling right along the dotted arcs and up along the solid arcs from one dotted arc to the next, starting as far to the left as possible. The above diagrams each have two sequences, indicated in dashed and dotted lines. The sequences in a diagram are linearly ordered by scanning from top to bottom and recording a sequence when one of its dotted lines is first encountered. So in the above diagrams the dashed sequences precede the dotted ones. One now obtains a Jones normal form for the element by working through the sequences in turn, writing out the labels from left to right, and then taking the corresponding monomial in the U_i:

$$x = (U_4)(U_1U_2) = x_{(4,1),(4,2)}, \quad y = (U_2U_3U_4)(U_1U_2) = x_{(2,1),(4,2)}.$$

We now present a proof that the Jones basis spans, adding slightly more detail than we found in the references. The extra detail will be used in the next section.

Definition 2.4 Given a word $w = U_{i_1} \dots U_{i_n}$ in the U_i, define the *terminus* to be the subscript of the final letter of the word appearing, i_n, and denote it by $t(w)$. Set $t(1) = \infty$ as a convention. Define the *index* of w to be the minimum subscript i_j appearing, and denote it by $i(w)$.

Remark 2.5 The notions of Jones normal form and index in $\mathrm{TL}_n(a)$ coincide with those of Kassel and Turaev [2008], under the bijection which sends the generator e_i used in their paper to the generator U_{n-i} used in this paper for $1 \leqslant i \leqslant n-1$.

The following two lemmas are an enhancement of [Kassel and Turaev 2008, Lemmas 5.25 and 5.26].

Lemma 2.6 *Any word $w \in \mathrm{TL}_n(a)$ is equal in $\mathrm{TL}_n(a)$ to a scalar multiple of a word w' in which*

(a) *$i(w) = i(w')$ and $U_{i(w)}$ appears exactly once in w';*

(b) *$t(w') = t(w)$.*

Point (a) occurs as [Kassel and Turaev 2008, Lemma 5.25], and the following is a simple extension of the proof that appears there. We have opted to give our proof in full because, as well as the minor extension of the proof, our notation differs from that of [Kassel and Turaev 2008] as in Remark 2.5.

Proof We proceed by reverse induction on the index $i(w)$ of w, which lies in the range $1 \leqslant i(w) \leqslant n-1$. If $i(w) = n-1$, then $w = U_{n-1}^i$ for some $i \geq 1$, so $w = a^{i-1}U_{n-1}$ is a scalar multiple of the word U_{n-1}. Since the words U_{n-1}^i and U_{n-1} have the same index and terminus, the result holds in this case.

Suppose that the claim holds for all words of index $> p$ and let w be a nonempty word of index p. Suppose that U_p appears in w at least twice. Then we may write $w = w_1 U_p w' U_p w_2$, where $i(w') = \ell > p$.

If $\ell > p+1$, then all letters of w' commute with U_p, so that

$$w = w_1 U_p w' U_p w_2 = w_1 w' U_p^2 w_2 = a w_1 w' U_p w_2.$$

Thus we have reduced the number of occurrences of U_p in w while preserving the (nonempty) final portion $U_p w_2$ of the word, so that the terminus remains unchanged.

If $\ell = p+1$, then by the induction hypothesis we may assume that U_{p+1} appears only once in w', so that $w' = w_3 U_{p+1} w_4$ where w_3, w_4 are words of index $\geq p+2$. Therefore w_3, w_4 commute with U_p, and consequently

$$\begin{aligned} w = w_1 U_p w' U_p w_2 &= w_1 U_p w_3 U_{p+1} w_4 U_p w_2 \\ &= w_1 w_3 U_p U_{p+1} U_p w_4 w_2 = w_1 w_3 U_p w_4 w_2 \\ &= w_1 w_3 w_4 U_p w_2. \end{aligned}$$

So again, we have reduced the number of occurrences of U_p in the word while preserving the final (nonempty) portion $U_p w_2$, and in particular preserving the terminus.

Repeating the process of reducing the number of occurrences of U_p while preserving the terminus, we find that w is a scalar multiple of a word w' of the required form. □

Lemma 2.7 *Any word $w \in \mathrm{TL}_n(a)$ is equivalent in $\mathrm{TL}_n(a)$ to a scalar multiple of a word w' such that*

(a) *w' is written in Jones normal form*;

(b) $t(w') \leqslant t(w)$;

(c) *if $t(w') < t(w)$ then $t(w') \leqslant t(w) - 2$.*

Proof As in the previous lemma, point (a) occurs as [Kassel and Turaev 2008, Lemma 5.26]. We refer the reader to that proof, with the following modifications:

- Invoke the bijection of generators of Remark 2.5. This amounts to replacing each occurrence of e_i with U_{n-i}, so for example the subscripts 1 and $n-1$ are interchanged, and inequalities are "reversed".
- Whenever the inductive hypothesis is used in [Kassel and Turaev 2008, Lemma 5.25], instead use the statement of the present lemma as a stronger inductive hypothesis.
- At the point where Lemma 5.25 of [Kassel and Turaev 2008] is used in their Lemma 5.26, use instead Lemma 2.6.

With these modifications in place, one can simply observe how the terminus changes in the proof of [Kassel and Turaev 2008, Lemma 5.26], to obtain the present strengthening of that result. □

2.2 Induced modules of Temperley–Lieb Algebras

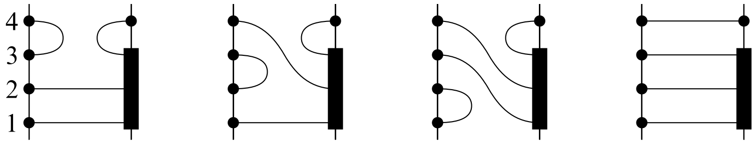
Definition 2.8 (the trivial module $\mathbb{1}$) The *trivial* module $\mathbb{1}$ of the Temperley–Lieb algebra $\mathrm{TL}_n(a)$ is the module consisting of R with the action of $\mathrm{TL}_n(a)$ in which all of the generators $U_1,\dots,U_{n-1}$ act as 0. We can regard $\mathbb{1}$ as either a left or right module over $\mathrm{TL}_n(a)$, and we will usually do that without indicating so in the notation.

Definition 2.9 (subalgebra convention) For $m \leqslant n$, we will regard $\mathrm{TL}_m(a)$ as the subalgebra of $\mathrm{TL}_n(a)$ generated by the elements $U_1,\dots,U_{m-1}$. We will often regard $\mathrm{TL}_n(a)$ as a left $\mathrm{TL}_n(a)$–module and a right $\mathrm{TL}_m(a)$–module, so that we obtain the left $\mathrm{TL}_n(a)$–module $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$.

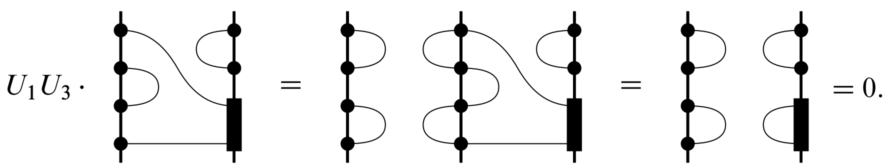
Remark 2.10 Elements of $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ can always be written as elementary tensors of the form $y \otimes 1$, since in this module $x \otimes r = rx \otimes 1$ for all $r \in R$.

The modules $\mathrm{TL}_n \otimes_{\mathrm{TL}_m} \mathbb{1}$ are an essential ingredient in the rest of this paper: they will be the building blocks of all the complexes we construct in order to prove our main results, in particular the complex of planar injective words $W(n)$. The rest of this section will study them in some detail, in particular finding a basis for them analogous to the Jones basis.

Remark 2.11 ($\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ via diagrams) The elements of $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ can be regarded as diagrams, just like the elements of $\mathrm{TL}_n(a)$, except that now the first m dots on the right are encapsulated within a *black box*, and if any cups can be absorbed into the black box, then the diagram is identified with 0. For example, some elements of $\mathrm{TL}_4(a) \otimes_{\mathrm{TL}_3(a)} \mathbb{1}$ are depicted as follows:



The structure of $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ as a left module for $\mathrm{TL}_n(a)$ is given by pasting diagrams on the left, and then simplifying, as in the following example for $n = 4$ and $m = 2$:



Definition 2.12 (the ideal I_m) Given $0 \leqslant m \leqslant n$, let I_m denote the left ideal of $\mathrm{TL}_n(a)$ generated by the elements $U_1,\dots,U_{m-1}$.

Lemma 2.13 *$\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ and $\mathrm{TL}_n(a)/I_m$ are isomorphic as left $\mathrm{TL}_n(a)$–modules via the maps*

$$\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1} \to \mathrm{TL}_n(a)/I_m, \quad y \otimes r \mapsto yr + I_m,$$
$$\mathrm{TL}_n(a)/I_m \to \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}, \quad y + I_m \mapsto y \otimes 1.$$

Proof Observe that the generators $U_1, \dots, U_{m-1}$ of the left ideal I_m in TL_n are precisely the generators of the subalgebra TL_m of TL_n. Thus the map $y \otimes r \mapsto yr + I_m$ is well defined because if $i = 1, \dots, m-1$ then elements of the form $yU_i \otimes r$ and $y \otimes U_i r$ both map to 0 in TL_n/I_m. And $y + I_m \mapsto y \otimes 1$ is well defined because elements of I_m are linear combinations of ones of the form $x \cdot U_i$ for $i = 1, \dots, m-1$, and $(x \cdot U_i) \otimes 1 = x \otimes (U_i \cdot 1) = x \otimes 0 = 0$ for $i = 1, \dots, m-1$. One can now check that the two maps are inverses of one another. □

Remark 2.14 Lemma 2.13 justifies the description of $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ in terms of diagrams with "black boxes" that we gave in Remark 2.11. Indeed, I_m is precisely the span of those diagrams which have a cup on the right between the dots i and $i+1$ for some $i = 1, \dots, m-1$. But these are precisely the diagrams which are made to vanish by having a cup fall into the black box. Thus $\mathrm{TL}_n(a)/I_m$ has basis given by the remaining diagrams, ie the ones that are not rendered 0 by the black box.

Lemma 2.15 *For $m \leqslant n$, the ideal I_m of $\mathrm{TL}_n(a)$ has basis consisting of those elements of $\mathrm{TL}_n(a)$ written in Jones normal form $x_{\underline{a},\underline{b}}$, which have terminus $b_1 \leqslant m-1$ (and $k \neq 0$).*

Proof Recall that words of the form $x_{\underline{a},\underline{b}}$ give a basis for TL_n. Then by definition any word $w \in I_m$ is of the form $w = x_{\underline{a},\underline{b}} v$ for $v \in \langle U_1, \dots, U_{m-1} \rangle$ and $v \neq e$. Then $t(w) \leqslant m-1$. Now apply Lemma 2.7 to w to complete the proof. □

Lemma 2.16 *For $m \leqslant n$, $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ has basis given by $x_{\underline{a},\underline{b}} \otimes \mathbb{1}$ such that the terminus $b_1 > m-1$.*

Proof From Lemma 2.13, $\mathrm{TL}_n \otimes_{\mathrm{TL}_m} \mathbb{1}$ is isomorphic to TL_n/I_m. Then elements of the form $x_{\underline{a},\underline{b}}$ give a basis for TL_n and elements of the form $x_{\underline{a},\underline{b}}$, which have terminus $b_1 \leqslant m-1$ give a basis for I_m by Lemma 2.15. Therefore a basis for the quotient is given by $x_{\underline{a},\underline{b}}$ such that the terminus $b_1 > m-1$, and under the isomorphism in Lemma 2.13 this gives the required basis. □

Example 2.17 The Jones basis of $\mathrm{TL}_3(a)$ is

$$1, \quad U_2, \quad U_1U_2, \quad U_1, \quad U_2U_1.$$

So $\mathrm{TL}_3(a) \otimes_{\mathrm{TL}_2(a)} \mathbb{1}$ has basis consisting of those elements whose terminus is strictly greater than 1, namely

$$1, \quad U_2, \quad U_1U_2.$$

(Recall that by convention the terminus of 1 is ∞.)

Lemma 2.18 *For $m \leqslant n$, suppose that $y \in \mathrm{TL}_n(a)$ and that $y \cdot U_{m-1}$ lies in I_{m-1}. Then $y \cdot U_{m-1}$ lies in I_{m-2}.*

Proof The product $y \cdot U_{m-1}$ is a linear combination of words ending with U_{m-1}, ie of words w with $t(w) = m-1$. By Lemma 2.7, this can be rewritten as a linear combination of Jones basis elements $x_{\underline{a},\underline{b}}$ whose terminus satisfies $t(x_{\underline{a},\underline{b}}) = m-1$ or $t(x_{\underline{a},\underline{b}}) \leqslant m-3$. Since $y \cdot U_{m-1} \in I_{m-1}$, this means that in fact no basis elements with terminus $m-1$ remain after cancellation, and therefore all remaining words have terminus $m-3$ or less, and so lie in I_{m-2}. □

2.3 Iwahori–Hecke algebras

Definition 2.19 (the Iwahori–Hecke algebra) Let $n \geqslant 0$ and let $q \in R^\times$. The *Iwahori–Hecke algebra* $\mathcal{H}_n(q)$ of type A_{n-1} is the algebra with generators

$$T_1, \dots, T_{n-1}$$

satisfying the relations

- $T_i T_j = T_j T_i$ for $i \neq j \pm 1$,
- $T_i T_j T_i = T_j T_i T_j$ for $i = j \pm 1$,
- $T_i^2 = (q-1)T_i + q$.

Definition 2.20 (from Iwahori–Hecke to Temperley–Lieb) Now suppose that there is $v \in R^\times$ such that $q = v^2$. Then there are two natural homomorphisms

$$\theta_1, \theta_2 \colon \mathcal{H}_n(q) \to \mathrm{TL}_n(v + v^{-1}),$$

defined by $\theta_1(T_i) = vU_i - 1$ and $\theta_2(T_i) = v^2 - vU_i$ for $i = 1, \dots, n-1$. They induce isomorphisms

$$\bar{\theta}_1 \colon \mathcal{H}_n(q)/I_1 \xrightarrow{\cong} \mathrm{TL}_n(v + v^{-1}) \quad \text{and} \quad \bar{\theta}_2 \colon \mathcal{H}_n(q)/I_2 \xrightarrow{\cong} \mathrm{TL}_n(v + v^{-1}),$$

where I_1 is the two-sided ideal generated by elements of the form

$$T_i T_j T_i + T_i T_j + T_j T_i + T_i + T_j + 1$$

for $i = j \pm 1$, and I_2 is the two-sided ideal generated by elements of the form

$$T_i T_j T_i - q T_i T_j - q T_j T_i + q^2 T_i + q^2 T_j - q^3$$

for $i = j \pm 1$. See [Fan and Green 1997; Kassel and Turaev 2008, Theorem 5.29; Halverson et al. 2009, Section 2.3], though unfortunately conventions change from author to author. Another standard convention of setting $a = -(v + v^{-1})$ can easily be accounted for by swapping v with $-v^{\pm 1}$.

We will take an agnostic approach to the homomorphisms θ_1 and θ_2. We will choose one of them and denote it by simply

$$\theta \colon \mathcal{H}_n(q) \to \mathrm{TL}_n(v + v^{-1}),$$

and denote by λ the constant term in $\theta(T_i)$, and by μ the coefficient of U_i in $\theta(T_i)$, so that

$$\theta(T_i) = \lambda + \mu U_i.$$

Then θ induces an isomorphism

$$\bar{\theta} \colon \mathcal{H}_n(q)/I \xrightarrow{\cong} \mathrm{TL}_n(v + v^{-1}),$$

where I is the two-sided ideal generated by elements of the form

$$T_i T_j T_i - \lambda T_i T_j - \lambda T_j T_i + \lambda^2 T_i + \lambda^2 T_j - \lambda^3$$

for $i = j \pm 1$. And moreover, the elements $\theta(T_i)$ act on the trivial module $\mathbb{1}$ as multiplication by λ.

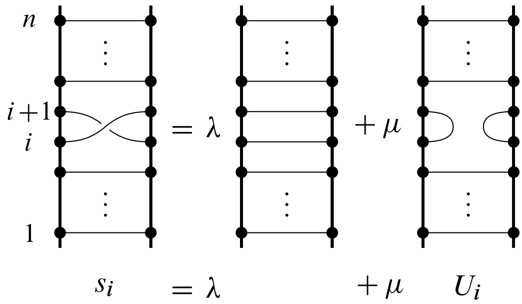


Figure 2: Smoothings of s_i.

Definition 2.21 Let $v \in R^\times$. We define $s_1, \dots, s_{n-1} \in \mathrm{TL}_n(v + v^{-1})$ by setting

$$s_i = \theta(T_i) = \lambda + \mu U_i,$$

and note that these elements satisfy the following properties:

- $s_i^2 = (v^2 - 1)s_i + v^2$ for all i.
- $s_i s_j = s_j s_i$ for $i \neq j \pm 1$.
- $s_i s_j s_i = s_j s_i s_j$ for $i = j \pm 1$.
- $s_i s_j s_i - \lambda s_i s_j - \lambda s_j s_i + \lambda^2 s_i + \lambda^2 s_j - \lambda^3 = 0$ for $i = j \pm 1$.
- s_i acts on $\mathbb{1}$ as multiplication by λ.

Remark 2.22 There is a homomorphism from (the group algebra of) the braid group into $\mathrm{TL}_n(v + v^{-1})$ given on generators by $s_i \mapsto s_i$. This is the content of the second and third bullet points above, together with the fact that the s_i are invertible, which follows from the first bullet point (and the fact that v is a unit). Diagrammatically, this homomorphism can be viewed as a *smoothing expansion* from braided diagrams to planar diagrams: take a braid diagram, and then smooth each crossing in turn in the two possible ways, using appropriate weightings for each smoothing. For example, we can visualise the image of s_i in $\mathrm{TL}_n(v + v^{-1})$ as the standard braid group generator crossing strand i over strand $i + 1$. There are two ways this crossing can be resolved to a planar diagram, and we equate s_i to the sum of these two states. They are the identity and U_i, as shown in Figure 2. The coefficient of the identity is λ and the coefficient of U_i is μ, simply because we defined $s_i = \lambda + \mu U_i$. Similarly, we consider the image of s_i^{-1} as strand i crossing under strand $i + 1$, and when this is smoothed the coefficient of the identity is λ^{-1} and the coefficient of U_i is μ^{-1}, precisely because one can verify that $s_i^{-1} = \lambda^{-1} + \mu^{-1} U_i$ in $\mathrm{TL}_n(v + v^{-1})$.

In principle we could describe how various Reidemeister moves affect the smoothing expansion, but it will not be necessary for the rest of the paper. Moreover, we will only encounter positive powers of s_i.

3 Inductive resolutions

In this section we prove the following two theorems, which we recall from the introduction.

Theorem A *Let R be a commutative ring and let a be a unit in R. Then*

$$\mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \quad \textit{and} \quad \mathrm{Ext}^d_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$$

both vanish for $d>0$.

Theorem F *Let R be a commutative ring and let $a\in R$. Let $0\leqslant m<n$. Then*

$$\mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1},\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_m(a)}\mathbb{1}) \quad \textit{and} \quad \mathrm{Ext}^d_{\mathrm{TL}_n(a)}(\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_m(a)}\mathbb{1},\mathbb{1})$$

vanish for $d>0$.

In fact for Theorem A we will prove the following stronger claim:

Claim 3.1 *Suppose that the parameter $a\in R$ is invertible. Then for any $0\leqslant m\leqslant n$, the groups*

$$\mathrm{Tor}_d^{\mathrm{TL}_n(a)}(\mathbb{1},\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_m(a)}\mathbb{1}) \quad \textit{and} \quad \mathrm{Ext}_d^{\mathrm{TL}_n(a)}(\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_m(a)}\mathbb{1},\mathbb{1})$$

both vanish for $d>0$.

The similarity between Theorem F and Claim 3.1 is now clear. Both will be proved by induction on m, the initial cases $m=0,1$ being immediate because then TL_m is the ground ring R so that $\mathrm{TL}_n\otimes_{\mathrm{TL}_m}\mathbb{1}\cong\mathrm{TL}_n$ is free. In order to produce an inductive proof, we construct resolutions of the modules $\mathrm{TL}_n\otimes_{\mathrm{TL}_m}\mathbb{1}$ whose terms are not free or projective or injective, but instead whose terms are the modules considered earlier in the induction, specifically $\mathrm{TL}_n\otimes_{\mathrm{TL}_{m-1}}\mathbb{1}$ and $\mathrm{TL}_n\otimes_{\mathrm{TL}_{m-2}}\mathbb{1}$. For this reason we refer to these resolutions as *inductive resolutions*. This approach is inspired by homological stability arguments, in which one considers complexes whose building blocks are induced up from the earlier objects in the sequence. The difference here is that our complexes are actual resolutions — they are acyclic rather than just acyclic up to a point — and because Shapiro's lemma is unavailable we do not change the algebra we are working over, rather we change the algebra from which we are inducing our modules.

3.1 The inductive resolutions

In this subsection we establish the resolutions $C(m)$ and $D(m)$ of $\mathrm{TL}_n\otimes_{\mathrm{TL}_m}\mathbb{1}$ required to prove Claim 3.1 and Theorem F above.

Definition 3.2 (the complex $C(m)$) Let $2\leqslant m\leqslant n$ and assume that a is invertible. We define a chain complex of left $\mathrm{TL}_n(a)$–modules as in Figure 3, left. The degree is indicated in the right-hand column. The differentials of $C(m)$ are all given by extending the algebra over which the tensor product is taken, by right multiplying in the first factor by the indicated element of $\mathrm{TL}_n(a)$, or by a combination of the two. So, for example, the differential originating in degree 1 sends $x\otimes r\in\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_{m-2}(a)}\mathbb{1}$ to $(x\cdot a^{-1}U_{m-1})\otimes r\in\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_{m-1}(a)}\mathbb{1}$. The complex is periodic of period 2 in degrees 1 and above, so that all entries are $\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_{m-2}(a)}\mathbb{1}$ and the boundary maps between them alternate between $a^{-1}U_{m-1}$ and $1-a^{-1}U_{m-1}$. The boundary maps are well defined because U_{m-1} commutes inside $\mathrm{TL}_n(a)$ with all elements of $\mathrm{TL}_{m-2}(a)$.

$$
\begin{array}{cc}
\vdots & \\
\big\downarrow{\scriptstyle (1-a^{-1}U_{m-1})} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1} & 3 \\
\big\downarrow{\scriptstyle a^{-1}U_{m-1}} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1} & 2 \\
\big\downarrow{\scriptstyle (1-a^{-1}U_{m-1})} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1} & 1 \\
\big\downarrow{\scriptstyle a^{-1}U_{m-1}} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-1}} \mathbb{1} & 0 \\
\big\downarrow{\scriptstyle 1} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m}} \mathbb{1} & -1
\end{array}
\qquad\qquad
\begin{array}{cc}
\vdots & \\
\big\downarrow{\scriptstyle (1-U_{m-1}U_m)} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1} & 3 \\
\big\downarrow{\scriptstyle U_{m-1}U_m} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1} & 2 \\
\big\downarrow{\scriptstyle (1-U_{m-1}U_m)} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1} & 1 \\
\big\downarrow{\scriptstyle U_{m-1}} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-1}} \mathbb{1} & 0 \\
\big\downarrow{\scriptstyle 1} & \\
\mathrm{TL}_n \otimes_{\mathrm{TL}_{m}} \mathbb{1} & -1
\end{array}
$$

Figure 3: The complexes $C(m)$ (left) and $D(m)$ (right).

Definition 3.3 (the complex $D(m)$) Let $2 \leqslant m < n$, and do not assume that a is invertible. We define a chain complex of left $\mathrm{TL}_n(a)$–modules as in Figure 3, right. The degree is indicated in the right-hand column. The differentials of $D(m)$ are all given by extending the algebra over which the tensor product is taken, by right multiplying in the first factor by the indicated element of $\mathrm{TL}_n(a)$, or by a combination of the two. So, for example, the differential originating in degree 1 sends $x \otimes r \in \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{m-2}(a)} \mathbb{1}$ to $x \cdot U_{m-1} \otimes r \in \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{m-1}(a)} \mathbb{1}$. The complex is periodic of period 2 in degrees 1 and above, so that in that range all terms are $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{m-2}(a)} \mathbb{1}$ and the boundary maps between them alternate between $U_{m-1}U_m$ and $(1-U_{m-1}U_m)$. The boundary maps are well defined because U_{m-1} and $U_{m-1}U_m$ commute inside $\mathrm{TL}_n(a)$ with all elements of $\mathrm{TL}_{m-2}(a)$. Observe that the condition $m < n$ is necessary in order to ensure that U_m is actually an element of $\mathrm{TL}_n(a)$.

Lemma 3.4 (1) *Let $2 \leqslant m \leqslant n$ and let a be invertible. Then $a^{-1}U_{m-1} \in \mathrm{TL}_n(a)$ is idempotent.*

(2) *Let $2 \leqslant m < n$ and let a be arbitrary. Then $U_{m-1}U_m \in \mathrm{TL}_n(a)$ is idempotent.*

Proof We calculate

$$(a^{-1}U_i)^2 = a^{-2}U_i^2 = a^{-2}aU_i = a^{-1}U_i,$$

$$U_{m-1}U_m \cdot U_{m-1}U_m = U_{m-1}U_mU_{m-1} \cdot U_m = U_{m-1}U_m. \qquad \square$$

From now on in this section, we will attempt to talk about $C(m)$ and $D(m)$ at the same time. When we refer to $C(m)$, the relevant assumptions should be understood, namely that $2 \leqslant m \leqslant n$ and that $a \in R$ is a unit. And when we refer to $D(m)$, the assumptions that $2 \leqslant m < n$ but $a \in R$ is arbitrary should be understood. We trust that this will not be confusing.

Lemma 3.5 *$C(m)$ and $D(m)$ are indeed chain complexes.*

Proof We give the proof for $C(m)$. The proof for $D(m)$ is similar. We must check that consecutive boundary maps of $C(m)$ compose to 0. In the case of the composite from degree 1 to -1, the composition is given by

$$x \otimes r \mapsto (x \cdot a^{-1} U_{m-1}) \otimes r = x \otimes (a^{-1} U_{m-1} \cdot r) = x \otimes 0 = 0;$$

this holds because the tensor product is over TL_m, which contains $a^{-1}U_{m-1}$. In the case of the remaining composites, this follows immediately from

$$(a^{-1}U_{m-1}) \cdot (1 - a^{-1}U_{m-1}) = 0 = (1 - a^{-1}U_{m-1}) \cdot (a^{-1}U_{m-1}),$$

which is a consequence of the fact that $a^{-1}U_{m-1}$ is idempotent (from Lemma 3.4). □

Lemma 3.6 *The complexes $C(m)$ and $D(m)$ are acyclic.*

Proof In degree -1 it is clear that the boundary map is surjective, for both $C(m)$ and $D(m)$.

In degree 0, we will give the proof for $C(m)$, the proof for $D(m)$ being similar. Suppose that $y \otimes 1 \in \mathrm{TL}_n \otimes_{\mathrm{TL}_{m-1}} \mathbb{1}$ lies in the kernel of the boundary map, or in other words that $y \otimes 1 \in \mathrm{TL}_n \otimes_{\mathrm{TL}_m} \mathbb{1}$ vanishes. This means that y lies in the left ideal generated by the elements $U_1, \ldots, U_{m-1}$. Since all but the last of these generators lie in TL_{m-1}, and we started with $y \otimes 1 \in \mathrm{TL}_n \otimes_{\mathrm{TL}_{m-1}} \mathbb{1}$, we may assume without loss that $y = y' \cdot U_{m-1}$ for some y'. But then

$$y \otimes 1 = y' \cdot U_{m-1} \otimes 1 = ay' \cdot (a^{-1}U_{m-1}) \otimes 1$$

does indeed lie in the image of the boundary map.

In degree 1, we give the proof for both complexes. First, for $C(m)$, suppose that $y \otimes 1 \in \mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1}$ lies in the kernel of the boundary map. It follows that $y \cdot (a^{-1}U_{m-1}) \otimes 1$ vanishes in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-1}} \mathbb{1}$, which means that $y \cdot (a^{-1}U_{m-1})$ lies in the left ideal I_{m-1} generated by $U_1, \ldots, U_{m-2}$. It follows from Lemma 2.18 that $y \cdot (a^{-1}U_{m-1})$ lies in the left ideal I_{m-2} generated by $U_1, \ldots, U_{m-3}$, so that in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1}$ the element $y \cdot (a^{-1}U_{m-1}) \otimes 1$ vanishes. Thus

$$y \otimes 1 = y \cdot (1 - a^{-1}U_{m-1}) \otimes 1$$

does indeed lie in the image of the boundary map. Second, for $D(m)$, suppose that $y \otimes 1 \in \mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1}$ lies in the kernel of the boundary map. Then, as for $C(m)$, $y \cdot U_{m-1}$ lies in I_{m-2}. So $y \cdot U_{m-1}U_m$ also lies in the left ideal I_{m-2} since U_m commutes with the generators of I_{m-2}. Thus $y \cdot U_{m-1}U_m \otimes 1$ vanishes in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1}$, so that $y \otimes 1 = y \cdot (1 - U_{m-1}U_m) \otimes 1$ does indeed lie in the image of the boundary map.

In degrees 2 and higher, acyclicity is an immediate consequence of the fact that $a^{-1}U_{m-1}$ and $U_{m-1}U_m$ are idempotents, by Lemma 3.4. □

Lemma 3.7 *The following complexes are acyclic*:

$$\mathbb{1} \otimes_{\mathrm{TL}_n(a)} C(m), \quad \mathbb{1} \otimes_{\mathrm{TL}_n(a)} D(m), \quad \mathrm{Hom}_{\mathrm{TL}_n(a)}(C(m), \mathbb{1}) \quad \textit{and} \quad \mathrm{Hom}_{\mathrm{TL}_n(a)}(D(m), \mathbb{1}).$$

$$
\begin{array}{cc}
\vdots & \\
\downarrow & \\
\mathbb{1} & 3 \\
\downarrow{\scriptstyle 0} & \\
\mathbb{1} & 2 \\
\downarrow{\scriptstyle 1} & \\
\mathbb{1} & 1 \\
\downarrow{\scriptstyle 0} & \\
\mathbb{1} & 0 \\
\downarrow{\scriptstyle 1} & \\
\mathbb{1} & -1
\end{array}
$$

Figure 4: The complex $\mathbb{1} \otimes C(m)$.

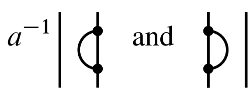
Proof We give the proof for $\mathbb{1} \otimes_{\mathrm{TL}_n} C(m)$, the proof for the other parts being similar. The terms of $C(m)$ have the form $\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-i}} \mathbb{1}$, where $i = 0, 1, 2$, depending on the degree. Thus $\mathbb{1} \otimes_{\mathrm{TL}_n} C(m)$ has terms of the form $\mathbb{1} \otimes_{\mathrm{TL}_n} (\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-i}} \mathbb{1}) \cong \mathbb{1} \otimes_{\mathrm{TL}_{m-i}} \mathbb{1} \cong \mathbb{1}$. Moreover, by tracing through this isomorphism, one sees that if a boundary map in $C(m)$ is labelled by an element $x \in \mathrm{TL}_n$, then the corresponding boundary map in $\mathbb{1} \otimes_{\mathrm{TL}_n} C(m)$ is simply the map $\mathbb{1} \to \mathbb{1}$ given by the action of x on $\mathbb{1}$. Thus $\mathbb{1} \otimes_{\mathrm{TL}_n} C(m)$ is nothing other than the complex in Figure 4. (The right-hand column indicates the degree.) This is visibly acyclic, and this completes the proof. $\square$

Remark 3.8 (representation theory and the inductive resolutions) Schur–Weyl duality relates representations of TL_n with representations of the quantum group $U_q(\mathfrak{sl}_2)$, and it is possible to use this to construct our inductive resolutions via the representation theory of $U_q(\mathfrak{sl}_2)$. We will try to describe this briefly. We are indebted to a referee for explaining this connection to us.

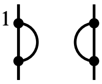
One instance of Schur–Weyl duality is the following. Let V denote the standard representation of $U_q(\mathfrak{sl}_2)$. Then there is an isomorphism $\mathrm{TL}_n \cong \mathrm{End}_{U_q(\mathfrak{sl}_2)}(V^{\otimes n})$, and more generally there are isomorphisms $\mathrm{TL}(n, m) \cong \mathrm{Hom}_{U_q(\mathfrak{sl}_2)}(V^{\otimes n}, V^{\otimes m})$ that assemble into a monoidal functor on the *Temperley–Lieb category* TL. (The objects of TL are the nonnegative integers, the morphism space $\mathrm{TL}(n, m)$ is the R–module spanned by planar diagrams with n marked points on the left and m marked points on the right, and composition is defined just like multiplication in TL_n.) See Webster [2017].

One can write down exact sequences of $U_q(\mathfrak{sl}_2)$–modules that, after applying Schur–Weyl duality, yield the inductive resolutions $C(m)$ and $D(m)$. We will not detail the construction of these sequences, except to say that each one relies on the construction of an appropriate splitting of some tensor power of V. The relevant splittings are constructed in each case as follows:

- In the case where a is invertible, the morphisms

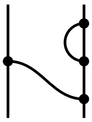
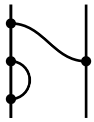


in TL compose to give the identity morphism in $\mathrm{TL}(0,0)$. (The two semicircles compose to the circle morphism from 0 to itself, and by the usual rule for composing diagrams, the circle morphism is a times the identity.) This then corresponds to a pair of maps $R = V^{\otimes 0} \to V^{\otimes 2}$ and $V^{\otimes 2} \to V^{\otimes 0} = R$ that compose to the identity, showing that $V^{\otimes 2}$ splits off a copy of R. Note that the map $V^{\otimes 2} \to V^{\otimes 2}$ that projects onto this copy of R is represented by the morphism

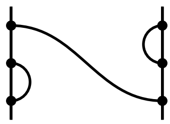
a^{-1} 

in TL. Compare this with the idempotent $a^{-1}U_{m-1}$ appearing in $C(n)$.

- When a is not invertible, we consider the morphisms

 and 

which compose to give the identity morphism in $\mathrm{TL}(1,1)$. These diagrams correspond to a pair of maps $V \to V^{\otimes 3} \to V$ that compose to the identity, showing that $V^{\otimes 3}$ splits off a copy of V. Observe that the map $V^{\otimes 3} \to V^{\otimes 3}$ that projects to this copy of V is represented by the morphism



which can be compared to the idempotent $U_{m-1}U_m$ appearing in $D(n)$.

3.2 The spectral sequence of a double complex

Since the spectral sequence of a particular kind of double complex is used several times during this paper, we introduce and discuss it in this subsection.

We begin with the homological version. Suppose we have a chain complex Q_* of left TL_n–modules, such as $C(m)$ or $D(m)$, or the complex of planar injective words $W(n)$ to be introduced later. Then we choose a projective resolution P of $\mathbb{1}$ as a right module over TL_n, and we consider the double complex $P_* \otimes_{\mathrm{TL}_n} Q_*$. This is a homological double complex in the sense that both differentials reduce the grading. Associated to this double complex are two spectral sequences, $\{{}^{\mathrm{I}}E^r\}$ and $\{{}^{\mathrm{II}}E^r\}$, which both converge to the homology of the totalisation, $H_*(\mathrm{Tot}(P_* \otimes_{\mathrm{TL}_n} Q_*))$ as in [Weibel 1994, Section 5.6]. The first spectral sequence has E^1–term given by ${}^{\mathrm{I}}E^1_{i,j} = H_j(P_i \otimes_{\mathrm{TL}_n} Q_*) \cong P_i \otimes_{\mathrm{TL}_n} H_j(Q_*)$, with $d^1 : {}^{\mathrm{I}}E^1_{i,j} \to {}^{\mathrm{I}}E^1_{i-1,j}$ induced by the differential $P_i \to P_{i-1}$. The isomorphism above holds because each P_i is projective and therefore flat. It follows that the E^2–term is

$${}^{\mathrm{I}}E^2_{i,j} = \mathrm{Tor}_i^{\mathrm{TL}_n}(\mathbb{1}, H_j(Q_*)).$$

The second spectral sequence has E^1–term given by ${}^{\mathrm{II}}E^1_{i,j} = H_j(P_* \otimes_{\mathrm{TL}_n} Q_i)$, ie

$$ {}^{\mathrm{II}}E^1_{i,j} = \mathrm{Tor}^{\mathrm{TL}_n}_j(\mathbb{1}, Q_i), $$

with $d^1 \colon {}^{\mathrm{II}}E^1_{i,j} \to {}^{\mathrm{II}}E^1_{i-1,j}$ induced by the boundary maps of Q_*.

We now consider the cohomological version. Suppose we have a chain complex Q_* of left TL_n–modules, again such as $C(m)$, $D(m)$ or $W(n)$ (the latter to be introduced later). Then we choose an injective resolution I^* of $\mathbb{1}$ as a left module over TL_n, and we consider the double complex $\mathrm{Hom}_{\mathrm{TL}_n}(Q_*, I^*)$. This is a cohomological double complex in the sense that both differentials increase the grading. Associated to this double complex are two spectral sequences, $\{{}^{\mathrm{I}}E_r\}$ and $\{{}^{\mathrm{II}}E_r\}$, both converging to the cohomology of the totalisation, $H^*(\mathrm{Tot}(\mathrm{Hom}_{\mathrm{TL}_n}(Q_*, I^*)))$ as in [Weibel 1994, Section 5.6]. The first spectral sequence has E_1–term given by ${}^{\mathrm{I}}E_1^{i,j} = H^j(\mathrm{Hom}_{\mathrm{TL}_n}(Q_*, I^i)) \cong \mathrm{Hom}_{\mathrm{TL}_n}(H_j(Q_*), I^i)$, with $d^1 \colon {}^{\mathrm{I}}E^1_{i,j} \to {}^{\mathrm{I}}E^1_{i+1,j}$ induced by the differential of I^*. The isomorphism above holds because each I^i is injective, so that the functor $\mathrm{Hom}_{\mathrm{TL}_n}(-, I^i)$ is exact. It follows that the E_2–term is

$$ {}^{\mathrm{I}}E_2^{i,j} = \mathrm{Ext}^i_{\mathrm{TL}_n}(H_j(Q_*), \mathbb{1}). $$

The second spectral sequence has E_1–term ${}^{\mathrm{II}}E_1^{i,j} = H^j(\mathrm{Hom}_{\mathrm{TL}_n}(Q_i, I^*))$, ie

$$ {}^{\mathrm{II}}E_1^{i,j} = \mathrm{Ext}^j_{\mathrm{TL}_n}(Q_i, \mathbb{1}), $$

with differential $d^1 \colon {}^{\mathrm{II}}E_1^{i,j} \to {}^{\mathrm{II}}E_1^{i+1,j}$ induced by the differential of Q_*.

3.3 Proof of Theorems A and F

We can now prove Claim 3.1 (which implies Theorem A) and Theorem F. The proofs of the two results will be almost identical except that the former uses the complex $C(m)$ and the latter uses the complex $D(m)$. Moreover, each result has a homological and cohomological part, referring to Tor and Ext, respectively. In each case the two parts are proved similarly, by using either the homological or cohomological spectral sequence from Section 3.2. We will therefore only prove the homological part of Claim 3.1, ie we will prove that $\mathrm{Tor}^{\mathrm{TL}_n}_*(\mathbb{1}, \mathrm{TL}_n \otimes_{\mathrm{TL}_m} \mathbb{1})$ vanishes in positive degrees, leaving the details of the other parts to the reader.

Proof of Claim 3.1, Tor case We prove the claim by fixing n and using strong induction on m in the range $n \geqslant m \geqslant 0$. As noted before, the initial cases $m = 0, 1$ of the induction are immediate since then TL_m is the ground ring and $\mathrm{TL}_n \otimes_{\mathrm{TL}_n} \mathbb{1} \cong \mathrm{TL}_n$ is free. We therefore fix m in the range $2 \leqslant m \leqslant n$.

We now employ the homological spectral sequences $\{{}^{\mathrm{I}}E^r\}$ and $\{{}^{\mathrm{II}}E^r\}$ of Section 3.2, in the case $Q = C(m)$. Then ${}^{\mathrm{I}}E^2_{i,j} = \mathrm{Tor}^{\mathrm{TL}_n}_i(\mathbb{1}, H_j(C(m))) = 0$ for all i and j, since $C(m)$ is acyclic by Lemma 3.6. Thus $\{{}^{\mathrm{I}}E^r\}$ converges to zero, and the same must therefore be true of $\{{}^{\mathrm{II}}E^r\}$, since both spectral sequences have the same target. In the second spectral sequence the E^1–page

$$ {}^{\mathrm{II}}E^1_{i,j} = \mathrm{Tor}^{\mathrm{TL}_n}_j(\mathbb{1}, C(m)_i) $$

is largely known to us. The bottom $j = 0$ row of ${}^{\mathrm{II}}E^1$ is precisely the complex $\mathbb{1} \otimes_{\mathrm{TL}_n} C(m)$, which is acyclic by Lemma 3.7. And when $i \geqslant 0$, the term $C(m)_i$ is either $\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-1}} \mathbb{1}$ or $\mathrm{TL}_n \otimes_{\mathrm{TL}_{m-2}} \mathbb{1}$, and our

$$\begin{array}{c|l|lllll}
j & & & & & & \\
\vdots & \vdots & & & & & \\
3 & \mathrm{Tor}_3^{\mathrm{TL}_n}(\mathbb{1}, C(m)_{-1}) & & & & & \\
2 & \mathrm{Tor}_2^{\mathrm{TL}_n}(\mathbb{1}, C(m)_{-1}) & \mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1}, C(m)_i) = 0 & & & & \\
1 & \mathrm{Tor}_1^{\mathrm{TL}_n}(\mathbb{1}, C(m)_{-1}) & & & & & \\
0 & \mathbb{1}\otimes_{\mathrm{TL}_n} C(m)_{-1} \leftarrow & \mathbb{1}\otimes_{\mathrm{TL}_n} C(m)_0 \leftarrow & \mathbb{1}\otimes_{\mathrm{TL}_n} C(1)_1 \leftarrow & \mathbb{1}\otimes_{\mathrm{TL}_n} C(m)_2 & \leftarrow & \cdots \\
\hline
 & -1 & 0 & 1 & 2 & \cdots & i
\end{array}$$

Figure 5: The page ${}^{\mathrm{II}}E^1$. The only differentials that affect the ${}^{\mathrm{II}}E^2$–page are shown on the $j=0$ row.

inductive hypothesis applies to these ($m-1<m$ and $m-2<m$) to show that ${}^{\mathrm{II}}E^1_{i,j} = \mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1}, C(m)_i) = 0$ when $j>0$. See Figure 5 for a visualisation of the E^1–page. Altogether, this tells us that ${}^{\mathrm{II}}E^2_{i,j}$ vanishes except for the groups

$${}^{\mathrm{II}}E^2_{-1,j} = \mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1}, C(m)_{-1}) = \mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1}, \mathrm{TL}_n \otimes_{\mathrm{TL}_m} \mathbb{1})$$

for $j>0$, which are concentrated in a single column and therefore not subject to any further differentials. Thus ${}^{\mathrm{II}}E^2 = {}^{\mathrm{II}}E^\infty$. But we know that ${}^{\mathrm{II}}E^\infty$ vanishes identically, so that the inductive hypothesis is proved, and so, therefore, is the proof of the homological part of Claim 3.1. □

4 Planar injective words

Throughout this section we will consider the Temperley–Lieb algebra $\mathrm{TL}_n(a) = \mathrm{TL}_n(v+v^{-1})$, where $v \in R^\times$. We will make use of the elements $s_1, \dots, s_{n-1}$ of Definition 2.21.

Definition 4.1 For $n \geqslant 0$ we define a chain complex $W(n)_*$ of left $\mathrm{TL}_n(a)$–modules as follows. For i in the range $-1 \leqslant i \leqslant n-1$, the degree-$i$ part of $W(n)_*$ is defined by

$$W(n)_i = \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{n-i-1}(a)} \mathbb{1},$$

and in all other degrees we set $W(n)_i = 0$. Note that

$$W(n)_{-1} = \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_n(a)} \mathbb{1} = \mathbb{1}.$$

For $i \geqslant 0$ the boundary map $d^i\colon W(n)_i \to W(n)_{i-1}$ is defined to be the alternating sum $\sum_{j=0}^i (-1)^j d^i_j$, where $d^i_j\colon W(n)_i \to W(n)_{i-1}$ is the map

$$d^i_j\colon \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{n-i-1}(a)} \mathbb{1} \to \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_{n-i}(a)} \mathbb{1}$$

defined by

$$d^i_j(x \otimes r) = (x \cdot s_{n-i+j-1} \cdots s_{n-i}) \otimes \lambda^{-j} r.$$

$$
\begin{array}{rcc}
 & \mathrm{TL}_n \otimes_{\mathrm{TL}_0} \mathbb{1} & n-1 \\
 {\scriptstyle d_0^{n-1} - d_1^{n-1} + \cdots + (-1)^{n-1} d_{n-1}^{n-1}} & \big\downarrow & \\
 & \mathrm{TL}_n \otimes_{\mathrm{TL}_1} \mathbb{1} & n-2 \\
 {\scriptstyle d_0^{n-2} - d_1^{n-2} + \cdots + (-1)^{n-2} d_{n-2}^{n-2}} & \big\downarrow & \\
 & \vdots & \\
 & \big\downarrow & \\
 & \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-3}} \mathbb{1} & 2 \\
 {\scriptstyle d_0^2 - d_1^2 + d_2^2} & \big\downarrow & \\
 & \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-2}} \mathbb{1} & 1 \\
 {\scriptstyle d_0^1 - d_1^1} & \big\downarrow & \\
 & \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-1}} \mathbb{1} & 0 \\
 & \big\downarrow & \\
 & \mathbb{1} & -1
\end{array}
$$

Figure 6: The complex $W(n)$.

In the expression $s_{n-i+j-1} \cdots s_{n-i}$, the indices decrease from left to right. Thus, for example, the product is $s_{n-i+1}s_{n-i}$ when $j=2$, it is s_{n-i} when $j=1$, and it is trivial (the unit element) when $j=0$ (the latter point can be regarded as a convention if one wishes). Recall that λ is indeed invertible since $\lambda=-1$ or v^2, and v is a unit. For notational purposes we will write $W(n)$ and only use a subscript when identifying a particular degree.

Observe that d_j is well defined because the elements $s_{n-i}, \dots, s_{n-i+j-1}$ all commute with all generators of TL_{n-i-1}. We have depicted $W(n)$ in Figure 6.

Lemma 4.2 *The boundary maps of $W(n)$ satisfy $d^{i-1} \circ d^i = 0$.*

Proof We will show that if $i \geqslant 1$ and $0 \leqslant j < k \leqslant i$, then the composite maps

$$d_j^{i-1} d_k^i, d_{k-1}^{i-1} d_j^i \colon W(n)_i \to W(n)_{i-2}$$

coincide. (Thus the d_j^i satisfy the semisimplicial identities, so $W(n)$ is a semisimplicial R–module.) The fact that $d \circ d$ vanishes then follows. We have

$$
\begin{aligned}
d_j^{i-1} d_k^i (x \otimes r) &= [x \cdot (s_{n-i+k-1} \cdots s_{n-i}) \cdot (s_{n-i+j} \cdots s_{n-i+1})] \otimes \lambda^{-(j+k)} r, \\
d_{k-1}^{i-1} d_j^i (x \otimes r) &= [x \cdot (s_{n-i+j-1} \cdots s_{n-i}) \cdot (s_{n-i+k-1} \cdots s_{n-i+1})] \otimes \lambda^{-(j+k-1)} r.
\end{aligned}
$$

Now

$$
\begin{aligned}
(s_{n-i+k-1} \cdots s_{n-i}) \cdot (s_{n-i+j} \cdots s_{n-i+1}) &= (s_{n-i+j-1} \cdots s_{n-i}) \cdot (s_{n-i+k-1} \cdots s_{n-i}) \\
&= (s_{n-i+j-1} \cdots s_{n-i}) \cdot (s_{n-i+k-1} \cdots s_{n-i+1}) \cdot s_{n-i}.
\end{aligned}
$$

Here, the first equality follows by taking the letters of the second parenthesis in turn, and "passing through" the first parenthesis, using a single braid relation, with the result that the letter's index is reduced by 1. Thus,

$$\begin{aligned} d^{i-1}_j d^i_k(x\otimes r) &= [x\cdot(s_{n-i+k-1}\cdots s_{n-i})\cdot(s_{n-i+j}\cdots s_{n-i+1})]\otimes \lambda^{-(j+k)}r \\ &= [x\cdot(s_{n-i+j-1}\cdots s_{n-i})\cdot(s_{n-i+k-1}\cdots s_{n-i+1})\cdot s_{n-i}]\otimes \lambda^{-(j+k)}r \\ &= [x\cdot(s_{n-i+j-1}\cdots s_{n-i})\cdot(s_{n-i+k-1}\cdots s_{n-i+1})]\otimes s_{n-i}\cdot(\lambda^{-(j+k)}r) \\ &= [x\cdot(s_{n-i+j-1}\cdots s_{n-i})\cdot(s_{n-i+k-1}\cdots s_{n-i+1})]\otimes \lambda^{-(j+k-1)}r \\ &= d^{i-1}_{k-1} d^i_j(x\otimes r), \end{aligned}$$

where the third equality holds because this computation takes place in $W(n)_{i-2}=\mathrm{TL}_n\otimes_{\mathrm{TL}_{n-i+1}}\mathbb{1}$ and $s_{n-i}\in\mathrm{TL}_{n-i+1}$. □

Remark 4.3 Let us explain the motivation for the definition of $W(n)$. Let $\mathfrak{S}_n$ denote the symmetric group on n letters. The *complex of injective words* is the chain complex $\mathscr{C}(n)$ of $\mathfrak{S}_n$–modules, concentrated in degrees -1 to $n-1$, that in degree i is the free R–module with basis given by tuples $(x_0,\dots,x_i)$, where $x_0,\dots,x_i\in\{1,\dots,n\}$ and no letter appears more than once. We allow the empty word $()$, which lies in degree -1. The differential of $\mathscr{C}(n)$ sends a word $(x_0,\dots,x_i)$ to the alternating sum $\sum_{j=0}^i(-1)^j(x_0,\dots,\widehat{x_j},\dots,x_i)$. A theorem of Farmer [1979] shows that the homology of $\mathscr{C}(n)$ vanishes in degrees $i\leqslant n-2$, and the same result has been proved since then by many authors [Maazen 1979; Björner and Wachs 1983; Kerz 2005; Randal-Williams 2013]. The complex of injective words has been used by several authors to prove homological stability for the symmetric groups [Maazen 1979; Kerz 2005; Randal-Williams 2013].

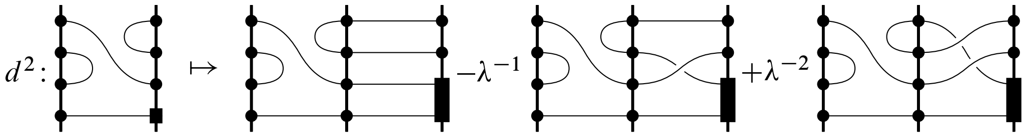
For this paragraph only, let us abuse our established notation and denote by $s_1,\dots,s_{n-1}\in\mathfrak{S}_n$ the elements defined by $s_i=(i\ i+1)$, the transposition of i with $i+1$. Then these elements satisfy the braid relations, ie the second and third identities of Definition 2.21. The complex of injective words $\mathscr{C}(n)$ can be rewritten in terms of the group ring $R\mathfrak{S}_n$ and the elements s_i. Indeed, it is shown in [Hepworth 2022] that $\mathscr{C}(n)_i\cong R\mathfrak{S}_n\otimes_{R\mathfrak{S}_{n-i-1}}\mathbb{1}$, where $\mathbb{1}$ is the trivial module of $R\mathfrak{S}_{n-i-1}$, and that under this isomorphism the differential $d^i\colon\mathscr{C}(n)_i\to\mathscr{C}(n)_{i-1}$ becomes the map

$$d^i\colon R\mathfrak{S}_n\otimes_{R\mathfrak{S}_{n-i-1}}\mathbb{1}\to R\mathfrak{S}_n\otimes_{R\mathfrak{S}_{n-i}}\mathbb{1}$$

defined by $d^i(x\otimes 1)=\sum_{j=0}^i(-1)^j x\cdot(s_{n-i+j-1}\cdots s_{n-i})\otimes 1$. (There are no constants λ in this expression). Comparing this description of $\mathscr{C}(n)$ with our definition of $W(n)$, we see that our complex of planar injective words is precisely analogous to the original complex of injective words, after systematically replacing the group algebras of symmetric groups with the Temperley–Lieb algebras. The lack of constants in the differential for $\mathscr{C}(n)$ is explained by the fact that the effect of s_i on $\mathbb{1}$ is multiplication by λ in the Temperley–Lieb setting, and multiplication by 1 in the symmetric group setting.

Since we regard the Temperley–Lieb algebra as the planar analogue of the symmetric group, we chose the name *planar injective words* for our complex $W(n)$. This seemed the least discordant way of giving our complex an appropriate name. See the next remark for a means of picturing the complex.

Remark 4.4 Let us describe a method for visualising $W(n)$. Recall from the diagrammatic description of $\mathrm{TL}_n(a) \otimes_{\mathrm{TL}_m(a)} \mathbb{1}$ when $m \leqslant n$ given in Remark 2.11 that elements of $W(n)_i$ can be regarded as diagrams where the first $n-i-1$ dots on the right are encapsulated within a black box, and if any cups can be absorbed into the black box, then the diagram is identified with 0. The differential $d^i : W(n)_i \to W(n)_{i-1}$ is then given by pasting special elements onto the right of a diagram, followed by taking their signed and weighted sum. These special elements each enlarge the black box by an extra strand, and plumb one of the free strands into the new space in the black box. Here is an example for $n = 4$ and $i = 2$:



The resulting diagrams can be simplified using the smoothing rules for diagrams with crossings described in Remark 2.22. We leave it to the reader to make this description as precise as they wish, and note here that this is where the notion of *braiding*, so often seen in homological stability arguments, fits into our setup.

Remark 4.5 Readers who are familiar with the theory will recognise that $W(n)$ is the chain complex associated to an augmented semisimplicial $\mathrm{TL}_n(a)$–module.

The main result about the complex of planar injective words is the following, which we recall from the introduction. It is analogous to the homological vanishing property of the complex of injective words first proved by Farmer [1979].

Theorem E *The homology of $W(n)$ vanishes in degrees $d \leqslant n-2$.*

The proof of Theorem E is the most technical part of this work, and will be given in Section 8.

The complex of injective words on n letters has rich combinatorial features: its Euler characteristic is the number of derangements of $\{1, \dots, n\}$; when one works over $\mathbb{C}$, its top homology has a description as a virtual representation that categorifies a well-known alternating sum formula for the number of derangements; and again when one works over $\mathbb{C}$, its top homology has a compact description in terms of Young diagrams and counts of standard Young tableaux. In the associated paper [Boyd and Hepworth 2021] we establish analogues of these for the complex of planar injective words. In particular we show that when the ring R is Noetherian the rank of $H_{n-1}(W(n))$ is the n^{th} *Fine number* [Deutsch and Shapiro 2001]. (The rank of the Temperley–Lieb algebra is the n^{th} *Catalan number*, which is the number of

Dyck paths of length $2n$. The n^{th} Fine number is the number of Dyck paths of length $2n$ whose first peak occurs at an even height, and as we explain in [Boyd and Hepworth 2021], it is an analogue of the number of derangements.) We also discover a new feature of the complex: the differentials have an alternative expression in terms not of the s_i but of the U_i. This expression demonstrates a connection with the *Jacobsthal numbers*, and we will briefly explain the result for the top differential below. The top homology of the Tits building is known as the *Steinberg module*. This inspires the name in the following definition.

Definition 4.6 We define the n^{th} *Fineberg module* to be the $\mathrm{TL}_n(a)$–module $\mathscr{F}_n(a) = H_{n-1}(W(n))$. We often suppress the a and simply write $\mathscr{F}_n$.

The Fineberg module is an important ingredient in the full statement of our stability result, Theorem 5.1. In order to detect the nonzero homology group appearing in Theorem C we need to study it in more detail using the connection with Jacobsthal numbers from [Boyd and Hepworth 2021].

The n^{th} *Jacobsthal number* J_n [Sloane 2000] is (among other things) the number of sequences $n > a_1 > a_2 > \cdots > a_r > 0$ whose initial term has the opposite parity to n. Some examples, when $n = 4$, are 3, 1, $3 > 2$, $3 > 1$ and $3 > 2 > 1$. (We allow the empty sequence, and say that by convention its initial term is $a_1 = 0$ and $r = 0$. Of course this only occurs when n is odd.) Another viewpoint of J_n in terms of compositions of n is given in [Boyd and Hepworth 2021].

Definition 4.7 Let $a = v + v^{-1}$, where $v \in R^\times$ is a unit. We define the *Jacobsthal* element in $\mathrm{TL}_n(a)$ by

$$\mathscr{J}_n = (-1)^{n-1} \sum_{\substack{n>a_1>\cdots>a_r>0 \\ n-a_1 \text{ odd}}} \left(\frac{\mu}{\lambda}\right)^r U_{a_1} \dots U_{a_r}.$$

Recall that we allow the empty sequence ($a_1 = 0$ and $r = 0$) when n is odd. This corresponds to a constant summand 1 in $\mathscr{J}_n$ for odd n. Note that the number of irreducible terms in $\mathscr{J}_n$ is J_n.

Example 4.8 In the cases $n = 1, 2, 3, 4$, and choosing $\theta = \theta_1$ so that $(\lambda, \mu) = (-1, v)$, we have

$$\mathscr{J}_1 = 1, \quad \mathscr{J}_2 = vU_1, \quad \mathscr{J}_3 = v^2U_2U_1 - vU_2 + 1, \quad \mathscr{J}_4 = v^3U_3U_2U_1 - v^2U_3U_2 - v^2U_3U_1 + vU_3 + vU_1.$$

Spencer [2022] has computed the Jacobsthal elements $\mathscr{J}_n$ up to $n = 9$.

Since $\mathscr{F}_n$ is the homology of $W(n)$ in the top degree, it is simply the kernel of the top differential $d^{n-1}\colon W(n)_{n-1} \to W(n)_{n-2}$. There are identifications

$$W(n)_{n-1} = \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_0(a)} \mathbb{1} \cong \mathrm{TL}_n(a) \quad \text{and} \quad W(n)_{n-2} \cong \mathrm{TL}_n(a) \otimes_{\mathrm{TL}_1(a)} \mathbb{1} \cong \mathrm{TL}_n(a).$$

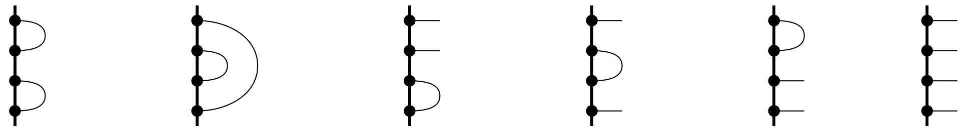
Proposition 4.9 [Boyd and Hepworth 2021, Theorem D] *Under the above identifications, the top differential of $W(n)$ is right multiplication by $\mathscr{J}_n$. In particular, there is an exact sequence*

$$0 \to \mathscr{F}_n(a) \to \mathrm{TL}_n(a) \xrightarrow{-\cdot \mathscr{J}_n} \mathrm{TL}_n(a).$$

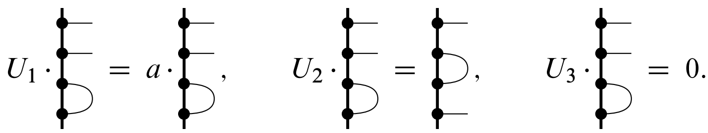
Remark 4.10 Definition 4.7 gives a different value for the element $\mathscr{J}_n$ than the one that appears in [Boyd and Hepworth 2021, Definition 8.1 and Theorem D]. This is because the proof of [Boyd and Hepworth 2021, Theorem D] contains a sign error: it assumes that $s_i = \lambda - \mu U_i$ rather than $s_i = \lambda + \mu U_i$ as it should have done. This error is fixed by replacing μ with $-\mu$ in the formula in [Boyd and Hepworth 2021, Definition 8.1]. It is possible to check Example 4.8 by hand to confirm that the signs in the present formula for $\mathscr{J}_n$ are the correct ones.

The Fineberg module $\mathscr{F}_n$ appears to be a new and interesting representation, and looks likely to be highly nontrivial for each choice of n. Let us illustrate this by computing $\mathscr{F}_2$, $\mathscr{F}_3$ and $\mathscr{F}_4$. We will continue with the choice $\theta = \theta_1$ so that $(\lambda, \mu) = (-1, v)$.

Our description will be phrased in terms of the cell modules of TL_n, which we describe briefly. A *half-diagram* (or *link state* in the language of [Ridout and Saint-Aubin 2014]) consists of a vertical line in the plane decorated with dots labelled $1, \dots, n$ from bottom to top, together with a collection of arcs in the plane, each of which either connects two dots, or is connected to a dot at one end, in such a way that each dot lies on precisely one arc. The arcs must lie to the right of the vertical line, they must be disjoint, and the half-diagrams are taken up to isotopy. Thus the half-diagrams on four dots are as follows:



The *cell module* $S(n, m)$ is the TL_n–module with R–basis consisting of the half-diagrams on n dots in which m arcs have free ends. The TL_n–module structure on $S(n, m)$ is obtained by pasting planar diagrams onto the left of half-diagrams and simplifying the result exactly as with composition in TL_n, with the extra condition that if pasting produces an arc with two free ends, then the resulting diagram is set to 0. In $S(4, 2)$, for example, we have



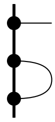
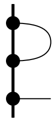
(The reader is reminded that we label the dots from bottom to top.) Observe that $S(n, n) = \mathbb{1}$ is the trivial module for each n, and that $S(n, m)$ is nonzero only when $n - m$ is even.

Example 4.11 (the Fineberg module $\mathscr{F}_2$) The module $\mathscr{F}_2$ is the kernel of the map $\mathrm{TL}_2 \to \mathrm{TL}_2$ given by $x \mapsto x \cdot \mathscr{J}_2$. Now $\mathscr{J}_2 = vU_1$ as in Example 4.8, so that $\mathscr{F}_2$ is the R–module of rank 1 spanned by the element $a - U_1$. This is a copy of the trivial module $\mathbb{1} = S(2, 2)$.

Example 4.12 (the Fineberg module $\mathscr{F}_3$) The module $\mathscr{F}_3$ is the kernel of the map $\mathrm{TL}_3 \to \mathrm{TL}_3$ given by $x \mapsto x \cdot \mathscr{J}_3$, where $\mathscr{J}_3 = v^2 U_2 U_1 - vU_2 + 1$ as in Example 4.8. Thus $\mathscr{F}_3$ is the R–module of rank 2 with basis elements

$$\alpha = U_1 U_2 - vU_1 \quad \text{and} \quad \beta = U_2 - vU_2 U_1.$$

One can now check that there is an isomorphism of TL_3–modules $\mathscr{F}_3 \cong S(3,1)$ given by

$\mathscr{F}_3 \xrightarrow{\cong} S(3,1), \qquad \alpha \mapsto$  and $\beta \mapsto$ .

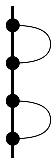
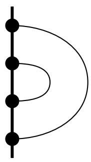
Example 4.13 (the Fineberg module $\mathscr{F}_4$) The module $\mathscr{F}_4$ is the kernel of the map $\mathrm{TL}_4 \to \mathrm{TL}_4$ given by $x \mapsto x \cdot \mathscr{J}_4$, where $\mathscr{J}_4 = v^3 U_3 U_2 U_1 - v^2 U_3 U_2 - v^2 U_3 U_1 + v U_3 + v U_1$ as in Example 4.8. It is now possible to check (at length) that $\mathscr{F}_4$ is a free R–module of rank 6 with basis

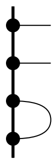
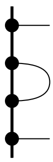
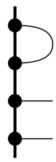
$$\begin{aligned}
A &= U_3 U_1 - a U_3 U_1 U_2, \\
B &= U_2 U_3 U_1 - a U_2 U_3 U_1 U_2, \\
X &= U_1 U_2 U_3 - U_3 U_1 U_2 - a U_1 U_2 + U_1, \\
Y &= U_2 U_3 - U_2 U_3 U_1 U_2 - a U_2 + U_2 U_1, \\
Z &= U_3 U_2 U_1 - U_3 U_1 U_2 - a U_3 U_2 + U_3, \\
P &= U_3 U_1 U_2 - U_1 - U_3 + a.
\end{aligned}$$

If we now define

$$M_0 = \operatorname{span}(A, B), \quad M_1 = \operatorname{span}(A, B, X, Y, Z), \quad M_2 = \operatorname{span}(A, B, X, Y, Z, P),$$

so that $M_0 \subseteq M_1 \subseteq M_2 = \mathscr{F}_4$, then one can check directly (by computing the effect of multiplying on the left by U_1, U_2, U_2) that M_0 and M_1 are submodules of $\mathscr{F}_4$, and, moreover, that we have isomorphisms

$M_0 \xrightarrow{\cong} S(4,0), \qquad A \mapsto$ , $B \mapsto$ ,

$M_1/M_0 \xrightarrow{\cong} S(4,2), \qquad X \mapsto$ , $Y \mapsto$ , $Z \mapsto$ ,

$M_2/M_1 \xrightarrow{\cong} \mathbb{1}, \qquad P \mapsto 1.$

Thus $\mathscr{F}_4$ has a filtration in which each of the three cell modules appears as precisely one of the filtration quotients. We emphasise that this result holds with *no* further assumptions on the ground ring R or on the parameter v.

5 Homological stability and stable homology

The aim of this section is to prove the following result. Theorem B is an immediate consequence, and Theorem C will be proved in the next section as a corollary of it.

Theorem 5.1 *Let R be a commutative ring, let $v \in R$ be a unit, and let $a = v + v^{-1}$. Then, for n odd,*

$$\mathrm{Tor}_i^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \cong \begin{cases} R & \text{if } i = 0, \\ 0 & \text{if } 1 \leqslant i \leqslant n-1, \\ \mathrm{Tor}_{i-n}^{\mathrm{TL}_n(a)}(\mathbb{1},\mathscr{F}_n(a)) & \text{if } i \geqslant n, \end{cases}$$

and, for n even and $i \neq n-1, n$,

$$\mathrm{Tor}_i^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \cong \begin{cases} R & \text{if } i = 0, \\ 0 & \text{if } 1 \leqslant i \leqslant n-2, \\ \mathrm{Tor}_{i-n}^{\mathrm{TL}_n(a)}(\mathbb{1},\mathscr{F}_n(a)) & \text{if } i \geqslant n+1, \end{cases}$$

while in degrees $n-1$ and n there is an exact sequence

$$0 \to \mathrm{Tor}_n^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \to \mathbb{1} \otimes_{\mathrm{TL}_n(a)} \mathscr{F}_n(a) \xrightarrow{\mathfrak{Q}_n} \mathbb{1} \to \mathrm{Tor}_{n-1}^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \to 0. \tag{5-1}$$

Analogous results hold for the Ext *groups. For n odd,*

$$\mathrm{Ext}^i_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \cong \begin{cases} R & \text{if } i = 0, \\ 0 & \text{if } 1 \leqslant i \leqslant n-1, \\ \mathrm{Ext}^{i-n}_{\mathrm{TL}_n(a)}(\mathscr{F}_n(a),\mathbb{1}) & \text{if } i \geqslant n, \end{cases}$$

and, for n even and $i \neq n-1, n$,

$$\mathrm{Ext}^i_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \cong \begin{cases} R & \text{if } i = 0, \\ 0 & \text{if } 1 \leqslant i \leqslant n-2, \\ \mathrm{Ext}^{i-n}_{\mathrm{TL}_n(a)}(\mathscr{F}_n(a),\mathbb{1}) & \text{if } i \geqslant n+1, \end{cases}$$

while in degrees $n-1$ and n there is an exact sequence

$$0 \to \mathrm{Ext}^{n-1}_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \to \mathbb{1} \xrightarrow{\mathfrak{Q}^n} \mathrm{Hom}_{\mathrm{TL}_n(a)}(\mathscr{F}_n(a),\mathbb{1}) \to \mathrm{Ext}^n_{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1}) \to 0. \tag{5-2}$$

The central maps $\mathfrak{Q}_n$ and $\mathfrak{Q}^n$ of (5-1) and (5-2), respectively, are described as follows. Regard $\mathscr{F}_n(a)$ as a left submodule of $\mathrm{TL}_n(a)$, as in Proposition 4.9. Then the maps are

$$\mathfrak{Q}_n\colon \mathbb{1} \otimes_{\mathrm{TL}_n(a)} \mathscr{F}_n(a) \to \mathbb{1}, \qquad x \otimes f \mapsto x \cdot f,$$
$$\mathfrak{Q}^n\colon \mathbb{1} \to \mathrm{Hom}_{\mathrm{TL}_n(a)}(\mathscr{F}_n(a),\mathbb{1}), \qquad x \mapsto (f \mapsto f \cdot x),$$

where $x \cdot f$ and $f \cdot x$ denote the action of $f \in \mathscr{F}_n(a) \subseteq \mathrm{TL}_n(a)$ on the right and left of $\mathbb{1}$, respectively.

In order to prove this theorem, we will use the complex of planar injective words $W(n)$ introduced in the previous section. Recall that the Fineberg module $\mathscr{F}_n$ appearing in the statement is the top homology group $H_{n-1}(W(n))$.

Lemma 5.2 *The homology groups of both complexes $\mathbb{1} \otimes_{\mathrm{TL}_n(a)} W(n)$ and $\mathrm{Hom}_{\mathrm{TL}_n(a)}(W(n),\mathbb{1})$ are concentrated in degree $n-1$, where in both cases they are given by $\mathbb{1}$ if n is even and 0 if n is odd.*

Proof We have $W(n)_i = \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$, and the boundary map $d^i\colon W(n)_i \to W(n)_{i-1}$ is given by $x \otimes r \mapsto x \cdot D_i \otimes r$, where $D_i = \sum_{j=0}^{i} (-1)^j s_{n-i+j-1} \cdots s_{n-i} \lambda^{-j}$.

By regarding $\mathbb{1}$ as both a left and right TL_n–module, we may regard $\mathbb{1} \otimes_{\mathrm{TL}_n} W(n)_i$ as a left TL_n–module. With this TL_n–module structure, we obtain $\mathbb{1} \otimes_{\mathrm{TL}_n} W(n)_i = \mathbb{1} \otimes_{\mathrm{TL}_n} (\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}) \cong \mathbb{1}$. Under these isomorphisms, the boundary map originating in degree i becomes the action on $\mathbb{1}$ of the element D_i. Similarly, $\mathrm{Hom}_{\mathrm{TL}_n}(W(n)_i, \mathbb{1}) = \mathrm{Hom}_{\mathrm{TL}_n}(\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}, \mathbb{1}) \cong \mathbb{1}$, and under these isomorphisms the boundary map originating in degree $i-1$ becomes the action of the element D_i on $\mathbb{1}$.

The action of $s_{n-i+j-1} \cdots s_{n-i}$ on $\mathbb{1}$ is simply multiplication by λ^j, with one factor of λ for each s term (recall $s_i = \mu U_i + \lambda$). Thus the action of D_i on $\mathbb{1}$ is nothing other than multiplication by $\sum_{j=0}^{i}(-1)^j$, which is 0 for i odd and 1 for i even.

So altogether $\mathbb{1} \otimes_{\mathrm{TL}_n} W(n)$ and $\mathrm{Hom}_{\mathrm{TL}_n}(W(n), \mathbb{1})$ are isomorphic to complexes with a copy of R in each degree $i = -1, \dots, n-1$, and with boundary maps alternating between the identity map and 0. In $\mathbb{1} \otimes_{\mathrm{TL}_n} W(n)$ the identity maps originate in even degrees, and in $\mathrm{Hom}_{\mathrm{TL}_n}(W(n), \mathbb{1})$ they originate in odd degrees. The claim now follows. $\square$

Proof of Theorem 5.1 We begin with the Tor case.

In degree $d = 0$ the theorem holds trivially. Recall that P_* is a projective resolution of $\mathbb{1}$ as a right TL_n–module. We use the two homological spectral sequences $\{{}^{\mathrm{I}}E^r\}$ and $\{{}^{\mathrm{II}}E^r\}$ associated to $W(n)$ as described in Section 3.2.

Let us consider $\{{}^{\mathrm{I}}E^r\}$. We have

$$ {}^{\mathrm{I}}E^2_{i,j} = \begin{cases} \mathrm{Tor}_i^{\mathrm{TL}_n}(\mathbb{1}, \mathscr{F}_n) & \text{if } j = n-1, \\ 0 & \text{if } j \neq n-1, \end{cases} $$

and consequently the spectral sequence converges to $\mathrm{Tor}_{*-n+1}^{\mathrm{TL}_n}(\mathbb{1}, \mathscr{F}_n)$ for $* = i + j$. The same is therefore true of $\{{}^{\mathrm{II}}E^r\}$.

Let us write $\varepsilon_n = H_{n-1}(\mathbb{1} \otimes_{\mathrm{TL}_n} W(n))$, so that, by Lemma 5.2, ε_n is trivial for n odd and $\mathbb{1}$ for n even. Since $\mathscr{F}_n$ consists of the cycles in $W(n)_{n-1}$, the map

$$ \mathbb{1} \otimes_{\mathrm{TL}_n} \mathscr{F}_n \to \mathbb{1} \otimes_{\mathrm{TL}_n} W(n)_{n-1} $$

again lands in the cycles, giving us a map

$$ \mathbb{1} \otimes_{\mathrm{TL}_n} \mathscr{F}_n \to H_{n-1}(\mathbb{1} \otimes_{\mathrm{TL}_n} W(n)) = \varepsilon_n. $$

When n is even and ε_n is identified with $\mathbb{1}$ as in the lemma, then this map simply becomes $\mathscr{Q}_n$ as described in the statement of the theorem.

We now know that $\{{}^{\mathrm{II}}E^r\}$ converges to $\mathrm{Tor}_{*-n+1}^{\mathrm{TL}_n}(\mathbb{1}, \mathscr{F}_n)$. Its E^1–page ${}^{\mathrm{II}}E^1_{i,j} = \mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1}, W(n)_i)$ is largely known to us. Indeed, when $j = 0$ the terms are $\mathrm{Tor}_0^{\mathrm{TL}_n}(\mathbb{1}, W(n)_i) = \mathbb{1} \otimes_{\mathrm{TL}_n} W(n)_i$, with d^1–maps between them induced by the boundary maps of $W(n)$. In other words, the $j = 0$ part of ${}^{\mathrm{II}}E^1_{i,j}$ is precisely the complex $\mathbb{1} \otimes_{\mathrm{TL}_n} W(n)$. When $0 \leqslant i \leqslant n-1$, the term $W(n)_i = \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$ satisfies $0 \leqslant n-i-1 < n$, so that, by Theorem F,

$$ {}^{\mathrm{II}}E^1_{i,j} = \mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1}, \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}) = 0 \quad \text{for } j > 0. $$

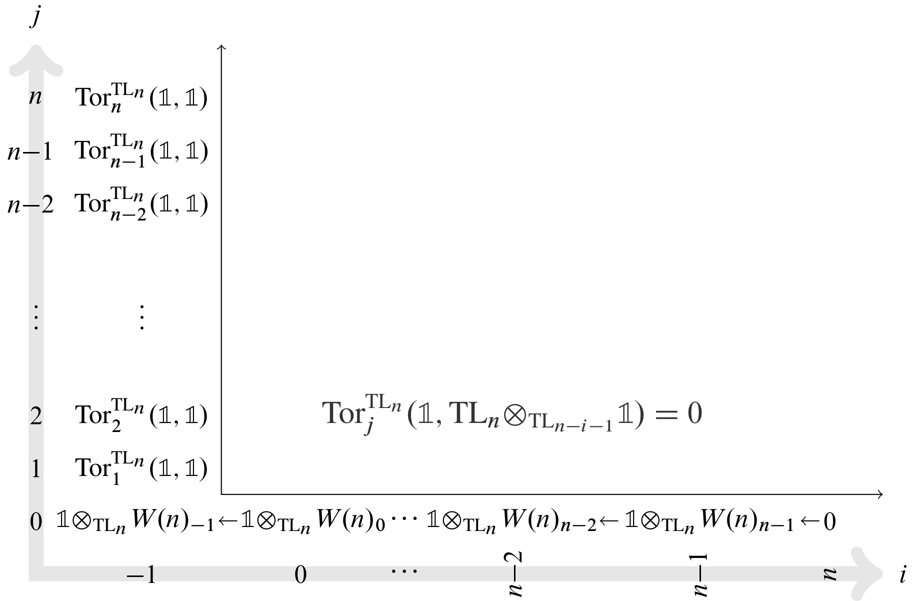


Figure 7: The page ${}^{\mathrm{II}}E^1$. The only differentials that affect the ${}^{\mathrm{II}}E^2$–page are shown on the $j=0$ row.

When $i=-1$, we have $W(n)_{-1}=\mathbb{1}$, so that ${}^{\mathrm{II}}E^1_{-1,j}=\mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})$ for $j>0$. This is depicted in Figure 7.

By the description in the previous paragraph, we can now identify ${}^{\mathrm{II}}E^2_{*,*}$. The only possible differentials are in the $j=0$ part, which is $\mathbb{1}\otimes_{\mathrm{TL}_n}W(n)$, and whose homology is ε_n concentrated in degree $n-1$. Thus ${}^{\mathrm{II}}E^2_{*,*}$ is zero except for the groups

$$ {}^{\mathrm{II}}E^2_{i,j}=\begin{cases}\mathrm{Tor}_j^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1}) & \text{if } i=-1 \text{ and } j>0,\\ \varepsilon_n & \text{if } i=n-1 \text{ and } j=0,\end{cases} $$

as depicted in Figure 8.

From the E^2–page onwards there is precisely one possible differential, namely $d^n\colon E^n_{n-1,0}\to E^n_{-1,n-1}$, which is a map $d^n\colon\varepsilon_n\to\mathrm{Tor}_{n-1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})$. It forms part of an exact sequence

$$0\to{}^{\mathrm{II}}E^\infty_{n-1,0}\to\varepsilon_n\xrightarrow{d^n}\mathrm{Tor}_{n-1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})\to{}^{\mathrm{II}}E^\infty_{-1,n-1}\to 0.$$

In ${}^{\mathrm{II}}E^\infty_{*,*}$, each total degree has only one nonzero group, except (possibly) for total degree $n-1$, where we have the two groups ${}^{\mathrm{II}}E^\infty_{-1,n}$ and ${}^{\mathrm{II}}E^\infty_{n-1,0}$. The relationship between the infinity page of a spectral sequence and the sequence's target now give us a short exact sequence

$$0\to{}^{\mathrm{II}}E^\infty_{-1,n}\to\mathrm{Tor}_0^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n)\to{}^{\mathrm{II}}E^\infty_{n-1,0}\to 0.$$

The last two exact sequences combine to give us

$$0\to{}^{\mathrm{II}}E^\infty_{-1,n}\to\mathrm{Tor}_0^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n)\to\varepsilon_n\to\mathrm{Tor}_{n-1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})\to{}^{\mathrm{II}}E^\infty_{-1,n-1}\to 0.$$

The leftmost term is ${}^{\mathrm{II}}E^\infty_{-1,n}={}^{\mathrm{II}}E^2_{-1,n}=\mathrm{Tor}_n^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})$. The only group in total degree $n-2$ is ${}^{\mathrm{II}}E^\infty_{-1,n-1}$, so it coincides with $\mathrm{Tor}_{(n-2)-n+1}^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n)=\mathrm{Tor}_{-1}^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n)=0$. Also, $\mathrm{Tor}_0^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n)=\mathbb{1}\otimes_{\mathrm{TL}_n}\mathscr{F}_n$. The

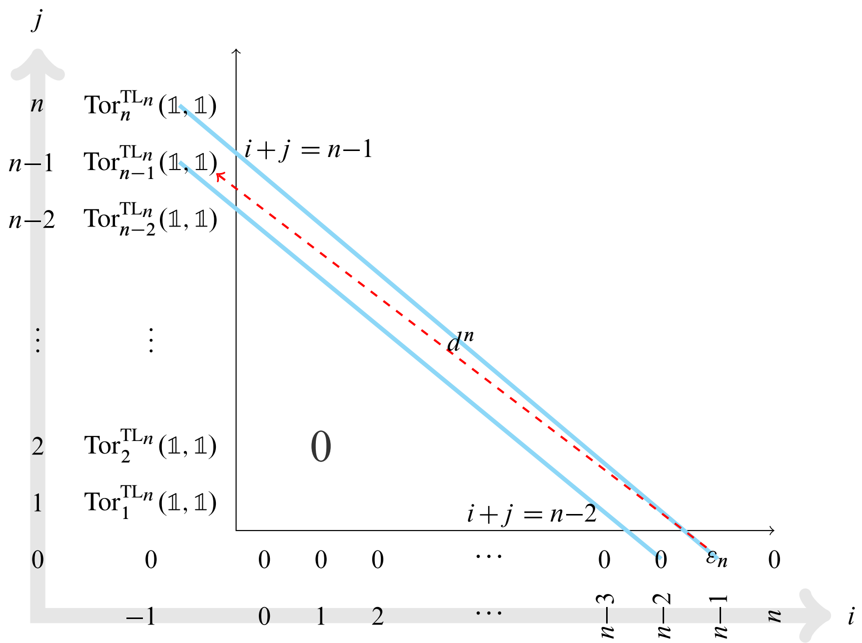


Figure 8: The page ${}^{\mathrm{II}}E^2$. This page stays constant until ${}^{\mathrm{II}}E^n$ where the only possible further differential lies: this is shown as the dashed arrow. The $i+j=n-1$ and $i+j=n-2$ diagonal lines are drawn alongside.

last exact sequence becomes

$$0 \to \mathrm{Tor}_n^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1}) \to \mathbb{1} \otimes_{\mathrm{TL}_n} \mathscr{F}_n \to \varepsilon_n \to \mathrm{Tor}_{n-1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1}) \to 0.$$

When n is even, we claim that the map $\mathbb{1} \otimes_{\mathrm{TL}_n} \mathscr{F}_n \to \varepsilon_n$ in this sequence is $\mathscr{Q}_n$. Let $\mathscr{F}_n[n-1]$ be the complex consisting of a copy of $\mathscr{F}_n$ concentrated in degree $n-1$. There is a natural inclusion of chain complexes $\mathscr{F}_n[n-1] \hookrightarrow W(n)$, and this leads to a map of double complexes and then of spectral sequences. The map $\mathbb{1} \otimes_{\mathrm{TL}_n} \mathscr{F}_n \to \varepsilon_n$ can be identified using this map of spectral sequences.

It follows from the sequence that in the case n odd, when $\varepsilon_n = 0$, the final term satisfies $\mathrm{Tor}_{n-1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1}) = 0$, and the first two terms satisfy

$$\mathrm{Tor}_n^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1}) \cong \mathbb{1} \otimes_{\mathrm{TL}_n} \mathscr{F}_n = \mathrm{Tor}_0^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n),$$

as required.

The previous discussion determines what happens in total degrees $n-1$ and $n-2$. In total degrees d other than $n-1$ and $n-2$, and when $j>0$, the only term on the E^∞–page is ${}^{\mathrm{II}}E^\infty_{-1,d+1} = \mathrm{Tor}_{d+1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})$, which must therefore equal $\mathrm{Tor}_{d-n+1}^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n)$. Thus $\mathrm{Tor}_d^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1}) \cong \mathrm{Tor}_{d-n}^{\mathrm{TL}_n}(\mathbb{1},\mathscr{F}_n)$ for $d \neq n, n-1$. This completes the proof.

For the Ext case we use the two cohomological spectral sequences associated to $W(n)$ as in Section 3.2, and then proceed dually to the above. We leave the details to the reader. □

6 Sharpness

We recall the statement of Theorem C from the introduction.

Theorem C *Let n be even and suppose that a is not a unit. Then* $\mathrm{Tor}_{n-1}^{\mathrm{TL}_n(a)}(\mathbb{1},\mathbb{1})$ *is nonzero.*

Let $\mathcal{I} \subseteq \mathrm{TL}_n$ denote the left ideal generated by all diagrams which have a cup on the right in positions other than 1, together with all multiples of a. Thus

$$\mathcal{I} = (\mathrm{TL}_n \cdot a) + (\mathrm{TL}_n \cdot U_2) + \cdots + (\mathrm{TL}_n \cdot U_{n-1}).$$

Lemma 6.1 *Let n be even or odd, and let $1 \leqslant p \leqslant n-1$. Then $U_p \cdot \mathcal{J}_n \in \mathcal{I}$.*

Proof Recall from Definition 4.7 that the monomials appearing in $\mathcal{J}_n$ are those of the form $U_{i_1} \dots U_{i_r}$, where $n-1 \geqslant i_1 > i_2 \cdots > i_r \geqslant 1$ and $i_1 \equiv n-1 \bmod 2$, and that such a monomial appears in $\mathcal{J}_n$ with coefficient $(-1)^{n-1}(\mu/\lambda)^r$. We write $\mathcal{J}_n = K_n + L_n$ where K_n is the part of $\mathcal{J}_n$ featuring monomials of the form $U_i U_{i-1} \dots U_1$ for $i \equiv n-1 \bmod 2$ in the range $1 \leqslant i \leqslant n-1$, and L_n is the part of $\mathcal{J}_n$ featuring the remaining monomials.

If $U_{i_1} \dots U_{i_r}$ is a monomial appearing in L_n, then it must either end in U_{i_r} for $n-1 \geqslant i_r > 1$ or end in a monomial of the form $U_{i_j} \cdot U_{i_{j-1}} \dots U_1 = (U_{i_{j-1}} \dots U_1) \cdot U_{i_j}$ for some $i_j \geqslant i_{j-1}+2$ and $i_{j-1} \geqslant 1$, and hence must lie in $\mathcal{I}$. Thus $L_n \in \mathcal{I}$, and to prove the lemma it will be sufficient to show that $U_p \cdot K_n \in \mathcal{I}$.

Now observe that

$$K_n = (-1)^{n-1} \sum_{\substack{0 \leqslant i \leqslant n-1 \\ i \equiv n-1 \bmod 2}} \left(\frac{\mu}{\lambda}\right)^i \cdot U_i U_{i-1} \dots U_1.$$

(In the case $i = 0$ the product $U_i \dots U_1$ is empty and therefore equal to 1. This term only appears in K_n when n is odd.) Suppose that $U_i \dots U_1$ is a monomial appearing in the above sum. Then

$$U_p \cdot (U_i \dots U_1) = \begin{cases} (U_p \dots U_1) \cdot (U_i \dots U_{p+2}) & \text{if } p \leqslant i-2, \\ U_{i-1} \cdots U_1 & \text{if } p = i-1, \\ U_i \cdots U_1 \cdot a & \text{if } p = i, \\ U_{i+1} \cdots U_1 & \text{if } p = i+1, \\ (U_i \dots U_1) \cdot U_p & \text{if } p \geqslant i+2. \end{cases}$$

Thus $U_p \cdot (U_i \dots U_1) \in \mathcal{I}$ except for the cases $i = p-1$ and $i = p+1$. When $p \equiv n-1 \bmod 2$ these exceptional cases never occur, since we have assumed $i \equiv n-1 \bmod 2$, and so $U_p \cdot K_n \in \mathcal{I}$, as required. And when $p \equiv n \bmod 2$, we can compute the contribution from the two exceptional cases to find that, modulo $\mathcal{I}$, $U_p \cdot \mathcal{J}_n$ is equal to

$$\begin{aligned} &(-1)^{n-1}\left(\frac{\mu}{\lambda}\right)^{p-1} U_p \cdot (U_{p-1} \dots U_1) + (-1)^{n-1}\left(\frac{\mu}{\lambda}\right)^{p+1} U_p \cdot (U_{p+1} \dots U_1) \\ &\qquad = (-1)^{n-1}\left(\frac{\mu}{\lambda}\right)^{p-1} \cdot (U_p \dots U_1) + (-1)^{n-1}\left(\frac{\mu}{\lambda}\right)^{p+1} \cdot (U_p \dots U_1) \\ &\qquad = (-1)^{n-1}\left(\frac{\mu}{\lambda}\right)^{p}\left[\left(\frac{\mu}{\lambda}\right)^{-1} + \left(\frac{\mu}{\lambda}\right)^{1}\right] \cdot (U_p \dots U_1) \in \mathcal{I}. \end{aligned}$$

Now, from Definition 2.20, either $(\mu,\lambda)=(v,-1)$ or $(\mu,\lambda)=(-v,v^2)$. In both cases the square bracket above evaluates to $-a$ (recall $a=v+v^{-1}$). Thus $U_p\cdot K_n$ is a multiple of a, and therefore in $\mathcal{I}$, as required. □

Lemma 6.2 *Let n be even. Let $x\in\mathcal{F}_n(a)$, so that $x\cdot\mathcal{J}_n=0$. Then the constant term of x is a multiple of a.*

Proof Let b be the constant term of x, so that x is equal to b plus a linear combination of left multiples of the elements $U_1,\dots,U_{n-1}$. Thus $x\cdot\mathcal{J}_n$ is equal to $b\cdot\mathcal{J}_n$ plus a linear combination of left multiples of $U_1\cdot\mathcal{J}_n,\dots,U_{n-1}\cdot\mathcal{J}_n$, all of which lie in $\mathcal{I}$ by Lemma 6.1. Thus $x\cdot\mathcal{J}_n=b\cdot\mathcal{J}_n$ modulo $\mathcal{I}$.

As an R–module, the quotient $\mathrm{TL}_n/\mathcal{I}$ is isomorphic to the direct sum of copies of R/aR, with one summand for each monomial whose Jones normal form ends with U_1. We have that

$$\mathcal{J}_n=(-1)^{n-1}\Big[\Big(\frac{\mu}{\lambda}\Big)U_1+\Big(\frac{\mu}{\lambda}\Big)^3U_3U_2U_1+\cdots\Big]\quad\text{in }\mathrm{TL}_n/\mathcal{I},$$

and it follows that

$$b\cdot\mathcal{J}_n=(-1)^{n-1}\Big[b\Big(\frac{\mu}{\lambda}\Big)U_1+b\Big(\frac{\mu}{\lambda}\Big)^3U_3U_2U_1+\cdots\Big]\quad\text{in }\mathrm{TL}_n/\mathcal{I},$$

so b must vanish in R/aR. □

Lemma 6.3 *Let n be even. Then the image of the map*

$$\mathbb{1}\otimes_{\mathrm{TL}_n(a)}\mathcal{F}_n(a)\to\mathbb{1},\qquad 1\otimes x\mapsto 1\cdot x,$$

is contained in the ideal generated by a.

Proof Since the elements U_p act on $\mathbb{1}$ as multiplication by 0, the map above simply sends $1\otimes x$ to the constant term of x. But the previous lemma tells us that the constant term of x is a multiple of a. □

Proof of Theorem C Let n be even. From Theorem 5.1, we have the (fairly short) exact sequence

$$0\to\mathrm{Tor}_n^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})\to\mathbb{1}\otimes_{\mathrm{TL}_n}\mathcal{F}_n\to\mathbb{1}\to\mathrm{Tor}_{n-1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})\to 0,$$

and the image of $\mathbb{1}\otimes_{\mathrm{TL}_n}\mathcal{F}_n\to\mathbb{1}$ is contained in the ideal generated by a, and in particular does not contain the element 1, so that $\mathrm{Tor}_{n-1}^{\mathrm{TL}_n}(\mathbb{1},\mathbb{1})\neq 0$. □

7 The case of $\mathrm{TL}_2(a)$

In this section we briefly consider the case $n=2$, and fully compute the Tor and Ext groups. We do this first by a straightforward computation using an explicit free resolution. Then, in order to illustrate the theory developed in the paper, we reprove the same result by explicitly computing the Fineberg module $\mathscr{F}_2$ and applying Theorem 5.1.

Proposition 7.1 *The homology and cohomology of* $\mathrm{TL}_2(a)$ *are*

$$\mathrm{Tor}_i^{\mathrm{TL}_2(a)}(\mathbb{1},\mathbb{1}) = \begin{cases} R & \text{if } i=0, \\ R/aR & \text{if } i>0,\ i \text{ odd}, \\ R_a & \text{if } i>0,\ i \text{ even}, \end{cases} \qquad \mathrm{Ext}^i_{\mathrm{TL}_2(a)}(\mathbb{1},\mathbb{1}) = \begin{cases} R & \text{if } i=0, \\ R_a & \text{if } i>0,\ i \text{ odd}, \\ R/aR & \text{if } i>0,\ i \text{ even}, \end{cases}$$

where R_a denotes the kernel of the map $R \xrightarrow{a} R$. This holds for any choice of ground ring R and any choice of parameter $a \in R$.

Proof We define a chain complex of left TL_2–modules as follows:

$$\begin{array}{cc}
\vdots & \\
\big\downarrow{\scriptstyle (a-U_1)} & \\
\mathrm{TL}_2 & 3 \\
\big\downarrow{\scriptstyle U_1} & \\
\mathrm{TL}_2 & 2 \\
\big\downarrow{\scriptstyle (a-U_1)} & \\
\mathrm{TL}_2 & 1 \\
\big\downarrow{\scriptstyle U_1} & \\
\mathrm{TL}_2 & 0 \\
\big\downarrow & \\
\mathbb{1} & -1
\end{array}$$

The degree is indicated in the right-hand column. The boundary maps are given by right multiplication by the indicated element of TL_2, except for the last, which is the map $\mathrm{TL}_2 \to \mathbb{1}$, $x \mapsto x \cdot 1$.

The composite of consecutive boundary maps is 0, due to the computation

$$U_1 \cdot (a - U_1) = 0 = (a - U_1) \cdot U_1,$$

and the fact that U_1 acts by 0 on $\mathbb{1}$. Moreover, this complex is acyclic, as one sees by considering the bases $1, U_1$ and $1, (a - U_1)$ of TL_2. Thus the nonnegative part of the complex above, which we denote by P_*, is a free resolution of the left TL_2–module $\mathbb{1}$. Thus $\mathrm{Tor}_*^{\mathrm{TL}_2}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_2}(\mathbb{1},\mathbb{1})$ are the homology of $\mathbb{1} \otimes_{\mathrm{TL}_2} P_*$ and the cohomology of $\mathrm{Hom}_{\mathrm{TL}_2}(P_*, \mathbb{1})$, respectively. Using the isomorphisms $\mathbb{1} \otimes_{\mathrm{TL}_2} \mathrm{TL}_2 \cong \mathbb{1}$ given by $a \otimes x \mapsto a \cdot x$, and $\mathrm{Hom}^{\mathrm{TL}_2}(\mathrm{TL}_2, \mathbb{1}) \cong \mathbb{1}$ given by $f \mapsto f(1)$ in every degree, and working out the induced boundary maps, we see that $\mathbb{1} \otimes_{\mathrm{TL}_2} P_*$ and $\mathrm{Hom}_{\mathrm{TL}_2}(P_*, \mathbb{1})$ are isomorphic to the complexes depicted below:

$$\begin{array}{ccc}
\vdots & \vdots & \vdots \\
\Big\downarrow a & \Big\uparrow a & \\
R & R & 3 \\
\Big\downarrow 0 & \Big\uparrow 0 & \\
R & R & 2 \\
\Big\downarrow a & \Big\uparrow a & \\
R & R & 1 \\
\Big\downarrow 0 & \Big\uparrow 0 & \\
R & R & 0
\end{array}$$

The homology and cohomology of these complexes are easily computed, and give the claim. □

Proposition 7.2 *When $n = 2$ the Fineberg module satisfies $\mathscr{F}_2(a) \cong \mathbb{1}$, and the map*

$$\mathbb{1} \otimes_{\mathrm{TL}_2(a)} \mathscr{F}_2(a) \to \varepsilon_2 \cong \mathbb{1}$$

is multiplication by a.

Proof We compute $\mathscr{F}_2$ explicitly in Example 4.11: $\mathscr{F}_2 \cong \langle a - U_1\rangle \cong \mathbb{1}$. The map $\mathbb{1} \otimes_{\mathrm{TL}_2} \mathscr{F}_2 \to \varepsilon_2 \cong \mathbb{1}$ is the composite map

$$\mathbb{1} \otimes_{\mathrm{TL}_2} \mathscr{F}_2 \to \mathbb{1} \otimes_{\mathrm{TL}_2} W(2)_1 = \mathbb{1} \otimes_{\mathrm{TL}_2} (\mathrm{TL}_2 \otimes_{\mathrm{TL}_0} \mathbb{1}) \cong \mathbb{1}.$$

Under the central equality the basis element $a - U_1$ of $\mathscr{F}_2 \subset W(2)_1$ gets mapped to $a - U_1 = a$ in the tensor product. Therefore the composite map is given by multiplication by a, as required. □

Corollary 7.3 *Suppose that $v \in R$ is a unit and that $a = v + v^{-1}$. Then the groups $\mathrm{Tor}_i^{\mathrm{TL}_2(a)}(\mathbb{1}, \mathbb{1})$ and $\mathrm{Ext}_i^{\mathrm{TL}_2(a)}(\mathbb{1}, \mathbb{1})$ are as described in Proposition 7.1.*

Proof In the light of Proposition 7.2, the exact sequence from Theorem 5.1

$$0 \to \mathrm{Tor}_2^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}) \to \mathbb{1} \otimes_{\mathrm{TL}_2} \mathscr{F}_2 \to \mathbb{1} \to \mathrm{Tor}_1^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}) \to 0$$

now becomes

$$0 \to \mathrm{Tor}_2^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}) \to \mathbb{1} \otimes_{\mathrm{TL}_2} \mathbb{1} \xrightarrow{a} \mathbb{1} \to \mathrm{Tor}_1^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}) \to 0,$$

from which one can compute $\mathrm{Tor}_2^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}) = R_a$ and $\mathrm{Tor}_1^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}) = R/aR$. For $i \geqslant 3$ we have the recursive formula

$$\mathrm{Tor}_i^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}) = \mathrm{Tor}_{i-2}^{\mathrm{TL}_2}(\mathbb{1}, \mathscr{F}_2) \cong \mathrm{Tor}_{i-2}^{\mathrm{TL}_2}(\mathbb{1}, \mathbb{1}),$$

which completes the proof. The Ext results similarly follow from Theorem 5.1. □

8 High acyclicity

In this final section we prove high connectivity of $W(n)$, Theorem E.

Theorem E *$H_d(W(n))$ vanishes in degrees $d \leqslant n-2$.*

8.1 A filtration

In this subsection we introduce a filtration of $W(n)$. We state a theorem relating the filtration quotients to $W(n-1)$ (the proof of which is the topic of the next three subsections) and therefore, by induction, prove Theorem E.

Definition 8.1 (the filtration) We define a filtration F of $W(n)$,

$$F^0 \subseteq F^1 \subseteq \cdots \subseteq F^n = W(n),$$

as follows:

- F^0 is defined to be the span of the elements of two kinds. We call elements of the first kind *basic elements* and these are of the form
$$x \otimes 1$$
in degrees i such that $-1 \leqslant i \leqslant n-2$, where x is represented by a monomial in the s_j not involving the letter s_1. Elements of the second kind are those of the form
$$x \cdot (s_1 \cdots s_{n-i-1}) \otimes 1$$
in degrees i such that $0 \leqslant i \leqslant n-1$, where again x is represented by a monomial not involving the letter s_1.
- F^k for $k \geqslant 1$ is defined to be the span of F^{k-1} together with terms of the form
$$x \cdot (s_1 \cdots s_{n-i-1+k}) \otimes 1$$
in degrees i such that $k \leqslant i \leqslant n-1$, where again x is represented by a monomial not involving s_1.

Remark 8.2 In the description of F^0, it is possible for the product $s_1 \cdots s_{n-i-1}$ to be empty, ie the unit element, if the final index $n-i-1$ is zero ($i = n-1$). In contrast, in the description of F^k for $k \geqslant 1$, the product $s_1 \cdots s_{n-i-1+k}$ is never empty. This is one reason why it is important for us to treat F^0 quite separately from the other F^k, as is done in the remainder of this paper.

In Theorem 8.7 we show that each F^k is a subcomplex of $W(n)$. The fact that $F^n = W(n)$ will follow from Lemma 8.25.

Definition 8.3 Recall that the *cone* on a chain complex X (or, more precisely, the cone on the identity map of X) is the chain complex CX defined by $(CX)_i = X_i \oplus X_{i-1}$, and with differential defined by

$$d^i_{CX}(x, y) = (d^i_X(x) + y, -d^{i-1}_X(y)).$$

The *suspension* of a chain complex X is the complex ΣX defined by

$$(\Sigma X)_i = X_{i-1}$$

and with the same differential as X. The *truncation to degree* p of a chain complex X is the chain complex $\tau_p X$ defined by

$$(\tau_p X)_i = \begin{cases} X_i & \text{if } i \leqslant p, \\ 0 & \text{if } i > p, \end{cases}$$

and with the same differential as X (in the relevant degrees).

Remark 8.4 Our definitions of cone and suspension do not seem to match up very well. However, we have chosen our conventions in order to make the proof of the next theorem as direct as possible, and we believe that our choices are the best fit for this purpose.

Definition 8.5 Define the *shift map* σ to be the map

$$\sigma\colon \mathrm{TL}_{n-1}(a) \to \mathrm{TL}_n(a)$$

which sends each U_i to U_{i+1} for $1 \leqslant i \leqslant n-2$, and hence each s_i to s_{i+1}.

Lemma 8.6 *Each F^k consists of $\mathrm{TL}_{n-1}(a)$–submodules of $W(n)$, where $\mathrm{TL}_{n-1}(a)$ acts via the shift map σ.*

Proof Definition 8.1 defines each F^k as the span of certain "base elements" of the form $y \otimes 1$, where $y \in \mathrm{TL}_n$ is represented by a monomial in the s_j subject to certain restrictions. Multiplying any such y on the left by any s_j for $1 < j \leqslant n-1$ does not affect whether it meets these restrictions. Since $s_j = \sigma(s_{j-1})$ for $1 < j \leqslant n-1$, this shows that the generators of TL_{n-1} send the base elements of each F^k to other base elements of F^k, and therefore F^k itself is stable under the action of TL_{n-1}. □

Here is the main result of this section.

Theorem 8.7 *Each F^k is a subcomplex of $W(n)$. We identify*

$$F^0 \cong C(W(n-1)).$$

And for $k \geqslant 1$,

$$F^k / F^{k-1} \cong \tau_{n-1} \Sigma^{k+1} W(n-1).$$

Corollary 8.8 (Theorem E) *For each $n \geqslant 0$ the complex $W(n)$ is $(n-2)$–acyclic, or in other words, its homology vanishes up to and including degree $n-2$.*

Proof We prove this by induction on $n \geqslant 0$. One can verify the claim directly in the case $n = 0$. Fix $n \geqslant 1$ and suppose that the theorem has been proved for the previous case. Now $W(n)$ has the filtration $F^0 \subseteq F^1 \subseteq \cdots \subseteq F^n$. We prove below that F^0 and all filtration quotients F^k/F^{k-1} are $(n-2)$–acyclic, and then it follows (for example by using the short exact sequences $0 \to F^{k-1} \to F^k \to F^k/F^{k-1} \to 0$, or by using the spectral sequence of the filtration) that the same holds for $W(n)$ itself.

Observe that $F^0 \cong C(W(n-1))$, being isomorphic to a cone, is acyclic. Next, for $k \geqslant 1$ we have $F^k/F^{k-1} \cong \tau_{n-1}\Sigma^{k+1}W(n-1)$. The induction hypothesis states that $W(n-1)$ is $(n-3)$–acyclic, so that $\Sigma^{k+1}W(n-1)$ is $(n-2+k)$–acyclic and in particular $(n-2)$–acyclic, so that $\tau_{n-1}\Sigma^{k+1}W(n-1)$ is also $(n-2)$–acyclic. This completes the proof. □

Remark 8.9 (intuitions and motivations) The complex of planar injective words $W(n)$ is an analogue of the complex of injective words $\mathscr{C}(n)$, and Theorem E is the analogue for $W(n)$ of the well-known vanishing result for the homology of $\mathscr{C}(n)$; see Remark 4.3.

Our starting point in proving Theorem E was Kerz's proof [2005] of the vanishing theorem for the homology of $\mathscr{C}(n)$. Kerz identifies within $\mathscr{C}(n)$ a subcomplex F^0 that is isomorphic to the cone $C(\mathscr{C}(n-1))$. This is then extended to a filtration $F^0 \subseteq F^1 \subseteq \cdots \subseteq F^{n-1} \subseteq \mathscr{C}(n)$ in which each subsequent filtration quotient F^k/F^{k-1} is isomorphic to a direct sum of copies of the suspension $\Sigma^{k+1}\mathscr{C}(n-k-1)$. (In fact Kerz does not explicitly mention filtrations, but this is one way of framing his proof.) This permits an inductive proof of high acyclicity as in Corollary 8.8.

Our proof of Theorem E began as an attempt to mimic Kerz's approach. There is an evident way to embed $W(n-1)$ into $W(n)$ — this is the span of the basic elements of F^0 — and this can be extended to an embedding of the cone $C(W(n-1))$ into $W(n)$ by considering the elements of the second kind in F^0. The remainder of our proof is the result of attempting to extend this embedding to a complete filtration of $W(n)$. At this stage the parallels with [Kerz 2005] begin to fail, but the Jones normal form gives us an extra tool. Using this we characterise the basis elements of $W(n)$ that are not in the image of the cone $C(W(n-1))$, and this characterisation gives a surprising separation into subcomplexes which "look like" suspended and truncated copies of $W(n-1)$ — we build our filtration such that these are our filtration quotients F^k/F^{k-1}.

The final three subsections prove Theorem 8.7, by first setting up the required chain map for F^0, then for F^k, and then in the final section proving these chain maps are isomorphisms.

8.2 Proofs for F^0

In this subsection we prove F^0 is a subcomplex of $W(n)$. We define a map from the cone $C(W(n-1))$ to F^0 and prove this is a well-defined chain map.

Lemma 8.10 *F^0 is a subcomplex of $W(n)$.*

Proof To prove the claim, we must take a generator of F^0 in degree i, and show that under the boundary map $d^i\colon W(n)_i \to W(n)_{i-1}$ this generator is mapped into F^0. Since d^i is the alternating sum $d^i_0 - d^i_1 + \cdots + (-1)^i d^i_i$, it will suffice to fix j in the range $0 \leqslant j \leqslant i$, and show that d^i_j sends our generator into F^0. Recall from Definition 4.1 that

$$d^i_j(y \otimes r) = y \cdot (s_{n-i+j-1} \cdots s_{n-i}) \otimes \lambda^{-j} r.$$

Generators of F^0 come in two kinds. The first kind are the basic elements $x\otimes 1$ in degrees $-1\leqslant i\leqslant n-2$, where x is represented by a monomial not featuring the letter s_1. The map d^i_j only introduces a letter s_1 in the case $i=n-1$, which is excluded here, so that $d^i_j(x\otimes 1)$ is again a basic element and therefore also lies in F^0.

The second kind of generators of F^0 are elements

$$x\cdot(s_1\cdots s_{n-i-1})\otimes 1$$

in degrees $0\leqslant i\leqslant n-1$, where x is represented by a monomial not involving s_1. In the case $j=0$,

$$d^i_0(x\cdot(s_1\cdots s_{n-i-1})\otimes 1)=x\cdot(s_1\cdots s_{n-i-1})\otimes 1,$$

but this lies in $W(n)_{i-1}=\mathrm{TL}_n\otimes_{\mathrm{TL}_{n-i}}\mathbb{1}$, hence is equal to $x\otimes\lambda^{n-i-1}$, and since x is represented by a monomial not involving s_1, this does indeed lie in F^0. (This argument includes the special case $i=n-1$, where the product $s_1\cdots s_{n-i-1}$ is empty, but this clearly creates no issues.) In the case $j\geqslant 1$,

$$\begin{aligned}d^i_j(x\cdot(s_1\cdots s_{n-i-1})\otimes 1)&=x\cdot(s_1\cdots s_{n-i-1})\cdot(s_{n-i+j-1}\cdots s_{n-i})\otimes\lambda^{-j}\\&=x\cdot(s_1\cdots s_{n-i-1})\cdot(s_{n-i+j-1}\cdots s_{n-i+1})\cdot s_{n-i}\otimes\lambda^{-j}\\&=x\cdot(s_{n-i+j-1}\cdots s_{n-i+1})\cdot(s_1\cdots s_{n-i-1})\cdot s_{n-i}\otimes\lambda^{-j}\\&=x\cdot(s_{n-i+j-1}\cdots s_{n-i+1})\cdot(s_1\cdots s_{n-i})\otimes\lambda^{-j}\\&=(x\cdot(s_{n-i+j-1}\cdots s_{n-i+1}))\cdot(s_1\cdots s_{n-(i-1)-1})\otimes\lambda^{-j},\end{aligned}$$

which lies in F^0 since $x\cdot(s_{n-i+j-1}\cdots s_{n-i+1})$ does not involve the letter s_1, so $d^i_j(x\cdot(s_1\cdots s_{n-i-1})\otimes 1)$ is a scalar multiple of a generator of F^0, and thus in F^0, as required. □

Definition 8.11 Define a map

$$\Phi^0\colon C(W(n-1))\to F^0$$

as follows. Recall that

$$C(W(n-1))_i=W(n-1)_i\oplus W(n-1)_{i-1}=(\mathrm{TL}_{n-1}(a)\otimes_{\mathrm{TL}_{n-i-2}(a)}\mathbb{1})\oplus(\mathrm{TL}_{n-1}(a)\otimes_{\mathrm{TL}_{n-i-1}(a)}\mathbb{1})$$

and that

$$F^0_i\subseteq W(n)_i=\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_{n-i-1}(a)}\mathbb{1}.$$

We define Φ^0 in degree i by the rule

$$\Phi^0_i(x\otimes\alpha,y\otimes\beta)=\xi_i(x\otimes\alpha)+\eta_i(y\otimes\beta),$$

where

$$\begin{aligned}\xi_i\colon W(n-1)_i&\to W(n)_i, & x\otimes\alpha&\mapsto\sigma(x)\otimes\lambda^{n-1}\alpha,\\ \eta_i\colon W(n-1)_{i-1}&\to W(n)_i, & y\otimes\beta&\mapsto\sigma(y)\cdot(s_1\cdots s_{n-i-1})\otimes\lambda^i\beta.\end{aligned}$$

It is simple to check that the image of both maps lies in F^0_i.

Lemma 8.12 *The maps ξ_i and η_i are well defined.*

Proof In the case of ξ_i this is simple to verify, as the map $\sigma\colon \mathrm{TL}_{n-1} \to \mathrm{TL}_n$ is in fact a map of right modules with respect to the map of algebras $\sigma\colon \mathrm{TL}_{n-i-2} \to \mathrm{TL}_{n-i-1}$.

In the case of η_i, the definition of $\eta_i(y \otimes \beta)$ as presented depends on y and β themselves, and we must check that it depends only on $y \otimes \beta$. Thus we must show that

$$\eta_i(ys_j \otimes \beta) = \eta_i(y \otimes \lambda\beta)$$

whenever $1 \leqslant j \leqslant n-i-2$. And indeed,

$$\begin{aligned} \eta_i(ys_j \otimes \beta) &= \sigma(ys_j)\cdot(s_1\cdots s_{n-i-1}) \otimes \lambda^i\beta \\ &= \sigma(y)\cdot s_{j+1}\cdot(s_1\cdots s_{n-i-1}) \otimes \lambda^i\beta \\ &= \sigma(y)\cdot(s_1\cdots s_{n-i-1})\cdot s_j \otimes \lambda^i\beta \\ &= \sigma(y)\cdot(s_1\cdots s_{n-i-1}) \otimes \lambda^{i+1}\beta \\ &= \eta_i(y \otimes \lambda\beta), \end{aligned}$$

where the third equality holds since $2 \leqslant j+1 \leqslant n-i-1$ (a simple way to see this is to draw the s_i as braids), and the fourth holds since $j \leqslant n-i-2$ and the tensor product is over TL_{n-i-1}. □

Lemma 8.13 *The ξ_i and η_i interact with the boundary maps of $W(n)$ in the following way:*

1. $d^i_j \circ \xi_i = \xi_{i-1} \circ d^i_j$ *for i in the range $-1 \leqslant i \leqslant n-2$ and j in the range $0 \leqslant j \leqslant i$.*
2. $d^i_0 \circ \eta_i = \xi_{i-1}$ *for i in the range $0 \leqslant i \leqslant n-1$.*
3. $d^i_{j+1} \circ \eta_i = \eta_{i-1} \circ d^{i-1}_j$ *for i in the range $0 \leqslant i \leqslant n-1$ and j in the range $0 \leqslant j \leqslant i-1$.*

Proof For the first point,

$$\begin{aligned} d^i_j(\xi_i(x \otimes \alpha)) = d^i_j(\sigma(x) \otimes \lambda^{n-1}\alpha) &= \sigma(x)\cdot(s_{n-i+j-1}\cdots s_{n-i}) \otimes \lambda^{-j}\lambda^{n-1}\alpha \\ &= \sigma(x\cdot(s_{n-i+j-2}\cdots s_{n-i-1})) \otimes \lambda^{-j}\lambda^{n-1}\alpha \\ &= \xi_{i-1}(x\cdot(s_{n-i+j-2}\cdots s_{n-i-1}) \otimes \lambda^{-j}\alpha) \\ &= \xi_{i-1}(x\cdot(s_{(n-1)-i+j-1}\cdots s_{(n-1)-i}) \otimes \lambda^{-j}\alpha) \\ &= \xi_{i-1}(d^i_j(x \otimes \alpha)). \end{aligned}$$

For the second point,

$$\begin{aligned} d^i_0(\eta_i(y \otimes \beta)) &= d^i_0(\sigma(y)\cdot(s_1\cdots s_{n-i-1}) \otimes \lambda^i\beta) \\ &= \sigma(y)\cdot(s_1\cdots s_{n-i-1}) \otimes \lambda^i\beta \\ &= \sigma(y) \otimes \lambda^{n-i-1}\lambda^i\beta \\ &= \sigma(y) \otimes \lambda^{n-1}\beta \\ &= \xi_{i-1}(y \otimes \beta), \end{aligned}$$

where the third equality holds because the terms lie in $W(n)_{i-1} = \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i}} \mathbb{1}$. For the third point,

$$\begin{aligned}
d^i_{j+1}\eta_i(y\otimes\beta) &= d^i_{j+1}(\sigma(y)\cdot(s_1\cdots s_{n-i-1})\otimes\lambda^i\beta)\\
&= \sigma(y)\cdot(s_1\cdots s_{n-i-1})\cdot(s_{n-i+(j+1)-1}\cdots s_{n-i})\otimes\lambda^{-j-1}\lambda^i\beta\\
&= \sigma(y)\cdot(s_1\cdots s_{n-i-1})\cdot(s_{n-i+j}\cdots s_{n-i+1})\cdot s_{n-i}\otimes\lambda^{i-j-1}\beta\\
&= \sigma(y)\cdot(s_{n-i+j}\cdots s_{n-i+1})\cdot(s_1\cdots s_{n-i})\otimes\lambda^{i-j-1}\beta\\
&= \sigma(y\cdot(s_{n-i+j-1}\cdots s_{n-i}))\cdot(s_1\cdots s_{n-(i-1)-1})\otimes\lambda^{i-1}\lambda^{-j}\beta\\
&= \eta_{i-1}(y\cdot(s_{n-i+j-1}\cdots s_{n-i})\otimes\lambda^{-j}\beta)\\
&= \eta_{i-1}(y\cdot(s_{(n-1)-(i-1)+j-1}\cdots s_{(n-1)-(i-1)})\otimes\lambda^{-j}\beta)\\
&= \eta_{i-1}(d^{i-1}_j(y\otimes\beta)),
\end{aligned}$$

where for the final equality we recall that the source of η_{i-1} is $W(n-1)_{i-2}$. □

Lemma 8.14 *Φ^0 is a chain map.*

Proof Referring to the definition of the differential on $C(W(n-1))$ (Definition 8.3), we see that in order to check that $d^i\circ\Phi^0_i = \Phi^0_{i-1}\circ d^i$, it is enough to show that $d^i\circ\xi_i(x\otimes\alpha) = \xi_{i-1}(d^i(x\otimes\alpha))$ and $d^i\circ\eta_i(y\otimes\beta) = \xi_{i-1}(y\otimes\beta) - \eta_{i-1}(d^{i-1}(y\otimes\beta))$. Using the previous lemma, for the first we have

$$\begin{aligned}
d^i\circ\xi_i(x\otimes\alpha) &= \sum_{j=0}^{i}(-1)^j d^i_j(\xi_i(x\otimes\alpha))\\
&= \sum_{j=0}^{i}(-1)^j \xi_{i-1}(d^i_j(x\otimes\alpha))\\
&= \xi_{i-1}\Bigl(\sum_{j=0}^{i}(-1)^j d^i_j(x\otimes\alpha)\Bigr) = \xi_{i-1}(d^i(x\otimes\alpha)).
\end{aligned}$$

And for the second we have

$$\begin{aligned}
d^i\circ\eta_i(y\otimes\beta) &= \sum_{j=0}^{i}(-1)^j d^i_j(\eta_i(y\otimes\beta))\\
&= d^i_0(\eta_i(y\otimes\beta)) - \sum_{j=0}^{i-1}(-1)^j d^i_{j+1}\eta_i(y\otimes\beta)\\
&= \xi_{i-1}(y\otimes\beta) - \sum_{j=0}^{i-1}(-1)^j \eta_{i-1}d^{i-1}_j(y\otimes\beta)\\
&= \xi_{i-1}(y\otimes\beta) - \eta_{i-1}\Bigl(\sum_{j=0}^{i-1}(-1)^j d^{i-1}_j(y\otimes\beta)\Bigr)\\
&= \xi_{i-1}(y\otimes\beta) - \eta_{i-1}(d^{i-1}(y\otimes\beta)).
\end{aligned}$$
□

8.3 Proofs for F^k, for $k \geqslant 1$

In this subsection we prove, for $k \geqslant 1$, that F^k is a subcomplex of $W(n)$. We define a map from $\tau_{n-1}\Sigma^{k+1}W(n-1)$ to F^k/F^{k-1} and prove this is a well-defined chain map. We start off with some elementary lemmas involving the s_j, which we require for later proofs.

Lemma 8.15 *Let $m \geqslant 1$ and $p \leqslant m$. Then*

$$s_1 \cdots s_m \cdots s_p = (s_m \cdots s_{p+1}) \cdot (s_1 \cdots s_m).$$

In the case $m = p$ the product $s_m \cdots s_{p+1}$ is empty and therefore equal to 1.

Lemma 8.16 *Let $p \geqslant 1$ and $q \geqslant r \geqslant 1$. The product $(s_1 \cdots s_p) \cdot (s_q \cdots s_r)$ can be described as follows:*

(1) *When $r - 1 \leqslant p \leqslant q - 1$,*

$$(s_1 \cdots s_p) \cdot (s_q \cdots s_r) = (s_q \cdots s_{r+1}) \cdot (s_1 \cdots s_{p+1}).$$

(2) *When $p = q$, $(s_1 \cdots s_p) \cdot (s_q \cdots s_r)$ is a linear combination of terms of the form*

$$(s_t \cdots s_{r+1}) \cdot (s_1 \cdots s_t) \quad \text{for } p \geqslant t \geqslant r+1,$$

as well as $s_1 \cdots s_r$ and $s_1 \cdots s_{r-1}$.

(3) *When $p \geqslant q + 1$,*

$$(s_1 \cdots s_p) \cdot (s_q \cdots s_r) = (s_{q+1} \cdots s_{r+1}) \cdot (s_1 \cdots s_p).$$

Proof When $r - 1 \leqslant p \leqslant q - 1$,

$$\begin{aligned}
(s_1 \cdots s_p) \cdot (s_q \cdots s_r) &= (s_1 \cdots s_p) \cdot (s_q \cdots s_{p+2}) \cdot (s_{p+1} \cdots s_r) \\
&= (s_q \cdots s_{p+2}) \cdot (s_1 \cdots s_p) \cdot (s_{p+1} \cdots s_r) \\
&= (s_q \cdots s_{p+2}) \cdot (s_1 \cdots s_{p+1} \cdots s_r) \\
&= (s_q \cdots s_{p+2}) \cdot (s_{p+1} \cdots s_{r+1}) \cdot (s_1 \cdots s_{p+1}) \\
&= (s_q \cdots s_{r+1}) \cdot (s_1 \cdots s_{p+1}),
\end{aligned}$$

where we used Lemma 8.15 to obtain the fourth equality.

When $p = q$, we claim that

$$(s_1 \cdots s_p) \cdot (s_q \cdots s_r) = (s_1 \cdots s_p) \cdot (s_p \cdots s_r)$$

is a linear combination of terms of the form $(s_t \cdots s_{r+1}) \cdot (s_1 \cdots s_t)$ for $p \geqslant t \geqslant r+1$, as well as $s_1 \cdots s_r$ and $s_1 \cdots s_{r-1}$. We will prove this claim by induction on the difference $p - r$. When $p - r = 0$,

$$(s_1 \cdots s_p) \cdot (s_p \cdots s_r) = s_1 \cdots s_p \cdot s_p.$$

Now, since s_p^2 is a linear combination of s_p and 1, this is a linear combination of $s_1\cdots s_p = s_1\cdots s_r$ and $s_1\cdots s_{p-1} = s_1\cdots s_{r-1}$, as required. Now let $p-r \geqslant 1$, and assume that the claim holds for all smaller values. Then

$$(s_1\cdots s_p)\cdot(s_p\cdots s_r) = (s_1\cdots s_{p-1})\cdot s_p^2\cdot(s_{p-1}\cdots s_r)$$

is a linear combination of

$$(s_1\cdots s_{p-1})\cdot s_p\cdot(s_{p-1}\cdots s_r) = s_1\cdots s_p\cdots s_r = (s_p\cdots s_{r+1})\cdot(s_1\cdots s_p)$$

(where we used Lemma 8.15) and

$$(s_1\cdots s_{p-1})\cdot(s_{p-1}\cdots s_r).$$

The former is $(s_t\cdots s_{r+1})\cdot(s_1\cdots s_t)$ in the case $t=p$, while the induction hypothesis tells us that the latter is a linear combination of $(s_t\cdots s_{r+1})\cdot(s_1\cdots s_t)$ for $p-1\geqslant t\geqslant r+1$, as well as $s_1\cdots s_r$ and $s_1\cdots s_{r-1}$. This completes the proof of the claim.

When $p\geqslant q+1$,

$$\begin{aligned}(s_1\cdots s_p)\cdot(s_q\cdots s_r) &= (s_1\cdots s_{q+1})\cdot(s_{q+2}\cdots s_p)\cdot(s_q\cdots s_r)\\ &= (s_1\cdots s_{q+1})\cdot(s_q\cdots s_r)\cdot(s_{q+2}\cdots s_p)\\ &= (s_1\cdots s_{q+1}\cdots s_r)\cdot(s_{q+2}\cdots s_p)\\ &= (s_{q+1}\cdots s_{r+1})\cdot(s_1\cdots s_{q+1})\cdot(s_{q+2}\cdots s_p) = (s_{q+1}\cdots s_{r+1})\cdot(s_1\cdots s_p)\end{aligned}$$

(where we again used Lemma 8.15 to obtain the fourth equality), as required. □

Lemma 8.17 *For $k\geqslant 1$, F^k is a subcomplex of $W(n)$.*

Proof We fix $k\geqslant 1$ and take a generator of F^k/F^{k-1} in degree i, where $k\leqslant i\leqslant n-1$, and show that the boundary map $d^i\colon W(n)_i\to W(n)_{i-1}$ sends our generator into F^k. Since d is the alternating sum $d_0^i - d_1^i + \cdots + (-1)^i d_i^i$, it will suffice to fix j in the range $0\leqslant j\leqslant i$, and show that d_j^i sends our generator into F^k. Recall from Definition 8.1 that our generator of F^k/F^{k-1} in degree i is $x\cdot(s_1\cdots s_{n-i-1+k})\otimes 1$, where x does not involve the letter s_1. Note that

$$n-i-1+k = (n-1)-i+k \geqslant (n-1)-(n-1)+1 = 1,$$

so that the product $(s_1\cdots s_{n-i-1+k})$ is not empty. We have

$$d_j^i(x\cdot(s_1\cdots s_{n-i-1+k})\otimes 1) = x\cdot(s_1\cdots s_{n-i-1+k})\cdot(s_{n-i-1+j}\cdots s_{n-i})\otimes\lambda^{-j},$$

where the factor $(s_{n-i-1+j}\cdots s_{n-i})$ can be empty, in the case $j=0$.

- First we consider the case $j=0$. We find that

$$d_0^i(x\cdot(s_1\cdots s_{n-i-1+k})\otimes 1) = x\cdot(s_1\cdots s_{n-i-1+k})\otimes 1 = x\cdot(s_1\cdots s_{n-(i-1)-1+(k-1)})\otimes 1$$

lies in F^{k-1}, and therefore in F^k, as required.

- Now we consider the case $1 \leqslant j \leqslant k-1$. Then $n-i-1+k \geqslant (n-i-1+j)+1$, so that the third item of Lemma 8.16 applies, and shows that

$$\begin{aligned} d^i_j(x\cdot(s_1\cdots s_{n-i-1+k})\otimes 1) &= x\cdot(s_1\cdots s_{n-i-1+k})\cdot(s_{n-i-1+j}\cdots s_{n-i})\otimes\lambda^{-j} \\ &= x\cdot(s_{n-i+j}\cdots s_{n-i+1})\cdot(s_1\cdots s_{n-i-1+k})\otimes\lambda^{-j} \\ &= x\cdot(s_{n-i+j}\cdots s_{n-i+1})\cdot(s_1\cdots s_{n-(i-1)-1+(k-1)})\otimes\lambda^{-j}. \end{aligned}$$

Since $n-i+1 \geqslant n-(n-1)+1 = 2$, the word $(s_{n-i+j}\cdots s_{n-i+1})$ does not involve s_1, and consequently the element above lies in F^{k-1}, and therefore in F^k.

- Now we consider the case $j = k$. Then $n-i-1+k = n-i-1+j$ and so the second item of Lemma 8.16 applies and shows that

$$d^i_k(x\cdot(s_1\cdots s_{n-i-1+k})\otimes 1) = x\cdot(s_1\cdots s_{n-i-1+k})\cdot(s_{n-i-1+k}\cdots s_{n-i})\otimes\lambda^{-k}$$

is a linear combination of terms

$$x\cdot(s_t\cdots s_{n-i+1})\cdot(s_1\cdots s_t)\otimes\lambda^{-k}$$

for t in the range

$$n-i+1 \leqslant t \leqslant n-i-1+k = n-(i-1)-1+(k-1),$$

together with

$$x\cdot(s_1\cdots s_{n-(i-1)-1})\otimes\lambda^{-k} \quad\text{and}\quad x\cdot(s_1\cdots s_{n-(i-1)-2})\otimes\lambda^{-k} = x\otimes\lambda^{-k}.$$

Now $(s_t\cdots s_{n-i+1})$ does not involve s_1, so the first of these terms lies in F^{k-1}, while the second and third lie in F^0. So altogether we have the required result.

- Now we consider the case $k+1 \leqslant j$. Here

$$n-i-1 \leqslant (n-i-1+k)+1 \leqslant n-i-1+j,$$

so that the first item of Lemma 8.16 applies and shows that

$$\begin{aligned} d^i_j(x\cdot(s_1\cdots s_{n-i-1+k})\otimes 1) &= x\cdot(s_1\cdots s_{n-i-1+k})\cdot(s_{n-i-1+j}\cdots s_{n-i})\otimes\lambda^{-j} \\ &= x\cdot(s_{n-i-1+j}\cdots s_{n-i+1})\cdot(s_1\cdots s_{n-i-1+k+1})\otimes\lambda^{-j} \\ &= x\cdot(s_{n-i-1+j}\cdots s_{n-i+1})\cdot(s_1\cdots s_{n-(i-1)-1+k})\otimes\lambda^{-j}. \end{aligned}$$

Since $(s_{n-i-1+j}\cdots s_{n-i+1})$ does not involve s_1, the element above lies in F^k, as required. $\square$

Definition 8.18 Define a map

$$\Psi^k\colon \tau_{n-1}\Sigma^{k+1}W(n-1) \to F^k/F^{k-1}$$

as follows. Note that for i in the range $k \leqslant i \leqslant n-1$,

$$[\tau_{n-1}\Sigma^{k+1}W(n-1)]_i = W(n-1)_{i-k-1} = \mathrm{TL}_{n-1}(a)\otimes_{\mathrm{TL}_{(n-1)-(i-k-1)-1}(a)}\mathbb{1} = \mathrm{TL}_{n-1}(a)\otimes_{\mathrm{TL}_{n-i-1+k}(a)}\mathbb{1},$$

while $(F^k/F^{k-1})_i$ is a quotient of $\mathrm{TL}_n(a)\otimes_{\mathrm{TL}_{n-i-1}(a)}\mathbb{1}$. Define the degree i part of Ψ to be the map

$$\Psi_i^k\colon \mathrm{TL}_{n-1}(a)\otimes_{\mathrm{TL}_{n-i-1+k}(a)}\mathbb{1}\to(F^k/F^{k-1})_i,\quad x\otimes\alpha\mapsto(-1)^{-i(k+1)}\sigma(x)\cdot(s_1\cdots s_{n-i-1+k})\otimes\lambda^i\alpha.$$

For later convenience, we will denote by ψ_i^k the map

$$\psi_i^k\colon x\otimes\alpha\mapsto\sigma(x)\cdot(s_1\cdots s_{n-i-1+k})\otimes\lambda^i\alpha,$$

so that $\Psi_i^k=(-1)^{-i(k+1)}\psi_i^k$.

Lemma 8.19 *The map ψ_i^k is well defined (and the same therefore holds for Ψ_i^k).*

Proof As presented above, the value of $\psi_i^k(x\otimes\alpha)$ depends on the choices of x and α, rather than on $x\otimes\alpha$. So to check that ψ_i^k is well defined, we must check that $\psi_i^k(xs_p\otimes\alpha)=\psi_i^k(x\otimes\lambda\alpha)$ whenever $p\leqslant(n-i-1+k)-1$. Let us write $q=n-i-1+k$, so that $p\leqslant q-1$. (In particular we are assuming that $q\geqslant 2$.) Now

$$\begin{aligned}\psi_i^k(xs_p\otimes\alpha)=\sigma(xs_p)\cdot(s_1\cdots s_q)\otimes\lambda^i\alpha&=\sigma(x)\cdot s_{p+1}\cdot(s_1\cdots s_q)\otimes\lambda^i\alpha\\&=\sigma(x)\cdot s_{p+1}\cdot(s_1\cdots s_{p-1})\cdot(s_ps_{p+1})\cdot(s_{p+2}\cdots s_q)\otimes\lambda^i\alpha\\&=\sigma(x)\cdot(s_1\cdots s_{p-1})\cdot(s_{p+1}s_ps_{p+1})\cdot(s_{p+2}\cdots s_q)\otimes\lambda^i\alpha.\end{aligned}$$

Recall from Definition 2.21 that

$$s_{p+1}s_ps_{p+1}=\lambda s_ps_{p+1}+\lambda s_{p+1}s_p-\lambda^2s_p-\lambda^2s_{p+1}+\lambda^3.$$

Now

$$\begin{aligned}(s_1\cdots s_{p-1})\cdot(s_ps_{p+1})\cdot(s_{p+2}\cdots s_q)&=(s_1\cdots s_q),\\(s_1\cdots s_{p-1})\cdot(s_{p+1}s_p)\cdot(s_{p+2}\cdots s_q)&=(s_{p+1}\cdots s_q)\cdot(s_1\cdots s_p),\\(s_1\cdots s_{p-1})\cdot s_p\cdot(s_{p+2}\cdots s_q)&=(s_{p+2}\cdots s_q)\cdot(s_1\cdots s_p),\\(s_1\cdots s_{p-1})\cdot s_{p+1}\cdot(s_{p+2}\cdots s_q),&=(s_{p+1}\cdots s_q)\cdot(s_1\cdots s_{p-1}),\\(s_1\cdots s_{p-1})\cdot 1\cdot(s_{p+2}\cdots s_q)&=(s_{p+2}\cdots s_q)\cdot(s_1\cdots s_{p-1}),\end{aligned}$$

so it follows that

$$\begin{aligned}&\psi_i^k(xs_p\otimes\alpha)\\&=\sigma(x)\cdot(s_1\cdots s_q)\otimes\lambda^{i+1}\alpha+\sigma(x)\cdot(s_{p+1}\cdots s_q)\cdot(s_1\cdots s_p)\cdot\otimes\lambda^{i+1}\alpha\\&\qquad-\sigma(x)\cdot(s_{p+2}\cdots s_q)\cdot(s_1\cdots s_p)\otimes\lambda^{i+2}\alpha-\sigma(x)\cdot(s_{p+1}\cdots s_q)\cdot(s_1\cdots s_{p-1})\otimes\lambda^{i+2}\alpha\\&\qquad\qquad+\sigma(x)\cdot(s_{p+2}\cdots s_q)\cdot(s_1\cdots s_{p-1})\otimes\lambda^{i+3}\alpha.\end{aligned}$$

Now $p<n-i-1+k$, which means that the final four terms above all lie in F^{k-1}, so that in F^k/F^{k-1} we have, as required,

$$\psi_i^k(xs_p\otimes\alpha)=\sigma(x)\cdot(s_1\cdots s_q)\otimes\lambda^{i+1}\alpha=\sigma(x)\cdot(s_1\cdots s_{n-i-1+k})\otimes\lambda^{i+1}\alpha=\psi_i^k(x\otimes\lambda\alpha).\qquad\square$$

Lemma 8.20 *Let $k \geqslant 1$ and let $k \leqslant i \leqslant n-1$. Then, for j in the range $j \geqslant k+1$,*

$$\psi^k_{i-1} \circ d^{i-k-1}_{j-k-1} = d^i_j \circ \psi^k_i.$$

Proof Let $x \otimes \alpha \in W(n-1)_{i-k-1} = \mathrm{TL}_{n-1} \otimes_{\mathrm{TL}_{n-i-1+k}} \mathbb{1}$. Then

$$\begin{aligned} d^i_j(\psi^k_i(x \otimes \alpha)) &= d^i_j(\sigma(x) \cdot (s_1 \cdots s_{n-i-1+k}) \otimes \lambda^i \alpha) \\ &= \sigma(x) \cdot (s_1 \cdots s_{n-i-1+k}) \cdot (s_{n-i+j-1} \cdots s_{n-i}) \otimes \lambda^{i-j} \alpha. \end{aligned}$$

Since $n-i-1 \leqslant (n-i-1+k)+1 \leqslant n-i+j-1$, we may apply the first part of Lemma 8.16 to obtain

$$\begin{aligned} d^i_j(\psi^k_i(x \otimes \alpha)) &= \sigma(x) \cdot (s_1 \cdots s_{n-i-1+k}) \cdot (s_{n-i+j-1} \cdots s_{n-i}) \otimes \lambda^{i-j} \alpha \\ &= \sigma(x) \cdot (s_{n-i+j-1} \cdots s_{n-i+1}) \cdot (s_1 \cdots s_{n-i+k}) \otimes \lambda^{i-j} \alpha \\ &= \sigma(x) \cdot (s_{n-i+j-1} \cdots s_{n-i+1}) \cdot (s_1 \cdots s_{n-(i-1)-1+k}) \otimes \lambda^{(i-1)} \lambda^{1-j} \alpha \\ &= \sigma(x \cdot (s_{n-i+j-2} \cdots s_{n-i})) \cdot (s_1 \cdots s_{n-(i-1)-1+k}) \otimes \lambda^{(i-1)} \lambda^{1-j} \alpha \\ &= \psi^k_{i-1}(x \cdot (s_{n-i+j-2} \cdots s_{n-i}) \otimes \lambda^{1-j} \alpha). \end{aligned}$$

In the last line of the above computation, $x \cdot (s_{n-i+j-2} \cdots s_{n-i}) \otimes \lambda^{1-j} \alpha$ is an element of

$$W(n-1)_{(i-1)-k-1} = \mathrm{TL}_{n-1} \otimes_{\mathrm{TL}_{n-i+k}} \mathbb{1},$$

so

$$\begin{aligned} x \cdot (s_{n-i+j-2} \cdots s_{n-i}) \otimes \lambda^{1-j} \alpha &= x \cdot (s_{n-i+j-2} \cdots s_{n-i+k}) \cdot (s_{n-i+k-1} \cdots s_{n-i}) \otimes \lambda^{1-j} \alpha \\ &= x \cdot (s_{n-i+j-2} \cdots s_{n-i+k}) \otimes \lambda^k \lambda^{1-j} \alpha \\ &= x \cdot (s_{n-i+j-2} \cdots s_{n-i+k}) \otimes \lambda^{-(j-k-1)} \alpha. \end{aligned}$$

Thus,

$$\begin{aligned} d^i_j(\psi^k_i(x \otimes \alpha)) &= \psi^k_{i-1}(x \cdot (s_{n-i+j-2} \cdots s_{n-i}) \otimes \lambda^{1-j} \alpha) \\ &= \psi^k_{i-1}(x \cdot (s_{n-i+j-2} \cdots s_{n-i+k}) \otimes \lambda^{-(j-k-1)} \alpha.) \\ &= \psi^k_{i-1}(x \cdot (s_{(n-1)-(i-k-1)+(j-k-1)-1} \cdots s_{(n-1)-(i-k-1)}) \otimes \lambda^{-(j-k-1)} \alpha) \\ &= \psi^k_{i-1}(d^{i-k-1}_{j-k-1}(x \otimes \alpha)), \end{aligned}$$

as required. □

Corollary 8.21 *Ψ^k is a chain map.*

Proof The boundary map of $\tau_{n-1} \Sigma^{k+1} W(n-1)$ is given in degree i by the boundary map

$$d^{i-k-1} \colon W(n-1)_{i-k-1} \to W(n-1)_{i-k-2},$$

which is itself given by the formula $\sum_{j=0}^{i-k-1} (-1)^j d^{i-k-1}_j$.

The boundary map of F^k / F^{k-1} is given in degree i by the boundary map of $W(n)$ in degree i, which is the alternating sum $\sum_{j=0}^{i} (-1)^j d^i_j$. However, the proof of Lemma 8.17 shows that $d^i_0, \dots, d^i_k$ all send F^k into F^{k-1}, and hence that they vanish in the quotient F^k/F^{k-1}. Thus the boundary map of F^k/F^{k-1} is $\sum_{j=k+1}^{i}(-1)^j d_j^i$. It follows that

$$\begin{aligned} d^i \circ \Psi_i^k &= \sum_{j=k+1}^{i} (-1)^j d_j^i \circ [(-1)^{-i(k+1)} \psi_i^k] = \sum_{j=k+1}^{i} (-1)^{j-i(k+1)} \psi_{i-1}^k \circ d_{j-k-1}^{i-k-1} \\ &= \sum_{j=0}^{i-k-1} (-1)^{j+(k+1)-i(k+1)} \psi_{i-1}^k \circ d_j^{i-k-1} \\ &= [(-1)^{-(i-1)(k+1)} \psi_{i-1}^k] \circ \sum_{j=0}^{i-k-1} (-1)^j \psi_{i-1}^k \circ d_j^{i-k-1} = \Psi_{i-1}^k \circ d^{i-k-1}. \end{aligned}$$ □

8.4 Proof of Theorem 8.7

In this subsection we prove Theorem 8.7, which in turn completes the proof of Theorem E.

We begin by finding a basis for each part of the filtration in terms of the Jones normal form. This is done in Lemma 8.25 below, after some preliminary work.

Lemma 8.22 *Any word in the s_i not containing s_1 is a linear combination of words in the U_i, none of which involve U_1. Conversely, any word in the U_i not containing U_1 is a linear combination of words in the s_i not containing s_1.*

Proof Recall from Definition 2.21 that $s_i = \lambda + \mu U_i$, where λ and μ are both units in the ground ring, so that $U_i = -\mu^{-1}\lambda + \mu^{-1}s_i$. The claim follows immediately. □

Lemma 8.23 *For $1 \leqslant p \leqslant n-1$, the word $s_1 \dots s_p$ written in terms of the U_i generators is equal to $\mu^p U_1 \dots U_p$, plus a linear combination of scalar multiples (by units) of words w in the U_i with the properties*

- *$i(w) \geqslant 2$ and $t(w) \leqslant p$, or*
- *$i(w) = 1$ and $t(w) < p$.*

In particular, only the summand $w = \mu^p U_1 \dots U_p$ satisfies $i(w) = 1$ and $t(w) = p$.

Proof Using $s_i = \lambda + \mu U_i$ and multiplying out brackets gives

$$\begin{aligned} s_1 \dots s_p &= \sum_{r=0}^{p} \sum_{(1 \leqslant i_1 \leqslant \dots \leqslant i_r \leqslant p)} \lambda^{p-r} \mu^r U_{i_1} U_{i_2} \dots U_{i_r} \\ &= \mu^p U_1 \dots U_p + \sum_{r=0}^{p-1} \sum_{(1 \leqslant i_1 \leqslant \dots \leqslant i_r \leqslant p)} \lambda^{p-r} \mu^r U_{i_1} U_{i_2} \dots U_{i_r}. \end{aligned}$$

If $r = 0$ the term is a scalar, which has index ∞ by convention (thus the first point is satisfied). Suppose $0 < r < p$. Then if $i_1 > 1$ it follows that $i(U_{i_1} \dots U_{i_p}) \geqslant 2$. Otherwise $i_1 = 1$ and, since $r < p$, there is some $j \geqslant 2$ such that $i_j \geqslant i_{j-1}+2$, so that $U_{i_1}\dots U_{i_r}$ can be written as a word with terminus i_{j-1}, and then the claim follows. Coefficients are given by powers of λ and μ, and products of these. The terms λ and μ are defined via the homomorphisms in Definition 2.20 and lie in the set $\{-1,\pm v, v^2\}$. Since v is a unit, it follows that all coefficients are units. $\square$

Lemma 8.24 *Let $k \geqslant 0$ and $-1 \leqslant i \leqslant n-1$, and consider elements $x_{\underline{a},\underline{b}} \otimes 1$, where $x_{\underline{a},\underline{b}}$ is in Jones normal form and satisfies either*

- *$i(x_{\underline{a},\underline{b}}) \geqslant 2$ and $t(x_{\underline{a},\underline{b}}) \geqslant n-i-1$, or*
- *$i(x_{\underline{a},\underline{b}}) = 1$ and $n-i-1 \leqslant t(x_{\underline{a},\underline{b}}) \leqslant n-i-1+k$.*

Then these elements all lie in F_i^k.

Proof The first type of element lies in F^0, since in this case $x_{\underline{a},\underline{b}}$ is a word in the various U_i not containing U_1, and by Lemma 8.22, this is a linear combination of words in the s_i not containing s_1 (a basic element). For the second type of element, from the definition of Jones normal form, $x_{\underline{a},\underline{b}}$ must end in a string $U_1 \dots U_{n-i-1+j}$ for $0 \leq j \leq k$. We proceed by induction on k.

Base case We start with the base case $k=0$, so the only option is that $j=0$, ie $x_{\underline{a},\underline{b}} = y_{\underline{a},\underline{b}}U_1\dots U_{n-i-1}$ for some $y_{\underline{a},\underline{b}}$ in Jones normal form with $i(y_{\underline{a},\underline{b}}) \geq 2$. We aim to show that in this case $x_{\underline{a},\underline{b}}$ lies in F_i^0. Compare $x_{\underline{a},\underline{b}}$ to $y_{\underline{a},\underline{b}}s_1\cdots s_{n-i-1}$, which does lie in F_i^0 by Definition 8.1. From Lemma 8.23, multiplying out the string $s_1\cdots s_{n-i-1}$ will result in $y_{\underline{a},\underline{b}}s_1\cdots s_{n-i-1}$ being written as a linear combination (up to scalar multiplication by units) of three types of elements, and we consider their image in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$:

(1) $y_{\underline{a},\underline{b}}U_1\dots U_{n-i-1}$. This is equal to $x_{\underline{a},\underline{b}}$ and appears as a single summand of $y_{\underline{a},\underline{b}}s_1\dots s_{n-i-1}$.

(2) $y_{\underline{a},\underline{b}}w$, where $i(w) \geqslant 2$ and $t(w) \leqslant n-i-1$. These are all basic elements since $i(y_{\underline{a},\underline{b}}) \geq 2$, and thus lie in F^0.

(3) $y_{\underline{a},\underline{b}}w$, where $i(w) = 1$ and $t(w) < n-i-1$. These are all zero in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$, due to the terminus.

So it follows that in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$, up to scalar multiplication by units $x_{\underline{a},\underline{b}}$ is equal to a linear combination of $y_{\underline{a},\underline{b}}s_1\cdots s_{n-i-1}$ and basic elements. Since this is a linear combination of elements in F_i^0, it follows that $x_{\underline{a},\underline{b}}$ lies in F_i^0, as required.

Inductive step Assume the lemma is true for $k-1$ and prove for k. Let $x_{\underline{a},\underline{b}} = y_{\underline{a},\underline{b}}U_1\dots U_{n-i-1+j}$ for some $y_{\underline{a},\underline{b}}$ in Jones normal form with

$$i(y_{\underline{a},\underline{b}}) \geq 2 \qquad \text{and} \qquad t(y_{\underline{a},\underline{b}}) > n-i-1+j \quad \text{for } 0 \leq j \leq k.$$

When $0 \leq j < k$, by the inductive hypothesis this element lies in $F_i^{k-1} \subset F_i^k$, and so we can restrict to the case where $j=k$, ie $x_{\underline{a},\underline{b}} = y_{\underline{a},\underline{b}}U_1\dots U_{n-i-1+k}$. We aim to show that, in this case, $x_{\underline{a},\underline{b}}$ lies in F_i^k. As in the base case, we compare $x_{\underline{a},\underline{b}}$ with $y_{\underline{a},\underline{b}}s_1\cdots s_{n-i-1+k}$, which lies in F_i^k by Definition 8.1. From Lemma 8.23, $y_{\underline{a},\underline{b}}s_1\cdots s_{n-i-1+k}$ is a linear combination (up to scalar multiplication by a unit) of three types of elements, which we evaluate in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$:

(1) $y_{\underline{a},\underline{b}} U_1 \dots U_{n-i-1+k}$. This is equal to $x_{\underline{a},\underline{b}}$ and appears as a single summand of $y_{\underline{a},\underline{b}} s_1 \cdots s_{n-i-1+k}$.

(2) $y_{\underline{a},\underline{b}} w$, where $i(w) \geqslant 2$ and $t(w) \leqslant n-i-1+k$. Since $i(y_{\underline{a},\underline{b}}) \geq 2$ they are all basic elements, and thus lie in $F^0 \subset F^k$.

(3) $y_{\underline{a},\underline{b}} w$, where $i(w) = 1$ and $t(w) < n-i-1+k$. Rewriting these in Jones normal form gives elements $y_{\underline{a},\underline{b}} w = z_{\underline{a},\underline{b}}$ such that $i(z_{\underline{a},\underline{b}}) = 1$ and $t(z_{\underline{a},\underline{b}}) \leq t(w) < n-i-1+k$ (by Lemmas 2.6 and 2.7). These are then Jones normal form elements ending in $U_1 \dots U_{n-i-1+j}$ for $0 \leq j \leq k-1$ so by the inductive hypothesis these lie in $F_i^{k-1} \subset F_i^k$.

Again it follows that in $\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$, the element $x_{\underline{a},\underline{b}}$ is, up to scalar multiplication by units, equal to a linear combination of $y_{\underline{a},\underline{b}} s_1 \cdots s_{n-i-k}$, elements in F_i^k (by the inductive hypothesis), and basic elements. Since this is a linear combination of elements in F_i^k, it follows that $x_{\underline{a},\underline{b}}$ lies in F_i^k, as required. □

Lemma 8.25 *Let $k \geqslant 0$ and $-1 \leqslant i \leqslant n-1$. Then F_i^k has basis consisting of elements $x_{\underline{a},\underline{b}} \otimes 1$, where $x_{\underline{a},\underline{b}}$ is in Jones normal form and satisfies either*

- *$i(x_{\underline{a},\underline{b}}) \geqslant 2$ and $t(x_{\underline{a},\underline{b}}) \geqslant n-i-1$, or*
- *$i(x_{\underline{a},\underline{b}}) = 1$ and $n-i-1 \leqslant t(x_{\underline{a},\underline{b}}) \leqslant n-i-1+k$.*

Proof This is a subset of the known basis for $\mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1} \supseteq F_i^k$, and by the previous lemma we know these elements lie in F_i^k, so it is enough to show that F_i^k is spanned by these elements. First of all, note that since $F_i^k \subseteq \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$, any word in F_i^k written in Jones normal form will vanish if $t(x_{\underline{a},\underline{b}}) \leqslant n-i-2$, therefore we will always have $t(x_{\underline{a},\underline{b}}) \geqslant n-i-1$. By definition F_i^k is spanned by elements of the form

- $x \otimes 1$, and
- $x \cdot (s_1 \cdots s_{n-i-1+k'}) \otimes 1$,

where x is a word in the various U_i with $i(x) \geqslant 2$ (ie containing no U_1) and $0 \leqslant k' \leqslant k$ (note that in the case $i = n-1$ and $k' = 0$ the two kinds coincide). The first kind is spanned by $x_{\underline{a},\underline{b}}$ such that $i(x_{\underline{a},\underline{b}}) \geqslant 2$, as described in the first bullet point in the statement of the lemma. From Lemma 8.23, expanding the product $(s_1 \cdots s_{n-i-1+k'})$ in the second kind gives a linear combination of words $x \cdot w \otimes 1$ such that $t(w) \leqslant n-i-1+k'$. Either $i(w)$ will be $\geqslant 2$ or $i(w) = 1$. In the first case, since $i(x) \geqslant 2$ it follows that $i(x \cdot w) \geqslant 2$ and so when written in Jones normal form this will remain the case, giving an element of the first type described in the lemma. In the second case, when $i(w) = 1$, since $i(x) \geqslant 2$ then either $i(x \cdot w) \geqslant 2$ and as in the previous sentence we are done, or $i(x \cdot w) = 1$ and, by Lemma 2.7, when written in Jones normal form the terminus $t(x \cdot w) = t(w) \leqslant n-i-1+k' \leqslant n-i-1+k$ will either remain the same or reduce. This puts us in the setting of the second bullet point in the statement of the lemma, and thus we have shown that the two types of elements span F_i^k. □

Proposition 8.26 *The map $\Phi^0\colon C(W(n-1)) \to F^0$ from Definition 8.11 is an isomorphism.*

Proof Recall that for $-1 \leqslant i \leqslant n-1$,

$$\Phi_i^0\colon (\mathrm{TL}_{n-1} \otimes_{\mathrm{TL}_{n-i-2}} \mathbb{1}) \oplus (\mathrm{TL}_{n-1} \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}) \to F_i^0$$

is given by

$$\Phi_i^0(x \otimes \alpha, y \otimes \beta) = \xi_i(x \otimes \alpha) + \eta_i(y \otimes \beta),$$

where

$$\xi_i(x \otimes \alpha) = \sigma(x) \otimes \lambda^{n-1}\alpha \quad \text{and} \quad \eta_i(y \otimes \beta) = \sigma(y) \cdot (s_1 \cdots s_{n-i-1}) \otimes \lambda^i \beta.$$

By Lemma 2.16, a basis for the left-hand side is given by elements of either the form $(x_{\underline{a},\underline{b}} \otimes 1, 0)$ such that $t(x_{\underline{a},\underline{b}}) > n-i-3$ or the form $(0, x_{\underline{a}',\underline{b}'} \otimes 1)$ such that $t(x_{\underline{a}',\underline{b}'}) > n-i-2$. Under the map Φ_i^0, the element $(x_{\underline{a},\underline{b}} \otimes 1, 0)$ is taken to a scalar multiple (by a unit) of $\sigma(x_{\underline{a},\underline{b}}) \otimes 1$, where $\sigma(x_{\underline{a},\underline{b}})$ is a Jones basis element with $i(\sigma(x_{\underline{a},\underline{b}})) \geqslant 2$ and $t(\sigma(x_{\underline{a},\underline{b}})) > n-i-2$. By Lemma 8.23, the element $(0, x_{\underline{a}',\underline{b}'} \otimes 1)$ is taken to a linear combination of scalar multiples (by units) of terms $\sigma(x_{\underline{a}',\underline{b}'}) \cdot w \otimes 1$ such that $t(w) \leqslant n-i-1$. Since $F_i^0 \subseteq \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$ the only nonzero terms in the image will occur when $t(w) = n-i-1$. We consider two cases: $i(w) \geqslant 2$ or $i(w) = 1$. By Lemma 2.7, converting to Jones normal form in the first case gives an element with index $i(\sigma(x_{\underline{a}',\underline{b}'}) \cdot w) > 2$ and terminus $t(\sigma(x_{\underline{a}',\underline{b}'}) \cdot w) = n-i-1$, or zero, since the terminus will either remain the same or reduce when converting. When $i(w) = 1$ and $t(w) = n-i-1$, by Lemma 8.23 it follows that $w = U_1 \ldots U_{n-i-1}$ and therefore the terms will be of the form $\sigma(x_{\underline{a}',\underline{b}'}) \cdot U_1 \ldots U_{n-i-1}$. These elements are already in Jones normal form, with index 1 and terminus $n-i-1$. Furthermore, all Jones basis elements with this index and terminus arise in this way. By Lemma 8.25 a basis for F_i^0 is given by elements $y_{\underline{a},\underline{b}} \otimes 1$, where $y_{\underline{a},\underline{b}}$ is in Jones normal form and satisfies

- $i(y_{\underline{a},\underline{b}}) \geqslant 2$ and $t(y_{\underline{a},\underline{b}}) \geqslant n-i-1$, or
- $i(y_{\underline{a},\underline{b}}) = 1$ and $t(y_{\underline{a},\underline{b}}) = n-i-1$.

By our analysis, all of these elements lie in the image of Φ_i^0, up to scalar multiplication by units; hence, Φ^0 is a bijection on bases and therefore an isomorphism. □

Lemma 8.27 *A basis for $(F^k/F^{k-1})_i$ is given by words $x_{\underline{a},\underline{b}}$ in Jones normal form such that $i(x_{\underline{a},\underline{b}}) = 1$ and $t(x_{\underline{a},\underline{b}}) = n-i-1+k$.*

Proof This is a direct consequence of taking the quotient of the bases for F^k and F^{k-1} given in Lemma 8.25. □

Proposition 8.28 *The map $\Psi^k\colon \tau_{n-1}\Sigma^{k+1}W(n-1) \to F^k/F^{k-1}$ defined in Definition 8.18 is an isomorphism.*

Proof For i in the range $k \leqslant i \leqslant n-1$, recall the map

$$\Psi_i^k\colon \mathrm{TL}_{n-1} \otimes_{\mathrm{TL}_{n-i-1+k}} \mathbb{1} \to (F^k/F^{k-1})_i, \quad x \otimes \alpha \mapsto (-1)^{-i(k+1)}\sigma(x) \cdot (s_1 \cdots s_{n-i-1+k}) \otimes \lambda^i \alpha.$$

By Lemma 2.16, a basis for the domain is given by $x_{\underline{a},\underline{b}}$ such that $t(x_{\underline{a},\underline{b}}) > (n-i-1+k)-1$. Note also that $x_{\underline{a},\underline{b}}$ does not contain the letter U_{n-1}. By Lemma 8.23, the image $\Psi_i^k(x_{\underline{a},\underline{b}})$ is a linear combination of scalar multiples (by units) of terms $\sigma(x_{\underline{a},\underline{b}})\cdot w$ such that $t(w) \leqslant n-i-1+k$. These terms are zero in $(F^k/F^{k-1})_i \subseteq \mathrm{TL}_n \otimes_{\mathrm{TL}_{n-i-1}} \mathbb{1}$ only when w cannot be written as a word with $t(w) < n-i-1$. Rewriting these elements in Jones normal form will maintain or decrease the terminus, and $i(\sigma(x_{\underline{a},\underline{b}})) \geqslant 2$, so $i(\sigma(x_{\underline{a},\underline{b}})\cdot w) = 1$ only when $i(w) = 1$. Therefore by Lemma 8.25, quotienting out by F^{k-1} leaves only the term for which $i(w) = 1$ and $t(w) = n-i-1+k$. In particular, by Lemma 8.23 this term is a scalar multiple (by a unit) of $\sigma(x_{\underline{a},\underline{b}})\cdot U_1 \dots U_{n-i-1+k}$.

Since $\sigma(x_{\underline{a},\underline{b}})$ has index $\geqslant 2$ and terminus $> n-i-1+k$, it follows that $\sigma(x_{\underline{a},\underline{b}})\cdot U_1 \dots U_{n-i-1+k}$ is in Jones normal form. From Lemma 8.27 this is a Jones basis element for F^k/F^{k-1} and all basis elements arise in this way. Therefore up to unit scalars, the map Ψ^k is a bijection on bases, and hence an isomorphism. □

9 Jones–Wenzl projectors and vanishing

This section relates our results with the existence of the *Jones–Wenzl projectors*, to strengthen our vanishing results when R is a field. This section is written such that the reader can read the introduction, the background on Temperley–Lieb algebras, and continue straight to this section. For the time being we make the substitutions $a \leftrightarrow \delta$ and $v \leftrightarrow q$, as is common in the recent literature concerning Jones–Wenzl projectors.

Throughout this section, we will consider a commutative ring R, a unit $q \in R$, the parameter $\delta = q + q^{-1}$, and we will work in $\mathrm{TL}_n(\delta)$. Recall that we show in Theorem A that, when δ is invertible, $\mathrm{Tor}_*^{\mathrm{TL}_n(\delta)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(\delta)}(\mathbb{1},\mathbb{1})$ vanish in every nonzero degree. In this section we investigate the case where $\delta = 0$ and R is a field using established results on Jones–Wenzl projectors. We prove the following theorem:

Theorem D *Let $n = 2k+1$, and let R be a field whose characteristic does not divide $\binom{k}{t}$ for any $0 \le t \le k$. Let q be a unit in R and assume that $\delta = q + q^{-1} = 0$. Then $\mathrm{Tor}_*^{\mathrm{TL}_n(0)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(0)}(\mathbb{1},\mathbb{1})$ vanish in positive degrees.*

For example, when $n = 3$, R is a field and $\delta = q + q^{-1}$ for $q \in R^\times$, then combining this theorem with Theorem A demonstrates that $\mathrm{Tor}_*^{\mathrm{TL}_3(\delta)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_3(\delta)}(\mathbb{1},\mathbb{1})$ vanish in positive degrees with no further condition on δ. If one wishes to show that $\mathrm{Tor}_*^{\mathrm{TL}_5(\delta)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_5(\delta)}(\mathbb{1},\mathbb{1})$ can be nonzero in positive degrees, then the only chance of this happening is in characteristic 2.

The theorem is in strict contrast to the n even case, where we show in Theorem A that for a general ring R and δ not invertible, $\mathrm{Tor}_{n-1}^{\mathrm{TL}_n(\delta)}(\mathbb{1},\mathbb{1}) = R/bR$ is nonzero for b some multiple of δ. Therefore, in the particular case where n is even, R is a field and $\delta = 0$, there can be no vanishing in all positive degrees.

9.1 Jones–Wenzl projectors

In this subsection we introduce the Jones–Wenzl projector and relate its existence to the projectivity of the trivial module $\mathbb{1}$. The original references are [Jones 1983; Wenzl 1987]; see also [Kauffman and Lins 1994; Lickorish 1992, Section 4].

Definition 9.1 Recall that $I_n \subseteq \mathrm{TL}_n$ is the two-sided ideal generated by the U_i for $i = 1, \dots, n-1$. Then, if it exists, the n^{th} *Jones–Wenzl projector* JW_n is the element of TL_n characterised by the properties

(i) $\mathrm{JW}_n \in 1 + I_n$, and

(ii) $I_n \cdot \mathrm{JW}_n = 0 = \mathrm{JW}_n \cdot I_n$.

Lemma 9.2 *If* JW_n *exists, it is unique.*

Proof Suppose a second element JW'_n in TL_n satisfies (i) and (ii) of Definition 9.1. Write $\mathrm{JW}_n = 1 + i$ and $\mathrm{JW}'_n = 1 + i'$ for $i, i' \in I_n$. Then $\mathrm{JW}_n \cdot i' = 0 = i \cdot \mathrm{JW}'_n$ by (ii). It follows that

$$\mathrm{JW}'_n = \mathrm{JW}'_n + i \cdot \mathrm{JW}'_n = (1 + i) \cdot \mathrm{JW}'_n = \mathrm{JW}_n \cdot \mathrm{JW}'_n,$$

and similarly that $\mathrm{JW}_n = \mathrm{JW}_n \cdot \mathrm{JW}'_n$. □

The Jones–Wenzl projector was first introduced by Jones [1983], was further studied by Wenzl [1987], and has since become important in representation theory, knot theory and the study of 3–manifolds. It is a key ingredient in the definition of the coloured Jones polynomial and SU(2) quantum invariants more generally, and is important in the study of tilting modules of (quantum) $\mathfrak{sl}_2$.

9.2 JW_n and projectivity of $\mathbb{1}$

We will now show that the Jones–Wenzl projector exists if and only if the trivial module $\mathbb{1}$ is projective. Thus, existence of JW_n implies the vanishing of $\mathrm{Tor}_*^{\mathrm{TL}_n(\delta)}(\mathbb{1}, \mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(\delta)}(\mathbb{1}, \mathbb{1})$ in positive degrees. Our own Theorem A implies that vanishing for δ invertible, while Theorem C proves nonvanishing for n even and δ not invertible. It turns out that there is a rich interplay between these two sources of (non)vanishing results.

Proposition 9.3 JW_n *exists if and only if* $\mathbb{1}$ *is a projective left* $\mathrm{TL}_n(\delta)$*–module, which is if and only if* $\mathbb{1}$ *is a projective right* $\mathrm{TL}_n(\delta)$*–module.*

Before proving the proposition we need the following.

Lemma 9.4 *In Definition 9.1, it is sufficient to replace* (ii) *with either*

(ii)′ $I_n \cdot \mathrm{JW}_n = 0$, *or*

(ii)″ $\mathrm{JW}_n \cdot I_n = 0$.

Proof Suppose $\mathrm{JW} \in \mathrm{TL}_n$ satisfies (i) and (ii)$'$. We have suggestively named this element, and will show it is in fact JW_n, by showing that JW also satisfies (ii)$''$ and hence (ii). Let $\mathrm{TL}_n \to \mathrm{TL}_n$, $d \mapsto \overline{d}$, be the antiautomorphism which reverses the order of letters in a monomial, ie $\overline{U_{i_1} \dots U_{i_n}} = U_{i_n} \dots U_{i_1}$. In diagrammatic terms, this map flips the diagram corresponding to the monomial in the left-to-right direction. Since JW satisfies (ii)$'$, it follows that $\overline{\mathrm{JW}}$ satisfies (ii)$''$. Then the argument of Lemma 9.2 can be repeated to show that $\mathrm{JW} = \overline{\mathrm{JW}}$, so that JW satisfies (ii)$'$ and (ii)$''$, hence it satisfies (ii) and $\mathrm{JW} = \mathrm{JW}_n$. □

Proof of Proposition 9.3 We prove the equivalence for left-modules.

If JW_n exists, then the maps $\mathbb{1} \to \mathrm{TL}_n$, $1 \mapsto \mathrm{JW}_n$, and $\mathrm{TL}_n \to \mathbb{1}$, $d \mapsto d \cdot 1$, are maps of left TL_n–modules composing to the identity. It follows that $\mathbb{1}$ is a direct summand of TL_n, and thus is projective.

Conversely, if $\mathbb{1}$ is a projective left TL_n–module, then the surjection $\mathrm{TL}_n \to \mathrm{TL}_n/I_n = \mathbb{1}$, regarded as a map of left TL_n–modules, has a splitting $s\colon \mathbb{1} \to \mathrm{TL}_n$, again a map of left TL_n–modules. By construction the element $s(1)$ then satisfies condition (i) of 9.1 and condition (ii)$'$ of 9.4, so that $\mathrm{JW}_n = s(1)$ exists, as required. □

9.3 Jones–Wenzl projectors and quantum binomial coefficients

Here we work in the Laurent polynomial ring $\mathbb{Z}[q, q^{-1}]$, and we set $\delta = q + q^{-1}$. For this section, let n and r be integers such that $n \geqslant r \geqslant 0$.

Definition 9.5 The *quantum integer* $[n]_q$ is defined to be

$$[n]_q = \frac{q^n - q^{-n}}{q - q^{-1}} = q^{n-1} + q^{n-3} + \dots + q^{-(n-3)} + q^{-(n-1)},$$

the *quantum factorial* $[n]_q!$ is defined by

$$[n]_q! = [n]_q[n-1]_q \cdots [1]_q,$$

and the *quantum binomial coefficient* $\begin{bmatrix} n \\ r \end{bmatrix}_q$ is then given by computing the normal binomial coefficient but replacing integers with quantum integers,

$$\begin{bmatrix} n \\ r \end{bmatrix}_q = \frac{[n]_q!}{[r]_q![n-r]_q!}.$$

The quantum binomial coefficients satisfy the recursion relations

$$\begin{bmatrix} n \\ r \end{bmatrix}_q = q^{n-r}\begin{bmatrix} n-1 \\ r-1 \end{bmatrix}_q + q^{-r}\begin{bmatrix} n-1 \\ r \end{bmatrix}_q \quad \text{and} \quad \begin{bmatrix} n \\ r \end{bmatrix}_q = q^{-(n-r)}\begin{bmatrix} n-1 \\ r-1 \end{bmatrix}_q + q^{r}\begin{bmatrix} n-1 \\ r \end{bmatrix}_q.$$

Either one of these relations gives an inductive proof that $\begin{bmatrix} n \\ r \end{bmatrix}_q$ lies in $\mathbb{Z}[q, q^{-1}]$.

Taken together, these relations give an inductive proof that $\begin{bmatrix} n \\ r \end{bmatrix}_q$ is invariant under inverting q, and consequently that it lies in $\mathbb{Z}[\delta]$. (Recall that $\delta = q + q^{-1}$.)

This means that we may evaluate $\begin{bmatrix} n \\ r \end{bmatrix}_q$ in any ring containing an element named δ, to obtain an element of that ring, which we continue to denote by $\begin{bmatrix} n \\ r \end{bmatrix}_q$.

The following result is proved by Webster [2017] using Schur–Weyl duality. For a purely diagrammatic approach see recent work of Spencer [2023].

Theorem 9.6 [Webster 2017, Theorem A.2; Spencer 2023, Section 10.3] *Let $R = \Bbbk$ be a field, let $q \in \Bbbk$ be nonzero, and set $\delta = q + q^{-1}$. The n^{th} Jones–Wenzl projector $\mathrm{JW}_n \in \mathrm{TL}_n(\delta)$ exists if and only if the quantum binomial coefficients $\left[{n \atop r}\right]_q$ are nonzero in $\Bbbk$ for all $0 \le r \le n$.*

Remark 9.7 Suppose that $R = \Bbbk$ is a field. Whenever $\Bbbk$, q and n satisfy the conditions of Theorem 9.6, we obtain the vanishing of $\mathrm{Tor}_*^{\mathrm{TL}_n(\delta)}(\mathbb{1}, \mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(\delta)}(\mathbb{1}, \mathbb{1})$ in positive degrees. (We will refer to this as simply "vanishing" for the present remark.)

- In the case of n even, Theorems A and C show that vanishing holds if and only if $\delta \neq 0$, and this is in fact stronger than the result obtained from Theorem 9.6. For example, if we take $n = 4$ then the $\left[{n \atop k}\right]_q$ take values 1, $\delta(\delta^2 - 2)$ and $(\delta^2 - 1)(\delta^2 - 2)$, so that Theorem 9.6 requires δ to avoid the values $0, \pm 1, \pm\sqrt{2}$. For n even, δ is always a factor of $\left[{n \atop 1}\right]_q$, so that Theorem A will always apply more generally than Theorem 9.6 in this case.
- In the case of n odd, the situation is more interesting. Theorem A demonstrates vanishing when $\delta \neq 0$. But if we take $n = 3$, for example, then the $\left[{n \atop k}\right]_q$ take values 1 and $\delta^2 - 1$, so that Theorem 9.6 demonstrates vanishing so long as $\delta \neq \pm 1$. Neither of these vanishing results implies the other, but taken together they demonstrate vanishing for *all* values of δ.

9.4 Identifying the quantum binomial coefficients

In this section, we identify the quantum binomial coefficients upon specialising $\delta = q + q^{-1} = 0$. The results are assembled in the following proposition.

Proposition 9.8 *When $\delta = q + q^{-1} = 0$, the quantum binomial coefficients have the following form:*

- *When n is even and r is odd,*
$$\begin{bmatrix} n \\ r \end{bmatrix}_q = 0.$$
- *When n and r are both even, let $n = 2a$ and $r = 2t$. Then*
$$\begin{bmatrix} n \\ r \end{bmatrix}_q = \binom{a}{t}.$$
- *When n is odd and r is even, let $n = 2a + 1$ and $r = 2t$. Then*
$$\begin{bmatrix} n \\ r \end{bmatrix}_q = (-1)^t \binom{a}{t}.$$
- *When n and r are both odd, let $n = 2a + 1$ and $r = 2t + 1$. Then*
$$\begin{bmatrix} n \\ r \end{bmatrix}_q = (-1)^{a-t} \binom{a}{t}.$$

Remark 9.9 Proposition 9.8 shows that the "quantum Pascal's triangle" with $\delta = 0$ looks like a Pascal's triangle in the even rows, with every coefficient separated by a zero, and a "doubled" Pascal's triangle with signs on the odd rows. This is shown in Figure 9.

$$
\begin{array}{l|ccccccccccccccccccc}
n=0 & & & & & & & & & & 1 & & & & & & & & & \\
n=1 & & & & & & & & & 1 & & 1 & & & & & & & & \\
n=2 & & & & & & & & 1 & & 0 & & 1 & & & & & & & \\
n=3 & & & & & & & 1 & & -1 & & -1 & & 1 & & & & & & \\
n=4 & & & & & & 1 & & 0 & & 2 & & 0 & & 1 & & & & & \\
n=5 & & & & & 1 & & 1 & & -2 & & -2 & & 1 & & 1 & & & & \\
n=6 & & & & 1 & & 0 & & 3 & & 0 & & 3 & & 0 & & 1 & & & \\
n=7 & & & 1 & & -1 & & -3 & & 3 & & 3 & & -3 & & -1 & & 1 & & \\
n=8 & & 1 & & 0 & & 4 & & 0 & & 6 & & 0 & & 4 & & 0 & & 1 & \\
n=9 & 1 & & 1 & & -4 & & -4 & & 6 & & 6 & & -4 & & -4 & & 1 & & 1
\end{array}
$$

Figure 9: The quantum binomial coefficients with $\delta = q + q^{-1} = 0$.

The proof of the four points in this proposition are given by applying a result of Désarménien [1983], which we recall below. This result is given not in terms of *quantum* binomials, but in terms of *Gaussian* binomials, so we recall these first.

Let p be an indeterminate. The *Gaussian binomial coefficients* are the quantities

$$\begin{bmatrix} n \\ r \end{bmatrix}_p^G = \frac{[n]_p^G!}{[r]_p^G![n-r]_p^G!}$$

defined in terms of the *Gaussian integers*

$$[n]_p^G = 1 + p + \cdots + p^{n-1}$$

and *Gaussian factorials*

$$[n]_p^G! = [n]_p^G[n-1]_p^G \cdots [1]_p^G.$$

The relation between the Gaussian and quantum binomial coefficients is

$$\begin{bmatrix} n \\ r \end{bmatrix}_q = q^{r^2-nr}\begin{bmatrix} n \\ r \end{bmatrix}_{q^2}^G.$$

Proposition 9.10 [Désarménien 1983, proposition 2.2] *Fix a $k \geq 0 \in \mathbb{N}$ and let Φ_k be the k^{th} cyclotomic polynomial. Let $n = ka + b$ and $r = kt + s$ with $0 \leq b, s \leq k-1$. Then the Gaussian binomial coefficient satisfies the congruence*

$$\begin{bmatrix} n \\ r \end{bmatrix}_p^G \equiv \binom{a}{t}\begin{bmatrix} b \\ s \end{bmatrix}_p^G \mod \Phi_k.$$

Proof of Proposition 9.8 Note that when $\delta = q + q^{-1} = 0$, rearranging this equation gives that $q^{\pm 2} = -1$. Recall that the parameter p in the Gaussian binomial coefficient is q^2, and so $p^2 = q^4 = 1$. We invoke Proposition 9.10 with $k = 2$. Then the cyclotomic polynomial $\Phi_2(p) = 1 + p = 1 + q^2 = 0$. Let $n = 2a + b$ and $r = 2t + s$ with $0 \leq b, s \leq 1$. Then the quantum binomial coefficient satisfies

$$\begin{bmatrix} n \\ r \end{bmatrix}_q = q^{r^2-nr}\begin{bmatrix} n \\ r \end{bmatrix}_{q^2}^G = q^{r^2-nr}\binom{a}{t}\begin{bmatrix} b \\ s \end{bmatrix}_{q^2}^G.$$

When n is even and r is odd,

$$\begin{bmatrix} b \\ s \end{bmatrix}_{q^2}^G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{q^2}^G = 0,$$

which gives the first case of the proposition. For all other cases, $\left[\begin{smallmatrix} b \\ s \end{smallmatrix}\right]_{q^2}^G = 1$ and so

$$\begin{bmatrix} n \\ r \end{bmatrix}_q = q^{r^2-nr}\binom{a}{t}.$$

Computing the coefficient q^{r^2-nr}, using $q^{\pm 2} = -1$, yields the result for the remaining three cases. □

9.5 Proof of Theorem D

Proof This proof puts together three previous results. By Proposition 9.3 we know that if JW_n exists then $\mathbb{1}$ is projective and it follows that $\mathrm{Tor}_i^{\mathrm{TL}(0)}(\mathbb{1},\mathbb{1})$ and $\mathrm{Ext}^*_{\mathrm{TL}_n(0)}(\mathbb{1},\mathbb{1})$ vanish for all $i > 0$. So it is enough to show that JW_n exists under the hypotheses of the theorem. Theorem 9.6 tells us that JW_n exists precisely when the quantum binomial coefficients $\left[\begin{smallmatrix} n \\ r \end{smallmatrix}\right]_q$ are nonzero for all $1 \le r \le n$. Finally, Proposition 9.8 explicitly describes these coefficients when $\delta = 0$. We see that for n even, there is always a quantum binomial coefficient $\left[\begin{smallmatrix} n \\ r \end{smallmatrix}\right]_q = 0$ and so we learn nothing new. However when $n = 2k+1$ is odd, the quantum binomial coefficients take values in the set

$$\left\{ \pm\binom{k}{t} \;\middle|\; 0 \le t \le k \right\}$$

and, up to sign, all values in this set are realised as some $\left[\begin{smallmatrix} n \\ r \end{smallmatrix}\right]_q$. The hypotheses of the theorem precisely say that these numbers are nonzero in R and so the result follows. □

References

[Björner and Wachs 1983] **A Björner**, **M Wachs**, *On lexicographically shellable posets*, Trans. Amer. Math. Soc. 277 (1983) 323–341 MR Zbl

[Bowman et al. 2023] **C Bowman**, **A Cox**, **A Hazi**, *Path isomorphisms between quiver Hecke and diagrammatic Bott–Samelson endomorphism algebras*, Adv. Math. 429 (2023) art. id. 109185 MR Zbl

[Boyd 2020] **R Boyd**, *Homological stability for Artin monoids*, Proc. Lond. Math. Soc. 121 (2020) 537–583 MR Zbl

[Boyd and Hepworth 2021] **R Boyd**, **R Hepworth**, *Combinatorics of injective words for Temperley–Lieb algebras*, J. Combin. Theory Ser. A 181 (2021) art. id. 105446 MR Zbl

[Boyd et al. 2021] **R Boyd**, **R Hepworth**, **P Patzt**, *The homology of the Brauer algebras*, Selecta Math. 27 (2021) art. id. 85 MR Zbl

[Brundan and Kleshchev 2009] **J Brundan**, **A Kleshchev**, *Blocks of cyclotomic Hecke algebras and Khovanov–Lauda algebras*, Invent. Math. 178 (2009) 451–484 MR Zbl

[Charney 1980] **R M Charney**, *Homology stability for* GL_n *of a Dedekind domain*, Invent. Math. 56 (1980) 1–17 MR Zbl

[Church 2012] **T Church**, *Homological stability for configuration spaces of manifolds*, Invent. Math. 188 (2012) 465–504 MR Zbl

[Désarménien 1983] **J Désarménien**, *Une autre interprétation du nombre de dérangements*, Sém. Lothar. Combin. 8 (1983) 11–16 Zbl

[Deutsch and Shapiro 2001] **E Deutsch**, **L Shapiro**, *A survey of the Fine numbers*, Discrete Math. 241 (2001) 241–265 MR Zbl

[Fan and Green 1997] **C K Fan**, **R M Green**, *Monomials and Temperley–Lieb algebras*, J. Algebra 190 (1997) 498–517 MR Zbl

[Farmer 1979] **F D Farmer**, *Cellular homology for posets*, Math. Japon. 23 (1979) 607–613 MR Zbl

[Galatius and Randal-Williams 2018] **S Galatius**, **O Randal-Williams**, *Homological stability for moduli spaces of high dimensional manifolds, I*, J. Amer. Math. Soc. 31 (2018) 215–264 MR Zbl

[Galatius et al. 2018] **S Galatius**, **A Kupers**, **O Randal-Williams**, *Cellular E_k–algebras*, preprint (2018) arXiv 1805.07184

[Graham and Lehrer 1996] **J J Graham**, **G I Lehrer**, *Cellular algebras*, Invent. Math. 123 (1996) 1–34 MR Zbl

[Halverson et al. 2009] **T Halverson**, **M Mazzocco**, **A Ram**, *Commuting families in Hecke and Temperley–Lieb algebras*, Nagoya Math. J. 195 (2009) 125–152 MR Zbl

[Harer 1985] **J L Harer**, *Stability of the homology of the mapping class groups of orientable surfaces*, Ann. of Math. 121 (1985) 215–249 MR Zbl

[Hatcher and Vogtmann 1998] **A Hatcher**, **K Vogtmann**, *Rational homology of* $\mathrm{Aut}(F_n)$, Math. Res. Lett. 5 (1998) 759–780 MR Zbl

[Hatcher and Vogtmann 2004] **A Hatcher**, **K Vogtmann**, *Homology stability for outer automorphism groups of free groups*, Algebr. Geom. Topol. 4 (2004) 1253–1272 MR Zbl

[Hatcher and Wahl 2010] **A Hatcher**, **N Wahl**, *Stabilization for mapping class groups of 3–manifolds*, Duke Math. J. 155 (2010) 205–269 MR Zbl

[Hepworth 2016] **R Hepworth**, *Homological stability for families of Coxeter groups*, Algebr. Geom. Topol. 16 (2016) 2779–2811 MR Zbl

[Hepworth 2022] **R Hepworth**, *Homological stability for Iwahori–Hecke algebras*, J. Topol. 15 (2022) 2174–2215 MR Zbl

[Jones 1983] **V F R Jones**, *Index for subfactors*, Invent. Math. 72 (1983) 1–25 MR Zbl

[Jones 1985] **V F R Jones**, *A polynomial invariant for knots via von Neumann algebras*, Bull. Amer. Math. Soc. 12 (1985) 103–111 MR Zbl

[van der Kallen 1980] **W van der Kallen**, *Homology stability for linear groups*, Invent. Math. 60 (1980) 269–295 MR Zbl

[Kassel and Turaev 2008] **C Kassel**, **V Turaev**, *Braid groups*, Graduate Texts in Math. 247, Springer (2008) MR Zbl

[Kauffman 1987] **L H Kauffman**, *State models and the Jones polynomial*, Topology 26 (1987) 395–407 MR Zbl

[Kauffman 1990] **L H Kauffman**, *An invariant of regular isotopy*, Trans. Amer. Math. Soc. 318 (1990) 417–471 MR Zbl

[Kauffman 2005] **L H Kauffman**, *Knot diagrammatics*, from "Handbook of knot theory" (W Menasco, M Thistlethwaite, editors), Elsevier, Amsterdam (2005) 233–318 MR Zbl

[Kauffman and Lins 1994] **L H Kauffman**, **S L Lins**, *Temperley–Lieb recoupling theory and invariants of 3–manifolds*, Ann. of Math. Stud. 134, Princeton Univ. Press (1994) MR Zbl

[Kerz 2005] **M C Kerz**, *The complex of words and Nakaoka stability*, Homology Homotopy Appl. 7 (2005) 77–85 MR Zbl

[Lickorish 1992] **W B R Lickorish**, *Calculations with the Temperley–Lieb algebra*, Comment. Math. Helv. 67 (1992) 571–591 MR Zbl

[Maazen 1979] **H Maazen**, *Homology stability for the general linear group*, PhD thesis, Utrecht University (1979) `https://dspace.library.uu.nl/handle/1874/237657`

[Nakaoka 1960] **M Nakaoka**, *Decomposition theorem for homology groups of symmetric groups*, Ann. of Math. 71 (1960) 16–42 MR Zbl

[Quillen 1972] **D Quillen**, *On the cohomology and K–theory of the general linear groups over a finite field*, Ann. of Math. 96 (1972) 552–586 MR Zbl

[Randal-Williams 2013] **O Randal-Williams**, *Homological stability for unordered configuration spaces*, Q. J. Math. 64 (2013) 303–326 MR Zbl

[Randal-Williams 2016] **O Randal-Williams**, *Resolutions of moduli spaces and homological stability*, J. Eur. Math. Soc. 18 (2016) 1–81 MR Zbl

[Randal-Williams 2021] **O Randal-Williams**, *A remark on the homology of Temperley–Lieb algebras*, preprint (2021) `https://www.dpmms.cam.ac.uk/~or257/notes/Descent.pdf`

[Randal-Williams and Wahl 2017] **O Randal-Williams**, **N Wahl**, *Homological stability for automorphism groups*, Adv. Math. 318 (2017) 534–626 MR Zbl

[Ridout and Saint-Aubin 2014] **D Ridout**, **Y Saint-Aubin**, *Standard modules, induction and the structure of the Temperley–Lieb algebra*, Adv. Theor. Math. Phys. 18 (2014) 957–1041 MR Zbl

[Sloane 2000] **N J A Sloane**, *Jacobsthal sequence (or Jacobsthal numbers): $a(n) = a(n-1) + 2 * a(n-2)$, with $a(0) = 0$, $a(1) = 1$; also $a(n) =$ nearest integer to $2^n/3$*, from "The on-line encyclopedia of integer sequences" (2000) A001045

[Spencer 2022] **R Spencer**, *Jacobsthal elements*, web resource (2022) `https://robertandrewspencer.com/research/jacobsthal`

[Spencer 2023] **R A Spencer**, *The modular Temperley–Lieb algebra*, Rocky Mountain J. Math. 53 (2023) 177–208 MR Zbl

[Sroka 2022] **R J Sroka**, *The homology of a Temperley–Lieb algebra on an odd number of strands* (2022) arXiv 2202.08799 To appear in Algebr. Geom. Topol.

[Szymik and Wahl 2019] **M Szymik**, **N Wahl**, *The homology of the Higman–Thompson groups*, Invent. Math. 216 (2019) 445–518 MR Zbl

[Temperley and Lieb 1971] **H N V Temperley**, **E H Lieb**, *Relations between the "percolation" and "colouring" problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the "percolation" problem*, Proc. Roy. Soc. Lond. Ser. A 322 (1971) 251–280 MR Zbl

[Webster 2017] **B Webster**, *The existence of Jones–Wenzl projectors* (2017) MR Zbl Appendix to B Elias and N Libedinsky, "Indecomposable Soergel bimodules for universal Coxeter groups", Trans. Amer. Math. Soc. 369 (2017) 3883–3910

[Weibel 1994] **C A Weibel**, *An introduction to homological algebra*, Cambridge Stud. Adv. Math. 38, Cambridge Univ. Press (1994) MR Zbl

[Wenzl 1987] **H Wenzl**, *On sequences of projections*, C. R. Math. Rep. Acad. Sci. Canada 9 (1987) 5–9 MR Zbl

[Westbury 1995] **B W Westbury**, *The representation theory of the Temperley–Lieb algebras*, Math. Z. 219 (1995) 539–565 MR Zbl

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge
Cambridge, United Kingdom

Current address: *School of Mathematics and Statistics, University of Glasgow*
Glasgow, United Kingdom

Institute of Mathematics, University of Aberdeen
Aberdeen, United Kingdom

rachael.boyd@glasgow.ac.uk, r.hepworth@abdn.ac.uk

https://www.maths.gla.ac.uk/~rboyd/, http://homepages.abdn.ac.uk/r.hepworth/pages/

Proposed: Nathalie Wahl — Received: 22 February 2022
Seconded: Ulrike Tillmann, Ciprian Manolescu — Revised: 8 July 2022

Guidelines for Authors

Submitting a paper to Geometry & Topology

Papers must be submitted using the upload page at the GT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Geometry & Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in GT is available on the GT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

GEOMETRY & TOPOLOGY

Volume 28 Issue 3 (pages 1005–1499) 2024