



Geometry & Topology

Volume 29 (2025)

Issue 5 (pages 2251–2782)

GEOMETRY & TOPOLOGY

msp.org/gt

MANAGING EDITORS

Robert Lipshitz University of Oregon
lipshitz@uoregon.edu

András I Stipsicz Alfréd Rényi Institute of Mathematics
stipsicz@renyi.hu

BOARD OF EDITORS

Mohammed Abouzaid	Stanford University abouzaid@stanford.edu	Mark Gross	University of Cambridge mgross@dpmms.cam.ac.uk
Dan Abramovich	Brown University dan_abramovich@brown.edu	Rob Kirby	University of California, Berkeley kirby@math.berkeley.edu
Ian Agol	University of California, Berkeley ianagol@math.berkeley.edu	Bruce Kleiner	NYU, Courant Institute bkleiner@cims.nyu.edu
Arend Bayer	University of Edinburgh arend.bayer@ed.ac.uk	Sándor Kovács	University of Washington skovacs@uw.edu
Mark Behrens	University of Notre Dame mbehren1@nd.edu	Urs Lang	ETH Zürich urs.lang@math.ethz.ch
Mladen Bestvina	University of Utah bestvina@math.utah.edu	Marc Levine	Universität Duisburg-Essen marc.levine@uni-due.de
Martin R Bridson	University of Oxford bridson@maths.ox.ac.uk	Ciprian Manolescu	University of California, Los Angeles cm@math.ucla.edu
Jim Bryan	University of British Columbia jbryan@math.ubc.ca	Haynes Miller	Massachusetts Institute of Technology hrm@math.mit.edu
Dmitri Burago	Pennsylvania State University burago@math.psu.edu	Tomasz Mrowka	Massachusetts Institute of Technology mrowka@math.mit.edu
Tobias H Colding	Massachusetts Institute of Technology colding@math.mit.edu	Aaron Naber	Northwestern University anaber@math.northwestern.edu
Simon Donaldson	Imperial College, London s.donaldson@ic.ac.uk	Peter Ozsváth	Princeton University petero@math.princeton.edu
Yasha Eliashberg	Stanford University eliash-gt@math.stanford.edu	Leonid Polterovich	Tel Aviv University polterov@post.tau.ac.il
Benson Farb	University of Chicago farb@math.uchicago.edu	Colin Rourke	University of Warwick gt@maths.warwick.ac.uk
David M Fisher	Rice University davidfisher@rice.edu	Roman Sauer	Karlsruhe Institute of Technology roman.sauer@kit.edu
Mike Freedman	Microsoft Research michaelf@microsoft.com	Stefan Schwede	Universität Bonn schwede@math.uni-bonn.de
David Gabai	Princeton University gabai@princeton.edu	Natasa Sesum	Rutgers University natasas@math.rutgers.edu
Stavros Garoufalidis	Southern U. of Sci. and Tech., China stavros@mpim-bonn.mpg.de	Gang Tian	Massachusetts Institute of Technology tian@math.mit.edu
Cameron Gordon	University of Texas gordon@math.utexas.edu	Ulrike Tillmann	Oxford University tillmann@maths.ox.ac.uk
Jesper Grodal	University of Copenhagen jg@math.ku.dk	Nathalie Wahl	University of Copenhagen wahl@math.ku.dk
Misha Gromov	IHÉS and NYU, Courant Institute gromov@ihes.fr	Anna Wienhard	Universität Heidelberg wienhard@mathi.uni-heidelberg.de

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

The subscription price for 2025 is US \$865/year for the electronic version, and \$1210/year (+\$75, 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/>

© 2025 Mathematical Sciences Publishers

Quantitative Thomas–Yau uniqueness

YANG LI

Under Floer-theoretic conditions, we obtain quantitative estimates on the closeness (Hausdorff distance, flat norm and F-metric) between two Lagrangians, depending on the smallness of Lagrangian angles. Some applications include a strong–weak uniqueness theorem for special Lagrangians, and a characterization of varifold convergence to special Lagrangians in terms of Lagrangian angles.

53C38, 53D12, 57R58

1 Background and introduction

1.1 Special Lagrangians

Inside an n -dimensional compact Kähler manifold (X, ω) with a nowhere-vanishing holomorphic volume form Ω such that $\omega^n/n! = e^{2\rho}(i^{n^2}/2^n)\Omega \wedge \bar{\Omega}$, an n -dimensional compact oriented submanifold L is called *special Lagrangian* if

$$(1) \quad \omega|_L = 0, \quad \text{Im } \Omega|_L = 0.$$

These sit at the crossroad of minimal surface theory and symplectic geometry:

- If the metric is Calabi–Yau (namely $\rho = 0$), then special Lagrangians are absolute volume minimizers inside their homology classes.
- Lagrangian submanifolds equipped with (unobstructed) brane structures define objects inside the Fukaya category.

Recall that the *Lagrangian angle* $\theta: L \rightarrow \mathbb{R}$ is defined by $\Omega|_L = e^{-\rho}e^{i\theta} \text{dvol}_L$, so special Lagrangians amount to the condition $\theta = 0$. More generally, a Lagrangian is called *quantitative almost calibrated* if $|\theta| \leq \frac{1}{2}\pi - \epsilon_0$ for some fixed small $\epsilon_0 > 0$. Such Lagrangians are important in the Thomas–Yau program; see Li [9] and Thomas and Yau [12]. One of its immediate consequences is the a priori homological mass bound

$$\text{Vol}(L) = \int_L \text{dvol} \leq \frac{1}{\sin \epsilon_0} \int_L \text{Re } \Omega.$$

1.2 Thomas–Yau uniqueness

As part of a wider program to relate the existence and uniqueness questions of special Lagrangian branes to the Fukaya category and stability condition, Thomas and Yau [12] proved a remarkable uniqueness theorem, which from the modern perspective reads:

Theorem 1.1 (Imagi [6], Imagi, Joyce and Oliveira dos Santos [7] and Thomas and Yau [12]) *Let L and L' be two compact embedded special Lagrangians with unobstructed brane structures, defining isomorphic objects in the derived Fukaya category. Then their supports coincide.*

The Floer degree of a transverse intersection point $p \in \text{CF}^*(L_1, L_2)$ between two Lagrangians L_1 and L_2 is

$$(2) \quad \mu_{L_1, L_2}(p) = \frac{1}{\pi} \left(\sum_1^n \alpha_i - \theta_{L_2}(p) + \theta_{L_1}(p) \right) \in \mathbb{Z},$$

where inside $T_p X \simeq \mathbb{C}^n$ we can put the tangent spaces into the standard form

$$T_p L_1 = \mathbb{R}^n, \quad T_p L_2 = (e^{i\alpha_1}, \dots, e^{i\alpha_n})\mathbb{R}^n \quad \text{for } 0 < \alpha_i < \pi.$$

The proof idea of Theorem 1.1 can be outlined as follows. Assume $L \neq L'$.

- By making C^∞ -small Hamiltonian perturbations of L and L' , we can replace L and L' by isomorphic objects L_1 and L_2 *intersecting transversely*. Moreover, either by a Morse-theoretic argument [12], or using real analyticity via a Łojasiewicz inequality [7], one can remove the degree-zero intersection points, to ensure the Floer cochain group $\text{CF}^0(L_1, L_2)$ is 0.
- Consequently, the Floer cohomology $\text{HF}^0(L, L) \simeq \text{HF}^0(L_1, L_2)$ is 0. In particular the unit of the Floer cohomology ring vanishes, which is impossible, because it violates Poincaré duality (alternatively, because the image of the unit under the open–closed map is the homology class $[L] \in H_n(X, \Lambda_{\text{nov}})$, which cannot be zero).

We take note of a few conceptual features:

- The proof of Thomas–Yau uniqueness is analogous to a *strong maximum principle* argument. The role of symplectic topology is to force the existence of a degree-zero intersection point, which is analogous to the existence of a maximum, and a local calculation concerning the intersection point results in a contradiction.
- The metric is not necessarily Calabi–Yau. The compactness of X can be replaced by any other standard settings where Floer theory makes sense.
- The known proofs rely essentially on smoothness assumptions of the Lagrangians.

Remark 1.2 (for the analysts) Setting up Floer cohomology (or more generally, the Fukaya category) requires a large baggage train, such as brane structure (spin structure, local system and bounding cochains), complicated perturbation schemes and algebraic machinery. For background see Joyce [8, Section 2.5] for a lightning-quick overview, and Auroux [5] for a gentle introduction on the Fukaya category.

The Thomas–Yau argument (as well as the rest of this paper) is *not sensitive to the details of the Fukaya category*. Knowledge on the setup is not essential; the only facts from Floer theory that one really needs for this paper are:

- Floer cohomology groups are invariant under Hamiltonian deformations of the Lagrangians.

- Nonzero Floer cohomology implies the existence of Lagrangian intersection points in the corresponding degree, provided the Lagrangian intersections are all transverse.
- If two Lagrangian branes are in the same derived Fukaya category class, then they live in the same homology class. (This follows from the property of the open–closed map.)

1.3 Quantitative Thomas–Yau uniqueness

A common theme in geometric analysis is *rigidity theorems*, eg a strong maximum principle naturally suggests a Harnack inequality. Analogously, we ask:

Question 1 Suppose L and L' have *small* Lagrangian angles (eg $\|\theta\|_{C^0} \ll 1$ or $\|\theta\|_{L^1} \ll 1$). Then do they have to be *uniformly* close to each other (eg in the Hausdorff distance for subsets, in the flat norm distance for integral currents or the F -metric for varifolds)?

Remark 1.3 A paper of Abouzaid and Imagi [1] studies symplectic topological consequences of special Lagrangians lying inside a small C^0 -neighbourhood of a given special Lagrangian, under additional hypotheses on the fundamental group and its representations.

Remark 1.4 New ideas are needed for this question: In a naive strategy, one takes a sequence of L_i and L'_i , with Lagrangian angles converging to zero, and attempts to use compactness theorems in geometric measure theory to extract subsequential limits L_∞ and L'_∞ , which should be special Lagrangian integral currents. To deduce that L_i is close to L'_i , one would like to prove $L_\infty = L'_\infty$. This would require a *singular Lagrangian* version of the Thomas–Yau uniqueness theorem, which is yet unknown. (The interested reader may see Li [9, section 5.7] for some heuristic ideas.)

Our main technical result, Theorem 2.1, gives one answer to the question above. Morally, it provides Floer-theoretic conditions such that if a quantitative almost-calibrated Lagrangian has small enough L^1 -norm on its Lagrangian angle, then it is close to a given special Lagrangian, with respect to several distance notions from geometric measure theory (Hausdorff distance, flat norm and F -metric), with quantitative estimates depending on explicit powers of $\|\theta\|_{L^1}$.

We now discuss qualitative consequences of Theorem 2.1, to be proved in Section 3. Theorem 2.1 can be interpreted as the quantitative version of a *strong–weak uniqueness* theorem.

Definition 1.5 Let L_i be a sequence of smooth embedded unobstructed Lagrangian branes which lie in a fixed nonzero derived Fukaya category class (so in particular the same rational homology class), and are all quantitatively almost calibrated. We say a Lagrangian C^∞ -submanifold (resp. Lagrangian integral current) L_∞ is a C^∞ -limit (resp. varifold limit) if the L_i converge to L_∞ in the C^∞ -topology (resp. in both the weak topology of varifolds and the flat norm topology of integral currents). Notice there is no requirement that L_∞ is equipped with any brane structure.

Theorem 1.6 (strong–weak uniqueness) *Fix a derived Fukaya category class, and let L_∞ be a C^∞ -limit and L'_∞ a varifold limit. Assume both are special Lagrangian $\theta_{L_\infty} = \theta_{L'_\infty} = 0$. Then $L_\infty = L'_\infty$ as integral currents.*

Provided a derived Fukaya category class admits a special Lagrangian C^∞ -limit, we have a satisfactory characterization of varifold topology convergence in terms of the Lagrangian angle function.

Theorem 1.7 *Fix a derived Fukaya category class, and assume there is a sequence of smooth embedded representatives L_i , converging in C^∞ topology to a special Lagrangian L_∞ . Let L'_i be a sequence of quantitatively almost-calibrated Lagrangian branes in the same derived Fukaya category class. Then L'_i converges in the varifold topology to L_∞ if and only if $\|\theta_{L'_i}\|_{L^1} \rightarrow 0$.*

Remark 1.8 Since in our setting $\|\theta\|_{L^\infty}$ and the mass of the Lagrangian both have uniform bounds, by interpolation $\|\theta\|_{L^1} \rightarrow 0$ is equivalent to $\|\theta\|_{L^p} \rightarrow 0$ for any fixed $p > 0$.

2 Proof of quantitative Thomas–Yau uniqueness

2.1 Statement of quantitative Thomas–Yau uniqueness

Our main technical result is a quantitative version of the Thomas–Yau uniqueness theorem.

Theorem 2.1 *Let L and L' be two compact smoothly embedded Lagrangians with unobstructed brane structures, in the same fixed homology classes in $H_n(X, \mathbb{Q})$, such that $\mathrm{HF}^0(L, L') \neq 0$ or $\mathrm{HF}^0(L', L) \neq 0$ holds. Assume the quantitative almost-calibrated condition $\|\theta_{L'}\|_{C^0} \leq \frac{1}{2}\pi - \epsilon_0$ for some fixed $\epsilon_0 > 0$, and moreover that there is some ϵ small enough that the Lagrangian angles satisfy*

- $\|\theta_L\|_{C^0} \leq \epsilon \ll 1$,
- $\mathrm{Vol}(\{\theta_{L'} > \epsilon\}) \ll \epsilon^n$.

Then the Hausdorff distance between L and L' is uniformly small:

$$(3) \quad \begin{cases} \sup_{p \in L} \mathrm{dist}(p, L') \leq C\epsilon^{1/(2n)}, \\ \sup_{p \in L'} \mathrm{dist}(p, L) \leq C\epsilon^{1/(4n^2)}. \end{cases}$$

The flat norm distance between L and L' is bounded uniformly by $C\epsilon^{1/(4n)+1/(4n^2)}$. The F -metric between L and L' is bounded by $C\epsilon^{1/(8n)}$. The constants depend on ϵ_0 , X and L , but not on the regularity bounds on L' , nor on the small ϵ .

A few explanations may clarify the significance of the assumptions:

- If L and L' are isomorphic nonzero objects in the derived Fukaya category, then $\mathrm{HF}^0(L, L') \neq 0$, $\mathrm{HF}^0(L', L) \neq 0$ and $[L] = [L'] \in H_n(X, \mathbb{Q})$.

- The weak L^1 -type bound $\text{Vol}(\{\theta_{L'} > \epsilon\}) \ll \epsilon^n$ is implied by $\|\theta_{L'}\|_{L^1(L')} \ll \epsilon^{n+1}$. This is much weaker than assuming C^0 -smallness of $\theta_{L'}$. This weakening is needed for Theorem 1.7, because if a sequence L'_i of quantitatively almost-calibrated Lagrangians converge to a special Lagrangian in the weak topology of varifolds (ie as Radon measures on the Grassmannian bundle), then $\|\theta_{L'_i}\|_{L^1} \rightarrow 0$ as $i \rightarrow \infty$, but $\|\theta_{L'_i}\|_{C^0}$ does not necessarily converge to zero.
- As we will indicate in a series of remarks, the embedded assumption on the Lagrangians in Theorem 2.1 can be relaxed to immersed Lagrangians with *connected domains* and transverse self-intersection points, with rather minor changes. The connected domain assumption is quite crucial to our arguments here (see Remark 2.3), and also appears in the literature on Thomas–Yau uniqueness for immersed Lagrangians [6]. On the other hand, the hypothesis of Theorem 2.1 is a little weaker than assuming L and L' lie in the same derived Fukaya category class, and does not make explicit use of holomorphic curves.

2.2 The small $\|\theta\|_{C^0}$ case

In the special case where $\|\theta\|_{C^0} \ll 1$ is small, we show that the set of points on L lying near L' is sufficiently dense. This argument contains the Floer-theoretic ingredient in our strategy. We do not yet use $[L] = [L'] \in H_n(X)$.

Lemma 2.2 *Assume $\|\theta\|_{C^0} \leq \epsilon \ll 1$ for both L and L' . For all $p \in L$, there is some $p' \in L \cap B(C_1 \epsilon^{1/2n})$ such that $\text{dist}(p', L') \leq C \epsilon^{(n+1)/(2n)}$. In particular $\text{dist}(p, L') \leq C \epsilon^{1/2n}$.*

Proof We perform a Hamiltonian perturbation of the L within the Weinstein tubular neighbourhood $T^*L \simeq NL$ of the Lagrangian L . For a smooth Hamiltonian function $h: X \rightarrow \mathbb{R}$, we get a Hamiltonian vector field X_h by $\iota_{X_h} \omega = -dh$ (equivalently $X_h = J \nabla h$), and for $\epsilon \|dh\|_{C^1} \ll 1$, the time ϵ flow sends L to $\varphi_\epsilon(L)$, whose Lagrangian angle is denoted by $\theta_{\varphi_\epsilon(L)}$. In our later applications, we can further assume that along L , the vector field X_h is orthogonal to the tangent spaces of L , or equivalently $X_h = J \nabla^L h$ where $\nabla^L h$ is the gradient of h on L . Alternatively, to the same effect, we can also choose the perturbed Lagrangian as the graph of ϵdh inside the Weinstein neighbourhood T^*L .

Notice the Lagrangian angle θ is a locally defined smooth function on the Grassmannian bundle of Lagrangian subspaces of the tangent space, and the flow φ_ϵ acting on the Grassmannian bundle is close to the identity, so by Taylor expansion

$$\theta_{\varphi_\epsilon(L)} = \theta_L + \epsilon \mathcal{L}_{X_h} \theta + O(\epsilon^2 \|X_h\|_{C^1}^2) = \theta_L + \epsilon \mathcal{L}_{X_h} \theta + O(\epsilon^2 \|dh\|_{C^1}^2).$$

The linearized operator $\mathcal{L}(h) = \mathcal{L}_{X_h} \theta$ can be computed in the case of a general weighting function ρ as follows, similar to the computation in Thomas–Yau [12, Lemma 2.3] in the Calabi–Yau metric case. We observe that by the closedness of the holomorphic volume form Ω ,

$$\mathcal{L}_{X_h} \Omega = -i \mathcal{L}_{J X_h} \Omega = i (d \iota_{-J X_h} \Omega + \iota_{-J X_h} d \Omega) = i d \iota_{-J X_h} \Omega.$$

By the definition of the Lagrangian angle $\Omega|_L = e^{-\rho} e^{i\theta} \text{dvol}_L$, and the assumption that $-JX_h = \nabla^L h$ is tangent to L , the left side converts to

$$e^{-\rho} e^{i\theta} (-\mathcal{L}_{X_h} \rho + i \mathcal{L}_{X_h} \theta) \text{dvol}_L + e^{-\rho} e^{i\theta} \mathcal{L}_{X_h} (\text{dvol}_L),$$

while the right side converts to

$$i d(e^{-\rho} e^{i\theta} \iota_{\nabla^L h} \text{dvol}_L) = i d(e^{-\rho} e^{i\theta} *_L dh) = i e^{i\theta} d(e^{-\rho} *_L dh) - e^{-\rho} e^{i\theta} d\theta \wedge *_L dh.$$

Now dividing both sides by $e^{i\theta}$ and then comparing the imaginary parts, we obtain the formula

$$\mathcal{L}(h) e^{-\rho} \text{dvol} = d(e^{-\rho} *_L dh),$$

namely that $\mathcal{L}(h)$ is the drift Laplacian

$$(4) \quad \mathcal{L}(h) = e^\rho \text{div}_L (e^{-\rho} \nabla^L h).$$

In the Calabi–Yau metric case ($\rho = 0$) this reduces to the Laplacian. By construction,

$$|\theta_{\varphi_\epsilon(L)} - \theta_L - \epsilon \mathcal{L}(h)| \leq C \epsilon^2 \|dh\|_{C^1}^2.$$

Given any point $p \in L$, we prescribe a smooth function $f : L \rightarrow \mathbb{R}$ such that $f = 4$ outside $B(p, \eta)$ for some $\eta \ll 1$ to be determined, the weighted integral $\int_L e^{-\rho} f \text{dvol}_L$ is 0, and $|f| + |df| \eta \leq C \eta^{-n}$ inside $B(p, \eta)$. We solve $\mathcal{L}(h) = f$ with $\int_L h e^{-\rho} \text{dvol}_L = 0$, to find the smooth function h with bound $|dh| + \eta |\nabla dh| \leq C \eta^{1-n}$. Using the smoothness of L , we regard its Weinstein neighbourhood as the cotangent bundle T^*L , which is isomorphic to the normal bundle NL via the metric. We can then extend h to a smooth function on X with the same bounds such that near L the function h is constant on the cotangent fibres. This ensures that the Hamiltonian vector field X_h equals $J\nabla^L h$ along L .

We choose $\eta = C_1 \epsilon^{1/2n}$ for some large constant C_1 independent of ϵ . Then the displacement of the time ϵ -flow is bounded by the C^0 norm of the vector field ϵX_h , which is of order $C \epsilon |dh| \leq C \epsilon \eta^{1-n} \leq C \epsilon^{(n+1)/(2n)} \ll 1$, so $\varphi_\epsilon(L)$ is well defined inside the Weinstein neighbourhood. Furthermore,

$$|\theta_{\varphi_\epsilon(L)} - \theta_L - \epsilon \mathcal{L}(h)| \leq C \epsilon^2 \|dh\|_{C^1}^2 \leq C (\epsilon \eta^{-n})^2 \ll \epsilon,$$

whence on $\varphi_\epsilon(L \setminus B(p, \eta))$,

$$\theta_{\varphi_\epsilon(L)} \geq \theta_L + \epsilon \mathcal{L}(h) - \epsilon \geq \epsilon \mathcal{L}(h) - 2\epsilon \geq 4\epsilon - 2\epsilon \geq 2\epsilon.$$

Without loss $\varphi_\epsilon(L)$ is transverse to L' by genericity. By formula (2), any Lagrangian intersection $q \in \text{CF}^*(\varphi_\epsilon(L), L')$ has Floer degree

$$\mu(q) \geq \frac{1}{\pi} (\theta_{\varphi_\epsilon(L)} - \theta_{L'}) \geq \frac{1}{\pi} \epsilon,$$

and hence degree-zero intersections cannot occur on $\varphi_\epsilon(L \setminus B(p, \eta))$. By the Hamiltonian invariance of Floer cohomology

$$\text{HF}^0(\varphi_\epsilon(L), L') = \text{HF}^0(L, L') \neq 0.$$

This forces there to be an intersection $q \in \varphi_\epsilon(L \cap B(p, \eta)) \cap L'$. Thus there is $p' \in L \cap B(p, \eta)$ whose distance to q is less than $C \epsilon |dh| \leq C \epsilon^{(n+1)/(2n)}$. □

Remark 2.3 If we replace the embedded Lagrangians by immersed Lagrangians with connected domain and transverse self intersections, we can still use the Floer cohomology of Akaho and Joyce [2]. The main caveat is that in order for the Hamiltonian function $h: L \rightarrow \mathbb{R}$ to extend to X , we need h to take the same value at the finitely many self-intersection points. These finitely many linear constraints are easy to meet, by relaxing $f = 4$ outside $B(p, \eta)$ to the more flexible condition $f \geq 4$ outside $B(p, \eta)$.

The connected domain hypothesis on L is important for solving the equation $\mathcal{L}(h) = f$. If we drop this hypothesis, then take for instance L the disjoint union of two special Lagrangians and L' to be one of the components, and the lemma would be false.

Remark 2.4 Lemma 2.2 easily gives a *new proof* to the original Thomas–Yau uniqueness theorem (Theorem 1.1), not relying on Morse theory or real analyticity. Indeed, by taking the $\epsilon \rightarrow 0$ limit, we deduce that each point on L has zero distance to L' , and hence $L \subset L'$. Since the roles of L and L' are symmetric in Theorem 1.1, we recover $L = L'$. The rest of this paper treats the additional difficulties caused by giving up quantitative regularity bounds on L' , and by relaxing the C^0 -smallness of $\theta_{L'}$ to merely smallness on most of the measure.

2.3 Generic perturbation technique

The technique of controlling the number of intersection points by utilizing a sufficiently rich family of perturbations originated from Arnold [4], who used it to study the dynamical growth of intersection points. Proposition 2.5 is a detailed exposition to clarify the dependence of constants. Seidel [11, Lecture 5] contains an application to bound the rank of Lagrangian Floer cohomology.

Let M be an m -dimensional compact manifold, N and N' be two compact submanifolds of complementary dimension and $U \subset N'$ be an open subset. Let V be a p -dimensional space of vector fields on M , which is surjective to the normal space TM/TN at every point of N . Fix a Euclidean metric on V , and consider the p -dimensional family of diffeomorphisms ϕ_t of M obtained by exponentiating the vector fields in $P = B(0, \epsilon) \subset V$ for some small ϵ . This induces a p -parameter deformation family $\mathcal{N} \rightarrow P$ for N whose fibres are $\phi_t(N)$ for $t \in P$. There is an evaluation map $\pi: \mathcal{N} \rightarrow M$, which by construction is surjective on tangent spaces.

Proposition 2.5 *There exists some $t \in P$ such that $\phi_t(N)$ intersects N' transversely, and the number of intersection points is*

$$|\phi_t(N) \cap U| \leq C\epsilon^{-n} \text{Vol}(U),$$

where the constant C depends only on M , V and the C^1 -regularity of N , but not on the small ϵ and the regularity bounds of N' .

Proof The transversality holds for a.e. $t \in P$ by Sard and Smale, so that $|\phi_t(N) \cap U|$ is a well defined integer for a.e. $t \in P$. By the coarea formula,

$$\int_P |\phi_t(N) \cap U| dt \leq \text{Vol}(\mathcal{N} \cap \pi^{-1}(U)),$$

where the volume is computed with respect to the restriction of the product metric on $M \times P$. Now $\pi: \mathcal{N} \cap \pi^{-1}(U) \rightarrow U$ is surjective on tangent spaces, with Jacobian factor bounded below by C^{-1} , so

$$\text{Vol}(\mathcal{N} \cap \pi^{-1}(U)) \leq C \int_U \text{Vol}(\mathcal{N} \cap \pi^{-1}(y)) \, dy.$$

For each $y \in U$, the contributions to $\mathcal{N} \cap \pi^{-1}(y)$ come from the deformations of the small local region $N \cap B(y, C\epsilon)$. The assumption that V is surjective to TM/TN at every point of N , together with the C^1 -regularity of N , allows us to find a codimension- n subspace $V_y \subset V$ such that the projection

$$\mathcal{N} \cap \pi^{-1}(y) \rightarrow P = B(0, \epsilon) \subset V \rightarrow V_y$$

exhibits $\mathcal{N} \cap \pi^{-1}(y)$ as part of a C^1 -graph over the ϵ -ball inside V_y , whose volume is bounded by $C\epsilon^{p-n}$. We deduce

$$\text{Vol}(\mathcal{N} \cap \pi^{-1}(y)) \leq C\epsilon^{p-n}.$$

Combining the above,

$$\int_P |\phi_t(N) \cap U| \, dt \leq C\epsilon^{p-n} \text{Vol}(U),$$

so we can select some $t \in P$ with

$$|\phi_t(N) \cap U| \leq C \text{Vol}(U)\epsilon^{p-n}(\text{Vol}(P))^{-1} \leq C \text{Vol}(U)\epsilon^{-n}. \quad \square$$

In our application, the idea is that under sufficiently generic Hamiltonian deformation, subsets with very small measure do not contribute to Lagrangian intersection points. This allows us to relax Lemma 2.2 to a weak L^1 -assumption on the Lagrangian angle of L' .

Proposition 2.6 *Assume $\|\theta_L\|_{C^0} \leq \epsilon$, while $\text{Vol}(\{\theta_{L'} > \epsilon\}) \ll \epsilon^{-n}$. Then there are constants C_1 and C independent of ϵ such that for all $p \in L$, there is some $p' \in L \cap B(C_1\epsilon^{1/2n})$ satisfying $\text{dist}(p', L') \leq C\epsilon^{(n+1)/(2n)}$.*

Proof We can find a large-dimensional vector space V of Hamiltonian vector fields which is surjective to the normal space TX/TL at every point of L . We can pick the Euclidean metric on V so that for any v in the unit ball of V , the induced vector field acting on the Grassmannian bundle over X has C^0 -norm $\ll 1$. In particular, the Hamiltonian diffeomorphisms ϕ_t obtained via exponentiating vector fields in $P = B(0, \epsilon) \subset V$ only change the Lagrangian angle function by an amount $\ll \epsilon$.

We now reexamine the argument of Lemma 2.2. For any $p \in L$, we can construct the C^1 -small Hamiltonian deformation $\varphi_\epsilon(L)$ of L such that on $\varphi_\epsilon(L)(L \setminus B(p, \eta))$, we have $\theta_{\varphi_\epsilon(L)} \geq 2\epsilon$. Thus for any $t \in P$, the Hamiltonian deformation $\phi_t\varphi_\epsilon(L)$ satisfies away from $B(p, \eta)$

$$\theta_{\phi_t\varphi_\epsilon(L)} \geq \theta_{\varphi_\epsilon(L)} - \frac{1}{2}\epsilon \geq \frac{3}{2}\epsilon.$$

Let $U = \{\theta_{L'} > \epsilon\} \subset L'$. Then according to Proposition 2.5, there is some $t \in P$ such that $\phi_t\varphi_\epsilon(L)$ is transverse to L' , and the number of intersection points is

$$|\phi_t\varphi_\epsilon(L) \cap U| \leq C\epsilon^{-n} \text{Vol}(U) \ll 1.$$

Thus $\phi_t \varphi_\epsilon(L) \cap U$ is in fact *empty*. This forces $\theta_{L'} \leq \epsilon$ at the Lagrangian intersections $\phi_t \varphi_\epsilon(L) \cap L'$, and we conclude as in Lemma 2.2. \square

Consequently, we get one half of the Hausdorff distance estimate:

Corollary 2.7 *Under the same conditions,*

$$(5) \quad \sup_{p \in L} \text{dist}(p, L') \leq C\epsilon^{1/2n}.$$

Remark 2.8 We have so far not used the condition $[L] = [L'] \in H_n(X, \mathbb{Q})$. Without this condition, the constant C in Corollary 2.7 really requires some regularity bound on L . For instance, we consider a surgery exact triangle $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_1[1]$, where L_2 is the Lagrangian connected sum of L_1 and L_3 with very small neck, and all Lagrangians angles can be made arbitrarily small in C^0 . The mere assumption that $\text{HF}^0(L_1, L_2) \neq 0$ cannot imply that $\sup_{p \in L_2} \text{dist}(p, L_1)$ is small. The proof breaks down because the C^1 -regularity bound of L_2 is highly degenerate in the neck region.

2.4 Monotonicity inequality

In minimal surface theory, the famous monotonicity formula says that for an n -dimensional submanifold N inside a smooth compact ambient manifold, if the mean curvature $\|\vec{H}\|_{L^\infty}$ is less than $+\infty$, then inside coordinate balls the volume ratio

$$e^{Cr} r^{-n} \text{Vol}(B(p, r))$$

increases with the radius r . One useful consequence is that the volume of $N \cap B(p, r)$ has a lower bound for $p \in N$. We only impose assumptions on the Lagrangian angle θ , but not on $|\vec{H}|$ directly. We shall see that the *volume lower bound* is still satisfied. Our style of argument is inspired by Neves [10, Section 3.3]; see also [9, Section 5.2].

Recall a special case of the optimal isoperimetric inequality of Almgren. Denote by ω_n the volume of the n -dimensional unit ball in \mathbb{R}^n .

Proposition 2.9 [3, Theorem 10] *Let T be an $(n-1)$ -dimensional integral current inside \mathbb{R}^N with $\partial T = 0$. Then there is an integral n -current Q inside \mathbb{R}^N with $\partial Q = T$ and*

$$\text{Mass}(Q) \leq n^{-n/(n-1)} \omega_n^{-1/(n-1)} \text{Mass}(T)^{n/(n-1)}.$$

Remark 2.10 The inequality is saturated by the n -dimensional unit ball; this sharpness of constant will be important for getting the sharp coefficient in Proposition 2.11(ii), which is essential for proving that the points on L' close to L occupy almost the full measure in L' as in Lemma 2.14. Moreover, if T is contained in some ball $B(R)$, then Q can also be taken inside $B(R)$, because there is a retraction of \mathbb{R}^N to $B(R)$ with Lipschitz constant 1.

We now work inside the Kähler manifold X . Around any fixed $p \in X$, we can take local holomorphic coordinates z_1, \dots, z_n such that for small $|z|$,

$$\omega = \frac{\sqrt{-1}}{2} \sum dz_i \wedge d\bar{z}_i + O(|z|), \quad e^{\rho(p)}\Omega = dz_1 \wedge \dots \wedge dz_n + O(|z|), \quad g = \sum |dz_i|^2 + O(|z|).$$

Beware that the smooth function ρ measures the failure of the metric to be Calabi–Yau.

Proposition 2.11 *Let L' be a smooth Lagrangian with $|\theta| \leq \frac{1}{2}\pi - \epsilon_0$, and assume there exists $p' \in L'$ with $|p'| \leq \lambda \ll 1$. We shall use $C(\epsilon_0)$ to denote constants depending only on ϵ_0 and the coordinate chart.*

- (i) *For the Euclidean balls $B_{\text{Eucl}}(r)$ with radius $r > \lambda$ contained in the chart, we have the volume lower bound $\text{Vol}(L' \cap B_{\text{Eucl}}(r)) \geq C(\epsilon_0)(r - \lambda)^n$.*
 - (ii) *Assume furthermore that $\text{Vol}(\{\theta_{L'} > \epsilon\} \cap L') \leq \lambda^n$ where $\lambda \geq \epsilon^2$. Then for the Euclidean balls $B_{\text{Eucl}}(r)$ with radius $r > C(\epsilon_0)\lambda$ contained in the chart, we have the sharper volume lower bound*
- (6)
$$\text{Vol}(L' \cap B_{\text{Eucl}}(r)) \geq \omega_n(r - C(\epsilon_0)\lambda)^n(1 - C(\epsilon_0)r).$$

Proof For a.e. $0 < r$ smaller than the radius of the coordinate ball (of order $O(1)$), the level set $L' \cap \partial B_{\text{Eucl}}(r)$ is smooth, and the isoperimetric inequality allows us to find $Q \subset B_{\text{Eucl}}(r)$ with $\partial Q = L' \cap \partial B_{\text{Eucl}}(r)$ and mass bound

$$\text{Mass}(Q) \leq n^{-n/(n-1)} \omega_n^{-1/(n-1)} \mathcal{H}^{n-1}(L' \cap \partial B_{\text{Eucl}}(r))^{n/(n-1)}.$$

Since $\partial(L' \cap B_{\text{Eucl}}(r)) = \partial Q$, the form $\text{Re } \Omega$ is closed and $\text{Re } \Omega|_Q \leq (1 + O(r))e^{-\rho(p)} \text{dvol}_Q$,

$$\int_{L' \cap B_{\text{Eucl}}(r)} \text{Re } \Omega = \int_Q \text{Re } \Omega \leq (1 + O(r))e^{-\rho(p)} \text{Mass}(Q).$$

Under the quantitative almost-calibrated condition $|\theta| \leq \frac{1}{2}\pi - \epsilon_0$,

$$\text{Vol}(L' \cap B_{\text{Eucl}}(r)) = \int_{L' \cap B_{\text{Eucl}}(r)} \text{dvol} \leq \frac{1}{\sin \epsilon_0} \int_{L' \cap B_{\text{Eucl}}(r)} e^{\rho} \text{Re } \Omega.$$

Combining the above,

$$\text{Vol}(L' \cap B_{\text{Eucl}}(r)) \leq C(\epsilon_0) \mathcal{H}^{n-1}(L' \cap \partial B_{\text{Eucl}}(r))^{n/(n-1)}.$$

The function $f(r) = \text{Vol}(L' \cap B_{\text{Eucl}}(r))$ is increasing in r . Using the coarea formula, we rewrite the differential inequality as

$$f(r) \leq C(\epsilon_0) f'(r)^{n/(n-1)}.$$

For $r > \lambda$, we have $f(r) > 0$, and the increasing function f satisfies

$$(f^{1/n})' \geq C(\epsilon_0),$$

whence $f \geq C(\epsilon_0)(r - \lambda)^n$.

Next we assume $\text{Vol}(\{\theta_{L'} > \epsilon\} \cap L') \leq \lambda^n$ and deduce the sharper volume lower bound. For $O(1) > r > C\lambda$ with C depending only on the volume lower bound constant, we have $f(r) > \lambda^n$. Notice that on $L' \cap \{\theta_{L'} \leq \epsilon\} \cap B_{\text{Eucl}}(r)$,

$$\int_{L' \cap B_{\text{Eucl}}(r) \cap \{\theta_{L'} \leq \epsilon\}} \text{dvol} \leq \frac{1}{\cos \epsilon} \int_{L' \cap B_{\text{Eucl}}(r) \cap \{\theta_{L'} \leq \epsilon\}} e^\rho \text{Re } \Omega \leq \frac{1}{\cos \epsilon} \int_{L' \cap B_{\text{Eucl}}(r)} e^\rho \text{Re } \Omega.$$

Hence

$$\begin{aligned} f(r) &\leq \lambda^n + \frac{1}{\cos \epsilon} \int_{L' \cap B_{\text{Eucl}}(r)} e^\rho \text{Re } \Omega \leq \lambda^n + \frac{1}{\cos \epsilon} (1 + O(r)) \text{Mass}(Q) \\ &\leq \lambda^n + \frac{1}{\cos \epsilon} (1 + O(r)) n^{-n/(n-1)} \omega_n^{-1/(n-1)} f'(r)^{n/(n-1)}. \end{aligned}$$

Now $1/\cos \epsilon = 1 + O(\epsilon^2) = 1 + O(\lambda) = 1 + O(r)$, so the $1/\cos \epsilon$ factor can be absorbed into the $1 + O(r)$ factor by changing the constant. Rearranging the terms,

$$f'(r) \geq n \omega_n^{1/n} (1 - O(r)) (f - \lambda^n)^{(n-1)/n}.$$

We deduce

$$\frac{d}{dr} (f - \lambda^n)^{1/n} \geq \omega_n^{1/n} (1 - O(r)),$$

and after integration

$$f \geq \omega_n (r - C\lambda)^n (1 - O(r)),$$

as required. □

Remark 2.12 In the context of special Lagrangians in \mathbb{C}^n , a standard way to deduce the monotonicity formula is to compare the volume of $L' \cap B(r)$ with the cone over $L' \cap \partial B(r)$. If we follow this strategy for smooth Lagrangians inside \mathbb{C}^n with $|\theta| \leq \epsilon$, then we would deduce

$$\text{Vol}(L' \cap B(r)) \leq \frac{1}{\cos \epsilon} \int_{L' \cap B(r)} \text{Re } \Omega \leq \frac{r}{n \cos \epsilon} \mathcal{H}^{n-1}(L' \cap B(r)),$$

whence $d/dr \log \text{Vol}(L' \cap B(r)) \geq n \cos \epsilon / r$. Unfortunately, this would only prove the monotonicity of $\text{Vol}(L' \cap B(r)) r^{-n \cos \epsilon}$, which is *not sufficient* to deduce the volume lower bound above.

We think it is an interesting question whether $e^{Cr} \text{Vol}(L' \cap B(r)) r^{-n}$ is monotone under the mere hypothesis that $\|\theta\|_{C^0}$ is small, with no assumption directly on the mean curvature.

2.5 Hausdorff distance bound

We still need to show that all the points of L' must stay close to L . The strategy is to first show that such points occupy almost the full measure of L' , and then use a monotonicity inequality argument to extend this to all points.

We record a weighted volume upper bound:

Lemma 2.13 *If $\text{Vol}(L' \cap \{\theta_{L'} > \epsilon\}) \leq \epsilon^n$, then $\int_{L'} e^{-\rho} \text{dvol}_{L'} \leq \int_{L'} \text{Re } \Omega / \cos \epsilon + C\epsilon^n$.*

Proof On the subset $L' \cap \{\theta_{L'} \leq \epsilon\}$, we integrate $e^{-\rho} \text{dvol}_{L'} \leq (1/\cos \epsilon) \text{Re } \Omega|_{L'}$. □

Lemma 2.14 *In the setting of Theorem 2.1,*

$$(7) \quad \int_{L' \cap \{\text{dist}(\cdot, L) \leq \epsilon^{1/4n}\}} e^{-\rho} \, \text{dvol}_{L'} \geq (1 - C\epsilon^{1/4n}) \int_L \text{Re } \Omega.$$

Proof We can take $\lambda = C\epsilon^{1/2n}$ such that by Corollary 2.7, any $p \in L$ lies within distance λ to L' . Our setting implies $\lambda \geq \epsilon^2$ and $\text{Vol}(\{\theta_{L'} > \epsilon\}) \leq \lambda^n$. By Proposition 2.11, and the fact that Euclidean balls and Riemannian balls only differ by relative error $O(r)$, we see that for $O(1) \geq r \geq C\lambda$,

$$\text{Vol}_g(L' \cap B_g(p, r)) \geq \omega_n(r - C\lambda)^n(1 - Cr) \quad \text{for all } p \in L.$$

The constants are uniform for $p \in L$, so after integration

$$\begin{aligned} \int_L e^{-\rho(p)} \text{Vol}_g(L' \cap B_g(p, r)) \, \text{dvol}_L(p) &\geq \omega_n(r - C\lambda)^n(1 - Cr) \int_L e^{-\rho} \, \text{dvol}_L \\ &\geq \omega_n(r - C\lambda)^n(1 - Cr) \int_L \text{Re } \Omega. \end{aligned}$$

The left side can be computed by Fubini's theorem as

$$\int_{L'} e^{-\rho(q)} \, \text{dvol}_{L'}(q) \int_{L \cap B_g(q, r)} e^{\rho(q) - \rho(p)} \, \text{dvol}_L(p).$$

The function ρ is smooth, so $|\rho(q) - \rho(p)| \leq Cr$. The Lagrangian L has fixed C^1 -regularity bound, so for small radius $r \ll 1$,

$$\int_{L \cap B_g(q, r)} \text{dvol}_L(p) \leq (1 + Cr)\omega_n r^n.$$

Combining the above,

$$\begin{aligned} \omega_n(r - C\lambda)^n(1 - Cr) \int_L \text{Re } \Omega &\leq \int_L e^{-\rho(p)} \text{Vol}_g(L' \cap B_g(p, r)) \, \text{dvol}_L(p) \\ &\leq (1 + Cr)\omega_n r^n \int_{L' \cap \{\text{dist}(\cdot, L) \leq r\}} e^{-\rho(q)} \, \text{dvol}_{L'}(q), \end{aligned}$$

whence for $C\lambda \leq r \leq O(1)$,

$$(8) \quad \int_{L' \cap \{\text{dist}(\cdot, L) \leq r\}} e^{-\rho(q)} \, \text{dvol}_{L'}(q) \geq (1 - Cr - C\frac{\lambda}{r}) \int_L \text{Re } \Omega.$$

The choice $r = \epsilon^{1/4n}$ yields the claim. □

Remark 2.15 In the above, we assumed L is embedded. If L is merely immersed with transverse self intersections, then the inequality

$$\int_{L \cap B_g(q, r)} \text{dvol}_L(p) \leq (1 + Cr)\omega_n r^n$$

will break down for q in the r -neighbourhood of the self-intersection points. Essentially the same proof would then give

$$\int_{L' \cap \{\text{dist}(\cdot, L) \leq \epsilon^{1/4n}\}} e^{-\rho} \, \text{dvol}_{L'} \geq (1 - C\epsilon^{1/4n}) \int_L \text{Re } \Omega - Cr^n.$$

Since $r^n \ll \epsilon^{1/4n}$, the r^n term can be absorbed.

We can finally prove the remaining half of the *Hausdorff distance bound*:

Proposition 2.16 *In the setting of Theorem 2.1,*

$$(9) \quad \sup_{p \in L'} \text{dist}(p, L) \leq C\epsilon^{1/(4n^2)}.$$

Proof The assumption that $[L] = [L'] \in H_n(X, \mathbb{Q})$ and the hypothesis that $\text{Vol}(L' \cap \{\theta_{L'} > \epsilon\}) \leq \epsilon^n$, together with Lemma 2.13, imply

$$\int_{L'} e^{-\rho} \, \text{dvol}_{L'} \leq \frac{1}{\cos \epsilon} \int_L \text{Re } \Omega + C\epsilon^n \leq (1 + C\epsilon^2) \int_L \text{Re } \Omega.$$

Combined with Lemma 2.14,

$$\int_{L' \cap \{\text{dist}(\cdot, L) \geq \epsilon^{1/4n}\}} e^{-\rho} \, \text{dvol}_{L'} \leq C\epsilon^{1/4n} \int_L \text{Re } \Omega.$$

Assume $p \in L'$ with $r_p := \text{dist}(p, L) > \epsilon^{1/4n}$, so that $L' \cap B(p, r_p - \epsilon^{1/4n})$ is in $L' \cap \{\text{dist}(\cdot, L) \geq \epsilon^{1/4n}\}$. From Proposition 2.11,

$$C(r_p - \epsilon^{1/4n})^n \leq \text{Vol}(L' \cap B(p, r_p - \epsilon^{1/4n})) \leq C\epsilon^{1/4n} \int_L \text{Re } \Omega.$$

Consequently $r_p \leq C\epsilon^{1/4n^2}$. □

2.6 Flat norm distance bound

Recall that the flat norm distance between L and L' is defined as

$$\inf\{\text{Mass}(T) + \text{Mass}(R) : L - L' = \partial T + R\},$$

where T and R are integral currents of dimension $n + 1$ and n , respectively. Intuitively, the Hausdorff distance bound is an L^∞ -type bound on the distance function, while the flat norm behaves like an L^1 -type bound. In the corollary below, we leverage the previous information to get a slightly better exponent than the obvious bound from $L^\infty \rightarrow L^1$, but we expect this is still not sharp.

Corollary 2.17 *In the setting of Theorem 2.1, there is an integral current T with $\partial T = L' - L$ such that $\text{Mass}(T) \leq C\epsilon^{1/(4n^2)+1/(4n)}$.*

Proof The Hausdorff bound shows that L' lies within a small C^0 -neighbourhood of L . Using the controlled C^1 -regularity of L , we can view the C^0 -neighbourhood as its normal bundle, and obtain a projection map $\pi: L' \rightarrow L$. Since L is homologous to L' by assumption, π has degree 1. We form an $(n+1)$ -dimensional integral current T from the union of the line segments joining the points $q \in L'$ to $\pi(q) \in L$. Up to the choice of orientation sign, $\partial T = L' - \pi_*(L') = L' - L$. The lengths of the segments are comparable to $\text{dist}(q, L)$. Thus

$$\text{Mass}(T) \leq C \int_{L'} \text{dist}(\cdot, L) \, \text{dvol}_{L'} = C \int_0^{\sup \text{dist}} dr \int_{L' \cap \{\text{dist}(\cdot, L) \geq r\}} \text{dvol}_{L'}.$$

Recall the weak L^1 -type estimate (8), which implies

$$\int_{L' \cap \{\text{dist}(\cdot, L) \geq r\}} \text{dvol}_{L'} \leq C \begin{cases} 1 & \text{if } r = O(\epsilon^{1/2n}), \\ \epsilon^{1/2n}/r & \text{if } \epsilon^{1/2n} \leq r \leq \epsilon^{1/4n}, \\ \epsilon^{1/4n} & \text{if } r > \epsilon^{1/4n}. \end{cases}$$

By the Hausdorff distance estimate, $\sup \text{dist}(\cdot, L) \leq C\epsilon^{1/4n^2}$. Combining the above gives the mass bound. □

Remark 2.18 If L is immersed rather than embedded, then π would be only Lipschitz near the self-intersection points, but the above argument is unaffected.

2.7 The F-metric bound

We view the C^0 neighbourhood of L as its normal bundle, which has a projection π to its zero section L . Normal geodesic flow allows us to identify the tangent spaces $T_q X$ with $T_{\pi(q)} X$. We now show that L' is graphical over L with small norm, away from a set with small measure. Denote by $\vec{T}(q)$ the unit n -vector $e_1 \wedge \cdots \wedge e_n$ on L' , where e_i is an oriented orthonormal frame of $T_q L'$, and by $\vec{T}(\pi(q))$ the analogous unit n -vector from the oriented orthonormal basis of $T_{\pi(q)} L$.

Proposition 2.19 *In the setting of Theorem 2.1, away from a subset $E \subset L$ with $\text{Vol}(E) + \text{Vol}(\pi^{-1}(E)) \leq C\epsilon^{1/4n}$, the projection $\pi: L' \setminus \pi^{-1}(E) \rightarrow L \setminus E$ is bijective, and*

$$\int_{L' \setminus \pi^{-1}(E)} |\vec{T}(q) - \vec{T}(\pi(q))|^2 \, \text{dvol}(q) \leq C\epsilon^{1/4n}.$$

Proof The projection $\pi: L' \rightarrow L$ is almost metric decreasing:

$$|d\pi|_{T_q L'}| \leq 1 + O(\text{dist}(q, \pi(q))).$$

As a general rule, the error coming from comparing ambient Kähler structures at q and $\pi(q)$ is bounded by $O(\text{dist}(q, \pi(q)))$. Let $E'_1 = \{\text{dist}(\cdot, L) \geq \epsilon^{1/4n}\} \subset L'$. Then E'_1 has measure bounded by $C\epsilon^{1/4n}$ by Lemma 2.13 and (7).

Since $\pi : L' \rightarrow L$ has degree 1, it is surjective. Let E be the subset of L with at least two preimages. The volume almost decreasing property of $d\pi$ together with the assumption $|\theta| \leq \epsilon$ on L imply

$$2 \int_E \operatorname{Re} \Omega = 2 \int_E e^{-\rho} \operatorname{dvol} (1 + O(\epsilon)) \leq (1 + O(\epsilon^{1/4n^2})) \int_{\pi^{-1}(E)} e^{-\rho} \operatorname{dvol}.$$

In particular,

$$(10) \quad \operatorname{Vol}(E) + \operatorname{Vol}(\pi^{-1}(E)) \leq C \int_{\pi^{-1}(E)} e^{-\rho} \operatorname{dvol} \leq C' \left(\int_{\pi^{-1}(E)} e^{-\rho} \operatorname{dvol} - \int_E \operatorname{Re} \Omega \right).$$

On the other hand, at the points $q \in L' \setminus \pi^{-1}(E) \cup E'_1$, the difference between the tangent planes is bounded:

$$|\vec{T}(q) - \vec{T}(\pi(q))|^2 \leq C(1 - |d\pi(\vec{T})| + O(\epsilon^{1/4n})).$$

Thus

$$\begin{aligned} \int_{L' \setminus \pi^{-1}(E)} |\vec{T}(q) - \vec{T}(\pi(q))|^2 \operatorname{dvol}(q) &\leq C\epsilon^{1/4n} + C \int_{L' \setminus \pi^{-1}(E)} (1 - |d\pi(\vec{T})|) \operatorname{dvol} \\ &\leq C\epsilon^{1/4n} + C' \int_{L' \setminus \pi^{-1}(E)} (1 - |d\pi(\vec{T})|) e^{-\rho} \operatorname{dvol}. \end{aligned}$$

Here the contribution from E'_1 is absorbed into $C\epsilon^{1/4n}$. From the pointwise inequality on $L' \setminus \pi^{-1}(E) \cup E'_1$,

$$|d\pi(\vec{T})| e^{-\rho} = (1 - O(\epsilon^{1/4n})) |d\pi(\vec{T})| |\operatorname{Re} \Omega| \geq \operatorname{Re} \Omega(\pi_* \vec{T}) (1 - O(\epsilon^{1/4n})),$$

the above is bounded by

$$C\epsilon^{1/4n} + C' \left(\int_{L' \setminus \pi^{-1}(E)} e^{-\rho} \operatorname{dvol} - \int_{L \setminus E} \operatorname{Re} \Omega \right).$$

The coefficient C' can be arranged as the same (large) constant appearing in (10). Adding the two contributions and invoking the weighted volume upper bound in Lemma 2.13,

$$\begin{aligned} \operatorname{Vol}(E) + \operatorname{Vol}(\pi^{-1}(E)) + \int_{L' \setminus \pi^{-1}(E)} |\vec{T}(q) - \vec{T}(\pi(q))|^2 \operatorname{dvol}(q) &\leq C\epsilon^{1/4n} + C' \left(\int_{L'} e^{-\rho} \operatorname{dvol} - \int_L \operatorname{Re} \Omega \right) \\ &\leq C\epsilon^{1/4n} + C' \left(\int_{L'} \operatorname{Re} \Omega - \int_L \operatorname{Re} \Omega \right) = C\epsilon^{1/4n}, \end{aligned}$$

as required. □

Remark 2.20 If L is immersed instead of embedded, then we include the $O(\epsilon^{1/4n^2})$ neighbourhood of the self intersections into E , and the rest of the arguments run almost verbatim.

Recall the distance between two varifolds L and L' is measured by *F-metric*

$$\sup \left\{ \left| \int_L f(p, T_p L) \operatorname{dvol}_L - \int_{L'} f(q, T_q L') \operatorname{dvol}_{L'} \right| \right\},$$

where f ranges over all functions on the Grassmannian bundle of n -dimensional tangent subspaces over X , with $|f| \leq 1$ and Lipschitz constant at most 1. Given a uniform upper bound on the mass, the F -metric induces the same topology as the weak topology on varifolds (because continuous functions can be approximated by Lipschitz functions).

Corollary 2.21 *In the setting above, the F -metric between L and L' is bounded by $C\epsilon^{1/8n}$.*

Proof Since $|f| \leq 1$, the integrals on E, E'_1 and $\pi^{-1}(E)$ are bounded by $C\epsilon^{1/4n}$. At $q \in L' \setminus \pi^{-1}(E) \cup E'_1$, by the Lipschitz bound on f ,

$$|f(q, T_q L') - f(\pi(q), T_{\pi(q)} L)| \leq |q - \pi(q)| + |\vec{T}(q) - \vec{T}(\pi(q))| \leq C\epsilon^{1/4n} + |\vec{T}(q) - \vec{T}(\pi(q))|.$$

The discrepancy between the volume forms $\pi^* \text{dvol}_L$ and $\text{dvol}_{L'}$ is bounded by

$$C|\vec{T}(q) - \vec{T}(\pi(q))|^2 \text{dvol}_{L'}.$$

Consequently, the F -metric is bounded by

$$C\epsilon^{1/4n} + \int_{L' \setminus \pi^{-1}(E)} |\vec{T}(q) - \vec{T}(\pi(q))| \text{dvol}_{L'} + C \int_{L' \setminus \pi^{-1}(E)} |\vec{T}(q) - \vec{T}(\pi(q))|^2 \text{dvol}_{L'}.$$

Using Proposition 2.19 and Cauchy–Schwarz, this is bounded by $C\epsilon^{1/8n}$. □

3 Consequences of Theorem 2.1

We now prove the consequences stated in the introduction.

Proof of Theorem 1.6 From the C^∞ -limit assumption, $L_i \rightarrow L_\infty$ with $\|\theta\|_{C^0} \rightarrow 0$. From the varifold limit assumption, $L'_i \rightarrow L'$. This implies that $\|\theta\|_{L^1} \rightarrow 0$ as follows. Since the Lagrangian angle is a continuous function on the open subset $\{\text{Re } \Omega > 0\}$ in the Grassmannian bundle, the weak topology convergence of L'_i implies

$$\int_{L'_i} |\theta| \text{dvol}_{L'_i} \rightarrow \int_{L'_\infty} |\theta| \text{dvol} = 0.$$

All Lagrangians lie in the same homology class, and all the approximants L_i and L'_i lie in the same derived category class. Thus Theorem 2.1 applies to L_i and L'_i for $i \gg 1$ and arbitrarily small ϵ , and the uniform C^∞ -regularity on L_i means that all the constants are uniform. The flat norm distance between L_i and L'_i is bounded by $C\epsilon^{1/(4n)+1/(4n^2)}$. Taking the limit as $i \rightarrow +\infty$, the flat norm distance between L_∞ and L'_∞ is bounded by $C\epsilon^{1/(4n)+1/(4n^2)}$, and letting $\epsilon \rightarrow 0$ gives $L_\infty = L'_\infty$. □

Proof of Theorem 1.7 As above, varifold convergence to L^∞ implies $\|\theta_{L'_i}\|_{L^1} \rightarrow 0$. Conversely, if $\|\theta_{L'_i}\|_{L^1} \rightarrow 0$, then Theorem 2.1 applies to L_i and L'_i for $i \gg 1$ and arbitrarily small ϵ , whence the F -metric between L_i and L'_i is bounded by $C\epsilon^{1/8n}$ for $i \gg 1$ depending on ϵ . In the $i \rightarrow +\infty$ limit, the F -metric distance between L'_i and L_∞ tends to zero. Since all Lagrangians have uniform mass bounds, the F -metric induces the same topology as the weak topology on varifolds, whence $L'_i \rightarrow L_\infty$ in the varifold topology. □

4 Open questions

We raise some natural questions:

Question 2 With the same assumptions on L and L' as in Theorem 2.1, what would be the *sharp exponent* for the Hausdorff distance bound in terms of ϵ ?

This question has many variants. For instance, one can replace $\text{Vol}(\{\theta_{L'} > \epsilon\}) \leq \epsilon^n$ with the smallness of $\|\theta_{L'}\|_{L^p(L')}$ (or $\|\theta_{L'}\|_{C^0(L')}$). We expect the exponents of ϵ in Theorem 2.1 to be not optimal. On the other hand, even in the simple setting of Calabi–Yau metrics, assuming L is a fixed smooth special Lagrangian and $\|\theta_{L'}\|_{L^1} < \epsilon$, we do not expect the Hausdorff distance to be uniformly bounded by $O(\epsilon)$. The reason is that the linearization of the special Lagrangian equation is the Poisson equation $\Delta u = f$, and the failure of the Sobolev embedding $W^{2,1} \rightarrow C^1$ means we cannot expect u to have C^1 -bounds proportional to $\|f\|_{L^1}$. Some nonlinear effects are necessary for a result like Theorem 2.1.

Question 3 Do the uniform constants really depend on the C^∞ -regularity bounds of L ?

The part of our arguments most sensitive to C^∞ -regularity bounds is Lemma 2.2, which involves solving the Poisson equation on L with bounds. If L has some mild degeneration, eg when it is approximately the desingularization of some special Lagrangian with certain isolated conical singularities, then we expect results similar to Theorem 2.1 would hold, with possibly different exponents for ϵ . On the other hand, it is not clear if we can drop all regularity bounds on L beyond the information of the homology class and the Lagrangian angles. A closely related question is whether *strong uniqueness* holds:

Question 4 Consider the variant of Theorem 1.6 where both L_∞ and L'_∞ are varifold limits. Assuming both are special Lagrangian $\theta_{L_\infty} = \theta_{L'_\infty} = 0$, do they coincide as integral currents?

Question 5 If L and L' are immersed Lagrangians defining the same object in the derived Fukaya category, not necessarily with connected domains, then is there still an analogue of the quantitative Thomas–Yau uniqueness theorem? Is there a corresponding strong–weak uniqueness theorem?

Acknowledgement This work was completed while the author was a Clay research fellow based at MIT. He thanks P Seidel for discussions, Y Imagi for comments and the referees for useful suggestions.

References

- [1] M Abouzaid, Y Imagi, *Nearby special Lagrangians*, preprint (2021) arXiv 2112.10385
- [2] M Akaho, D Joyce, *Immersed Lagrangian Floer theory*, J. Differential Geom. 86 (2010) 381–500 MR Zbl
- [3] F Almgren, *Optimal isoperimetric inequalities*, Indiana Univ. Math. J. 35 (1986) 451–547 MR Zbl

- [4] **VI Arnol'd**, *Dynamics of complexity of intersections*, Bol. Soc. Brasil. Mat. 21 (1990) 1–10 MR Zbl
- [5] **D Auroux**, *A beginner's introduction to Fukaya categories*, from “Contact and symplectic topology” (F Bourgeois, V Colin, A Stipsicz, editors), Bolyai Soc. Math. Stud. 26, János Bolyai Math. Soc., Budapest (2014) 85–136 MR
- [6] **Y Imagi**, *A uniqueness theorem for gluing calibrated submanifolds*, Comm. Anal. Geom. 23 (2015) 691–715 MR
- [7] **Y Imagi, D Joyce, J Oliveira dos Santos**, *Uniqueness results for special Lagrangians and Lagrangian mean curvature flow expanders in \mathbb{C}^m* , Duke Math. J. 165 (2016) 847–933 MR
- [8] **D Joyce**, *Conjectures on Bridgeland stability for Fukaya categories of Calabi–Yau manifolds, special Lagrangians, and Lagrangian mean curvature flow*, EMS Surv. Math. Sci. 2 (2015) 1–62 MR Zbl
- [9] **Y Li**, *Thomas–Yau conjecture and holomorphic curves*, EMS Surv. Math. Sci. (online publication February 2025)
- [10] **A Neves**, *Recent progress on singularities of Lagrangian mean curvature flow*, from “Surveys in geometric analysis and relativity” (HL Bray, WP Minicozzi, editors), Adv. Lect. Math. 20, International, Somerville, MA (2011) 413–438 MR
- [11] **P Seidel**, *Categorical dynamics*, lecture notes, MIT (2012) Available at <https://math.mit.edu/~seidel/talks/mordell.pdf>
- [12] **R P Thomas, S-T Yau**, *Special Lagrangians, stable bundles and mean curvature flow*, Comm. Anal. Geom. 10 (2002) 1075–1113 MR Zbl

*Department of Pure Mathematics and Mathematical Statistics, Cambridge University
Cambridge, United Kingdom*

y1454@cam.ac.uk

Proposed: Richard P Thomas
Seconded: Mark Gross, Jim Bryan

Received: 24 September 2022
Revised: 29 October 2024

Knot surgery formulae for instanton Floer homology, I: The main theorem

ZHENKUN LI

FAN YE

We prove an integral surgery formula for framed instanton homology $I^\sharp(Y_m(K))$ for any knot K in a 3-manifold Y with $[K] = 0 \in H_1(Y; \mathbb{Q})$ and $m \neq 0$. Although the statement is similar to Ozsváth–Szabó’s integral surgery formula for Heegaard Floer homology, the proof is new and based on sutured instanton homology SHI and the octahedral lemma in the derived category. As byproducts, we obtain a formula computing instanton knot homology of the dual knot analogous to Eftekhary’s and Hedden–Levine’s work, and also an exact triangle between $I^\sharp(Y_m(K))$, $I^\sharp(Y_{m+k}(K))$ and k copies of $I^\sharp(Y)$ for any $m \neq 0$ and large k . In the proof of the formula, we discover many new exact triangles for sutured instanton homology and relate some surgery cobordism map to the sum of bypass maps, which are of independent interest. In a companion paper, we derive many applications and computations based on the integral surgery formula.

57K10, 57K14, 57K16, 57K18, 57K31

1. Introduction	2269
2. Basic setup	2274
3. Integral surgery formulae	2293
4. Dehn surgery and bypass maps	2304
5. Some exactness by diagram chasing	2308
6. Some technical constructions	2312
7. The map in the third exact triangle	2323
References	2341

1 Introduction

The framed instanton homology $I^\sharp(Y)$ for a closed 3-manifold Y was introduced by Kronheimer and Mrowka [2011] and has been conjectured to be isomorphic to the hat version of Heegaard Floer homology, $\widehat{HF}(Y)$. This conjecture is still widely open and, due to the computational difficulty of instanton Floer homology, not many examples are known. In recent years, many people have done computations of the framed instanton homology of special families of 3-manifolds; see for example [Lidman et al. 2022] and [Baldwin and Sivek 2023; 2021]. So far, most results have focused on computing the dimension of framed instanton Floer homology and many techniques only work for S^3 or rational homology spheres,

but a general structural theorem that relates the framed instanton homology of Dehn surgeries to the information from the knot complement still remains elusive.

In [Li and Ye 2021], we proved a large surgery formula for framed instanton homology, which led to a series of applications in computing the framed instanton homology and studying the representations of the fundamental groups of Dehn surgeries of some families of knots. However, in that work, the Dehn surgery slope must be large (at least $2g + 1$, where g is the Seifert genus of the knot), and thus still not much is known about the framed instanton homology of small Dehn surgery slopes. In this paper, we further prove an integral surgery formula for rationally null-homologous knots, inspired by Ozsváth and Szabó’s surgery formula [2008; 2011] for Heegaard Floer homology. For simplicity, in the introduction we present only the discussions and results for (integral) null-homologous knots (eg knots in S^3) and leave the general setup to Section 3.3.

First let us recall the results from [Li and Ye 2021]. Suppose $K \subset Y$ is a null-homologous knot. Let $Y \setminus N(K)$ be the knot complement, and let Γ_μ be the union of two oppositely oriented meridians of the knot on $\partial(Y \setminus N(K))$. Let $\text{SHI}(-Y \setminus N(K), -\Gamma_\mu)$ be the corresponding sutured instanton homology introduced by Kronheimer and Mrowka [2010], where the minus sign denotes orientation-reversal for technical needs; note that $\text{SHI}(-M, -\gamma) \cong \text{SHI}(M, \gamma)$ and in particular, $I^\sharp(-Y_{-m}(K)) \cong I^\sharp(Y_{-m}(K))$. A Seifert surface of K induces a \mathbb{Z} -grading on $\text{SHI}(-Y \setminus N(K), -\Gamma_\mu)$. In [Li and Ye 2021], we constructed a set of differentials on $\text{SHI}(-Y \setminus N(K), -\Gamma_\mu)$,

$$d_j^i : \text{SHI}(-Y \setminus N(K), -\Gamma_\mu, i) \rightarrow \text{SHI}(-Y \setminus N(K), -\Gamma_\mu, j)$$

for any gradings $i \neq j \in \mathbb{Z}$. We then constructed bent complexes

$$A_s = \left(\text{SHI}(-Y \setminus N(K), -\Gamma_\mu), \sum_{s \leq i < j} d_j^i + \sum_{s \geq i > j} d_j^i \right),$$

$$B^+ = \left(\text{SHI}(-Y \setminus N(K), -\Gamma_\mu), \sum_{i < j} d_j^i \right) \quad \text{and} \quad B^- = \left(\text{SHI}(-Y \setminus N(K), -\Gamma_\mu), \sum_{i > j} d_j^i \right).$$

From [Li and Ye 2021], the homologies of these complexes are related to the Dehn surgeries of K by

$$(1-1) \quad H(B^+) \cong H(B^-) \cong I^\sharp(-Y),$$

$$(1-2) \quad I^\sharp(Y_{-m}(K)) \cong \bigoplus_{s=\lfloor (1-m)/2 \rfloor}^{\lfloor (m-1)/2 \rfloor} H(A_s) \quad \text{for any integer } m \geq 2g(K) + 1.$$

To state the integral surgery formula, we introduce more notation. For $s \in \mathbb{Z}$, let B_s^\pm be identical copies of B^\pm . Define chain maps

$$\pi^{\pm, s} : A_s \rightarrow B_s^\pm$$

as follows. For $x \in (\text{SHI}(-Y \setminus N(K), -\Gamma_\mu), i)$,

$$\pi^{+, s}(x) = \begin{cases} x & \text{if } i \geq s, \\ 0 & \text{if } i < s, \end{cases} \quad \text{and} \quad \pi^{-, s}(x) = \begin{cases} x & \text{if } i \leq s, \\ 0 & \text{if } i > s. \end{cases}$$

Let π^\pm denote the direct sum of all $\pi^{\pm,s}$. While this slightly abuses the notation, we also use it to denote the induced maps on the homologies. The main result of the paper is the following.

Theorem 1.1 (integral surgery formula) *Suppose $K \subset Y$ is a null-homologous knot. Let A_s, B_s^\pm and π^\pm be defined as above. Then for any $m \in \mathbb{Z} \setminus \{0\}$, there exists an isomorphism*

$$\Xi_m : \bigoplus_{s \in \mathbb{Z}} H(B_s^+) \xrightarrow{\cong} \bigoplus_{s \in \mathbb{Z}} H(B_{s+m}^-)$$

as the direct sum of isomorphisms

$$\Xi_{m,s} : H(B_s^+) \xrightarrow{\cong} H(B_{s+m}^-)$$

such that

$$I^\#(-Y_{-m}(K)) \cong H\left(\text{Cone}\left(\pi^- + \Xi_m \circ \pi^+ : \bigoplus_{s \in \mathbb{Z}} H(A_s) \rightarrow \bigoplus_{s \in \mathbb{Z}} H(B_s^-)\right)\right).$$

Remark 1.2 As an analog of the surgery formula in Heegaard Floer homology, the map π^- is related to the vertical projection map v , and the map $\Xi_m \circ \pi^+$ is related to the map h , which is defined by the composition of the horizontal projection and a chain homotopy equivalence between $C\{j \geq 0\}$ to $C\{i \geq 0\}$. Here we bend the horizontal part of the hook complex to become vertical, so the differentials go upwards. The homotopy equivalence in Heegaard Floer homology depends on many auxiliary choices; cf the construction before [Hedden and Levine 2024, Lemma 2.16]. The same situation applies to Ξ_m . Hence we only state the existence of the isomorphism.

Remark 1.3 The hypothesis of Theorem 1.1 excludes the case where $m = 0$. This is due to the sign ambiguity in the definition of sutured instanton homology. The original version of sutured instanton homology defined by Kronheimer and Mrowka [2010] was only well-defined up to isomorphisms, and then Baldwin and Sivek [2015] proved that they are well-defined up to a scalar in \mathbb{C} . As a result, all related maps are only well-defined up to scalars. When $m \neq 0$, the maps $\pi^{+,s}$ and $\Xi_m \circ \pi^{-,s}$ have distinct target spaces, namely B_s and B_{s+m} . As a result, the scalar ambiguity for individual maps does not influence the dimension of the homology of the mapping cone. However, when $m = 0$, different scalars would indeed make differences. For an example of this subtlety, see the end of Section 4.

Remark 1.4 We also obtain a formula computing instanton knot homology $\text{KHI}(-Y_{-m}(K), K_{-m})$ of the dual knot K_{-m} inside the resulting manifold in Theorem 3.22, which is analogous to the results by Eftekhary [2018, Proposition 1.5] for knot Floer homology $\widehat{\text{HF}}K$ and by Hedden and Levine [2024]. In this formula, we may assume $m = 0$ because the scalar issue in Remark 1.3 does not appear.

With the isomorphisms in (1-2) and (1-1), we can truncate the above formula for $I^\#(Y_{-m}(K))$ to obtain the following exact triangle.

Corollary 1.5 (generalized surgery exact triangle) *Suppose $K \subset Y$ is a null-homologous knot, and m is a fixed nonzero integer. Then for any sufficiently large integer k , there exists an exact triangle*

$$(1-3) \quad \begin{array}{ccc} I^\sharp(-Y_{-m-k}(K)) & \xrightarrow{\quad\quad\quad} & \bigoplus_{i=1}^k I^\sharp(-Y) \\ & \swarrow \quad \quad \quad \searrow & \\ & I^\sharp(-Y_{-m}(K)) & \end{array}$$

Remark 1.6 The analogous result of the exact triangle (1-3) in Heegaard Floer theory was proved by Ozsváth and Szabó [2008] using twisted coefficients, which is a crucial step towards proving the integral surgery formula in their setup. The proof cannot be applied to instanton theory directly. Thus, in this paper, we adopt a reversed approach: we use sutured instanton theory to prove Theorem 1.1 and derive Corollary 1.5 as a direct application. The strategy to prove Theorem 1.1 can be found in Sections 3.1 and 3.2.

The analogs of $\pi^{\pm,s}$ in Heegaard Floer theory can be interpreted as cobordism maps associated to some particular spin^c structures. In instanton theory, there is a decomposition of cobordism maps along basic classes. However, currently such a decomposition is only known to exist for cobordisms whose first Betti number is zero. So for the moment let us assume the ambient 3-manifold Y is a rational homology sphere. For any integer m , there is a natural cobordism W_m from $-Y_{-m}^3(K)$ to $-Y^3$. From [Baldwin and Sivek 2023, Section 1.2], there exists a decomposition of the cobordism map $I^\sharp(W_m)$ along basic classes

$$I^\sharp(W_m) = \sum_{s \in \mathbb{Z}} I^\sharp(W_m, [s]),$$

where $[s] \in H^2(W)$ denotes the class that satisfies the equality

$$[s]([\bar{S}]) = 2s - m.$$

We make the following conjecture.

Conjecture 1.7 Suppose that $K \subset Y$ is a null-homologous knot. Suppose that $b_1(Y) = 0$ and $m \in \mathbb{Z}$ with $m \geq 2g(K) + 1$. Let $A_s, B_s^-, \pi^{\pm,s}, W_m, I^\sharp(W_m, [s])$ be defined as above. Then for any $s \in [(m-1)/2], [(1-m)/2] \cap \mathbb{Z}$, there are commutative diagrams

$$\begin{array}{ccc} H(A_s) \hookrightarrow I^\sharp(-Y_{-m}(K)) & & H(A_s) \hookrightarrow I^\sharp(-Y_{-m}(K)) \\ \downarrow \pi^{-,s} & \downarrow I^\sharp(W_m, [s]) & \downarrow \pi^{+,s} \\ H(B_s) \xrightarrow{\cong} I^\sharp(-Y) & & H(B_s) \xrightarrow{\cong} I^\sharp(-Y) \end{array} \quad \begin{array}{ccc} H(A_s) \hookrightarrow I^\sharp(-Y_{-m}(K)) & & H(A_s) \hookrightarrow I^\sharp(-Y_{-m}(K)) \\ \downarrow \pi^{+,s} & \downarrow I^\sharp(W_m, [s+m]) & \downarrow \pi^{-,s} \\ H(B_s) \xrightarrow{\cong} I^\sharp(-Y) & & H(B_s) \xrightarrow{\cong} I^\sharp(-Y) \end{array}$$

Remark 1.8 In Heegaard Floer theory, the large surgery formula in [Ozsváth and Szabó 2004, Theorem 4.1] states that the homology of the bent complex A_s is isomorphic to the Heegaard Floer homology of $Y_{-m}(K)$ together with a spin^c structure specified by s . In instanton theory, we do not have the spin^c

structures in the construction of instanton Floer homology but a similar decomposition was introduced in [Li and Ye 2022; 2024]. However, involving the spin^c -type decomposition in the statement of Conjecture 1.7 would make the statement more complicated. So here we only write the top horizontal map in each commutative diagram as an inclusion.

The obstacle to obtaining a decomposition of the instanton cobordism map, in general, is one of the difficulties in exporting the original proof of the integral surgery formula in Heegaard Floer theory to the instanton setup. To overcome this problem, we need to work with a suitable setup for which some kind of decompositions do exist. A good candidate is sutured instanton theory. In sutured instanton theory, properly embedded surfaces induce \mathbb{Z} -gradings on the homology, and bypass maps relating different sutures are homogeneous with respect to such gradings. We have already used this setup to construct spin^c -like decompositions for the framed instanton homology of Dehn surgeries of knots, construct bent complexes in instanton theory, and establish a large surgery formula in our previous work [Li and Ye 2022; 2024; 2021].

In this paper, to prove the integral surgery formula, we further study the relations between different sutures on the knot complement and establish some new exact triangles and commutative diagrams that may be of independent interest. Then these new and old algebraic structures relating to different sutures enable us to apply the octahedral lemma to prove the desired integral surgery formula. It is worth mentioning that ultimately the whole proof in the current paper depends only on some most fundamental properties of Floer theory: the surgery exact triangle, the functoriality of the cobordism maps, and the adjunction inequality. This implies that the existence of the surgery formula is an inherent property of Floer theory.

The surgery formula developed in the current paper is a powerful tool to study the Dehn surgeries along knots. It enables us to do explicit computations in many cases, even when the ambient 3-manifold has a positive first Betti number. In a companion paper [Li and Ye 2025], we will use the surgery formula and the techniques developed in this paper to derive many new applications and computations. We sketch the results as follows.

- (1) We study the behavior of the integral surgery formula under the connected sum with a core knot in a lens space (whose complement is a solid torus) and then derive a rational surgery formula for framed instanton homology.
- (2) We study the 0-surgery on a knot K inside S^3 . We derive a formula computing the nonzero grading part of $I^\sharp(S_0(K))$ with respect to the grading induced by the Seifert surface.
- (3) We study nonzero integral surgeries on Borromean knots inside $Y = \#^{2g} S^1 \times S^2$, which gives nontrivial circle bundles over surfaces. In this case the bent complexes A_s and B_s^\pm can be computed directly and the maps π^\pm between them can be fixed with the help of the $\Lambda^* H_1(Y; \mathbb{C})$ -action. Moreover, we show $\dim_{\mathbb{C}} I^\sharp(Y) = \dim_{\mathbb{F}_2} \widehat{\text{HF}}(Y)$ for most Seifert fibered manifolds Y with nonzero orbifold degrees.

- (4) We study a family of alternating knots. Using an inductive argument by the oriented skein relation, we can describe their bent complexes explicitly and then the surgery formula applies routinely.
- (5) Using the same technique as above, we also study the nonzero integral surgery of twisted Whitehead doubles. The results for Whitehead doubles can also tell us the framed instanton Floer homology of the splicing of two knot complements in S^3 , where one knot is a twist knot, ie the Whitehead double of the unknot.
- (6) We study almost L -space knots, ie a non- L -space knot K such that there exists $n \in \mathbb{N}_+$ with $\dim I^\#(S_n^3(K)) = n + 2$; see [Baldwin and Sivek 2024] for the results in Heegaard Floer theory. We prove a genus-one almost L -space knot is either the figure-eight or the mirror of the knot 5_2 . We also show that almost L -space knots of genus at least 2 are fibered and strongly quasi-positive.

Organization The paper is organized as follows. In Section 2, we introduce basic setup, the notation in sutured instanton homology, and deal with the scalar ambiguity mentioned in Remark 2.4. We also present some algebraic lemmas including the octahedral lemma in the derived category that are used in latter sections. In Section 3, we present the strategy to prove the integral surgery formula. We first restate the integral surgery formula using sutured instanton homology, and explain how to apply the octahedral lemma to prove it. Then we explain how to translate the integral surgery formula from the language of sutured instanton theory to the language of bent complexes, which coincides with the discussions in the introduction. All the rest of the sections are devoted to prove the three exact triangles and three commutative diagrams that are involved in the octahedral lemma, ie equations (3-2) to (3-7). In Section 4, we study the relation between the (-1) -Dehn surgery map associated to a curve intersecting the suture twice and the two natural bypass maps associated to that curve. This helps us to prove equations (3-2) and (3-5). In Section 5, equations (3-3) and (3-6) and part of equation (3-4) are proved. The last two sections are devoted to prove equations (3-4) and (3-7), which is the most technical part of the paper. In Section 6 we prove some technical lemmas that are finally used in Section 7 to finish the proof.

Acknowledgements The authors thank John A Baldwin and Steven Sivek for the discussion on the proof of Proposition 4.1, and thank Zekun Chen and Linsheng Wang for the discussion on homological algebra. The authors thank Ciprian Manolescu and Jacob Rasmussen, and the anonymous referee for helpful comments. The authors also thank Sudipta Ghosh, Jianfeng Lin, Yi Xie and Ian Zemke for valuable discussions. Ye is also grateful to Yi Liu for inviting him to BICMR, Peking University when he was writing the early version of this paper.

2 Basic setup

2.1 Conventions

If it is not mentioned, all manifolds are smooth, oriented and connected. Homology groups and cohomology groups are with \mathbb{Z} coefficients. We write \mathbb{Z}_n for $\mathbb{Z}/n\mathbb{Z}$ and \mathbb{F}_2 for the field with two elements.

A knot $K \subset Y$ is called *null-homologous* if it represents the trivial homology class in $H_1(Y; \mathbb{Z})$, while it is called *rationally null-homologous* if it represents the trivial homology class in $H_1(Y; \mathbb{Q})$.

For any oriented 3-manifold M , we write $-M$ for the manifold obtained from M by reversing the orientation. For any surface S in M and any suture $\gamma \subset \partial M$, we write S and γ for the same surface and suture in $-M$, without reversing their orientations. For a knot K in a 3-manifold Y , we write $(-Y, K)$ for the induced knot in $-Y$ with induced orientation, called the *mirror knot* of K . The corresponding balanced sutured manifold is $(-Y \setminus N(K), -\gamma_K)$.

2.2 Sutured instanton homology

For any *balanced sutured manifold* (M, γ) [Juhász 2006, Definition 2.2], Kronheimer and Mrowka [2010, Section 7] constructed an isomorphism class of \mathbb{C} -vector spaces $\text{SHI}(M, \gamma)$. Later, Baldwin and Sivek [2015, Section 9] dealt with the naturality issue and constructed (untwisted and twisted versions of) projectively transitive systems related to $\text{SHI}(M, \gamma)$. We will use the twisted version, which we write as $\underline{\text{SHI}}(M, \gamma)$ and call *sutured instanton homology*.

In this paper, when considering maps between sutured instanton homology, we can regard them as linear maps between actual vector spaces, at the cost that equations (or commutative diagrams) between maps only hold up to a nonzero scalar due to the projectivity. A more detailed discussion on the projectivity can be found in the next subsection.

Moreover, there is a relative \mathbb{Z}_2 -grading on $\underline{\text{SHI}}(M, \gamma)$ obtained from the construction of sutured instanton homology, which we consider as a *homological grading* and use to take Euler characteristic.

Definition 2.1 Suppose K is a knot in a closed 3-manifold Y . Let $Y(1) := Y \setminus B^3$ and let δ be a simple closed curve on $\partial Y(1) \cong S^2$. Let $Y \setminus N(K)$ be the knot complement and let Γ_μ be two oppositely oriented meridians of K on $\partial(Y \setminus N(K)) \cong T^2$. Define

$$I^\sharp(Y) := \underline{\text{SHI}}(Y(1), \delta) \quad \text{and} \quad \underline{\text{KHI}}(Y, K) := \underline{\text{SHI}}(Y \setminus N(K), \Gamma_\mu).$$

Remark 2.2 By the naturality results, we should specify the places of the removing ball, the neighborhood of the knot, and the sutures to define $I^\sharp(Y)$ and $\underline{\text{KHI}}(Y, K)$. These data can be fixed by choosing a basepoint in Y or K . For simplicity, we omit those choices in the notation.

From now on, we will suppose $K \subset Y$ is a rationally null-homologous knot and fix some notation. Let μ be the meridian of K and pick a longitude λ (such that $\lambda \cdot \mu = 1$) to fix a framing of K . We will always assume $Y \setminus N(K)$ is irreducible, but many results still hold due to the following connected sum formula of sutured instanton homology [Li 2020, Remark 1.6]:

$$\underline{\text{SHI}}(Y' \# Y \setminus N(K), \gamma) \cong I^\sharp(Y') \otimes \underline{\text{SHI}}(Y \setminus N(K), \gamma).$$

Given coprime integers r and s , let $\Gamma_{r/s}$ be the suture on $\partial(Y \setminus N(K))$ consists of two oppositely oriented, simple closed curves of slope $-r/s$, with respect to the chosen framing (ie the homology of the curves are $\pm(-r\mu + s\lambda) \in H_1(\partial N(K))$). Pick S to be a minimal genus Seifert surface of K , with genus $g = g(S)$. Note that ∂S may have multiple components.

Convention For a fixed pair (λ, μ) as above, we write $p = \partial S \cdot \mu$ and $q = \partial S \cdot \lambda$. Note that when an orientation of the knot K is chosen, the orientation of S is induced by the knot. The orientation of μ is chosen such that $p > 0$, and the orientation of λ is chosen such that $\lambda \cdot \mu = 1$. Note that p is the order of $[K] \in H_1(Y)$, ie p is the minimal positive integer satisfying $p[K] = 0 \in H_1(Y)$. When K is null-homologous, we always choose the Seifert framing $\lambda = \partial S$. In such a case, we have $(p, q) = (1, 0)$.

Remark 2.3 The meanings of p and q above are different from our previous papers [Li and Ye 2022; 2021]. Before, we used $\hat{\mu}$ and $\hat{\lambda}$ to denote the meridian of the knot K and the preferred framing. In particular, the framing is fixed by [Li and Ye 2022, Definition 4.2]. In that case, we assume that ∂S is connected, and hence it is the same as the homological longitude (with the notation λ in previous papers, while we use μ to denote the homological meridian). Also, the numbers p and q in this paper should be q and q_0 in previous papers.

For simplicity, we use the bold symbol of the suture to represent the sutured instanton homology of the balanced sutured manifold with the reversed orientation:

$$\mathbf{\Gamma}_{r/s} := \underline{\text{SHI}}(-(Y \setminus N(K)), -\Gamma_{r/s}).$$

When $(r, s) = (\pm 1, 0)$, we write $\mathbf{\Gamma}_{r/s} = \mathbf{\Gamma}_\mu$. When $s = \pm 1$, we write $\mathbf{\Gamma}_n = \mathbf{\Gamma}_{n/1} = \mathbf{\Gamma}_{(-n)/(-1)}$. We also write $\mathbf{\Gamma}_\mu$ and $\mathbf{\Gamma}_n$ for the corresponding sutured instanton homologies.

Remark 2.4 Strictly speaking, the sutures corresponding to $(r, s) = (1, 0)$ and $(-1, 0)$ are not identical because the orientations are opposite. Since both sutures are on $\partial(Y \setminus N(K))$ of the same slope, they are isotopic. Moreover, we can choose a canonical isotopy by rotating the suture along the direction specified by the orientation of the knot. Due to discussion in Heegaard Floer theory [Sarkar 2015; Zemke 2015] and the conjectural relation between Heegaard Floer theory and instanton theory [Kronheimer and Mrowka 2010], it is expected that rotating the suture back to the original position induces a nontrivial isomorphism of the sutured instanton homology. So we pick the canonical isotopy to be the minimal rotation of the suture. Hence we can abuse notation and write $\mathbf{\Gamma}_\mu$ for both sutures. The same discussion also applies to the relation between $\mathbf{\Gamma}_{n/1}$ and $\mathbf{\Gamma}_{(-n)/(-1)}$.

We always assume that ∂S has minimal intersections with $\mathbf{\Gamma}_{r/s}$, ie $|\partial S \cap \gamma| = 2|rp - sq|$. When the intersection number $rp - sq$ is odd, then S induces a \mathbb{Z} -grading on $\mathbf{\Gamma}_{r/s}$. When $rp - sq$ is even, we need to perform either a positive stabilization or negative stabilization on S to induce a \mathbb{Z} -grading, and the

two gradings are related by an overall grading shift of 1. To get rid of stabilizations, we can equivalently regard that, in this case, the surface S induces a $(\mathbb{Z} + \frac{1}{2})$ -grading. We write the graded part of $\Gamma_{r/s}$ as

$$(\Gamma_{r/s}, i) := \underline{\text{SHI}}(-Y \setminus N(K)), -\Gamma_{r/s}, S, i$$

with $i \in \mathbb{Z}$ or $i \in \mathbb{Z} + \frac{1}{2}$, depending on the parity of the intersection number. From the construction of the grading in [Li 2021b], we have the following vanishing theorem due to the adjunction inequality.

Lemma 2.5 We have $(\Gamma_{r/s}, i) = 0$ when

$$|i| > \frac{|rp - sq| - \chi(S)}{2}.$$

Proof This follows from [Li and Ye 2022, Theorem 2.21(1)] (which is ultimately based on [Kronheimer and Mrowka 2010, Proposition 7.5]) and a direct computation in the new notation. \square

The bypass exact triangle for sutured instanton homology was introduced by Baldwin and Sivek in [2022b, Section 4]. In [Li and Ye 2022, Section 4.2], we applied the triangle to sutures on knot complements and computed the grading shifts. We restate the results in the notation introduced before.

Lemma 2.6 For any $n \in \mathbb{Z}$, there are two graded bypass exact triangles

$$\begin{array}{ccc}
 (\Gamma_n, i + \frac{1}{2}p) & \xrightarrow{\psi_{+,n+1}^n} & (\Gamma_{n+1}, i) \\
 & \swarrow \psi_{+,n}^\mu & \searrow \psi_{+,\mu}^{n+1} \\
 & & (\Gamma_\mu, i - \frac{1}{2}(np - q))
 \end{array}
 \qquad
 \begin{array}{ccc}
 (\Gamma_n, i - \frac{1}{2}p) & \xrightarrow{\psi_{-,n+1}^n} & (\Gamma_{n+1}, i) \\
 & \swarrow \psi_{-,n}^\mu & \searrow \psi_{-,\mu}^{n+1} \\
 & & (\Gamma_\mu, i + \frac{1}{2}(np - q))
 \end{array}$$

where the maps are homogeneous with respect to the homological \mathbb{Z}_2 -gradings.

Proof This is [Li and Ye 2022, Proposition 4.14] in the new notation. The idea of the proof can be found in [loc. cit., Lemma 3.18] (see also [loc. cit., Remark 3.19]). Roughly, we perturb the surface S by stabilizations so that its boundary is disjoint from the bypass arc. Then the grading shifts are obtained by counting the number of positive or negative stabilizations.

Unlike the setup in [loc. cit., Section 4], here K is not necessarily a dual knot of the Dehn surgery on a null-homologous knot, so we adopt the remarks in the beginning of [loc. cit., Section 5]. For example, when n is large enough so that $np - q \geq 0$ and [loc. cit., Proposition 4.14(1)] applies, we have

$$\hat{i}_{\max}^n = \frac{np - q - \chi(S)}{2}, \quad \hat{i}_{\min}^n = -\frac{np - q - \chi(S)}{2}, \quad \hat{i}_{\max}^\mu = \frac{p - \chi(S)}{2}, \quad \hat{i}_{\min}^\mu = -\frac{p - \chi(S)}{2},$$

where we omit $\lceil \cdot \rceil$ since we think about $\mathbb{Z} + \frac{1}{2}$ if necessary. Then

$$\hat{i}_{\min}^{n+1} - \hat{i}_{\min}^n = -\frac{p}{2} \quad \text{and} \quad \hat{i}_{\max}^{n+1} - \hat{i}_{\max}^\mu = \frac{np - q}{2}.$$

An easy way to understand the grading shift was described in [Li and Ye 2022, Remark 4.15]. Note that the grading shift of a map between two spaces equals half of the intersection number between ∂S and the curve corresponding to the third space up to the sign, while the sign depends on the choice of the sign in the bypass map. For example, we have $\partial S \cdot \mu = p$, so the grading shifts of $\psi_{\pm, n+1}^n$ are $\mp p/2$. \square

Remark 2.7 The reason to use balanced sutured manifolds with reversed orientation is because of the above bypass exact triangles.

Remark 2.8 If we do not mention gradings, the above triangles and the results in the rest of this subsection (except Corollary 2.9 and Lemma 2.19, since the statements involve gradings) also hold without the assumption that K is rationally null-homologous since the proofs only involve the neighborhood of $\partial(-Y \setminus N(K))$.

Corollary 2.9 For any sufficiently large integer n , the following properties for restrictions of maps hold:

- (1) The map $\psi_{+, n+1}^n|_{(\Gamma_n, i)}$ is an isomorphism when

$$i < \frac{np - q + \chi(S)}{2}.$$

- (2) The map $\psi_{-, n+1}^n|_{(\Gamma_n, i)}$ is an isomorphism when

$$i > -\frac{np - q + \chi(S)}{2}$$

- (3) For any grading i such that

$$-\frac{np - q + \chi(S)}{2} < i < \frac{(n - 2)p - q + \chi(S)}{2},$$

there is an isomorphism

$$(\psi_{+, n+1}^n)^{-1} \circ \psi_{-, n+1}^n : (\Gamma_n, i) \xrightarrow{\cong} (\Gamma_n, i + p).$$

- (4) The map $\psi_{-, 1-n}^{-n}|_{(\Gamma_{-n}, i)}$ is an isomorphism when

$$i < \frac{(n - 2)p + q + \chi(S)}{2}.$$

- (5) The map $\psi_{+, 1-n}^{-n}|_{(\Gamma_{-n}, i)}$ is an isomorphism when

$$i > -\frac{(n - 2)p + q + \chi(S)}{2}$$

- (6) For any grading i such that

$$-\frac{(n - 2)p + q + \chi(S)}{2} < i < \frac{(n - 4)p + q + \chi(S)}{2},$$

there is an isomorphism

$$(\psi_{+, 1-n}^{-n})^{-1} \circ \psi_{-, 1-n}^{-n} : (\Gamma_{-n}, i) \xrightarrow{\cong} (\Gamma_{-n}, i + p).$$

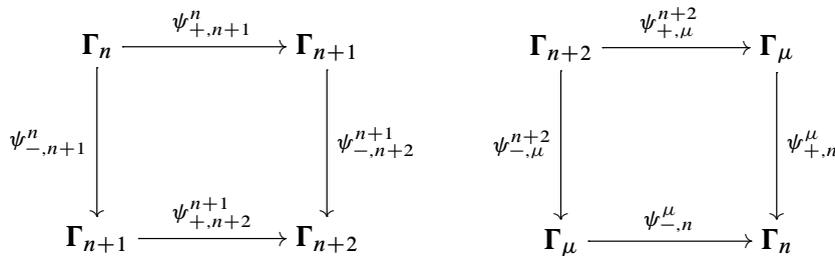
Proof It is a combination of Lemmas 2.5 and 2.6. \square

Definition 2.10 The maps in Lemma 2.6 are called *bypass maps*. The ones with subscripts + and − are called *positive* and *negative bypass maps*, respectively. We will use ± to denote one of the bypass maps. For any integer n and any positive integer k , define

$$\Psi_{\pm, n+k}^n := \psi_{\pm, n+k}^{n+k-1} \circ \cdots \circ \psi_{\pm, n+1}^n : \Gamma_n \rightarrow \Gamma_{n+k}.$$

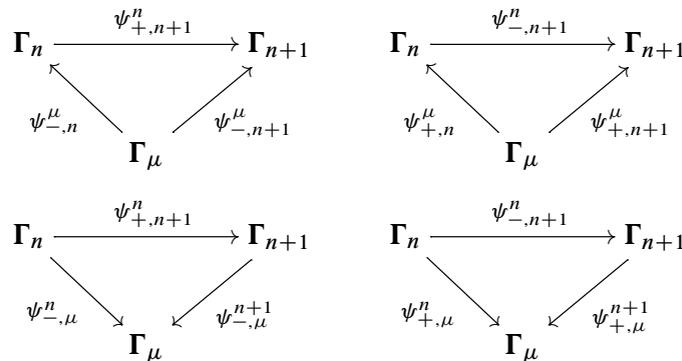
In [Li and Ye 2022, Section 4.4], we proved many commutative diagrams for bypass maps, which we restate as follows by notation introduced before.

Lemma 2.11 For any $n \in \mathbb{Z}$, we have the following commutative diagrams up to scalars:



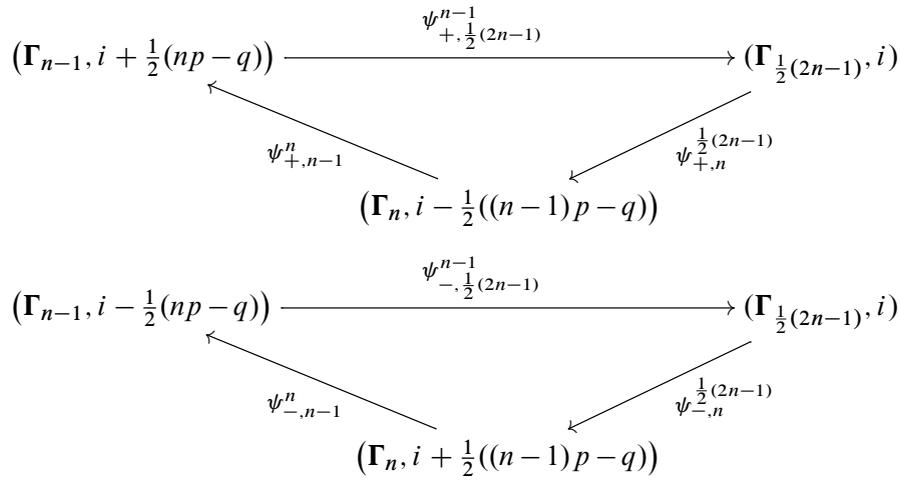
Proof The first diagram follows from [Li and Ye 2022, Lemma 4.33]. Note that the proof only used the functionality of the contact gluing map and did not depend on the assumption that K is rationally null-homologous. The second diagram is obtained from the first diagram by changing the choice of the framed knot. Explicitly, let K' be the dual knot corresponding to Γ_{n+1} . Let $\mu' = -(n + 1)\mu + \lambda$ denote its meridian. Then $\lambda' = -\mu$ is a framing of K' . Applying the first diagram to K' , we will obtain the second diagram for the original K . Note that the sign of the bypass map may switch when we regard it as the bypass map for the original knot. That is the reason for the signs in the second diagram. This can be double-checked by keeping track of the grading shifts. \square

Lemma 2.12 For any $n \in \mathbb{Z}$, we have the following commutative diagrams up to scalars:

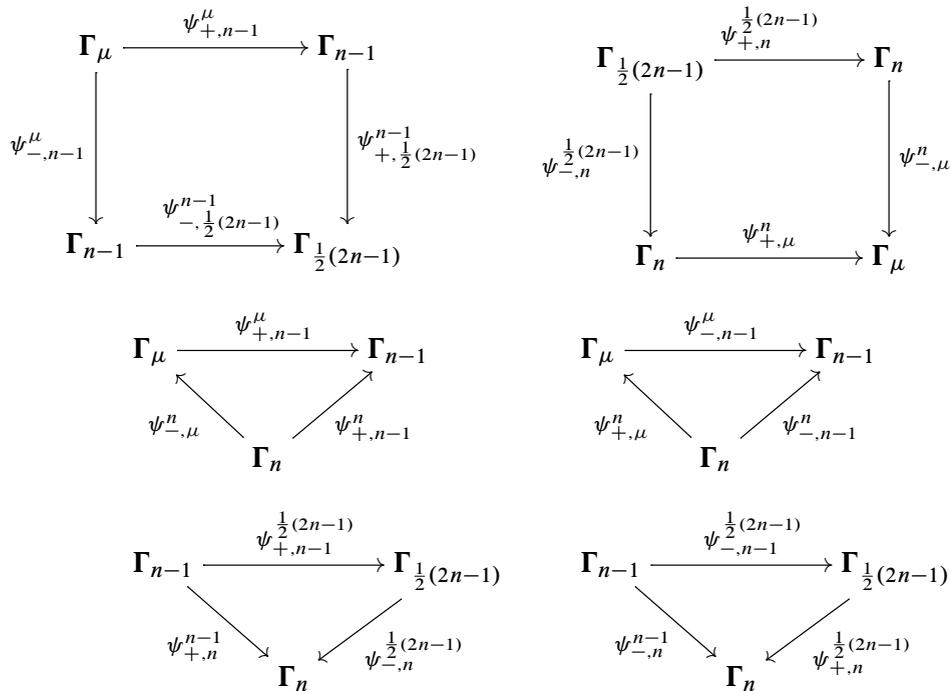


There are more bypass triangles involving more complicated sutures, which are obtained from changing the choice of the framed knot as in the proof of Lemma 2.11.

Lemma 2.13 For a knot $K \subset Y$ and $n \in \mathbb{Z}$, there are two graded bypass exact triangles:



Lemma 2.14 For a knot $K \subset Y$ and $n \in \mathbb{Z}$, there are commutative diagrams up to scalars



Remark 2.15 The choices of positive and negative bypass maps in Lemma 2.14 seem to be different from Lemmas 2.11 and 2.12. But indeed they are the same up to changing the framed knot. In particular, the grading shifts match. Similar to the second diagram in Lemma 2.11, we always use the notation of the bypass maps for the original knot, while the signs may change if the maps are regarded as the bypass maps of the dual knot.

Suppose α is a connected nonseparating simple closed curve on $\partial(Y \setminus N(K))$. We can push α into the interior of $Y \setminus N(K)$. For any fixed suture on $\partial(Y \setminus N(K))$ and a closure of the sutured manifold, the push-off of α is inside the closure, which is a closed 3-manifold. We can then take the framing on α induced by the surface $\partial(Y \setminus N(K))$ and there is an exact triangle associated to the instanton Floer homology of the (-1) -, 0 - and ∞ -surgeries along the push-off of α . Since the push-off of α is disjoint from $\partial(Y \setminus N(K))$, the exact triangle descends to one between corresponding sutured instanton Floer homologies.

According to [Baldwin and Sivek 2016a, Section 4], when α intersects the suture at two points, the 0 -surgery along the push-off of α (with framing induced by $\partial(Y \setminus N(K))$) corresponds to a 2-handle attachment along α . Note that attaching a 2-handle along $\alpha \subset \partial(Y \setminus N(K))$ will change the 3-manifold from $Y \setminus N(K)$ to $Y_\alpha(K) \setminus B^3$, where $Y_\alpha(K)$ is the Dehn surgery along K with slope specified by α . We write

$$Y_{r/s} := I^\sharp(-Y_{-r/s}(K)),$$

and in particular

$$Y_n := I^\sharp(-Y_{-n}(K)) \quad \text{and} \quad Y := I^\sharp(-Y).$$

Lemma 2.16 [Li and Ye 2022, Lemma 3.21] *For any $n \in \mathbb{Z}$, we have the following exact triangles:*

$$\begin{array}{ccc}
 \Gamma_n & \xrightarrow{H_n} & \Gamma_{n+1} \\
 & \swarrow G_n & \searrow F_{n+1} \\
 & Y &
 \end{array}
 \qquad
 \begin{array}{ccc}
 \Gamma_\mu & \xrightarrow{A_{n-1}} & \Gamma_{n-1} \\
 & \swarrow C_n & \searrow B_{n-1} \\
 & Y_n &
 \end{array}$$

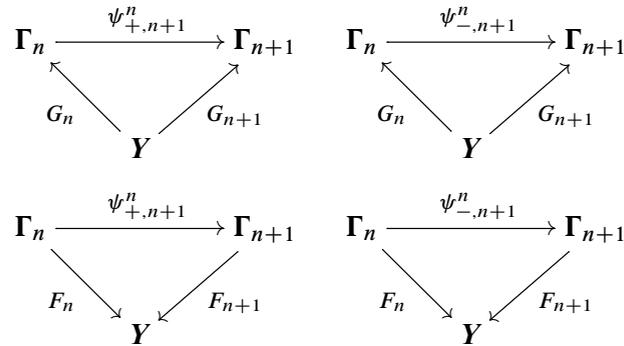
Proof To obtain the first exact triangle, we can take the sutured manifold $(-(Y \setminus N(K)), -\Gamma_n)$, and take a meridian $\alpha \subset \partial(Y \setminus N(K))$. As explained before the lemma, there is a surgery exact triangle associated to the sutured instanton Floer homology of the three sutured manifolds obtained by taking (-1) -, 0 -, and ∞ -surgeries [Scaduto 2015, Theorem 2.1]; see also [Floer 1990] for the original construction and [Baldwin and Sivek 2022b, Proof of Theorem 1.21, especially (16)–(19)] for the resolution of the subtlety of the bundle data.

The ∞ -surgery will keep the manifold $(-(Y \setminus N(K)), -\Gamma_n)$. The (-1) -surgery changes the framing and hence we obtain $(-(Y \setminus N(K)), -\Gamma_{n+1})$. The 0 -surgery, as discussed above, gives rise to the manifold $Y_\alpha(K) \setminus B^3$, which is $Y \setminus B^3$ since α is the meridian. Hence we obtain the expected triangle. The second exact triangle in the statement of the lemma is obtained similarly by taking α to be a curve on $\partial(Y \setminus N(K))$ having slope $-n$ instead of a meridian. □

Remark 2.17 From [Baldwin and Sivek 2016a, Section 4], we know the 0 -surgery corresponds to a 2-handle attachment and a 1-handle attachment. Hence Y in the above lemma is indeed $\text{KHI}(-Y, U)$, where U is the unknot in $-Y$ bounding an embedded disk. By [loc. cit., Section 4], a 1-handle attachment does not change the closure of the balanced sutured manifold, and then there is a canonical identification between $\text{KHI}(-Y, U)$ and $I^\sharp(-Y)$. Hence we can abuse the notation. The same discussion also applies to Y_n .

Furthermore, we proved the following properties in [Li and Ye 2022]. Note that the assumption that K is the dual knot of a null-homologous knot in that paper is inessential by remarks in the beginning of [Li and Ye 2022, Section 5]. The inequalities of the gradings are from Corollary 2.9.

Lemma 2.18 [Li and Ye 2022, Lemmas 3.21 and 4.9] *For any $n \in \mathbb{Z}$, we have the following commutative diagrams up to scalars:*



Lemma 2.19 [Li and Ye 2022, Lemma 4.17, Proposition 4.26, Lemma 4.29 and Proposition 4.32] *Let F_n and G_n be defined as in Lemma 2.16. Then for any sufficiently large integer n , we have the following properties:*

- (1) *The map G_{n-1} is zero and F_n is surjective. Moreover, for any grading*

$$-\frac{np - q + \chi(S)}{2} < i_0 < \frac{np - q + \chi(S)}{2} - p + 1,$$

the restriction of the map

$$F_n : \bigoplus_{i=0}^{p-1} (\Gamma_n, i_0 + i) \rightarrow Y$$

is an isomorphism.

- (2) *The map F_{-n+1} is zero and G_{-n} is injective. Moreover, for any grading*

$$-\frac{(n-2)p + q + \chi(S)}{2} < i_0 < \frac{(n-2)p + q + \chi(S)}{2} - p + 1,$$

the map

$$\text{Proj} \circ G_{-n} : Y \rightarrow \bigoplus_{i=0}^{p-1} (\Gamma_{-n}, i_0 + i)$$

is an isomorphism, where

$$\text{Proj} : \Gamma_{-n} \rightarrow \bigoplus_{i=0}^{p-1} (\Gamma_{-n}, i_0 + i)$$

is the projection.

The following lemma is a special case of Proposition 4.1, which we will prove later.

Lemma 2.20 *For any $n \in \mathbb{Z}$, let the maps H_n and $\psi_{\pm, n+1}^n$ be defined as in Lemmas 2.16 and 2.6, respectively. Then there exist $c_1, c_2 \in \mathbb{C} \setminus \{0\}$ such that*

$$H_n = c_1 \psi_{+, n+1}^n + c_2 \psi_{-, n+1}^n.$$

2.3 Fixing the scalars

By construction, sutured instanton homology forms a projectively transitive system, which means all the spaces and maps between spaces are well-defined only up to nonzero scalars. When the balanced sutured manifold is obtained from a framed knot as in the last subsection, we can make some canonical choices to reduce the projective ambiguities.

Suppose $K \subset Y$ is a framed knot with the meridian μ and the framing λ . Fix a knot complement $Y \setminus N(K)$ and the suture Γ_μ . We fix a special choice of a (marked odd) closure of $(Y \setminus N(K), \Gamma_\mu)$ following the construction in [Kronheimer and Mrowka 2010, formula (18)].

Let F be a closed, oriented, connected surface of genus at least 2. Suppose c_0 is a nonseparating curve in F . Let $c = \text{pt} \times c_0 \subset S^1 \times F$ and let (μ_c, λ_c) be the meridian and the longitude of c (the latter comes from the surface framing). Let

$$(2-1) \quad (\bar{Y}_\mu, R) = (S^1 \times F \setminus N(c) \cup_{(\mu_c, \lambda_c) \sim (\lambda, \mu)} Y \setminus N(K), \text{pt}' \times F),$$

where pt' is a point different from pt . We can pick $\alpha = S^1 \times \text{pt}''$ and $\eta \subset R$ be a curve intersecting $\text{pt}' \times c_0$ once. Since $\alpha \cdot R = 1$ and $\eta \cdot R = 0$, the pair $(\bar{Y}_\mu, \alpha \cup \eta)$ defines an instanton Floer homology in the setting of [Kronheimer and Mrowka 2010, Section 7.1]. Moreover, $\mathcal{D}_\mu = (Y, R, \eta, \alpha)$ forms a marked odd closure in [Baldwin and Sivek 2015, Definition 9.2], which was used in the naturality result [loc. cit., Theorem 9.17]. The reason for $g(F) \geq 2$ is to apply the naturality result; cf [loc. cit., Remark 9.4].

Similarly, for $(Y \setminus N(K), \Gamma_n)$ and $(Y \setminus N(K), \Gamma_{(2n-1)/2})$, we fix closures \mathcal{D}_n and $\mathcal{D}_{(2n-1)/2}$ as in (2-1), except replacing the gluing map $(\mu_c, \lambda_c) \sim (\lambda, \mu)$ by

$$(\mu_c, \lambda_c) \sim (-\mu, -n\mu + \lambda) \quad \text{and} \quad (\mu_c, \lambda_c) \sim (-n\mu + \lambda, (1 - 2n)\mu + 2\lambda),$$

respectively.

For the sutured manifold $(Y \setminus B^3, \delta)$, we regard it as $(Y \setminus N(U), \Gamma_{\mu, U})$ by Remark 2.17, where U is the unknot and $\Gamma_{\mu, U}$ is meridian suture on the unknot complement. Then we apply the above construction to obtain a special closure of $(Y \setminus B^3, \delta)$. We reverse the orientations of the chosen closures when the orientations of the sutured manifolds are reversed. Note that we do not choose canonical closures for $(Y_n(K) \setminus B^3, \delta)$ since we only care about the dimension of its framed instanton homology.

After fixing the choices of closures, we can view $\mathbf{\Gamma}_n$ and \mathbf{Y} as actual vector spaces, and then the elements in them are well-defined. Strictly speaking, we also need to choose extra data such as the metric and the

perturbation on the closure to define the instanton Floer homology of the closure, but different choices of metrics and perturbations now induce a transitive system of vector spaces, from which we can construct an actual vector space. So we omit the discussion on those extra data.

The construction of bypass maps and surgery maps may not be realized as cobordism maps between the chosen closures, but the construction of the projectively transitive system (cf [Baldwin and Sivek 2015, Definition 9.18]) guarantees the existence of such maps up to scalars. Now we make (noncanonical) choices of the maps to get rid of the scalar ambiguities in the commutative diagrams mentioned in the last subsection.

We first assume that $I^\sharp(Y) \neq 0$. When $I^\sharp(Y) = 0$, the first exact triangle in Lemma 2.16 is trivial for any n and hence the maps F_n and G_n that play an important role in later sections are both zero. This makes fixing the scalars a somewhat straightforward job: we just need to fix the scalars for the bypass maps. Also, it is worth mentioning that, the Euler characteristic result in [Scaduto 2015, Corollary 1.4] implies that $I^\sharp(Y) \neq 0$ for any rational homology sphere Y and, up to author's knowledge, there is no known closed oriented 3-manifold Y with $I^\sharp(Y) = 0$.

To help us fixing the scalars, suppose the maps F_n and G_n are defined as in the proof of Lemma 2.16, and we define

$$(2-2) \quad n_G = \min\{n \in \mathbb{Z} \mid G_n = 0\} \quad \text{and} \quad n_F = \max\{n \in \mathbb{Z} \mid F_n = 0\}.$$

We have the following basic properties for these indices.

Lemma 2.21 *Assume $I^\sharp(Y) \neq 0$. Suppose n_G and n_F are defined as in equation (2-2). Then we have*

$$-\infty < n_F \leq n_G < \infty.$$

Moreover, we have $G_n = 0$ if and only if $n \geq n_G$, and $F_n = 0$ if and only if $n \leq n_F$.

Proof The fact that $-\infty < n_F < \infty$ and $-\infty < n_G < \infty$ follow from Lemma 2.19 and the fact that they fits into an exact triangle as in Lemma 2.16. Next, the commutative diagrams in Lemma 2.18 imply that $G_{n+1} = 0$ whenever $G_n = 0$, and thus we know $G_n = 0$ if and only if $n \geq n_G$. The argument for F_n is similar. Finally, by definition we know $G_{n_G} = 0$ and hence from the exact triangle we know that

$$\text{Im } F_{n+1} = \ker G_{n_G} = I^\sharp(Y) \neq 0.$$

Hence we conclude that $n_F \leq n_G$. □

By Lemma 2.19, we can pick a sufficiently large integer n_0 such that $-n_0 < n_F \leq n_G$. Pick arbitrary representatives of the maps

$$G_{-n_0}, \quad \psi_{+,\mu}^{-n_0}, \quad \psi_{-,\mu}^{-n_0}, \quad \psi_{+,-n_0}^\mu, \quad \psi_{-,-n_0}^\mu,$$

and also pick arbitrary representatives of the maps

$$\psi_{+,n+1}^n, \psi_{+,(2n-1)/2}^{n-1} \quad \text{for all } n \in \mathbb{Z}.$$

Now we explain how to fix the scalars for maps with $n \geq n_0$. Note that we have already chosen a representative for $\psi_{+,-n_0+1}^{-n_0}$ and G_{-n_0} . From Lemma 2.18, we have

$$G_{-n_0+1} \doteq \psi_{+,-n_0+1}^{-n_0} \circ G_{-n_0},$$

where \doteq means commutative up to a nonzero scalar. We can choose a representative of G_{-n_0+1} to obtain an equality

$$G_{-n_0+1} = \psi_{+,-n_0+1}^{-n_0} \circ G_{-n_0}.$$

We then choose a representative of $\psi_{-,-n_0+1}^{-n_0}$ so that

$$G_{-n_0+1} = \psi_{-,-n_0+1}^{-n_0} \circ G_{-n_0}.$$

Next, we pick representatives of the maps $\psi_{-,n+1}^n$ inductively, with the base case $n = -n_0$ constructed above, so that

$$\psi_{-,n+1}^n \circ \psi_{+,n}^{n-1} = \psi_{+,n+1}^n \circ \psi_{-,n}^{n-1} \quad \text{for all } n \geq -n_0 + 1.$$

If the compositions happen to be zero, we could pick an arbitrary representative since the diagram will be trivially satisfied. We will discuss the ambiguity arising from the possibility that $\psi_{+,n+1}^n \circ \psi_{-,n}^{n-1} = 0$ more carefully later.

Similarly, we pick the maps $\psi_{+,\mu}^n$ inductively to satisfy the commutative diagram

$$\psi_{+,\mu}^n \circ \psi_{-,n}^{n-1} = \psi_{+,\mu}^{n-1}.$$

We can choose representatives of $\psi_{-, \mu}^n$, $\psi_{+,n}^\mu$ and $\psi_{-,n}^\mu$ in a similar manner.

Furthermore, the representatives of

$$\psi_{-,n}^{(2n-1)/2}, \quad \psi_{+,n-1}^n, \quad \psi_{-,n-1}^n, \quad \psi_{-, (2n-1)/2}^{n-1}, \quad \psi_{+,n}^{(2n-1)/2}$$

can be chosen according to Lemma 2.14. As mentioned in Remark 2.15, we always use the notation of the bypass maps for the original knot even though we consider some dual knots in the proofs. Hence here we first fix the knot $K \subset Y$ and then fix the representatives, while we do not fix the representatives by any commutative diagrams for the dual knot in the proofs.

Next, we deal with maps G_n and F_n in Lemma 2.16. We choose representatives of the maps G_n inductively so that

$$G_{n+1} = \psi_{+,n+1}^n \circ G_n \quad \text{for all } n \geq -n_0.$$

We pick arbitrary representatives of the map F_{n_F+1} and then pick F_n inductively so that

$$F_{n+1} \circ \psi_{+,n+1}^n = F_n \quad \text{for all } n \geq n_F + 1.$$

We can then use induction to prove the following two equalities.

- $G_{n+1} = \psi_{-,n+1}^n \circ G_n$ for all $n \geq -n_0$.
- $F_{n+1} \circ \psi_{-,n+1}^n = c \cdot F_n$ for a nonzero scalar c that is independent of n with $n \geq n_F + 1$.

We verify the equality for G first. The base case $n = n_0$ is by construction. Assuming we have already established the equality for n , from Lemmas 2.11 and 2.18, we have

$$\begin{aligned} G_{n+2} &= \psi_{+,n+2}^{n+1} \circ G_{n+1} \\ &= \psi_{+,n+2}^{n+1} \circ \psi_{-,n+1}^n \circ G_n \quad \text{by inductive hypothesis} \\ &= \psi_{-,n+2}^{n+1} \circ \psi_{+,n+1}^n \circ G_n \quad \text{by Lemma 2.11} \\ &= \psi_{-,n+2}^{n+1} \circ G_{n+1}. \end{aligned}$$

The argument for F_n is similar, once we take $c \neq 0$ to be the complex number such that

$$(2-3) \quad F_{n_F+2} \circ \psi_{-,n_F+2}^{n_F+1} = c \cdot F_{n_F+1}.$$

The remaining issues are summarized as follows:

- (i) When choosing representatives of $\psi_{-,n+1}^n$, we use the commutative diagram

$$\psi_{-,n+1}^n \circ \psi_{+,n}^{n-1} \doteq \psi_{+,n+1}^n \circ \psi_{-,n}^{n-1}.$$

However, when $\psi_{+,n+1}^n \circ \psi_{-,n}^{n-1} = 0$, there is no unique choice of $\psi_{-,n+1}^n$ and one might worry that different choices of $\psi_{-,n+1}^n$ may affect the commutative diagrams in Lemma 2.18.

- (ii) By Proposition 4.1, we can assume that the map H_n in the exact triangle in Lemma 2.16 has the form

$$H_n = \psi_{+,n+1}^n - c_n \cdot \psi_{-,n+1}^n.$$

We want to pin down the values of c_n .

- (iii) We want to get rid of the scalar c in equation (2-3).

We treat these issues in several different cases.

Case 1 ($n_F + 3 \leq n_G$) In this case, we know that

$$\psi_{+,n+1}^n \circ \psi_{-,n}^{n-1} \neq 0$$

for any integer n . Indeed, if $n + 1 < n_G$, we know from Lemma 2.18 that

$$\psi_{+,n+1}^n \circ \psi_{-,n}^{n-1} \circ G_{n-1} = G_{n+1} \neq 0.$$

If $n + 1 \geq n_G \geq n_F + 3$, instead of the above equation involving G , we have

$$F_{n+1} \circ \psi_{+,n+1}^n \circ \psi_{-,n}^{n-1} = c \cdot F_{n-1} \neq 0.$$

This implies that inductively we can fix a unique representative of $\psi_{-,n+1}^n$ for all $n \geq -n_0 + 1$.

For any integer n with $-n_0 \leq n < n_G - 1$, we have $G_{n+1} \neq 0$. Then we take an element $\alpha \in Y$ so that $G_{n+1}(\alpha) \neq 0$. Then we can solve the scalar c_n as follows:

$$\begin{aligned} 0 &= H_n \circ G_n(\alpha) = (\psi_{+,n+1}^n - c_n \cdot \psi_{-,n+1}^n) \circ G_n(\alpha) \\ &= \psi_{+,n+1}^n \circ G_n(\alpha) - c_n \cdot \psi_{-,n+1}^n \circ G_n(\alpha) \\ &= (1 - c_n) \cdot G_{n+1}(\alpha). \end{aligned}$$

Hence we conclude that $c_n = 1$. In particular, we can take $n = n_F + 1 < n_G - 1$. Note that $F_n \neq 0$ so we can take $x \in \Gamma_n$ so that $F_n(x) \neq 0$. Then we have

$$\begin{aligned} 0 &= F_{n+1} \circ H_n(x) = F_{n+1} \circ (\psi_{+,n+1}^n - \psi_{-,n+1}^n)(x) \\ &= F_{n+1} \circ \psi_{+,n+1}^n(x) - F_{n+1} \circ \psi_{-,n+1}^n(x) \\ &= (1 - c) \cdot F_n(x). \end{aligned}$$

This implies that $c = 1$ as well. Now for any $n \geq n_G - 1 \geq n_F + 2$, we can take $x \in \Gamma_n$ such that $F_n(x) \neq 0$. we have

$$0 = F_{n+1} \circ H_n(x) = F_{n+1} \circ (\psi_{+,n+1}^n - c_n \cdot \psi_{-,n+1}^n)(x) = (1 - c_n) \cdot F_n(x).$$

Hence we conclude that $c_n = 1$. In summary, in Case 1, we have the following:

- We can fix a unique representative of $\psi_{-,n+1}^n$ for any $n \geq -n_0$.
- We have $c_n = 1$ for any $n \geq -n_0$.
- We have $c = 1$.

Case 2 ($n_G = n_F + 2$) Some arguments in Case 1 still apply. We summarize as follows:

- For $n < n_G - 1$, we have $G_{n+1} \neq 0$ so there is a unique choice of $\psi_{-,n+1}^n$.
- For $n \geq n_G = n_F + 2$, we have $F_{n-1} \neq 0$ so again there is a unique choice of $\psi_{-,n+1}^n$.
- For $n < n_G - 1$, we have $G_{n+1} \neq 0$ so $c_n = 1$ in the expression of H_n .
- For $n \geq n_G - 1 = n_F + 1$, we have $F_n \neq 0$ so $c_n = c^{-1}$.

To resolve the issue (i), the only nonfixed index is $n = n_F + 1 = n_G - 1$. In the case that

$$\psi_{+,n+1}^n \circ \psi_{-,n}^{n-1} \neq 0,$$

there is a unique choice of $\psi_{-,n+1}^n$ so that

$$\psi_{-,n+1}^n \circ \psi_{+,n}^{n-1} = \psi_{+,n+1}^n \circ \psi_{-,n}^{n-1}.$$

Otherwise, we just fix any representative of $\psi_{-,n+1}^n$.

To resolve issues (ii) and (iii), we rescale the maps F_n according to the grading on Γ_n . To do this, for an integer n and a grading i , define

$$j(n, i) = \left\lfloor \frac{i}{p} - \frac{n}{2} \right\rfloor.$$

It is straightforward that

$$j\left(n + 1, i + \frac{1}{2}p\right) = j(n, i) \quad \text{and} \quad j\left(n + 1, i - \frac{1}{2}p\right) = j(n, i) - 1.$$

Hence we define

$$(2-4) \quad \tilde{F}_n = \sum_i c^{j(n,i)} \cdot F_n \circ \text{Proj}_n^i,$$

where the map

$$\text{Proj}_n^i : \Gamma_n \rightarrow (\Gamma_n, i)$$

is the projection. Equivalently, if we have an element $x \in (\Gamma_n, i)$, we take

$$\tilde{F}_n(x) = c^{j(n,i)} \cdot F_n(x).$$

We then need to verify the following two equalities for \tilde{F} :

- $\tilde{F}_{n+1} \circ \psi_{+,n+1}^n = \tilde{F}_n.$
- $\tilde{F}_{n+1} \circ \psi_{-,n+1}^n = \tilde{F}_n.$

For the first one, note that $\psi_{+,n+1}^n$ increases the grading by $p/2$ from Lemma 2.6. So assuming $x \in (\Gamma_n, i)$,

$$\tilde{F}_{n+1} \circ \psi_{+,n+1}^n(x) = c^{j(n+1, i+\frac{1}{2}p)} \cdot F_{n+1}(x) \circ \psi_{+,n+1}^n = c^{j(n,i)} \cdot F_n(x) = \tilde{F}_n(x).$$

On the other hand, the map $\psi_{-,n+1}^n$ decreases the grading by $p/2$ from Lemma 2.6, and so for $x \in (\Gamma_n, i)$,

$$\tilde{F}_{n+1} \circ \psi_{-,n+1}^n(x) = c^{j(n+1, i-\frac{1}{2}p)} \cdot F_{n+1}(x) \circ \psi_{-,n+1}^n = c^{j(n,i)-1} \cdot c \cdot F_n(x) = \tilde{F}_n(x).$$

Hence we can use \tilde{F}_n instead of F_n to get rid of the scalar c . However, changing F_n to \tilde{F}_n , we might break the exact triangle in Lemma 2.16. Hence we need to alter the definition of H_n as well. We define

$$(2-5) \quad \tilde{H}_n = \psi_{+,n+1}^n - \psi_{-,n+1}^n,$$

and it remains to establish the exact triangle

$$\begin{array}{ccc} \Gamma_n & \xrightarrow{\tilde{H}_n} & \Gamma_{n+1} \\ & \swarrow G_n & \searrow \tilde{F}_{n+1} \\ & Y & \end{array}$$

When $n < n_G - 2 = n_F$, the triangle automatically holds. Indeed, for such n , we have $c_n = 1$ so $\tilde{H}_n = H_n$; the fact that $F_{n+1} = 0$ implies that $\tilde{F}_{n+1} = 0$, so the new exact triangle is exactly the original one in Lemma 2.16.

When $n = n_F$, we still have $c_n = 1$ and hence $\tilde{H}_n = H_n$. So the exactness at Γ_n is from Lemma 2.16. By construction we have $\text{Im } \tilde{F}_{n+1} = \text{Im } F_{n+1}$ for any n so we conclude the exactness at Y . This also implies that

$$\dim \ker \tilde{F}_{n+1} = \dim \ker F_{n+1} = \dim \text{Im } H_n.$$

Hence to show the exactness at Γ_{n+1} , it remains to show that

$$\text{Im } H_n \subset \ker \tilde{F}_{n+1}.$$

For any $x \in \Gamma_n$, we have

$$\begin{aligned} \tilde{F}_{n+1} \circ H_n(x) &= \tilde{F}_{n+1} \circ (\psi_{+,n+1}^n - \psi_{-,n+1}^n)(x) = \tilde{F}_{n+1} \circ \psi_{+,n+1}^n(x) - \tilde{F}_{n+1} \circ \psi_{-,n+1}^n(x) \\ &= \tilde{F}_n(x) - \tilde{F}_n(x) = 0. \end{aligned}$$

Hence we are done.

Finally, we verify the exact triangle for $n \geq n_F + 1$. Define a homomorphism

$$\iota_n: \Gamma_n \rightarrow \Gamma_n \quad \text{by} \quad \iota_n(x) = \sum_i c^{-j(n,i)} \text{Proj}_n^i(x).$$

It is clear that ι is an isomorphism as its inverse is

$$\iota_n^{-1}(x) = \sum_i c^{j(n,i)} \text{Proj}_n^i(x),$$

and from the construction of \tilde{F} in equation (2-4), we have

$$\iota_{n+1}(\ker \tilde{F}_{n+1}) = \ker F_{n+1}.$$

From the fact that $H_n = \psi_{+,n+1}^n - c^{-1} \cdot \psi_{-,n+1}^n$ (we have $c_n = c^{-1}$), $\tilde{H}_n = \psi_{+,n+1}^n - \psi_{-,n+1}^n$, and that $\psi_{\pm,n+1}^n$ shifts the grading by $\mp p/2$, we conclude that

$$\iota_{n+1}(\text{Im } \tilde{H}_n) = \text{Im } H_n.$$

As a result, we conclude the exactness at Γ_{n+1} . The exactness at Y holds as above, since we still have

$$\text{Im } \tilde{F}_{n+1} = \text{Im } F_{n+1} = \ker G_n.$$

Dimension counting similar to the above argument then implies that

$$\dim \ker \tilde{H}_n = \dim \text{Im } G_n.$$

We then verify that

$$\text{Im } G_n \subset \ker \tilde{H}_n.$$

For any $\alpha \in Y$, we have

$$\begin{aligned} \tilde{H}_n \circ G_n(\alpha) &= (\psi_{+,n+1}^n - \psi_{-,n+1}^n) \circ G_n(\alpha) = \psi_{+,n+1}^n \circ G_n(\alpha) - \psi_{-,n+1}^n \circ G_n(\alpha) \\ &= G_{n+1}(\alpha) - G_{n+1}(\alpha) \\ &= 0. \end{aligned}$$

In summary, in Case 2, we do the following (extra things):

- We choose representatives of $\psi_{-,n+1}^n$ for all $n \geq -n_0 + 1$ so that

$$\psi_{-,n+1}^n \circ \psi_{+,n}^{n-1} = \psi_{+,n+1}^n \circ \psi_{-,n}^{n-1}.$$

- We define new maps \tilde{F}_n for all n so that

$$(2-6) \quad \tilde{F}_n = \tilde{F}_{n+1} \circ \psi_{\pm,n+1}^n.$$

- We define new maps \tilde{H}_n for all n so that we have the exact triangle

$$(2-7) \quad \begin{array}{ccc} \Gamma_n & \xrightarrow{\tilde{H}_n} & \Gamma_{n+1} \\ & \swarrow G_n & \searrow \tilde{F}_{n+1} \\ & & Y \end{array}$$

Case 3 ($n_G = n_F + 1$ or $n_G = n_F$) The situation and argument are similar to those in Case 2. We summarize the differences here:

- In Case 3, the composition $\psi_{+,n+1}^n \circ \psi_{-,n}^{n-1}$ could only be zero when $n_G - 1 \leq n \leq n_F + 1$. If the composition is indeed 0, we choose an arbitrary representative of the map $\psi_{-,n+1}^n$.
- We still define the maps \tilde{F}_n as in equation (2-4) and the maps \tilde{H}_n as in equation (2-5), and can verify equation (2-6) and the exact triangle (2-7) as in Case 2.

From Lemma 2.21, we must have $n_G \geq n_F$, so the above three cases cover all situations.

Finally, we could extend the choice of representatives for all relevant maps for the indices $n < -n_0$. Note that when $n < -n_0$, we have that G_n is injective and $F_n = 0$; hence we do not need to worry about the issues (i), (ii) and (iii).

Convention From now on, we write the maps \tilde{H}_n and \tilde{F}_n simply as H_n and F_n , respectively. From the above discussion, when $K \subset Y$ is a fixed rationally null-homologous knot, we can assume the first commutative diagram in Lemma 2.11, and all commutative diagrams in Lemmas 2.12, 2.14 and 2.18 hold without introducing scalars.

2.4 Algebraic lemmas

In this subsection, we introduce some lemmas in homological algebra. All graded vector spaces in this subsection are finite-dimensional and over \mathbb{C} , and all maps are complex linear maps. For convenience, we will switch freely between long exact sequences and exact triangles.

From Section 2.2, we know the sutured instanton homology is usually $\mathbb{Z} \oplus \mathbb{Z}_2$ -graded, where we regard the \mathbb{Z}_2 -grading as a homological grading. Many results in this subsection come from properties of the derived category of vector spaces over \mathbb{C} , for which people usually consider cochain complexes. However, for a \mathbb{Z}_2 -graded space there is no difference between the chain complex and the cochain complex. Hence by saying a *complex* we mean a \mathbb{Z}_2 -graded (co)chain complex, though all results apply to \mathbb{Z} -graded cochain complexes verbatim.

For a complex C and an integer n , we write C^n for its grading n part (under the natural map $\mathbb{Z} \rightarrow \mathbb{Z}_2$). With this notation, we suppose the differential d_C on C sends C^n to C^{n+1} . For any integer k , we write $C\{k\}$ for the complex obtained from C by the grading shift $C\{k\}^n = C^{n+k}$. We write $H(C, d_C)$ or $H(C)$ for the homology of a complex C with differential d_C . A \mathbb{Z}_2 -graded vector space is regarded as a complex with the trivial differential.

For a chain map $f: C \rightarrow D$, we write $\text{Cone}(f)$ for the *mapping cone* of f , ie the complex consisting of the space $D \oplus C\{1\}$ and the differential

$$d_{\text{Cone}(f)} := \begin{bmatrix} d_D & -f \\ 0 & -d_C \end{bmatrix}.$$

Then there is a long exact sequence

$$\dots \rightarrow H(C) \xrightarrow{f} H(D) \xrightarrow{i} H(\text{Cone}(f)) \xrightarrow{p} H(C)\{1\} \rightarrow \dots,$$

where i sends $x \in D$ to $(x, 0)$ and p sends $(x, y) \in D \oplus C\{1\}$ to $-y$. If the differentials of C and D are trivial, then we know

$$(2-8) \quad H(\text{Cone}(f)) \cong \ker(f) \oplus \text{coker}(f).$$

Remark 2.22 Our definitions about mapping cones follow from [Weibel 1994], which are different from those in [Ozsváth and Szabó 2008; 2011].

Note that the derived category is a triangulated category, so it satisfies the octahedral lemma; see for example [Weibel 1994, Proposition 10.2.4].

Lemma 2.23 (octahedral lemma) *Suppose X, Y, Z, X', Y', Z' are \mathbb{Z}_2 -graded vector spaces satisfying the long exact sequences*

$$\begin{aligned} \dots \rightarrow X \xrightarrow{f} Y \xrightarrow{h} Z' \rightarrow X\{1\} \rightarrow \dots, \\ \dots \rightarrow X \xrightarrow{g \circ f} Z \xrightarrow{h'} Y' \xrightarrow{l'} X\{1\} \rightarrow \dots, \\ \dots \rightarrow Y \xrightarrow{g} Z \rightarrow X' \xrightarrow{l} Y\{1\} \rightarrow \dots. \end{aligned}$$

Then we have the fourth long exact sequence

$$\dots \rightarrow Z' \xrightarrow{\psi} Y' \xrightarrow{\phi} X' \xrightarrow{h\{1\} \circ l} Z'\{1\} \rightarrow \dots$$

such that the diagram

(2-9)

commutes, where we omit the grading shifts and the notation for the maps h, l and j . We can also write (2-9) in another form, so that there is enough room to write the maps:

(2-10)

The maps ψ and ϕ in (2-10) can be written explicitly as follows. By the long exact sequences in the assumption of Lemma 2.23, we know that Z', X', Y' are chain homotopic to the mapping cones $\text{Cone}(f), \text{Cone}(g), \text{Cone}(g \circ f)$, respectively. Under such homotopies, we can write

$$\begin{aligned} \psi: Y \oplus X\{1\} &\rightarrow Z \oplus X\{1\}, & \psi(y, x) &\mapsto (g(y), x), \\ \phi: Z \oplus X\{1\} &\rightarrow Z \oplus Y\{1\}, & \phi(z, x) &\mapsto (z, f\{1\}(x)). \end{aligned}$$

However, the chain homotopies are not canonical, and hence the maps ψ and ϕ are also not canonical. Thus, usually we cannot identify them with other given maps. Fortunately, with an extra \mathbb{Z} -grading, we may identify $H(\text{Cone}(\phi))$ with $H(\text{Cone}(\phi'))$ for another map $\phi': Y' \rightarrow X'$.

First, we introduce the following lemma to deal with the projectivity of maps (ie maps well-defined only up to scalars). Note that the \mathbb{Z} -grading in the following lemma is not the homological grading used before.

Lemma 2.24 *Suppose X and Y are \mathbb{Z} -graded vector spaces and suppose $f, g: X \rightarrow Y$ are homogeneous maps with different grading shifts k_1 and k_2 . Then $\text{Cone}(f + g)$ is isomorphic to $\text{Cone}(c_1 f + c_2 g)$ for any $c_1, c_2 \in \mathbb{C} \setminus \{0\}$.*

Proof For simplicity, we can suppose $k_1 = 0$ and $k_2 = 1$. The proof for the general case is similar. For $i, j \in \mathbb{Z}$, we write (X, i) and (Y, j) for grading summands of X and Y , respectively. Suppose T is an automorphism of $X \oplus Y$ that acts by

$$\frac{c_2^i}{c_1^{i+1}} \text{Id} \quad \text{on } (X, i) \quad \text{and} \quad \frac{c_2^j}{c_1^j} \text{Id} \quad \text{on } (Y, j).$$

Then T is an isomorphism between $\text{Cone}(f + g)$ and $\text{Cone}(c_1 f + c_2 g)$. □

Then we state the lemma that relates the map ϕ in Lemma 2.23 to another map ϕ' constructed explicitly.

Lemma 2.25 *Suppose X, Y, Z, X', Y' are $\mathbb{Z} \oplus \mathbb{Z}_2$ -graded vector spaces satisfying the horizontal exact sequences*

$$\begin{array}{ccccc} Z & \xrightarrow{h'} & Y' & \xrightarrow{l'} & X\{1\} \\ \downarrow = & & \phi \downarrow \Downarrow & \phi' = a' + b' & \downarrow f\{1\} = a + b \\ Z & \xrightarrow{\phi \circ h' = \phi' \circ h'} & X' & \xrightarrow{l} & Y\{1\} \end{array}$$

where the shift $\{1\}$ is for the \mathbb{Z}_2 -grading. Suppose $\phi: Y' \rightarrow X'$ satisfies the two commutative diagrams and suppose $\phi': Y' \rightarrow X'$ satisfies the left commutative diagram. Suppose l and l' are homogeneous with respect to the \mathbb{Z} -grading. Suppose $f\{1\} = a + b$ and $\phi' = a' + b'$ are sums of homogeneous maps with different grading shifts with respect to the \mathbb{Z} -grading. Moreover, suppose the diagrams

$$\begin{array}{ccc} Y' & \xrightarrow{l'} & X\{1\} \\ \downarrow a' & & \downarrow a \\ X' & \xrightarrow{l} & Y\{1\} \end{array} \quad \begin{array}{ccc} Y' & \xrightarrow{l'} & X\{1\} \\ \downarrow b' & & \downarrow b \\ X' & \xrightarrow{l} & Y\{1\} \end{array}$$

hold up to scalars. Then there is an isomorphism between $H(\text{Cone}(\phi))$ and $H(\text{Cone}(\phi'))$.

Proof Since ϕ and ϕ' share the same domain and codomain, it suffices to show that they have the same rank. Fix inner products on Y' and X' such that we have orthogonal decompositions

$$Y' = \text{Im}(h') \oplus Y'' \quad \text{and} \quad X' = \text{Im}(\phi \circ h') \oplus X''.$$

By commutativity, we know both ϕ and ϕ' send $\text{Im}(h')$ onto $\text{Im}(\phi \circ h')$. Hence if we choose bases with respect to the decompositions so that linear maps are represented by matrices (we use row vectors), then

$$\phi = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \quad \text{and} \quad \phi' = \begin{bmatrix} A' & B' \\ 0 & C' \end{bmatrix},$$

where $A = A': \text{Im}(h') \rightarrow \text{Im}(\phi \circ h')$ has full row-rank. Then it suffices to show C and C' have the same row-rank.

By the exactness at Y' and X' , we know the restriction of l' on Y'' is an isomorphism between Y'' and $\text{Im}(l')$, and the restriction of l on X'' is an isomorphism between X'' and $\text{Im}(l)$. By commutativity, we know that both a and b send $\text{Im}(l')$ to $\text{Im}(l)$ and

$$\text{row rank}(C) = \text{rank}(f\{1\} | \text{Im}(l')) \quad \text{and} \quad \text{row rank}(C') = \text{rank}((c_1a + c_2b) | \text{Im}(l'))$$

for some $c_1, c_2 \in \mathbb{C} \setminus \{0\}$. Since l and l' are homogeneous, there exist induced \mathbb{Z} -gradings on $\text{Im}(l)$ and $\text{Im}(l')$. The maps a and b are still homogeneous with different grading shifts with respect to these induced gradings. Then we can apply Lemma 2.24 to obtain that the ranks of the restrictions of $f\{1\} = a + b$ and $c_1a + c_2b$ on $\text{Im}(l')$ are the same. \square

3 Integral surgery formulae

3.1 A formula for framed instanton homology

In this subsection, we propose an integral surgery formula based on sutured instanton homology, and we package it into the language of bent complexes in a later subsection.

Suppose $K \subset Y$ is a framed rationally null-homologous knot, and we adopt the notation introduced in Section 2.2. Define

$$\pi_{m,k}^\pm := \Psi_{\pm, m-1+2k}^{m+k} \circ \psi_{\mp, m+k}^{(2m+2k-1)/2}: \Gamma_{(2m+2k-1)/2} \rightarrow \Gamma_{m+2k-1}$$

and write $\pi_{m,k}^{\pm, i}$ as the restriction of $\pi_{m,k}^\pm$ on $(\Gamma_{(2m+2k-1)/2}, i)$. From Lemmas 2.13 and 2.6, we verify that $\pi_{m,k}^\pm$ shifts grading by $\pm(mp - q)/2$, and then the integral surgery formula can be stated as follows.

Theorem 3.1 *Suppose m is a fixed integer such that $mp - q \neq 0$. Then for any sufficiently large integer k , there exists an exact triangle*

$$\begin{array}{ccc} \Gamma_{(2m+2k-1)/2} & \xrightarrow{\pi_{m,k}^+ + \pi_{m,k}^-} & \Gamma_{m+2k-1} \\ & \searrow & \swarrow \\ & Y_m & \end{array}$$

Hence we have an isomorphism

$$Y_m \cong H(\text{Cone}(\pi_{m,k}^+ + \pi_{m,k}^-)).$$

Remark 3.2 Let μ and λ represent the meridian and the longitude of the knot K , respectively. Then, $mp - q \neq 0$ is equivalent to the fact that $-m\mu + \lambda$ is not isotopic to a connected component of the boundary of the Seifert surface. Specifically, if K is null-homologous, we must have $m \neq 0$.

In the rest of this subsection and in the next subsection, we state the strategy to prove Theorem 3.1, and defer the proofs of some propositions to the remaining sections. An important step is to apply the octahedral axiom mentioned in Section 2.4 to the following diagram:

(3-1)

$$\begin{array}{ccccc}
 & & \Gamma_{(2m+2k-1)/2} & & \\
 & \swarrow & \dashrightarrow & \searrow & \\
 Y_m & \dashrightarrow & & \dashrightarrow & \Gamma_{m-1+2k} \\
 \downarrow & \swarrow & & \swarrow & \uparrow \\
 \Gamma_\mu & \dashrightarrow & & \dashrightarrow & \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} \\
 & \searrow & \Gamma_{m-1} & \swarrow & \\
 & & & &
 \end{array}$$

To obtain the dotted exact triangle, we need to establish the following three exact triangles:

(3-2)

$$\begin{array}{ccc}
 I^\#(-S_m^3(K)) & & \\
 \downarrow & \swarrow & \\
 \Gamma_\mu & & \Gamma_{m-1}
 \end{array}$$

(3-3)

$$\begin{array}{ccc}
 & \Gamma_{(2m+2k-1)/2} & \\
 \swarrow & & \searrow \\
 \Gamma_\mu & \dashrightarrow & \Gamma_{m-1+k} \oplus \Gamma_{m-1+k}
 \end{array}$$

(3-4)

$$\begin{array}{ccc}
 & \Gamma_{m-1+2k} & \\
 & \uparrow & \\
 & \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} & \\
 \swarrow & & \searrow \\
 \Gamma_{m-1} & &
 \end{array}$$

and establish the following commutative diagram:

(3-5)

$$\begin{array}{ccc}
 \Gamma_\mu & \dashrightarrow & \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} \\
 \searrow & & \swarrow \\
 & \Gamma_{m-1} &
 \end{array}$$

The octahedral lemma then implies the existence of the dotted triangle and ensures that all diagrams in (3-1) other than exact triangles commute.

We will then use Lemma 2.25 to identify the map coming from the octahedral lemma with $\pi_{m,k}^+ + \pi_{m,k}^-$. We also require the following two extra diagrams to commute, where the maps other than $\pi_{m,k}^+ + \pi_{m,k}^-$ come from (3-1):

$$(3-6) \quad \begin{array}{ccc} \Gamma_{(2m+2k-1)/2} & \xrightarrow{\pi_{m,k}^+ + \pi_{m,k}^-} & \Gamma_{m-1+2k} \\ & \searrow & \uparrow \\ & & \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} \end{array}$$

$$(3-7) \quad \begin{array}{ccc} & \Gamma_{(2m+2k-1)/2} & \\ & \swarrow \quad \searrow & \\ \Gamma_\mu & & \Gamma_{m-1+2k} \\ & \searrow & \swarrow \\ & \Gamma_{m-1} & \end{array}$$

Indeed, by applying Lemma 2.25, it suffices to prove some weaker commutative diagrams involving $\pi_{m,k}^\pm$ separately.

3.2 A strategy of the proof

In this subsection, we provide more details of the strategy mentioned in Section 3.1. For simplicity, we fix the scalar ambiguities of commutative diagrams as in Section 2.3. To write down the maps, we redraw the octahedral diagram (3-1) as follows:

$$(3-8) \quad \begin{array}{ccccc} & & Y_m & \xrightarrow{\quad} & \Gamma_\mu \\ & & \swarrow \psi & & \swarrow l' \\ \Gamma_{m-1} & & & & \Gamma_{(2m+2k-1)/2} \\ \uparrow \psi_{+,m-1}^\mu + \psi_{-,m-1}^\mu & & \searrow (\Psi_{+,m-1+k}^{m-1}, \Psi_{-,m-1+k}^{m-1}) & & \downarrow \psi_{+,m-1}^\mu + \psi_{-,m-1}^\mu \\ \Gamma_\mu & \xrightarrow{\quad} & \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} & \xrightarrow{h'} & \Gamma_{m-1} \\ & \searrow (\psi_{-,m-1+k}^\mu; \psi_{+,m-1+k}^\mu) & & \searrow \phi & \downarrow l \\ & & & & \Gamma_{m-1+2k} \\ & & & & \uparrow l \\ & & & & \Psi_{-,m-1+2k}^{m-1+k} \\ & & & & -\Psi_{+,m-1+2k}^{m-1+k} \end{array}$$

where

$$h' = \psi_{-, (2m+2k-1)/2}^{m+k-1} - \psi_{+, (2m+2k-1)/2}^{m+k-1}$$

The reader can compare (3-8) with (2-9) and (2-10). We omit the term corresponding to $Z'\{1\}$ because there is not enough room, and the maps involving it are not important in our proof.

The first exact sequence of (3-8),

$$(3-9) \quad \Gamma_\mu \xrightarrow{\psi_{+,m-1}^\mu + \psi_{-,m-1}^\mu} \Gamma_{m-1} \rightarrow Y_m \rightarrow \Gamma_\mu,$$

follows from the second exact triangle in Lemma 2.16. Though the map A_{m-1} may not be the same as the sum $\psi_{+,m-1}^\mu + \psi_{-,m-1}^\mu$, we can use the following proposition and Lemma 2.24 (another special case of Proposition 4.1) to identify $\text{Cone}(A_{m-1})$ with $\text{Cone}(\psi_{+,m-1}^\mu + \psi_{-,m-1}^\mu)$. Here we use the assumption that $mp - q \neq 0$.

Proposition 3.3 *Suppose A_{n-1} is the map in Lemma 2.16. For any integer n , there exist scalars $c_1, c_2 \in \mathbb{C} \setminus \{0\}$ such that*

$$A_{n-1} = c_1 \psi_{+,n-1}^\mu + c_2 \psi_{-,n-1}^\mu.$$

The exactness at

$$\Gamma_{m-1+k} \oplus \Gamma_{m-1+k}$$

in the second and the third exact sequences are both special cases of the following proposition, which will be proved in Section 5.1 by diagram chasing.

Proposition 3.4 *Fixing the scalars as in Section 2.3, and given $n \in \mathbb{Z}$ and $k_0 \in \mathbb{N}_+$, for any c_1, c_2, c_3, c_4 satisfying the equation*

$$c_1 c_3 = -c_2 c_4,$$

the sequence

$$\Gamma_n \xrightarrow{(c_1 \Psi_{+,n+k_0}^n, c_2 \Psi_{-,n+k_0}^n)} \Gamma_{n+k_0} \oplus \Gamma_{n+k_0} \xrightarrow{c_3 \Psi_{-,n+2k_0}^{n+k_0} + c_4 \Psi_{+,n+2k_0}^{n+k_0}} \Gamma_{n+2k_0}$$

is exact.

Remark 3.5 The exactness at the direct summand for the second exact sequence (the one involving Γ_μ) might not be as clear from Proposition 3.4. Explicitly, we apply the proposition to the dual knot K' corresponding to Γ_{m+k} with framing $\lambda' = -\mu$ and $n = 0, k_0 = 1$.

The exactness at Γ_μ and $\Gamma_{(2m+2k-1)/2}$ in the second exact sequence of (3-8),

$$(3-10) \quad \Gamma_\mu \xrightarrow{(\psi_{-,m-1+k}^\mu, \psi_{+,m-1+k}^\mu)} \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} \xrightarrow{\psi_{-, \frac{1}{2}(2m+2k-1)}^{m+k-1} - \psi_{+, \frac{1}{2}(2m+2k-1)}^{m+k-1}} \Gamma_{\frac{1}{2}(2m+2k-1)} \xrightarrow{l'} \Gamma_\mu,$$

will also be proved by diagram chasing. We can explicitly construct the map l' by the composition of bypass maps

$$l' := \psi_{-, \mu}^{m+k} \circ \psi_{+, m+k}^{\frac{1}{2}(2m+2k-1)} = \psi_{+, \mu}^{m+k} \circ \psi_{-, m+k}^{\frac{1}{2}(2m+2k-1)},$$

where the last equation follows from Lemma 2.14 and the conventions in Section 2.3. The following proposition will be proved in Section 5.2 by diagram chasing.

Proposition 3.6 Suppose l' is constructed as above. For any $c_1, c_2, c_3, c_4 \in \mathbb{C} \setminus \{0\}$, the sequence

$$\begin{array}{ccc} \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} & \xrightarrow{c_3 \psi_{-, \frac{1}{2}(2m+2k-1)}^{m+k-1} + c_4 \psi_{+, \frac{1}{2}(2m+2k-1)}^{m+k-1}} & \Gamma_{\frac{1}{2}(2m+2k-1)} \xrightarrow{l'} \Gamma_{\mu} \xrightarrow{(c_1 \psi_{-, m-1+k}^{\mu}, c_2 \psi_{+, m-1+k}^{\mu})} \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} \end{array}$$

is exact.

Remark 3.7 In the proof of [Li and Ye 2021, Theorem 3.23], we obtained a long exact sequence

$$\Gamma_{\mu} \xrightarrow{(\psi_{+, n-1}^{\mu}, \psi_{-, n-1}^{\mu})} \Gamma_{n-1} \oplus \Gamma_{n-1} \rightarrow \Gamma_{\frac{1}{2}(2n-1)} \rightarrow \Gamma_{\mu}$$

by the octahedral lemma. However, we did not know the two maps involving $\Gamma_{\frac{1}{2}(2n-1)}$ explicitly. Thus, the second exact sequence here is stronger than the one from the octahedral lemma.

Remark 3.8 The reason that Proposition 3.6 holds for any choices of c_1, c_2, c_3, c_4 is because

$$\begin{aligned} \ker((c_1 \psi_{-, m-1+k}^{\mu}, c_2 \psi_{+, m-1+k}^{\mu})) &= \ker(c_1 \psi_{-, m-1+k}^{\mu}) \cap \ker(c_2 \psi_{+, m-1+k}^{\mu}), \\ \text{Im}(c_3 \psi_{-, \frac{1}{2}(2m+2k-1)}^{m+k-1} + c_4 \psi_{+, \frac{1}{2}(2m+2k-1)}^{m+k-1}) &= \text{Im}(c_3 \psi_{-, \frac{1}{2}(2m+2k-1)}^{m+k-1}) + \text{Im}(c_4 \psi_{+, \frac{1}{2}(2m+2k-1)}^{m+k-1}), \end{aligned}$$

where the right-hand sides of the equations are independent of scalars.

The exactness at Γ_{m-1} and Γ_{m-1+2k} in the third exact sequence of (3-8),

$$(3-11) \quad \Gamma_{m-1} \xrightarrow{(\Psi_{+, m-1+k}^{m-1}, \Psi_{-, m-1+k}^{m-1})} \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} \xrightarrow{\Psi_{-, m-1+2k}^{m-1+k} - \Psi_{+, m-1+2k}^{m-1+k}} \Gamma_{m-1+2k} \xrightarrow{l} \Gamma_{m-1},$$

is harder to prove since the map l cannot be constructed by bypass maps. We expect that there are many equivalent constructions of l , and we will use the one for which the exactness is easiest to prove. Even so, we only prove the exactness under the assumption that k is large. See Section 7.2 for more details.

Proposition 3.9 Suppose $c_1, c_2, c_3, c_4 \in \mathbb{C} \setminus \{0\}$ and suppose k_0 is a large integer. For any $n \in \mathbb{Z}$, there exists a map l such that the sequence

$$\Gamma_{n+k_0} \oplus \Gamma_{n+k_0} \xrightarrow{c_3 \Psi_{-, n+2k_0}^{n+k_0} + c_4 \Psi_{+, n+2k_0}^{n+k_0}} \Gamma_{n+2k_0} \xrightarrow{l} \Gamma_n \xrightarrow{(c_1 \Psi_{+, n+k_0}^n, c_2 \Psi_{-, n+k_0}^n)} \Gamma_{n+k_0} \oplus \Gamma_{n+k_0}$$

is exact.

Remark 3.10 In the first arXiv version of this paper, we only proved Proposition 3.9 for knots in S^3 because we had to use the fact that $\dim_{\mathbb{C}} I^{\#}(-S^3) = 1$ and S^3 has an orientation-reversing involution. The construction of l for knots in general 3-manifolds is inspired by the original proof for S^3 , and the proof in Section 7 is a generalization of the previous proof.

Remark 3.11 For the same reason as in Remark 3.8, the coefficients in Proposition 3.9 are not important.

Then we consider the commutative diagrams mentioned in Section 3.1. By Lemmas 2.6 and 2.12, we have

$$(\Psi_{+,m-1+k}^{m-1}, \Psi_{-,m-1+k}^{m-1}) \circ (\psi_{+,m-1}^\mu + \psi_{-,m-1}^\mu) = (\psi_{-,m-1+k}^\mu, \psi_{+,m-1+k}^\mu),$$

which verifies the commutative diagram in the assumption of the octahedral axiom.

Define

$$\phi' := \pi_{m,k}^+ + \pi_{m,k}^- = \Psi_{+,m-1+2k}^{m+k} \circ \psi_{-,m+k}^{\frac{1}{2}(2m+2k-1)} + \Psi_{-,m-1+2k}^{m+k} \circ \psi_{+,m+k}^{\frac{1}{2}(2m+2k-1)}.$$

By Lemmas 2.13 and 2.14 with $n = m + k$, we have

$$\begin{aligned} \phi' \circ h' &= (\Psi_{+,m-1+2k}^{m+k} \circ \psi_{-,m+k}^{\frac{1}{2}(2m+2k-1)} + \Psi_{-,m-1+2k}^{m+k} \circ \psi_{+,m+k}^{\frac{1}{2}(2m+2k-1)}) \\ &\quad \circ (\psi_{-, \frac{1}{2}(2m+2k-1)}^{m+k-1} - \psi_{+, \frac{1}{2}(2m+2k-1)}^{m+k-1}) \\ &= \Psi_{-,m-1+2k}^{m+k} \circ \psi_{+,m+k}^{\frac{1}{2}(2m+2k-1)} \circ \psi_{-, \frac{1}{2}(2m+2k-1)}^{m+k-1} - \Psi_{+,m-1+2k}^{m+k} \circ \psi_{-,m+k}^{\frac{1}{2}(2m+2k-1)} \circ \psi_{+, \frac{1}{2}(2m+2k-1)}^{m+k-1} \\ &= \Psi_{-,m-1+2k}^{m+k} \circ \psi_{-,m+k}^{m+k-1} - \Psi_{+,m-1+2k}^{m+k} \circ \psi_{+,m+k}^{m+k-1} \\ &= \Psi_{-,m-1+2k}^{m-1+k} - \Psi_{+,m-1+2k}^{m-1+k}. \end{aligned}$$

This verifies the second commutative diagram mentioned in Section 3.1.

Finally, we state a weaker version of the third commutative diagram mentioned in Section 3.1, which is enough to apply Lemma 2.25. The following proposition will be proved in Section 7.4.

Proposition 3.12 *Suppose l' and l are the maps in Propositions 3.6 and 3.9. Then there are two commutative diagrams up to scalars:*

$$\begin{array}{ccc} \Gamma_{\frac{1}{2}(2m+2k-1)} & \xrightarrow{l'} & \Gamma_\mu \\ \downarrow \pi_{m,k}^+ & & \downarrow \psi_{+,m-1}^\mu \\ \Gamma_{m-1+2k} & \xrightarrow{l} & \Gamma_{m-1} \end{array} \quad \begin{array}{ccc} \Gamma_{\frac{1}{2}(2m+2k-1)} & \xrightarrow{l'} & \Gamma_\mu \\ \downarrow \pi_{m,k}^- & & \downarrow \psi_{-,m-1}^\mu \\ \Gamma_{m-1+2k} & \xrightarrow{l} & \Gamma_{m-1} \end{array}$$

Proof of Theorem 3.1 We verified all assumptions of the octahedral lemma (Lemma 2.23) for the diagram (3-8). Hence, there exists a map ϕ such that

$$Y_m \cong H(\text{Cone}(\phi)).$$

We also verified all assumptions of Lemma 2.25 for $\phi' = \pi_{m,k}^+ + \pi_{m,k}^-$. Thus, we have

$$H(\text{Cone}(\phi')) \cong H(\text{Cone}(\phi)) \cong Y_m.$$

Then the desired triangle in the theorem holds. □

3.3 Reformulation by bent complexes

In this subsection, we restate Theorem 3.1 using the language of bent complexes introduced in [Li and Ye 2021]. Suppose K is a rationally null-homologous knot in a closed 3-manifold Y . We continue to adopt the notation and conventions from Sections 2.2 and 2.3.

Putting bypass triangles in Lemma 2.6 for different n together, we obtain the diagram

$$(3-12) \quad \begin{array}{ccccccc} \cdots & \longleftarrow & \Gamma_{n+1} & \xleftarrow{\psi_{+,n+1}^n} & \Gamma_n & \xleftarrow{\psi_{+,n}^{n-1}} & \Gamma_{n-1} & \xleftarrow{\psi_{+,n-1}^{n-2}} & \Gamma_{n-2} & \longleftarrow & \cdots \\ & & \searrow & & \nearrow & & \searrow & & \nearrow & & \\ & & \Gamma_\mu & & \Gamma_\mu & & \Gamma_\mu & & \Gamma_\mu & & \\ & & \swarrow & & \nwarrow & & \swarrow & & \nwarrow & & \\ \cdots & \longrightarrow & \Gamma_{n-2} & \xrightarrow{\psi_{-,n-2}^\mu} & \Gamma_{n-1} & \xrightarrow{\psi_{-,n-1}^\mu} & \Gamma_n & \xrightarrow{\psi_{-,n}^\mu} & \Gamma_{n+1} & \longrightarrow & \cdots \end{array}$$

where the \mathbb{Z} -grading shift of $\psi_{\pm,k}^\mu \circ \psi_{\pm,\mu}^k$ is $\pm p$ for any $k \in \mathbb{Z}$. From (3-12), we constructed in [Li and Ye 2021, Section 3.4] two spectral sequences $\{E_{r,+}, d_{r,+}\}_{r \geq 1}$ and $\{E_{r,-}, d_{r,-}\}_{r \geq 1}$ from Γ_μ to Y , where $d_{r,\pm}$ is roughly

$$(3-13) \quad \psi_{\pm,\mu}^k \circ (\Psi_{\pm,k+r}^k)^{-1} \circ \psi_{\pm,k+r}^\mu \quad \text{for any } k \in \mathbb{Z}.$$

The composition with the inverse map is well-defined on the r^{th} page, and the independence of k (and hence n in (3-13)) follows from Lemma 2.12. The \mathbb{Z} -grading shift of $d_{r,\pm}$ is $\pm rp$. By fixing an inner product on Γ_μ , we then lifted those spectral sequences to two differentials d_+ and d_- on Γ_μ such that

$$H(\Gamma_\mu, d_+) \cong H(\Gamma_\mu, d_-) \cong Y.$$

In such a way, the inverses of $\Psi_{\pm,k+r}^k$ are also well-defined, which we will use freely later.

Then we propose an integral surgery formula for Y_m using differentials d_+ and d_- on Γ_μ . To state the formula, we introduce the following notation.

Definition 3.13 [Li and Ye 2021, Construction 3.27 and Definition 5.12] For any integer s , define the complexes

$$\begin{aligned} B^\pm(s) &:= \left(\bigoplus_{k \in \mathbb{Z}} (\Gamma_\mu, s + kp), d_\pm \right), \\ B^+(\geq s) &:= \left(\bigoplus_{k \geq 0} (\Gamma_\mu, s + kp), d_+ \right), \\ B^-(\leq s) &:= \left(\bigoplus_{k \leq 0} (\Gamma_\mu, s + kp), d_- \right). \end{aligned}$$

Furthermore, define

$$I^+(s): B^+(\geq s) \rightarrow B^+(s) \quad \text{and} \quad I^-(s): B^-(\leq s) \rightarrow B^-(s)$$

to be the inclusion maps. We also use the same notation for the induced map on homology.

Remark 3.14 By Lemma 2.5, we know that the nontrivial gradings of Γ_μ are finite. Then, for any sufficiently large integer s_0 satisfying

$$s - s_0 p \leq -\frac{p - \chi(S)}{2} \quad \text{and} \quad s + s_0 p \geq \frac{p - \chi(S)}{2},$$

we have

$$B^+(s) = B^+(\geq s - s_0 p) \quad \text{and} \quad B^-(s) = B^-(\leq s + s_0 p).$$

In such a case, $I^+(s - s_0 p)$ and $I^-(s + s_0 p)$ are identities.

By splitting the diagram (3-12) into \mathbb{Z} -gradings, we can calculate homologies of the complexes defined in Definition 3.13.

Proposition 3.15 Suppose $n \in \mathbb{N}_+$ and i is a grading. Fix an inner product on Γ_n . If $i > \frac{1}{2}(p - \chi(S)) - np$, then there exists a canonical isomorphism

$$H(B^+(\geq i)) \cong \left(\Gamma_n, i + \frac{(n-1)p - q}{2} \right).$$

If $i < -(p - \chi(S))/2 + np$, then there exists a canonical isomorphism

$$H(B^-(\leq i)) \cong \left(\Gamma_n, i - \frac{(n-1)p - q}{2} \right).$$

Proof The proof mirrors that of [Li and Ye 2021, Lemma 5.13]. Following the notation in [Li and Ye 2021, (3.9) and (3.10)], if

$$i > \hat{i}_{\max}^\mu - nq = \frac{p - \chi(S)}{2} - np,$$

then $\Gamma_0^{i,+} = 0$ (the corresponding grading summand of Γ_0), and the isomorphism follows from the convergence theorem of the unrolled spectral sequence [Li and Ye 2021, Theorem 2.4]; see also [Boardman 1999, Theorem 6.1]. Note that the unrolled spectral sequence induces a filtration on Γ_n , and the homology is canonically isomorphic to the direct sum of all associated graded objects of the filtration. Then we use the inner product to identify the direct sum with the total space Γ_n . The other statement holds for the same reason. \square

Definition 3.16 [Li and Ye 2021, Construction 3.27 and Definition 5.12] For any integer s , define the *bent complex*

$$A(s) := \left(\bigoplus_{k \in \mathbb{Z}} (\Gamma_\mu, s + kp), d_s \right),$$

where for any element $x \in (\Gamma_\mu, s + kp)$,

$$d_s(x) = \begin{cases} d_+(x) & \text{if } k > 0, \\ d_+(x) + d_-(x) & \text{if } k = 0, \\ d_-(x) & \text{if } k < 0. \end{cases}$$

Define

$$\pi^+(s): A(s) \rightarrow B^+(s) \quad \text{and} \quad \pi^-(s): A(s) \rightarrow B^-(s)$$

by

$$\pi^+(s)(x) = \begin{cases} x & \text{if } k \geq 0, \\ 0 & \text{if } k < 0, \end{cases} \quad \text{and} \quad \pi^-(s)(x) = \begin{cases} x & \text{if } k \leq 0, \\ 0 & \text{if } k > 0, \end{cases}$$

where $x \in (\Gamma_\mu, s + kp)$. Define

$$\pi^\pm: \bigoplus_{s \in \mathbb{Z}} A(s) \rightarrow \bigoplus_{s \in \mathbb{Z}} B^\pm(s)$$

by putting $\pi^\pm(s)$ together for all s . We also use the same notation for the induced map on homology.

Remark 3.17 Similar to Remark 3.14, according to Lemma 2.5, there are only finitely many the nontrivial gradings of Γ_μ . Then, for any sufficiently large integer s_0 such that $s_0 \geq (p - \chi(S))/2$, we have

$$A(s_0) = B^-(s_0) \quad \text{and} \quad A(-s_0) = B^+(-s_0).$$

In such a case, $\pi^-(s_0)$ and $\pi^+(-s_0)$ are identities.

Now, we state the integral surgery formula in the above setup.

Theorem 3.18 Suppose m is a fixed integer such that $mp - q \neq 0$. Then there exists an isomorphism

$$\Xi_m: \bigoplus_{s \in \mathbb{Z}} H(B^+(s)) \xrightarrow{\cong} \bigoplus_{s \in \mathbb{Z}} H(B^-(s + mp - q))$$

as the direct sum of isomorphisms

$$\Xi_{m,s}: H(B^+(s)) \xrightarrow{\cong} H(B^-(s + mp - q))$$

so that

$$Y_m \cong H\left(\text{Cone}\left(\pi^- + \Xi_m \circ \pi^+: \bigoplus_{s \in \mathbb{Z}} H(A(s)) \rightarrow \bigoplus_{s \in \mathbb{Z}} H(B^-(s))\right)\right).$$

Proof According to Remark 3.17, we only need to consider the maps $\pi^\pm(s)$ for $|s|$ less than a fixed integer. For such values of s , we can apply the following proposition.

Proposition 3.19 [Li and Ye 2021, Proposition 3.28] Fix $m, s \in \mathbb{Z}$ such that $|s| \leq \frac{1}{2}(p - \chi(S))$. For any large integer k , fix inner products on $\Gamma_{\frac{1}{2}(2m+2k-1)}$ and Γ_{m-1+2k} . Then there exist $s_1, s_2^+, s_2^-, s_3^+, s_3^- \in \mathbb{Z}$ such that the diagram

$$\begin{CD} H(A(s)) @>\pi^\pm(s)>> H(B^\pm(s)) \\ @V\cong VV @VV\cong V \\ (\Gamma_{\frac{1}{2}(2m+2k-1)}, s_1) @>\pi_{m,k}^{\pm, s_1}>> (\Gamma_{m-1+2k}, s_3^\pm) \end{CD}$$

commutes, where the maps $\pi_{m,k}^{\pm, s_1}$, defined in Section 3.1, factor through (Γ_{m+k}, s_2^\pm) .

Remark 3.20 The maps $\pi^\pm(s)$ factor through $I^\pm(s)$ constructed in Definition 3.13. We write

$$\pi^\pm(s) = I^\pm(s) \circ \pi^{\pm, \prime}(s).$$

This corresponds to the factorization about (Γ_{m+k}, s_2^\pm) in Proposition 3.19 (we fix an inner product on Γ_{m+k} to apply Proposition 3.15), ie the following diagrams commute:

$$\begin{CD} H(A(s)) @>\pi^{+, \prime}(s)>> H(B^+(\geq s)) @>I^+(s)>> H(B^+(s)) \\ @V\cong VV @VV\cong V @VV\cong V \\ (\Gamma_{\frac{1}{2}(2m+2k-1)}, s_1) @>\psi_{-, m+k}^{\frac{1}{2}(2m+2k-1)}>> (\Gamma_{m+k}, s_2^+) @>\Psi_{+, m-1+2k}^{m+k}>> (\Gamma_{m-1+2k}, s_3^+) \\ \\ H(A(s)) @>\pi^{-, \prime}(s)>> H(B^-(\leq s)) @>I^-(s)>> H(B^-(s)) \\ @V\cong VV @VV\cong V @VV\cong V \\ (\Gamma_{\frac{1}{2}(2m+2k-1)}, s_1) @>\psi_{+, m+k}^{\frac{1}{2}(2m+2k-1)}>> (\Gamma_{m+k}, s_2^-) @>\Psi_{-, m-1+2k}^{m+k}>> (\Gamma_{m-1+2k}, s_3^-) \end{CD}$$

From the calculation in [Li and Ye 2021, Remark 3.29] (we replace n and l there by $m+k$ and $k-1$, and note that there is a typo about sign in the first arXiv version of [Li and Ye 2021]), the difference of the grading shifts is

$$s_3^+ - s_3^- = (m+k - (k-1) - 1)p - q = mp - q.$$

Note that the notation in this paper and [Li and Ye 2021] are different (cf Remark 2.3).

Then we can construct the isomorphism

$$\Xi_{m,s}: H(B^+(s)) \xrightarrow{\cong} H(B^-(s + mp - q))$$

for $|s| \leq \frac{1}{2}(p - \chi(S))$ by identifying both $H(B^+(s))$ and $H(B^-(s + mp - q))$ with (Γ_{m-1+2k}, s_3^+) for a sufficiently large k . A priori, this isomorphism depends on inner products on

$$\Gamma_\mu, \Gamma_{\frac{1}{2}(2m+2k-1)}, \Gamma_{m-1+2k} \text{ and } \Gamma_{m+k}.$$

For other s , we can take any isomorphism $\Xi_{m,s}$ since the choice does not affect the computation of the mapping cone.

Consequently, we obtain

$$H(\text{Cone}(\pi^- + \Xi_m \circ \pi^+)) \cong H(\text{Cone}(\pi_{m,k}^- + \pi_{m,k}^+)) \cong Y_m,$$

where the last isomorphism comes from Theorem 3.1. □

Remark 3.21 Theorem 3.18 is slightly weaker than Theorem 3.1. Indeed, when we use the integral surgery formula to calculate surgeries on the Borromean knot in the companion paper [Li and Ye 2025], we have to study the $H_1(Y)$ action on sutured instanton homology, where $Y = \#^{2n} S^1 \times S^2$ is the ambient manifold of the knot. This action vanishes on Γ_μ so vanishes on the bent complex. But it is nonvanishing on Γ_{m+k} and Γ_{m-1+2k} and we use this information to realize the computation. This issue for the bent complex might be resolved by introducing some E_0 -pages for differentials d_+ and d_- such that the action is nontrivial on E_0 -pages.

3.4 A formula for instanton knot homology

The third exact sequence (3-11) implies

$$\Gamma_{m-1} \cong H(\text{Cone}(\Psi_{-,m-1+2k}^{m-1+k} - \Psi_{+,m-1+2k}^{m-1+k} : \Gamma_{m-1+k} \oplus \Gamma_{m-1+k} \rightarrow \Gamma_{m-1+2k}))$$

for any sufficiently large integer k . Since there are two copies Γ_{m-1+k} , we can always regard the grading shifts of the maps $\Psi_{-,m-1+2k}^{m-1+k}$ as different ones by rescaling the grading of the first summand from i to $2i - 1$ and the second summand from i to $2i$. Hence we do not need the assumption $mp - q \neq 0$ as in the previous mapping cone formula in Theorem 3.1. By Lemma 2.24, we can replace the minus sign with any coefficient.

In this subsection, we restate this result in the language of bent complexes. The formula is inspired by Eftekhary’s formula [2018, Proposition 1.5] for knot Floer homology $\widehat{\text{HF}}K$; see also Hedden-Levine’s work [Hedden and Levine 2024]. Since m can be any integer, we replace $m - 1$ by m .

Theorem 3.22 Suppose $m, j \in \mathbb{Z}$. Define

$$j^+ = j - \frac{(m-1)p - q}{2} \quad \text{and} \quad j^- = j + \frac{(m-1)p - q}{2}.$$

Then there exists an isomorphism

$$\Xi'_{m,j} : H(B^+(j^+)) \xrightarrow{\cong} H(B^-(j^-))$$

such that

$$(\Gamma_m, j) \cong H(\text{Cone}(I^-(j^-) + \Xi'_{m,j} \circ I^+(j^+) : H(B^-(\leq j^-)) \oplus H(B^+(\geq j^+)) \rightarrow H(B^-(j^-)))).$$

Proof As mentioned before, we have

$$\Gamma_m \cong H(\text{Cone}(\Psi_{-,m+2k}^{m+k} - \Psi_{+,m+2k}^{m+k})) \cong H(\text{Cone}(\Psi_{-,m+2k}^{m+k} + \Psi_{+,m+2k}^{m+k}))$$

for any sufficiently large integer k .

Since bypass maps are homogeneous, the above mapping cone splits into \mathbb{Z} -gradings (or $(\mathbb{Z} + \frac{1}{2})$ -gradings). Hence we can use it to calculate (Γ_m, j) . By Lemma 2.6, the corresponding spaces are

$$(\Gamma_{m+k}, j - \frac{1}{2}kp) \oplus (\Gamma_{m+k}, j + \frac{1}{2}kp) \quad \text{and} \quad (\Gamma_{m+2k}, j).$$

From Proposition 3.15 with $i = j \pm \frac{1}{2}kp$, by fixing an inner product on Γ_{m+k} , we know that

$$\begin{aligned} (\Gamma_{m+k}, j - \frac{1}{2}kp) &\cong H(B^-(\leq j^-)) \quad \text{for } j - \frac{1}{2}kp < -\frac{1}{2}(p - \chi(S)) + (m+k)p, \\ (\Gamma_{m+k}, j + \frac{1}{2}kp) &\cong H(B^+(\geq j^+)) \quad \text{for } j + \frac{1}{2}kp > \frac{1}{2}(p - \chi(S)) - (m+k)p. \end{aligned}$$

Since m is fixed, when k is sufficiently large, we know that any j with (Γ_m, j) nontrivial (ie with $|j| \leq (|mp - q| - \chi(S))/2$ by Lemma 2.5) satisfies the above inequalities. By Proposition 3.15 again (fixing an inner product on Γ_{m+2k}) and Remark 3.14, for k sufficiently large, we know that

$$(\Gamma_{m+2k}, j) \cong H(B^-(j^-)) \cong H(B^+(j^+))$$

for such j . By unpackaging the construction of differentials d_+ and d_- in [Li and Ye 2021, Section 3.4], we know that the restrictions of maps $\Psi_{-,m+2k}^m$ and $\Psi_{+,m+2k}^m$ on the corresponding gradings coincide with the maps induced by the inclusions $I^-(j^-)$ and $I^+(j^+)$ under the canonical isomorphisms, respectively.

For $|j| \leq (|mp - q| - \chi(S))/2$, let

$$\Xi'_{m,j}: H(B^+(j^+)) \xrightarrow{\cong} H(B^-(j^-))$$

be the isomorphism obtained from identifying both spaces to the corresponding grading summand of Γ_{m+2k} . Note that it depends on inner products on Γ_μ, Γ_{m+k} and Γ_{m+2k} . For other j , we can take any isomorphism $\Xi'_{m,j}$ since the choice does not affect the computation of the mapping cone. Then we know that

$$\begin{aligned} (\Gamma_m, j) &\cong H(\text{Cone}(\Psi_{-,m+2k}^{m+k} + \Psi_{+,m+2k}^{m+k} \mid (\Gamma_{m+k}, j + \frac{1}{2}kp) \oplus (\Gamma_{m+k}, j - \frac{1}{2}kp))) \\ &\cong H(\text{Cone}(I^-(j^-) + \Xi'_{m,j} \circ I^+(j^+))). \end{aligned} \quad \square$$

4 Dehn surgery and bypass maps

In this section, we prove a generalization of Lemma 2.20 and Proposition 3.3.

Suppose (M, γ) is a balanced sutured manifold and $\alpha \subset \partial M$ is a connected simple closed curve that intersects the suture γ twice. There are two natural bypass arcs associated to α , each of which intersects the suture at three points and induces a bypass triangle (cf [Baldwin and Sivek 2022b, Section 4])

$$\begin{array}{ccc} \underline{\text{SHI}}(-M, -\gamma) & \xrightarrow{\psi_\pm} & \underline{\text{SHI}}(-M, -\gamma_2) \\ & \swarrow \quad \searrow & \\ & \underline{\text{SHI}}(-M, -\gamma_3^\pm) & \end{array}$$

where γ_2 and γ_3^\pm are the sutures coming from bypass attachments. Note that the horizontal arrows involve the same set of balanced sutured manifolds but have different maps. Let (M_0, γ_0) be obtained from

(M, γ) by attaching a contact 2-handle along α . From [Baldwin and Sivek 2016b, Section 3.3], it has been shown that a closure of $(-M_0, -\gamma_0)$ coincides with a closure of the sutured manifold obtained from $(-M, -\gamma)$ by 0-surgery along α with respect to the surface framing. Hence there is also a surgery exact triangle (cf [Li and Ye 2022, Lemma 3.21])

$$\begin{array}{ccc} \underline{\text{SHI}}(-M, -\gamma) & \xrightarrow{H_\alpha} & \underline{\text{SHI}}(-M, -\gamma_2) \\ & \swarrow & \searrow \\ & \underline{\text{SHI}}(-M_0, -\gamma_0) & \end{array}$$

The map H_α is related to the bypass maps ψ_\pm as follows:

Proposition 4.1 *There exist $c_1, c_2 \in \mathbb{C} \setminus \{0\}$ such that*

$$H_\alpha = c_1 \psi_+ + c_2 \psi_-.$$

Remark 4.2 The proof of Proposition 4.1 was developed through the discussions with John A Baldwin and Steven Sivek.

Proof of Proposition 4.1 Let $A \subset \partial M$ be a tubular neighborhood of $\alpha \subset \partial M$. Push the interior of A into the interior of M to make it a properly embedded surface. By a standard argument in [Honda 2000], we can assume that a collar of ∂M is equipped with a product contact structure such that γ is (isotopic to) the dividing set, α is a Legendrian curve, A is in the contact collar, and A is a convex surface with Legendrian boundary that separates a standard contact neighborhood of α off M . The convex decomposition of M along A yields two pieces

$$M = M' \cup_A V,$$

where M' is diffeomorphic to M and V is the contact neighborhood of α . It is straightforward to check that, after rounding the corners, the contact structure near the boundary of M' is still a product contact structure with $\partial M'$ being a convex boundary. Let γ' be the dividing set on $\partial M'$. Also, after rounding the corners, with the contact structure on $V \cong S^1 \times D^2$, we can suppose ∂V is a convex surface with dividing set being the union of two connected simple closed curves on ∂V of slope -1 . When viewing V as the complement of an unknot in S^3 , the dividing set coincides with the suture $\Gamma_1 \subset V$, so from now on we call it Γ_1 . By the construction of the gluing map in [Li 2021a], there exists a map

$$G_1 : \underline{\text{SHI}}(-M', -\gamma') \otimes \underline{\text{SHI}}(-V, -\Gamma_1) \rightarrow \underline{\text{SHI}}(-M, -\gamma).$$

As in [Li 2021a], the map G_1 comes from attaching contact handles to $(M', \gamma') \sqcup (V, \Gamma_1)$ to recover the gluing along A . From [Li 2021b, Proposition 1.4], we know that

$$\underline{\text{SHI}}(-V, -\Gamma_1) \cong \mathbb{C}.$$

Note that M' and M are equipped with the product contact structure near their boundaries. From the functoriality of the contact gluing map in [Li 2021a], we know that G_1 is an isomorphism. Now both the

(-1)-surgery along a push-off of α and the bypass attachments can be thought of as happening in the piece V . Note that the result of both (-1)-surgery and the bypass attachments for Γ_1 is Γ_2 . Hence we have the commutative diagram

$$(4-1) \quad \begin{array}{ccc} \underline{\text{SHI}}(-M', -\gamma') \otimes \underline{\text{SHI}}(-V, -\Gamma_1) & \xrightarrow{G_1} & \underline{\text{SHI}}(-M, -\gamma) \\ \text{Id} \otimes \widehat{H}_\alpha \downarrow & & H_\alpha \downarrow \\ \underline{\text{SHI}}(-M', -\gamma') \otimes \underline{\text{SHI}}(-V, -\Gamma_2) & \xrightarrow{G_2} & \underline{\text{SHI}}(-M, -\gamma_2) \end{array}$$

where \widehat{H}_α denotes the surgery map for the manifold V and G_2 is the gluing map obtained by attaching the same set of contact handles as G_1 . A similar commutative diagram holds when replacing H_α and \widehat{H}_α by ψ_\pm and

$$\widehat{\psi}_\pm : \underline{\text{SHI}}(-V, -\Gamma_1) \rightarrow \underline{\text{SHI}}(-V, -\Gamma_2)$$

in (4-1), respectively.

Since G_1 is an isomorphism, to obtain a relation between H_α and ψ_\pm , it suffices to understand the relation between \widehat{H}_α and $\widehat{\psi}_\pm$. From [Li 2021b, Proposition 1.4], we know that

$$\underline{\text{SHI}}(-V, -\Gamma_2) \cong \mathbb{C}^2.$$

Moreover, the meridian disk of V induces a $(\mathbb{Z} + \frac{1}{2})$ grading on $\underline{\text{SHI}}(-V, -\Gamma_2)$ and we have

$$\underline{\text{SHI}}(-V, -\Gamma_2) \cong \underline{\text{SHI}}(-V, -\Gamma_2, \frac{1}{2}) \oplus \underline{\text{SHI}}(-V, -\Gamma_2, -\frac{1}{2}),$$

with

$$\underline{\text{SHI}}(-V, -\Gamma_2, \frac{1}{2}) \cong \underline{\text{SHI}}(-V, -\Gamma_2, -\frac{1}{2}) \cong \mathbb{C}.$$

Let

$$\mathbf{1} \in \underline{\text{SHI}}(-V, -\Gamma_1) \cong \mathbb{C}$$

be a generator. In [Li 2021b, Section 4.3] it is shown that

$$\widehat{\psi}_-(\mathbf{1}) \in \underline{\text{SHI}}(-V, -\Gamma_2, \frac{1}{2}) \quad \text{and} \quad \widehat{\psi}_+(\mathbf{1}) \in \underline{\text{SHI}}(-V, -\Gamma_2, -\frac{1}{2})$$

are nonzero. Also, when viewing V as the complement of the unknot U , there is an exact triangle

$$(4-2) \quad \begin{array}{ccc} \underline{\text{SHI}}(-V, -\Gamma_1) & \xrightarrow{\widehat{H}_\alpha} & \underline{\text{SHI}}(-V, -\Gamma_2) \\ & \swarrow G_1 & \nwarrow F_2 \\ & I^\#(-S^3) & \end{array}$$

as in Lemma 2.16. Comparing the dimensions of the spaces in (4-2), we have $G_1 = 0$ and \widehat{H}_α is injective. From the fact that $\tau_I(U) = 0$, we know from [Ghosh et al. 2024, Corollary 3.5] that

$$F_2 | \underline{\text{SHI}}(-V, -\Gamma_2, \frac{1}{2}) \neq 0 \quad \text{and} \quad F_2 | \underline{\text{SHI}}(-V, -\Gamma_2, -\frac{1}{2}) \neq 0.$$

By the exactness in (4-2), we have $\ker(F_2) = \text{Im}(\widehat{H}_\alpha)$ and then $\widehat{H}_\alpha(\mathbf{1})$ is not in $\underline{\text{SHI}}(-V, -\Gamma_2, \pm\frac{1}{2})$, ie it is a linear combination of generators of $\underline{\text{SHI}}(-V, -\Gamma_2, \pm\frac{1}{2})$. Hence we know that there are $c_1, c_2 \in \mathbb{C} \setminus \{0\}$ such that

$$\widehat{H}_\alpha(\mathbf{1}) = c_1 \widehat{\psi}_+(\mathbf{1}) + c_2 \widehat{\psi}_-(\mathbf{1}).$$

Then the proposition follows from the commutative diagram (4-1). □

In Remark 1.3, we discussed the ambiguity arising from scalars. It is worth mentioning that such ambiguity already exists in instanton theory. For example, if M is the complement of a knot $K \subset S^3$ and γ consists of two meridians of the knot, which we denote by Γ_μ , we can choose α to be a curve on $\partial(S^3 \setminus N(K))$ of slope $-n$. Then we have a surgery triangle

$$\begin{array}{ccc} \underline{\text{SHI}}(-M, -\Gamma_\mu) & \xrightarrow{H_n} & \underline{\text{SHI}}(-M, -\Gamma_{n-1}) \\ & \swarrow & \searrow \\ & I^\#(-S^3_{-n}(K)) & \end{array}$$

Note that this triangle is not the one from Floer’s original exact triangle, but rather one with a slight modification on the choice of 1-cycles inside the 3-manifold that represents the second Stiefel–Whitney class of the relevant $\text{SO}(3)$ -bundle; see [Baldwin and Sivek 2021, Section 2.2] for more details. Floer’s original exact triangle, on the other hand, yields a different triangle

$$\begin{array}{ccc} \underline{\text{SHI}}(-S^3 \setminus N(K), -\Gamma_\mu) & \xrightarrow{H'_n} & \underline{\text{SHI}}(-S^3 \setminus N(K), -\Gamma_{n-1}) \\ & \swarrow & \searrow \\ & I^\#(-S^3_{-n}(K), \mu) & \end{array}$$

where $\mu \subset -S^3_{-n}(K)$ denotes a meridian of the knot. The difference between H_α and H'_α is that they come from the same cobordism but the $\text{SO}(3)$ -bundles over the cobordism are different. The local argument to prove Proposition 4.1 works for both H_α and H'_α . Hence there exists nonzero complex numbers c_1, c_2, c'_1, c'_2 such that

$$H_\alpha = c_1 \psi_{+,n}^\mu + c_2 \psi_{-,n-1}^\mu \quad \text{and} \quad H'_\alpha = c'_1 \psi_{+,n}^\mu + c'_2 \psi_{-,n-1}^\mu,$$

where the maps

$$\psi_{\pm,n-1}^\mu : \underline{\text{SHI}}(-S^3 \setminus N(K), -\Gamma_\mu) \rightarrow \underline{\text{SHI}}(-S^3 \setminus N(K), -\Gamma_{n-1})$$

are the two related bypass maps. When $n \neq 0$, these two bypass maps have different grading-shifting behavior, so by Lemma 2.24, a different choice of nonzero coefficients does not change the dimensions of the kernel and cokernel of the map. Hence we conclude that for $n \neq 0$,

$$I^\#(-S^3_{-n}(K), \mu) \cong I^\#(-S^3_{-n}(K)).$$

However, when $n = 0$, the two bypass maps $\psi_{\pm, n-1}^\mu$ both preserve gradings, making the coefficients significant, ie $I^\#(-S_0^3(K), \mu)$ and $I^\#(-S_0^3(K))$ might have different dimensions for different choices of coefficients. Indeed, it is observed by Baldwin and Sivek [2021] that for what they called W-shaped knots (which is clearly a nonempty class, eg the figure-eight knot [Baldwin and Sivek 2022a, Proposition 10.4]), these two framed instanton homologies have dimensions that differ by 2.

5 Some exactness by diagram chasing

5.1 At the direct summand

In this subsection, we prove Proposition 3.4 by diagram chasing. We restate the result in Proposition 5.1. We also adopt the conventions for scalars from Section 2.3, and this together with Lemma 2.11 implies that

$$\Psi_{+, n+2k_0}^{n+k_0} \circ \Psi_{-, n+k_0}^n = \Psi_{-, n+2k_0}^{n+k_0} \circ \Psi_{+, n+k_0}^n$$

for any n and k_0 .

Proposition 5.1 *Given $n \in \mathbb{Z}$ and $k_0 \in \mathbb{N}_+$, for any c_1, c_2, c_3, c_4 satisfying*

$$c_1 c_3 = -c_2 c_4,$$

the sequence

$$\Gamma_n \xrightarrow{(c_1 \Psi_{+, n+2k_0}^n, c_2 \Psi_{-, n+k_0}^n)} \Gamma_{n+k_0} \oplus \Gamma_{n+k_0} \xrightarrow{c_3 \Psi_{-, n+2k_0}^{n+k_0} + c_4 \Psi_{+, n+2k_0}^{n+k_0}} \Gamma_{n+2k_0}$$

is exact.

Proof For simplicity, we only prove the proposition for $n = 0$. The proof for any general n is similar (replacing all Γ_m below by Γ_{n+m} and modifying the notation for bypass maps). Also, we only prove the case when

$$c_1 = c_2 = c_3 = 1 \quad \text{and} \quad c_4 = -1.$$

The proof for general scalars can be obtained similarly.

We prove the proposition by induction on k_0 . We will use the exactness in Lemma 2.6 and the commutative diagrams in Lemmas 2.12 and 2.11 many times. For simplicity, we will use them without naming them.

First, we assume $k_0 = 1$. The proposition reduces to

$$\ker(\psi_{-, 2}^1 - \psi_{+, 2}^1) = \text{Im}((\psi_{+, 1}^0, \psi_{-, 1}^0)).$$

The commutative diagram in Lemma 2.11 implies

$$\ker(\psi_{-, 2}^1 - \psi_{+, 2}^1) \supset \text{Im}((\psi_{+, 1}^0, \psi_{-, 1}^0)).$$

We then prove

$$\ker(\psi_{-,2}^1 - \psi_{+,2}^1) \subset \text{Im}((\psi_{+,1}^0, \psi_{-,1}^0)).$$

Suppose

$$(x_1, x_2) \in \ker(\psi_{-,2}^1 - \psi_{+,2}^1), \text{ ie, } \psi_{-,2}^1(x_1) - \psi_{+,2}^1(x_2) = 0.$$

Then we have

$$\psi_{+,\mu}^1(x_1) = \psi_{+,\mu}^2 \circ \psi_{-,2}^1(x_1) = \psi_{+,\mu}^2 \circ \psi_{+,2}^1(x_2) = 0.$$

By exactness, there exists $y \in \Gamma_0$ such that $\psi_{+,1}^0(y) = x_1$. Then

$$\psi_{+,2}^1 \circ \psi_{-,1}^0(y) = \psi_{-,2}^1 \circ \psi_{+,1}^0(y) = \psi_{-,2}^1(x_1) \quad \text{and} \quad \psi_{+,2}^1(x_2 - \psi_{-,1}^0(y)) = 0.$$

By exactness, there exists $z \in \Gamma_\mu$ such that

$$\psi_{+,1}^\mu(z) = x_2 - \psi_{-,1}^0(y).$$

Let $y' = y + \psi_{+,0}^\mu(z)$. Then

$$\psi_{+,1}^0(y') = \psi_{+,1}^0(y) = x_1 \quad \text{and} \quad \psi_{-,1}^0(y') = \psi_{-,1}^0(y) + \psi_{-,1}^0 \circ \psi_{+,0}^\mu(z) = \psi_{-,1}^0(y) + \psi_{+,1}^\mu(z) = x_2,$$

which concludes the proof for $k_0 = 1$.

Suppose the proposition holds for $k_0 = k$. We prove it also holds for $k_0 = k + 1$. The proof is similar to the case for $k_0 = 1$. Again by Lemma 2.11, we have

$$\ker(\Psi_{-,2k+2}^{k+1} - \Psi_{+,2k+2}^{k+1}) \supset \text{Im}((\Psi_{+,k+1}^0, \Psi_{-,k+1}^0)).$$

Then we prove

$$\ker(\Psi_{-,2k+2}^{k+1} - \Psi_{+,2k+2}^{k+1}) \subset \text{Im}((\Psi_{+,k+1}^0, \Psi_{-,k+1}^0)).$$

Suppose

$$(x_1, x_2) \in \ker(\Psi_{-,2k+2}^{k+1} - \Psi_{+,2k+2}^{k+1}), \quad \text{ie } \Psi_{-,2k+2}^{k+1}(x_1) - \Psi_{+,2k+2}^{k+1}(x_2) = 0.$$

Then we have

$$\psi_{+,\mu}^{k+1}(x_1) = \psi_{+,\mu}^{2k+2} \circ \Psi_{-,2k+2}^{k+1}(x_1) = \psi_{+,\mu}^{2k+2} \circ \Psi_{+,2k+2}^{k+1}(x_2) = 0.$$

By exactness, there exists $y_1 \in \Gamma_k$ such that $\psi_{+,k+1}^k(y_1) = x_1$. By a similar reason, there exists $y_2 \in \Gamma_k$ such that $\psi_{-,k+1}^k(y_2) = x_2$. The goal is to prove

$$\Psi_{-,2k}^k(y'_1) = \Psi_{+,2k}^k(y'_2)$$

for some modifications y'_1 and y'_2 of y_1 and y_2 as for y' in the case of $k_0 = 1$. Then the induction hypothesis will imply that there exists $w \in \Gamma_0$ such that

$$\Psi_{+,k}^0(w) = y'_1 \quad \text{and} \quad \Psi_{-,k}^0(w) = y'_2.$$

Hence we will have

$$\Psi_{+,k+1}^0(w) = \psi_{+,k+1}^k(y'_1) = x_1 \quad \text{and} \quad \Psi_{-,k+1}^0(w) = \psi_{-,k+1}^k(y'_2) = x_2.$$

This will conclude the proof for $k_0 = k + 1$.

Now we start to construct y'_1 . We have

$$\begin{aligned} \psi_{+,2k+2}^{2k+1}(\Psi_{+,2k+1}^{k+1}(x_2) - \Psi_{-,2k+1}^k(y_1)) &= \psi_{+,2k+2}^{2k+1} \circ \Psi_{+,2k+1}^{k+1}(x_2) - \psi_{+,2k+2}^{2k+1} \circ \Psi_{-,2k+1}^k(y_1) \\ &= \Psi_{+,2k+2}^{k+1}(x_2) - \psi_{+,2k+2}^{2k+1} \circ \Psi_{-,2k+1}^k(y_1) \\ &= \Psi_{-,2k+2}^{k+1}(x_2) - \Psi_{+,2k+2}^{k+1}(x_1) \\ &= 0. \end{aligned}$$

By exactness, there exists $z_1 \in \Gamma_\mu$ such that

$$\psi_{+,2k+1}^\mu(z_1) = \Psi_{+,2k+1}^{k+1}(x_2) - \Psi_{-,2k+1}^k(y_1).$$

Let $y'_1 = y_1 + \psi_{+,k}^\mu(z_1)$. Then

$$\begin{aligned} \psi_{+,k+1}^k(y'_1) &= \psi_{+,k+1}^k(y_1) = x_1, \\ \Psi_{-,2k+1}^k(y'_1) &= \Psi_{-,2k+1}^k(y_1) + \Psi_{-,2k+1}^k \circ \psi_{+,k}^\mu(z_1) \\ &= \Psi_{-,2k+1}^k(y_1) + \psi_{+,2k+1}^\mu(z_1) \\ &= \Psi_{+,2k+1}^{k+1}(x_2). \end{aligned}$$

Then we start to construct y'_2 . We have

$$\begin{aligned} \psi_{-,2k+1}^{2k}(\Psi_{-,2k}^k(y'_1) - \Psi_{+,2k}^k(y_2)) &= \Psi_{-,2k+1}^k(y'_1) - \psi_{-,2k+1}^{2k} \circ \Psi_{+,2k}^k(y_2) \\ &= \Psi_{-,2k+1}^k(y'_1) - \Psi_{-,2k+1}^{k+1}(x_2) \\ &= 0. \end{aligned}$$

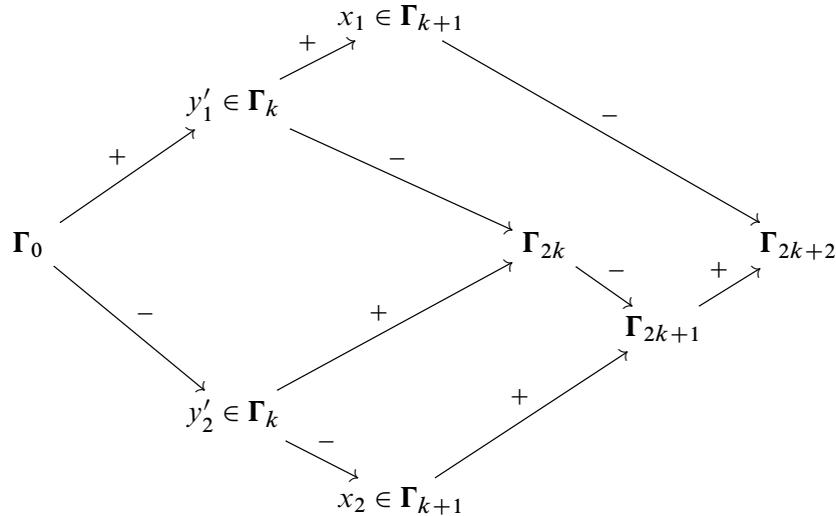
By exactness, there exists $z_2 \in \Gamma_\mu$ such that

$$\psi_{-,2k}^\mu(z_2) = \Psi_{-,2k}^k(y'_1) - \Psi_{+,2k}^k(y_2).$$

Let $y'_2 = y_2 + \psi_{-,k}^\mu(z_2)$. Then

$$\begin{aligned} \psi_{-,k+1}^k(y'_2) &= \psi_{-,k+1}^k(y_2) = x_2, \\ \Psi_{+,2k}^k(y'_2) &= \Psi_{+,2k}^k(y_2) + \Psi_{+,2k}^k \circ \psi_{-,k}^\mu(z_2) \\ &= \Psi_{+,2k}^k(y_2) + \psi_{-,2k}^\mu(z_2) \\ &= \Psi_{-,2k}^k(y'_1). \end{aligned}$$

Then we have the commutative diagrams



By the induction hypothesis, there exists $w \in \Gamma_0$ such that

$$\Psi_{+,k}^0(w) = y'_1 \quad \text{and} \quad \Psi_{-,k}^0(w) = y'_2,$$

which concludes the proof for $k_0 = k + 1$. □

Remark 5.2 By similar arguments, we can prove that the sequence

$$\Gamma_n \xrightarrow{(c_1 \Psi_{+,n+k_1}^n, c_2 \Psi_{-,n+k_2}^n)} \Gamma_{n+k_1} \oplus \Gamma_{n+k_2} \xrightarrow{c_3 \Psi_{-,n+k_1+k_2}^{n+k_1} + c_4 \Psi_{+,n+k_1+k_2}^{n+k_2}} \Gamma_{n+k_1+k_2},$$

is exact for any $k_1, k_2 \in \mathbb{N}_+$, where the scalars satisfy $c_1 c_3 = -c_2 c_4$.

5.2 The second exact triangle

In this subsection, we prove Proposition 3.6 by diagram chasing. For convenience, we restate it as follows, which is a little stronger than the previous version. Replacing the original knot in the proposition by the dual knot in the Dehn filling of slope $-(m+k)\mu + \lambda$ with framing $-\mu$ and setting $n = -1$ will recover Proposition 3.6.

Proposition 5.3 Suppose

$$l' = \psi_{+,n-1}^\mu \circ \psi_{+,\mu}^{n+1} = \psi_{-,n-1}^\mu \circ \psi_{-,\mu}^{n+1}.$$

Then for any $c_1, c_2, c_3, c_4 \in \mathbb{C} \setminus \{0\}$, the sequence

$$\Gamma_n \oplus \Gamma_n \xrightarrow{c_3 \psi_{-,n+1}^n + c_4 \psi_{+,n+1}^n} \Gamma_{n+1} \xrightarrow{l'} \Gamma_{n-1} \xrightarrow{(c_1 \psi_{-,n-1}^{n-1}, c_2 \psi_{+,n-1}^{n-1})} \Gamma_n \oplus \Gamma_n$$

is exact.

Proof We adopt the conventions from Section 2.3. We will use Lemmas 2.6, 2.11 and 2.12 without mentioning them. We prove the exactness at Γ_{n-1} first. We have

$$\psi_{\pm,n}^{n-1} \circ l' = \psi_{\pm,n}^{n-1} \circ \psi_{\pm,n-1}^{\mu} \circ \psi_{\pm,\mu}^{n+1} = 0.$$

Hence

$$\ker((c_1 \psi_{-,n}^{n-1}, c_2 \psi_{+,n}^{n-1})) \supset \text{Im}(l').$$

Next we prove

$$\ker((c_1 \psi_{-,n}^{n-1}, c_2 \psi_{+,n}^{n-1})) \subset \text{Im}(l').$$

Suppose

$$x \in \ker((c_1 \psi_{-,n}^{n-1}, c_2 \psi_{+,n}^{n-1})) = \ker(\psi_{-,n}^{n-1}) \cap \ker(\psi_{+,n}^{n-1}).$$

By exactness, there exists $y \in \Gamma_{\mu}$ such that $\psi_{+,n-1}^{\mu}(y) = x$. Then we have

$$\psi_{+,n}^{\mu}(y) = \psi_{-,n}^{n-1} \circ \psi_{+,n-1}^{\mu}(y) = \psi_{-,n}^{n-1}(x) = 0.$$

By exactness, there exists $z \in \Gamma_{n+1}$ such that $\psi_{+,\mu}^{n+1}(z) = y$. Thus, we have $l'(z) = x$, which concludes the proof for the exactness at Γ_{n-1} .

Next we prove the exactness at Γ_{n+1} . By exactness, we have

$$\ker(l') \supset \text{Im}(c_3 \psi_{-,n+1}^n + c_4 \psi_{+,n+1}^n) = \text{Im}(\psi_{-,n+1}^n) + \text{Im}(\psi_{+,n+1}^n).$$

Suppose $x \in \ker(l')$. If $\psi_{+,\mu}^{n+1}(x) = 0$, then by the exactness we know $x \in \text{Im}(\psi_{+,n+1}^n)$. If $\psi_{+,\mu}^{n+1}(x) \neq 0$, then by the exactness, there exists $y \in \Gamma_n$ such that

$$\psi_{+,\mu}^n(y) = \psi_{+,\mu}^{n+1}(x).$$

Then we know

$$x - \psi_{-,n+1}^n(y) \in \ker(\psi_{+,\mu}^{n+1}) = \text{Im}(\psi_{+,n+1}^n).$$

Thus, we have

$$x \in \text{Im}(\psi_{-,n+1}^n) + \text{Im}(\psi_{+,n+1}^n),$$

which concludes the proof of the exactness at Γ_{n+1} . \square

6 Some technical constructions

6.1 Filtrations

In this subsection, we study some filtrations on Y and Γ_{μ} that will be important in later sections. We continue to adopt conventions from Section 2.3. In particular, $K \subset Y$ is a rationally null-homologous knot and S is a rational Seifert surface of K .

Lemma 6.1 *The maps G_n in Lemma 2.16 lead to a filtration on Y : for a sufficiently large integer n_0 ,*

$$0 = \ker G_{-n_0} \subset \cdots \subset \ker G_n \subset \ker G_{n+1} \subset \cdots \subset \ker G_{n_0} = Y.$$

Proof It follows from Lemma 2.19 that when n_0 is sufficiently large we have

$$0 = \ker G_{-n_0} \quad \text{and} \quad \ker G_{n_0} = Y.$$

It follows from Lemma 2.18 that for any $n \in \mathbb{Z}$,

$$\ker G_n \subset \ker G_{n+1}. \quad \square$$

Lemma 6.2 *For any $n \in \mathbb{Z}$, the map G_n induces an isomorphism*

$$G_n: (\ker G_{n+1} / \ker G_n) \xrightarrow{\cong} \ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n.$$

Proof Suppose $x \in \ker G_{n+1}$. Then from Lemma 2.18 we know that

$$\psi_{\pm,n+1}^n \circ G_n(x) = G_{n+1}(x) = 0.$$

Hence we have

$$G_n(\ker G_{n+1}) \subset \ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n.$$

Clearly G_n is injective on $\ker G_{n+1} / \ker G_n$ so it suffices to show that the image is $\ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n$. To achieve this, for any element $x \in \ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n$, Lemma 2.20 implies that

$$x \in \ker H_n = \text{Im } G_n.$$

As a result, there exists $\alpha \in Y$ such that

$$x = G_n(\alpha).$$

Again from Lemma 2.18 we know that

$$G_{n+1}(\alpha) = \psi_{+,n}^{n+1} \circ G_n(\alpha) = \psi_{+,n}^{n+1}(x) = 0.$$

This implies that $\alpha \in \ker G_{n+1}$. □

Lemma 6.3 *For any $n \in \mathbb{Z}$, the maps $\psi_{\pm,n}^\mu$ induce isomorphisms*

$$\begin{aligned} \psi_{+,n}^\mu: (\text{Im } \psi_{+,\mu}^{n+2} / \text{Im } \psi_{+,\mu}^{n+1}) &\xrightarrow{\cong} \ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n, \\ \psi_{-,n}^\mu: (\text{Im } \psi_{-,\mu}^{n+2} / \text{Im } \psi_{-,\mu}^{n+1}) &\xrightarrow{\cong} \ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n. \end{aligned}$$

Proof We only prove the lemma for positive bypasses. The proof for the negative bypasses is similar. Let $u \in \text{Im } \psi_{+,\mu}^{n+2}$. By Lemmas 2.6 and 2.12, we have

$$\psi_{+,n+1}^n \circ \psi_{+,\mu}^\mu(u) = 0 \quad \text{and} \quad \psi_{-,n+1}^n \circ \psi_{+,\mu}^\mu(u) = \psi_{+,n+1}^\mu(u) = 0.$$

Hence we know

$$\psi_{+,n}^\mu(\text{Im } \psi_{+,n}^{n+2}) \subset \ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n.$$

Since $\ker \psi_{+,n}^\mu = \text{Im } \psi_{+,n}^{n+1}$, the map $\psi_{+,n}^\mu$ is injective on $\text{Im } \psi_{+,n}^{n+2} / \text{Im } \psi_{+,n}^{n+1}$. To show it is surjective as well, pick $x \in \ker \psi_{+,n+1}^n \cap \ker \psi_{-,n+1}^n$. Note that $x \in \ker \psi_{+,n+1}^n = \text{Im } \psi_{+,n}^\mu$ implies that there exists $u \in \Gamma_\mu$ such that $\psi_{+,n}^\mu(u) = x$. Lemma 2.12 then implies that

$$\psi_{+,n+1}^\mu(u) = \psi_{-,n+1}^n \circ \psi_{+,n}^\mu(u) = \psi_{-,n+1}^n(x) = 0.$$

As a result, $u \in \ker \psi_{+,n+1}^\mu = \text{Im } \psi_{+,n}^{n+2}$. □

Corollary 6.4 (1) For any $n \in \mathbb{Z}$, there is a canonical isomorphism

$$(\ker G_{n+1} / \ker G_n) \cong (\text{Im } \psi_{+,n}^{n+2} / \text{Im } \psi_{+,n}^{n+1}) \cong (\text{Im } \psi_{-,n}^{n+2} / \text{Im } \psi_{-,n}^{n+1}).$$

(2) For sufficiently large n_0 , there exists a (noncanonical) isomorphism

$$Y \cong (\text{Im } \psi_{+,n_0}^{n_0} / \text{Im } \psi_{+,n_0}^{-n_0}) \cong (\text{Im } \psi_{-,n_0}^{n_0} / \text{Im } \psi_{-,n_0}^{-n_0}).$$

Definition 6.5 For any integer $n \in \mathbb{Z}$ and any grading i , define the map F_n^i as the restriction

$$F_n^i = F_n | (\Gamma_n, i),$$

where F_n is the map from Lemma 2.16.

Lemma 6.6 Suppose $n_0 \in \mathbb{Z}$ is small enough that $F_{n_0} = 0$ (cf Lemma 2.19). Then for any integer $n \geq n_0$ and any grading i , we have

$$\psi_{\pm,\mu}^n(\ker F_n^i) = \text{Im}(\text{Proj}_\mu^{i \mp \frac{1}{2}((n-1)p-q)} \circ \psi_{\pm,\mu}^{n_0}),$$

where

$$\text{Proj}_\mu^{i \mp \frac{1}{2}((n-1)p-q)} : \Gamma_\mu \rightarrow \left(\Gamma_\mu, i \mp \frac{(n-1)p-q}{2} \right)$$

is the projection.

Proof We only prove the lemma for positive bypasses and the proof for negative bypasses is similar. First, suppose

$$u \in \text{Im}(\text{Proj}_\mu^{i - \frac{1}{2}((n-1)p-q)} \circ \psi_{+,n_0}^{n_0}) = \text{Im } \psi_{+,n_0}^{n_0} \cap \left(\Gamma_\mu, i - \frac{(n-1)p-q}{2} \right).$$

Pick $x \in (\Gamma_{n_0}, i - \frac{1}{2}(n - n_0)p)$ such that

$$\psi_{+,n_0}^{n_0}(x) = u.$$

Taking $y = \Psi_{-,n}^{n_0}(x)$, we know from Lemma 2.6 that $y \in (\Gamma_n, i)$, from Lemma 2.18 that $F_n(y) = F_{n_0}(x) = 0$, and from Lemma 2.12 that $\psi_{+,n}^n(y) = u$. As a result, we conclude that $u \in \psi_{+,n}^n(\ker F_n^i)$.

Second, suppose $u \in \psi_{+,\mu}^n(\ker F_n^i)$ is nonzero. Pick $x_1 \in \ker F_n^i$ such that

$$\psi_{+,\mu}^n(x_1) = u.$$

By Lemmas 2.5 and 2.6, the fact that $\psi_{+,\mu}^n(x_1) = u \neq 0$ implies that

$$(6-1) \quad -\frac{p - \chi(S)}{2} \leq i - \frac{(n-1)p - q}{2} \leq \frac{p - \chi(S)}{2}.$$

Pick a sufficiently large integer k and then take

$$x_2 = \Psi_{-,n+k}^n(x_1) \quad \text{and} \quad x_3 = \Psi_{+,2n+2k-n_0}^{n+k}(x_2).$$

By Lemma 2.18 we have

$$F_{2n+2k-n_0}(x_3) = F_{n+k}(x_2) = F_n(x_1) = 0.$$

Note that the grading j of x_3 equals

$$(6-2) \quad j = i + \frac{kp}{2} - \frac{(n+k-n_0)p}{2} = i - \frac{(n-n_0)p}{2}.$$

Combining (6-1) and (6-2), we obtain

$$\frac{(n_0 - 2)p - q - \chi(S)}{2} \leq j \leq \frac{n_0 p - q - \chi(S)}{2}.$$

Note that we pick k to be a sufficiently large integer. In particular, we can assume

$$\begin{aligned} \frac{-(2n + 2k - n_0)p + q + \chi(S)}{2} + 1 &\leq \frac{(n_0 - 2)p - q - \chi(S)}{2}, \\ \frac{n_0 p - q - \chi(S)}{2} &\leq \frac{(2n + 2k - n_0)p - q + \chi(S)}{2}. \end{aligned}$$

Thus j is in this range as well, and Lemma 2.19 implies that $F_{2n+2k-n_0}$ is injective on the grading j . Hence $x_3 = 0$. Then the following lemma (Lemma 6.7) applies to $(x, y) = (x_2, 0)$, and there exists $x_4 \in \Gamma_{n_0}$ such that

$$\Psi_{-,n+k}^{n_0}(x_4) = x_2.$$

Thus by Lemma 2.12,

$$u = \psi_{+,\mu}^n(x_1) = \psi_{+,\mu}^{n+k}(x_2) = \psi_{+,\mu}^{n_0}(x_4) \in \text{Im}(\text{Proj}_\mu^{i-\frac{1}{2}((n-1)p-q)} \circ \psi_{+,\mu}^{n_0}). \quad \square$$

Lemma 6.7 Suppose $n \in \mathbb{Z}$ and $k_1, k_2 \in \mathbb{N}_+$. Suppose $x \in \Gamma_{n+k_1}$ and $y \in \Gamma_{n+k_2}$ are such that

$$\Psi_{+,n+k_1+k_2}^{n+k_1}(x) = \Psi_{-,n+k_1+k_2}^{n+k_2}(y).$$

Then there exists $z \in \Gamma_n$ such that

$$\Psi_{-,n+k_1}^n(z) = x \quad \text{and} \quad \Psi_{+,n+k_2}^n(z) = y.$$

Proof This is a restatement of Remark 5.2. The proof is similar to that of Proposition 5.1. □

6.2 Tau invariants in a general 3-manifold

Definition 6.8 An element $\alpha \in Y$ is called *homogeneous* if there exists an $n \in \mathbb{Z}$ and a grading i such that $\alpha \in \text{Im } F_n^i$. Note that from Lemma 2.18 and Corollary 2.9, we know that

$$\alpha \in \text{Im } F_n^i \implies \alpha \in \text{Im } F_{n+1}^{i \pm \frac{1}{2}p}.$$

For a homogeneous element $\alpha \in Y$, we pick a sufficiently large n_0 and define

$$\begin{aligned} \tau^+(\alpha) &:= \max_i \{i \mid F_{n_0}(x) = \alpha \text{ for some } x \in (\Gamma_{n_0}, i)\} - \frac{(n_0 - 1)p - q}{2}, \\ \tau^-(\alpha) &:= \min_i \{i \mid F_{n_0}(x) = \alpha \text{ for some } x \in (\Gamma_{n_0}, i)\} + \frac{(n_0 - 1)p - q}{2}, \\ \tau(\alpha) &:= 1 + \frac{\tau^-(\alpha) - \tau^+(\alpha) + q}{p} = \frac{\min - \max}{p} + n_0. \end{aligned}$$

We will prove the independence of these τ invariants about n_0 later in Lemma 6.12.

Remark 6.9 Here we fix the knot $K \subset Y$ and define the tau invariants for a homogeneous element $\alpha \in I^\sharp(Y)$. The reason we go in this order is because

- (1) currently the definition of homogeneous elements depends on the choice of the knot, and
- (2) in this paper we only focus on the Dehn surgeries of a fixed knot.

Remark 6.10 The normalization $\mp \frac{1}{2}((n_0 - 1)p - q)$ comes from the grading shifts of $\psi_{\pm, \mu}^{n_0}$ in Lemma 2.6. When K is a knot inside $Y = S^3$, we have that $\tau^\pm(\alpha)$ is equal to the tau invariant $\tau_I(K)$ defined in [Ghosh et al. 2024], where α is the unique generator of $I^\sharp(-S^3) \cong \mathbb{C}$ up to a scalar. Then $\tau(\alpha) = 1 - 2\tau_I(K)$.

Lemma 6.11 *We have the following properties.*

- (1) Suppose n_1, n_2 are two integers and i_1, i_2 are two gradings such that there exist $x_1 \in (\Gamma_{n_1}, i_1)$ and $x_2 \in (\Gamma_{n_2}, i_2)$ with

$$F_{n_1}(x_1) = F_{n_2}(x_2) \neq 0.$$

Then there exists an integer N such that

$$i_2 = i_1 - \frac{(n_2 - n_1)p}{2} + Np,$$

ie when we send x_1 and x_2 into the same Γ_{n_3} with $n_3 > n_1, n_2$ by bypass maps, then the difference of the expected gradings of the images is divisible by p (the grading-shifts of the bypass maps $\psi_{\pm, n+1}^n$ are $\mp p/2$).

- (2) Suppose we have an integer n_1 , a grading i_1 and an element $x_1 \in (\Gamma_{n_1}, i_1)$. Then for any integer $n_2 \geq n_1$ and grading i_2 such that there exists an integer $N \in [0, n_2 - n_1]$ with

$$i_2 = i_1 - \frac{(n_2 - n_1)p}{2} + Np,$$

there exists an element $x_2 \in (\Gamma_{n_2}, i_2)$ such that

$$F_{n_1}(x_1) = F_{n_2}(x_2).$$

- (3) Suppose $n \in \mathbb{Z}$ and for $1 \leq j \leq l$ we have a grading i_j and an element $x_j \in (\Gamma_n, i_j)$ such that $F_n(x_1), \dots, F_n(x_l)$ are linearly independent. Then the element

$$\alpha = \sum_{j=1}^l F_n(x_j)$$

is homogeneous if and only if for any $1 \leq j \leq l$, we have

$$i_j \equiv i_1 \pmod{p}.$$

Proof (1) Take n_0 a sufficiently large integer. For $j = 1, 2$, take $i'_j \in (-\frac{1}{2}p, \frac{1}{2}p]$ to be the unique grading such that there exists an integer N_j with

$$i'_j = i_j - \frac{(n_0 - n_j)p}{2} + N_j p.$$

Take

$$x'_j = \Psi_{+,n_0}^{n_j+N_j} \circ \Psi_{-,n_j+N_j}^{n_j}(x_j).$$

From Lemma 2.18 we know that

$$x'_j \in (\Gamma_{n_0}, i'_j) \quad \text{and} \quad F_{n_0}(x'_1) = F_{n_1}(x_1) = F_{n_2}(x_2) = F_{n_0}(x'_2).$$

By Lemma 2.19, we know that $x'_1 = x'_2$ and in particular, $i'_1 = i'_2$. As a result, we can take $N = N_1 - N_2$, and then it is straightforward to verify that

$$i_2 = i_1 - \frac{(n_2 - n_1)p}{2} + Np.$$

- (2) We can take

$$x_2 = \Psi_{-,n_2}^{n_1+N} \circ \Psi_{+,n_1+N}^{n_1}(x_1).$$

Then it follows from Lemma 2.6 that $x_2 \in (\Gamma_{n_2}, i_2)$, and it follows from Lemma 2.18 that

$$F_{n_2}(x_2) = F_{n_1}(x_1).$$

- (3) The proof is similar to that of (1). □

Lemma 6.12 For a homogeneous element α , we have the following:

- (1) $\tau^\pm(\alpha)$ and hence $\tau(\alpha)$ are well-defined, ie they are independent of the choice of the large integer n_0 .
- (2) We have $\tau(\alpha) \in \mathbb{Z}$.

(3) For any integer n and grading i , the following two statements are equivalent.

- (a) There exists $x \in (\Gamma_n, i)$ such that $F_n(x) = \alpha$.
- (b) We have $n \geq \tau(\alpha)$ and there exists $N \in \mathbb{Z}$ such that $N \in [0, n - \tau(\alpha)]$ and

$$i = \frac{1}{2}(\tau^+(\alpha) + \tau^-(\alpha) - (n - \tau(\alpha))p) + Np.$$

(4) We have

$$\tau^+(\alpha) \geq -\frac{p - \chi(S)}{2} \quad \text{and} \quad \tau^-(\alpha) \leq \frac{p - \chi(S)}{2}.$$

Proof (1) Suppose α is a homogeneous element. Then by definition there exists $x \in (\Gamma_n, i)$ for some integer n and grading i such that

$$F_n(x) = \alpha.$$

Then for sufficiently large n_0 , we can take

$$y = \psi_{+,n_0}^n(x)$$

and then Lemma 2.18 implies that

$$F_{n_0}(y) = \alpha,$$

and hence $\tau^\pm(\alpha)$ exists.

To show that the value of $\tau^\pm(\alpha)$ is independent of n_0 as long as it is sufficiently large, a combination of Lemmas 2.5 and 2.6 implies that the map

$$\psi_{-,n_0+1}^{n_0} : (\Gamma_{n_0}, i) \rightarrow (\Gamma_{n_0+1}, i + \frac{1}{2}p)$$

is an isomorphism for any $i > g - \frac{1}{2}(n_0p - q - 1)$. Then Lemma 2.18 implies that τ^+ is well-defined. The argument for τ^- is similar.

(2) It follows directly from Lemma 6.11 part (1).

(3) We first establish the following claim.

Claim *There exists an element*

$$z \in (\Gamma_{\tau(\alpha)}, \frac{1}{2}(\tau^+(\alpha) + \tau^-(\alpha))) \quad \text{such that} \quad F_{\tau(\alpha)}(z) = \alpha.$$

Proof Suppose $n_0 \in \mathbb{Z}$ is sufficiently large and

$$x_\pm \in \left(\Gamma_{n_0}, \tau^\pm(\alpha) \pm \frac{(n_0 - 1)p - q}{2} \right)$$

such that $F_{n_0}(x_\pm) = \alpha$. Note that the existence of x_\pm follows from the definition of $\tau^\pm(\alpha)$. Let

$$x'_\pm = \Psi_{\pm, 2n_0 - \tau(\alpha)}^n(x_\pm).$$

It follows from Lemma 2.6 that

$$x'_\pm \in \left(\Gamma_{2n_0 - \tau(\alpha)}, \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} \right).$$

From Lemma 2.18 we know that

$$F_{2n_0-\tau(\alpha)}(x'_+) = \alpha = F_{2n_0-\tau(\alpha)}(x'_-).$$

By Lemma 2.19 this implies that

$$x'_+ = x'_-.$$

Hence Lemma 6.7 applies and there exists $z \in \Gamma_{\tau(\alpha)}$ such that

$$\Psi_{\mp, n}^{\tau(\alpha)}(z) = x_{\pm}.$$

Again Lemma 2.6 implies that z is in the grading

$$z \in \left(\Gamma_{\tau(\alpha)}, \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} \right),$$

and Lemma 2.18 implies

$$F_{\tau(\alpha)}(z) = \alpha. \quad \square$$

Now if an integer n and a grading i satisfy statement (b), then (a) is a direct consequence of the above claim and Lemma 6.11(2).

It remains to show that (a) implies (b). Suppose there exists $x \in (\Gamma_n, i)$ such that $F_n(x) = \alpha$. From the above claim, we already know that there exists

$$z \in \left(\Gamma_{\tau(\alpha)}, \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} \right) \quad \text{such that} \quad F_{\tau(\alpha)}(z) = \alpha.$$

Hence Lemma 6.11(1) implies that there exists $N \in \mathbb{Z}$ such that

$$i = \frac{\tau^+(\alpha) + \tau^-(\alpha) - (n - \tau(\alpha))p}{2} + Np.$$

If $N > n - \tau(\alpha)$, we can take a sufficiently large n_0 and

$$x' = \Psi_{-, n_0}^n(x).$$

It follows from Lemma 2.6 that

$$x' \in (\Gamma_{n_0}, i') \quad \text{with} \quad i' > \tau^+(\alpha) + \frac{(n_0 - 1)p - q}{2}.$$

Then Lemma 2.18 implies that

$$F_{n_0}(x') = \alpha,$$

which contradicts the definition of τ^+ in Definition 6.8. Similarly, if $N < 0$ we can take

$$x' = \Psi_{+, n_0}^n(x),$$

which would be an element contradicting the definition of τ^- . When $n < \tau(\alpha)$ we have $n - \tau(\alpha) < 0$, so there is always a contradiction by the above argument. This concludes (b).

(4) It follows from the definition of τ^{\pm} and Lemma 2.19 that F_{n_0} is an isomorphism when restricted to the direct sum of p consecutive middle gradings of Γ_{n_0} when n_0 is large. □

Lemma 6.13 For any $n \in \mathbb{Z}$ we have that

$$\text{Im } F_n = \text{Span}\{\alpha \in Y \mid \alpha \text{ is homogeneous and } \tau(\alpha) \leq n\}.$$

Proof Suppose $\alpha \in \text{Im } F_n$. Let

$$\alpha = \sum_i \alpha_i, \quad \text{where } \alpha_i \in \text{Im } F_n^i \text{ is homogeneous.}$$

From Lemma 6.12 we know that $\tau(\alpha_i) \leq n$ for all i . On the other hand, suppose

$$\alpha = \sum_i \alpha_i, \quad \text{where } \tau(\alpha_i) \leq n \text{ for all } i.$$

By Lemma 6.12(3), we can pick $z_i \in \Gamma_{\tau(\alpha_i)}$ such that

$$F_{\tau(\alpha_i)}(z_i) = \alpha_i.$$

Then from Lemma 2.18 we know

$$\alpha = F_n \left(\sum_i \Psi_{+,n}^{\tau(\alpha_i)}(z_i) \right). \quad \square$$

6.3 A basis for framed instanton homology

We pick a basis \mathfrak{B} for Y as follows. First

$$\mathfrak{B} = \bigcup_{n \in \mathbb{Z}} \mathfrak{B}_n.$$

To construct the set \mathfrak{B}_n , first, let $\mathfrak{B}_n = \emptyset$ if $F_n = 0$. By Lemma 2.19 this means $\mathfrak{B}_n = \emptyset$ for all small enough n . Write

$$\mathfrak{B}_{\leq n} = \bigcup_{k \leq n} \mathfrak{B}_k.$$

We pick the set \mathfrak{B}_n inductively. Note that we have taken $\mathfrak{B}_n = \emptyset$ for n with $F_n = 0$. Suppose we have already constructed the set $\mathfrak{B}_{\leq n-1}$ that consists of homogeneous elements and is a basis of $\text{Im } F_{n-1}$. We pick the set \mathfrak{B}_n such that \mathfrak{B}_n consists of homogeneous elements with $\tau = n$, and the set

$$\mathfrak{B}_{\leq n} = \mathfrak{B}_{\leq n-1} \cup \mathfrak{B}_n$$

forms a basis of $\text{Im } F_n$. Note that Lemma 6.13 implies that \mathfrak{B}_n exists and

$$|\mathfrak{B}_n| = \dim_{\mathbb{C}}(\text{Im } F_n / \text{Im } F_{n-1}).$$

For any $n, k \in \mathbb{Z}$ such that $k \leq n - 2$, define maps

$$\eta_{\pm,k}^n : \mathfrak{B}_n \rightarrow \Gamma_k$$

as follows: for any $\alpha \in \mathfrak{B}_n \subset \text{Im } F_n$, since α is homogeneous and $\tau(\alpha) = n$, we can pick

$$z \in \left(\Gamma_n, \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} \right)$$

by Lemma 6.12(3) such that $F_n(z) = \alpha$. Then define

$$\eta_{\pm,k}^n(\alpha) = \psi_{\pm,k}^\mu \circ \psi_{\pm,\mu}^n(z).$$

Lemma 6.14 Suppose $n, k \in \mathbb{Z}$ are such that $k \leq n - 2$.

- (1) The maps $\eta_{\pm, k}^n$ are all well-defined.
- (2) We have $\eta_{+, n-2}^n = c_n \cdot \eta_{-, n-2}^n$ for some scalar $c_n \in \mathbb{C} \setminus \{0\}$.
- (3) Elements in $\text{Im } \eta_{\pm, k}^n \subset \Gamma_k$ are linearly independent.
- (4) $\text{Im } \eta_{\pm, n-2}^n$ forms a basis for $\ker \psi_{+, n-1}^{n-2} \cap \ker \psi_{-, n-1}^{n-2}$.
- (5) For any $\alpha \in \mathfrak{B}_n$, we have

$$\eta_{\pm, k}^n(\alpha) \in \left(\Gamma_k, \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} \mp \frac{(n-2-k)p}{2} \right).$$

- (6) We have

$$\psi_{\mp, k}^{k-1} \circ \eta_{\pm, k-1}^n = \eta_{\pm, k}^n \quad \text{and} \quad \psi_{\pm, k}^{k-1} \circ \eta_{\pm, k-1}^n = 0.$$

Proof (1) We only work with $\eta_{+, k}^n$, and the arguments for $\eta_{-, k}^n$ are similar. Suppose there are $z_1, z_2 \in (\Gamma_n, i)$ such that $F_n(z_1) = F_n(z_2) = \alpha$, where $i = \frac{1}{2}(\tau^+(\alpha) + \tau^-(\alpha))$. Then

$$z_1 - z_2 \in \ker F_n^i,$$

and by Lemma 6.6 we have

$$\psi_{+, \mu}^n(z_1 - z_2) \in \psi_{+, \mu}^n(\ker F_n^i) \subset \text{Im } \psi_{+, \mu}^{n_0} \subset \text{Im } \psi_{+, \mu}^{k+1}.$$

Here $n_0 \in \mathbb{Z}$ is a small enough integer. As a result,

$$\eta_{+, k}^n(\alpha) = \psi_{+, k}^\mu \circ \psi_{+, \mu}^n(z_1) = \psi_{+, k}^\mu \circ \psi_{+, \mu}^n(z_2)$$

is well-defined.

- (2) This follows directly from Lemma 2.11. Note that in Section 2.3 we do not fix the scalars of the second commutative diagram of Lemma 2.11, and hence a nonzero coefficient c_n would possibly arise.

- (3) We only work with $\eta_{+, k}^n$ and the arguments for $\eta_{-, k}^n$ are similar. Suppose

$$\mathfrak{B}_n = \{\alpha_1, \dots, \alpha_l\},$$

where $l = |\mathfrak{B}_n| = \dim_{\mathbb{C}}(\text{Im } F_n / \text{Im } F_{n-1})$.

Suppose there exists $\lambda_1, \dots, \lambda_l$ such that

$$\sum_{j=1}^l \lambda_j \cdot \eta_{+, k}^n(\alpha_j) = 0.$$

Pick $z_j \in (\Gamma_n, \frac{1}{2}(\tau^+(\alpha_j) + \tau^-(\alpha_j)))$ such that $F_n(z_j) = \alpha_j$. Then we have

$$\psi_{+, k}^\mu \circ \psi_{+, \mu}^n \left(\sum_{j=1}^l \lambda_j z_j \right) = 0.$$

As a result, there exists $x \in \Gamma_{k+1}$ such that

$$\psi_{+,\mu}^{k+1}(x) = \psi_{+,\mu}^n \left(\sum_{j=1}^l \lambda_j z_j \right).$$

Note that, from Lemma 2.12, we know that

$$\psi_{+,\mu}^n \circ \Psi_{-,n}^{k+1}(x) = \psi_{+,\mu}^{k+1}(x) = \psi_{+,\mu}^n \left(\sum_{j=1}^l \lambda_j z_j \right),$$

so as a result there exists $y \in \Gamma_{n-1}$ such that

$$\sum_{j=1}^l \lambda_j z_j = \Psi_{-,n}^{k+1}(x) + \psi_{+,\mu}^{n-1}(y).$$

Hence by Lemma 2.18 we have

$$\sum_{j=1}^l \lambda_j \alpha_j = F_n \left(\sum_{j=1}^l \lambda_j z_j \right) = F_n \circ \Psi_{-,n}^{k+1}(x) + F_n \circ \psi_{+,\mu}^{n-1}(y) = F_{k+1}(x) + F_{n-1}(y) \subset \text{Im } F_{n-1}.$$

Since α_j form a basis of \mathfrak{B}_n , the sum cannot be in $\text{Im } F_{n-1}$ except $\lambda_i = 0$ for all i .

(4). For $\alpha \in \mathfrak{B}_n$, pick $z \in (\Gamma_n, \frac{1}{2}(\tau^+(\alpha) + \tau^-(\alpha)))$ such that $F_n(z) = \alpha$. Then by definition

$$\eta_{+,n-2}^n(\alpha) = \psi_{+,n-2}^\mu \circ \psi_{+,\mu}^n(z).$$

Now we can compute

$$\psi_{+,n-1}^{n-2} \circ \eta_{+,n-2}^n(\alpha) = \psi_{+,n-1}^{n-2} \circ \psi_{+,n-2}^\mu \circ \psi_{+,\mu}^n(z) = 0,$$

and by Lemma 2.12

$$\psi_{-,n-1}^{n-2} \circ \eta_{+,n-2}^n(\alpha) = \psi_{-,n-1}^{n-2} \circ \psi_{+,n-2}^\mu \circ \psi_{+,\mu}^n(z) = \psi_{+,n-1}^\mu \circ \psi_{+,\mu}^n(z) = 0.$$

Hence

$$\eta_{+,n-2}^n(\alpha) \in \ker \psi_{+,n-1}^{n-2} \cap \ker \psi_{-,n-1}^{n-2}.$$

Then (4) follows from (3), Lemma 6.2, and $\text{Im } F_n = \ker G_{n-1}$.

(5) It follows directly from the construction of $\eta_{\pm,k}^n$ and Lemma 2.6.

(6) It follows from the construction of $\eta_{\pm,k}^n$, the commutativity in Lemma 2.12 and the exactness in Lemma 2.6. □

Convention We can define

$$\tilde{\eta}_{+,k}^n = \eta_{+,k}^n \quad \text{and} \quad \tilde{\eta}_{-,k}^n = c_n \cdot \eta_{-,k}^n \quad \text{such that} \quad \tilde{\eta}_{+,n-2}^n = \tilde{\eta}_{-,n-2}^n,$$

and the new maps satisfy all properties in Lemma 6.14 except (2). We will use $\eta_{+,k}^n$ to denote $\tilde{\eta}_{+,k}^n$ in latter sections.

7 The map in the third exact triangle

In this section, we construct the map l in Propositions 3.9 and 3.12 and show it satisfies the exactness and the commutative diagram. We continue to adopt conventions from Section 2.3. We restate the propositions as follows, and no longer use the notation l, l' for maps.

Proposition 7.1 *Suppose $n \in \mathbb{Z}$ is fixed and $k \in \mathbb{Z}$ is sufficiently large. Then there is an exact triangle*

$$\begin{array}{ccc}
 \Gamma_n & \xrightarrow{\Phi_{n+k}^n} & \Gamma_{n+k} \oplus \Gamma_{n+k} \\
 & \swarrow \Phi_n^{n+2k} & \nwarrow \Phi_{n+2k}^{n+k} \\
 & \Gamma_{n+2k} &
 \end{array}$$

where two of the maps are already constructed:

$$\begin{aligned}
 \Phi_{n+k}^n &:= (\Psi_{+,n+k}^n, \Psi_{-,n+k}^n): \Gamma_n \rightarrow \Gamma_{n+k} \oplus \Gamma_{n+k}, \\
 \Phi_{n+2k}^{n+k} &:= \Psi_{-,n+2k}^{n+k} - \Psi_{+,n+2k}^{n+k}: \Gamma_{n+k} \oplus \Gamma_{n+k} \rightarrow \Gamma_{n+2k}.
 \end{aligned}$$

Proposition 7.2 *Suppose $n \in \mathbb{Z}$ is fixed and $k \in \mathbb{Z}$ is sufficiently large. Suppose Φ_{n-1}^{n+2k-1} is constructed in Proposition 7.1. Then, there are two commutative diagrams up to scalars:*

$$\begin{array}{ccc}
 \Gamma_{\frac{1}{2}(2n+2k+1)} & \xrightarrow{\psi_{+,\mu}^{n+k+1} \circ \psi_{-,n+k+1}^{\frac{1}{2}(2n+2k+1)}} & \Gamma_\mu \\
 \downarrow \Psi_{+,n+2k}^{n+k+1} \circ \psi_{-,n+k+1}^{\frac{1}{2}(2n+2k+1)} & & \downarrow \psi_{+,\mu}^\mu \\
 \Gamma_{n+2k} & \xrightarrow{\Phi_n^{n+2k}} & \Gamma_n
 \end{array}
 \qquad
 \begin{array}{ccc}
 \Gamma_{\frac{1}{2}(2n+2k+1)} & \xrightarrow{\psi_{-,\mu}^{n+k+1} \circ \psi_{+,n+k+1}^{\frac{1}{2}(2n+2k+1)}} & \Gamma_\mu \\
 \downarrow \Psi_{-,n+2k+1}^{n+k+1} \circ \psi_{+,n+k+1}^{\frac{1}{2}(2n+2k+1)} & & \downarrow \psi_{-,\mu}^\mu \\
 \Gamma_{n+2k} & \xrightarrow{\Phi_n^{n+2k}} & \Gamma_n
 \end{array}$$

7.1 Characterizations of the kernel and the image

Before constructing Φ_n^{n+2k} , we characterize the spaces $\ker \Phi_{n+k}^n$ and $\text{Im } \Phi_{n+2k}^{n+k}$. These results will motivate the construction of Φ_n^{n+2k} to ensure that

$$\text{Im } \Phi_n^{n+2k} = \ker \Phi_{n+k}^n \quad \text{and} \quad \ker \Phi_n^{n+2k} = \text{Im } \Phi_{n+2k}^{n+k}.$$

Since Φ_{n+k}^n and Φ_{n+2k}^{n+k} are constructed using bypass maps, it suffices to consider their restrictions on each grading.

Lemma 7.3 *Suppose $n \in \mathbb{Z}$ is fixed and $k \in \mathbb{Z}$ is sufficiently large. Let*

$$\text{Proj}_n^i: \Gamma_n \rightarrow (\Gamma_n, i)$$

be the projection. Then we have

$$\ker \Phi_{n+k}^n \cap (\Gamma_n, i) = \text{Im}(\text{Proj}_n^i \circ G_n).$$

Proof We need to apply Lemma 2.20. Following conventions in Section 2.3, we have

$$(7-1) \quad H_n = \psi_{+,n+1}^n - \psi_{-,n+1}^n.$$

Suppose $x \in \text{Im}(\text{Proj}_n^i \circ G_n)$. Pick $\alpha \in Y$ and $y \in \Gamma_n$ such that

$$G_n(\alpha) = x + y,$$

where $\text{Proj}_n^i(y) = 0$. When k is sufficiently large, we know from Lemma 2.19 that

$$G_{n+k} \equiv 0.$$

In particular, from Lemma 2.18,

$$\Psi_{\pm,n+k}^n(x) + \Psi_{\pm,n+k}^n(y) = G_{n+k}(\alpha) = 0.$$

Since the maps $\Psi_{\pm,n+k}^n$ are homogeneous, we know that

$$\Psi_{\pm,n+k}^n(x) = 0,$$

which implies that $x \in \ker \Phi_{n+k}^n \cap (\Gamma_n, i)$.

Next, suppose $x \in \ker \Phi_{n+k}^n \cap (\Gamma_n, i)$. We take $x_n^i = x$ and we will pick $x_n^j \in (\Gamma_n, j)$ for all $j \neq i$ such that

$$\sum_j x_n^j \in \ker H_n = \text{Im } G_n.$$

We will use the notation x_a^b to denote an element in (Γ_a, b) . Recall that from Lemma 2.6, the grading shifts of $\psi_{\pm,n+1}^n$ are $\mp \frac{1}{2}p$. Take

$$x_{n+k-1}^{i+\frac{1}{2}(k-1)p} = \Psi_{-,n+k-1}^n(x) \text{ and } x_{n+k-1}^{i+\frac{1}{2}(k+1)p} = 0.$$

Since $x \in \ker \Phi_{n+k}^n \cap (\Gamma_n, i)$ we know that

$$(7-2) \quad \psi_{-,n+k}^{n+k-1}(x_{n+k-1}^{i+\frac{1}{2}(k-1)p}) = \Psi_{-,n+k}^n(x) = 0 = \psi_{+,n+k}^{n+k-1}(x_{n+k-1}^{i+\frac{1}{2}(k+1)p}).$$

Hence, from Lemma 6.7, there exists

$$x_{n+k-2}^{i+\frac{1}{2}kp} \in (\Gamma_{n+k-2}, i + \frac{1}{2}kp)$$

such that

$$\psi_{-,n+k-2}^{n+k-1}(x_{n+k-2}^{i+\frac{1}{2}kp}) = x_{n+k-1}^{i+\frac{1}{2}(k+1)p} = 0 \text{ and } \psi_{+,n+k-1}^{n+k-2}(x_{n+k-2}^{i+\frac{1}{2}kp}) = x_{n+k-1}^{i+\frac{1}{2}(k-1)p}.$$

Then we can take

$$x_{n+k-2}^{i+\frac{1}{2}(k+2)p} = 0 \text{ and } x_{n+k-2}^{i+\frac{1}{2}(k-2)p} = \Psi_{-,n+k-2}^n(x).$$

We can apply the same argument and use Lemma 6.7 to find

$$x_{n+k-3}^{i+\frac{1}{2}(k+3)p}, \quad x_{n+k-3}^{i+\frac{1}{2}(k+1)p}, \quad x_{n+k-3}^{i+\frac{1}{2}(k-1)p}, \quad x_{n+k-3}^{i+\frac{1}{2}(k-3)p} \in \Gamma_{n+k-3}$$

such that the $\psi_{\pm, n+k-2}^{n+k-3}$ send them to corresponding elements in Γ_{n+k-2} . Repeating this argument, we can obtain elements

$$x_n^{i+pj} \in (\Gamma_n, i + pj) \quad \text{for } j \in [1, k] \cap \mathbb{Z}$$

such that

$$x_n^i = x, \quad \psi_{-, n+1}^n(x_n^{i+pk}) = 0, \quad \psi_{+, n+1}^n(x_n^{i-pk}) = 0,$$

and for any $j \in [1, k-1] \cap \mathbb{Z}$ we have

$$\psi_{-, n+1}^n(x_n^{i+pj}) = \psi_{+, n+1}^n(x_n^{i+p(j+1)}).$$

Note that we obtain the above x_n^{i+pj} for $j \in [1, k] \cap \mathbb{Z}$ essentially from the fact that $\Psi_{-, n+k}^n(x) = 0$ as in equation (7-2). However, $x \in \ker \Phi_{n+k}^n$ so we have $\Psi_{+, n+k}^n(x) = 0$ as well. A similar argument as above then yields

$$x_n^{i+pj} \in (\Gamma_n, i + pj) \quad \text{for } j \in [-k, -1] \cap \mathbb{Z}.$$

Together with $x_n^i = x$, we obtain x_n^{i+pj} for all $j \in [-k, k] \cap \mathbb{Z}$.

It is then straightforward to check that

$$y = \sum_{j=-k}^k x_n^{i+pj} \in \ker(\psi_{+, n+1}^n - \psi_{-, n+1}^n) = \ker H_n = \text{Im } G_n. \quad \square$$

Lemma 7.4 Suppose $\alpha \in Y$ is a homogeneous element and

$$\alpha = \sum_{j=1}^l \lambda_j \cdot \alpha_j,$$

where $\lambda_j \neq 0$ and $\alpha_j \in \mathfrak{B}$ for $1 \leq j \leq l$. Let n be an integer, i be a grading and k be a sufficiently large integer. For an element $x \in (\Gamma_{n+2k}, i)$ such that $F_{n+2k}(x) = \alpha$, the following is true.

(1) We have

$$\tau^+(\alpha) = \min_{1 \leq j \leq l} \{\tau^+(\alpha_j)\} \quad \text{and} \quad \tau^-(\alpha) = \max_{1 \leq j \leq l} \{\tau^-(\alpha_j)\}.$$

(2) We have $x \in \text{Im } \Phi_{n+2k}^{n+k}$ if and only if for any $1 \leq j \leq l$, at least one of the following inequalities holds:

$$i \geq \tau^-(\alpha_j) - \frac{(n-1)p - q}{2},$$

$$i \leq \tau^+(\alpha_j) + \frac{(n-1)p - q}{2}.$$

(3) If $x \notin \text{Im } \Phi_{n+2k}^{n+k}$ then there exist $j, N \in \mathbb{Z}$ such that $1 \leq j \leq l, 0 \leq N \leq \tau(\alpha_j) - n - 2$ and

$$i = \tau^+(\alpha_j) + \frac{(n-1)p - q}{2} + (N + 1)p.$$

Proof (1) We only demonstrate the proof of the result for τ^+ and the proof for τ^- is similar. First, we make the following two claims.

Claim 1 For any homogeneous elements (not necessarily elements in \mathfrak{B}) α_1 and α_2 such that $\alpha_1 + \alpha_2$ is also homogeneous, if $\tau^+(\alpha_1) > \tau^+(\alpha_2)$, then $\tau^+(\alpha_1 + \alpha_2) = \tau^+(\alpha_2)$.

To prove Claim 1, let n_0 be sufficiently large. From Lemma 6.11(3) we know that

$$(7-3) \quad \tau^+(\alpha_1) \equiv \tau^+(\alpha_2) \equiv \tau^+(\alpha_1 + \alpha_2) \pmod{p}.$$

Assume $\tau^+(\alpha_1 + \alpha_2) > \tau^+(\alpha_2)$. Let

$$\tau^+ = \min\{\tau^+(\alpha_1), \tau^+(\alpha_1 + \alpha_2)\} > \tau^+(\alpha_2).$$

We claim that there exist

$$x_1, x_3 \in \left(\Gamma_{n_0}, \tau^+ + \frac{(n_0 - 1)p - q}{2} \right)$$

such that

$$F_{n_0}(x_1) = \alpha_1 \quad \text{and} \quad F_{n_0}(x_3) = \alpha_1 + \alpha_2.$$

We prove only the existence of x_1 , and the argument for the existence of x_3 is similar. By Definition 6.8, we know that

$$\tau^+(\alpha_1) + \frac{(n_0 - 1)p - q}{2} = \frac{\tau^+(\alpha_1) + \tau^-(\alpha_1) - (n_0 - \tau(\alpha_1))p}{2} + (n_0 - \tau(\alpha_1))p.$$

Taking

$$N = (n_0 - \tau(\alpha_1)) - \frac{1}{p}(\tau^+(\alpha_1) - \tau^+),$$

we know that

$$(7-4) \quad \tau^+ + \frac{(n_0 - 1)p - q}{2} = \frac{\tau^+(\alpha_1) + \tau^-(\alpha_1) - (n_0 - \tau(\alpha_1))p}{2} + Np.$$

Equation (7-3) implies that $N \in \mathbb{Z}$. The definition of τ^+ makes sure that $N \leq n_0 - \tau(\alpha_1)$. The fact that n_0 is sufficiently large and Lemma 6.12(4) implies that $N \geq 0$. Hence Lemma 6.12(3) implies the existence of x_1 such that

$$x_1 \in \left(\Gamma_{n_0}, \tau^+ + \frac{(n_0 - 1)p - q}{2} \right) \quad \text{and} \quad F_{n_0}(x_1) = \alpha_1.$$

Now the existence of x_1 and x_3 implies that

$$F_n(x_3 - x_1) = \alpha_2,$$

which contradicts the definition of $\tau^+(\alpha_2)$.

Claim 2 Suppose $\alpha_1, \dots, \alpha_u \in \mathfrak{B}$ are pairwise distinct elements in \mathfrak{B} such that

$$\tau^+(\alpha_1) = \tau^+(\alpha_2) = \dots = \tau^+(\alpha_u) = \tau^+.$$

Suppose that

$$\alpha' = \sum_{i=1}^u \lambda_i \cdot \alpha_i$$

and suppose it is homogeneous. Then $\tau^+(\alpha') = \tau^+$.

To prove Claim 2, assume that $\tau^+(\alpha') > \tau^+$. Without loss of generality, assume that $\lambda_1 \neq 0$ and

$$\tau^-(\alpha_1) = \min_{1 \leq j \leq u} \{\tau^-(\alpha_j)\}.$$

Then a similar argument as in the proof of Claim 1 implies that

$$\tau^-(\alpha') \leq \tau^-(\alpha_1).$$

Note that we have assumed $\tau^+(\alpha') > \tau^+ = \tau^+(\alpha_1)$. Hence by Definition 6.8, $\tau(\alpha') < \tau(\alpha_1)$, which contradicts the construction of the set \mathfrak{B} .

Now we prove part (1). Suppose $\alpha_1, \dots, \alpha_l \in \mathfrak{B}$ are pairwise distinct elements in \mathfrak{B} . Let

$$\alpha = \sum_{j=1}^l \lambda_j \cdot \alpha_j.$$

We want to show that

$$\tau^+(\alpha) = \min_{1 \leq j \leq l} \{\tau^+(\alpha_j)\}.$$

To do this, relabel the elements α_j if necessary so that

$$\tau^+(\alpha_1) = \tau^+(\alpha_2) = \dots = \tau^+(\alpha_u) < \tau^+(\alpha_{u+1}) \leq \tau^+(\alpha_{u+2}) \leq \dots \leq \tau^+(\alpha_l).$$

Since α is homogeneous, from Lemma 6.11(3) we know that the sum

$$\sum_{j=1}^v \lambda_j \cdot \alpha_j$$

is also homogeneous for any $v = 1, \dots, l$. Applying Claim 2, we conclude that

$$\tau^+\left(\sum_{j=1}^u \lambda_j \cdot \alpha_j\right) = \tau^+(\alpha_1).$$

Hence we can apply Claim 1 repeatedly to conclude that

$$\tau^+\left(\sum_{j=1}^l \lambda_j \cdot \alpha_j\right) = \tau^+(\alpha_1) = \min_{1 \leq j \leq l} \{\tau^+(\alpha_j)\}.$$

(2) If $x \in \text{Im } \Phi_{n+2k}^{n+k}$, then there exists $y \in (\Gamma_{n+k}, i - \frac{1}{2}kp)$ and $z \in (\Gamma_{n+k}, i + \frac{1}{2}kp)$ such that

$$x = \Psi_{-,n+2k}^{n+k}(y) + \Psi_{+,n+2k}^{n+k}(z).$$

By assumption,

$$F_{n+2k}(x) = \alpha = \sum_{j=1}^l \lambda_j \cdot \alpha_j,$$

with $\lambda_j \neq 0$ and α homogeneous. By Lemma 2.18 we have

$$\alpha = F_{n+k}(y + z).$$

Since \mathfrak{B} forms a basis for Y , we can write

$$F_{n+k}(y) = \sum_{j=1}^l \lambda'_j \cdot \alpha_j \quad \text{and} \quad F_{n+k}(z) = \sum_{j=1}^l \lambda''_j \cdot \alpha_j,$$

where $l = |\mathfrak{B}|$. Then for any $1 \leq j \leq l$, at least one of λ'_j and λ''_j is nonzero. Since both $F_{n+k}(y)$ and $F_{n+k}(z)$ are homogeneous, from part (1) we know

$$\begin{aligned} i - \frac{kp}{2} &\geq \tau^-(\alpha_j) - \frac{(n+k-1)p-q}{2} && \text{when } \lambda'_j \neq 0, \\ i + \frac{kp}{2} &\leq \tau^+(\alpha_j) + \frac{(n+k-1)p-q}{2} && \text{when } \lambda''_j \neq 0. \end{aligned}$$

Conversely, suppose for any $1 \leq j \leq l$ at least one of the inequalities

$$i \geq \tau^-(\alpha_j) - \frac{(n-1)p-q}{2} \quad \text{or} \quad i \leq \tau^+(\alpha_j) + \frac{(n-1)p-q}{2}$$

holds. We need to show that $x \in \text{Im } \Phi_{n+2k}^{n+k}$. We deal with three cases.

Case 1 The grading i satisfies

$$i \geq \frac{(n+2k-2)p-q+\chi(S)}{2}.$$

We want to argue that

$$\Psi_{-,n+2k}^{n+k} : \left(\Gamma_{n+k}, i - \frac{kp}{2} \right) \rightarrow (\Gamma_{n+2k}, i)$$

is surjective and hence conclude that $x \in \text{Im } \Phi_{n+2k}^{n+k}$. To do this, note that $\Psi_{-,n+2k}^{n+k}|_{(\Gamma_{n+k}, i - \frac{1}{2}kp)}$ is the composition of maps $\psi_{-,n+k+j+1}^{n+k+j}|_{(\Gamma_{n+k+j}, i - \frac{1}{2}kp + \frac{1}{2}jp)}$ for $j = 0, 1, \dots, k-1$. With the assumption of Case 1, we have

$$i - \frac{kp}{2} + \frac{jp}{2} \geq \frac{(n+k+j-2)p-q+\chi(S)}{2} > -\frac{(n+k+j)p-q+\chi(S)}{2}.$$

(Note that since k is sufficiently large this is a very loose inequality.) Then Corollary 2.9(2) applies and we conclude that $\psi_{-,n+k+j+1}^{n+k+j}|_{(\Gamma_{n+k+j}, i - \frac{1}{2}kp + \frac{1}{2}jp)}$ is an isomorphism for all $j = 0, 1, \dots, k-1$. Hence we conclude Case 1.

Case 2 The grading i satisfies

$$i \leq -\frac{(n+2k-2)p-q+\chi(S)}{2}.$$

The argument is similar to that for Case 1, except for using $\Psi_{+,n+2k}^{n+k}$ instead of $\Psi_{-,n+2k}^{n+k}$.

Case 3 The grading i satisfies

$$|i| < \frac{(n+2k-2)p-q+\chi(S)}{2}.$$

Under the assumption of Case 3, Lemma 2.19(1) implies that F_{n+2k} is injective when restricted to (Γ_{n+2k}, i) .

Now, for each $j = 1, 2, \dots, l$, if we have

$$i \geq \tau^-(\alpha_j) - \frac{(n-1)p - q}{2},$$

we claim that there exists $y_j \in (\Gamma_{n+k}, i - \frac{1}{2}kp)$ such that

$$F_{n+k}(y_j) = \lambda_j \cdot \alpha_j.$$

If instead

$$i \leq \tau^+(\alpha_j) + \frac{(n-1)p - q}{2},$$

we claim that there exists $z_j \in (\Gamma_{n+k}, i + \frac{1}{2}kp)$ such that

$$F_{n+k}(z_j) = \lambda_j \cdot \alpha_j.$$

We will verify the existence of y_j or z_j in a moment, but for now let y be the sum of all the y_j and z be the sum of all the z_j . Then from Lemma 2.18 it is straightforward to check that

$$F_{n+2k}(\Psi_{-,n+2k}^{n+k}(y) + \Psi_{+,n+2k}^{n+k}(z)) = \alpha = F_{n+2k}(x).$$

Since in Case 3 the restriction of F_{n+2k} on (Γ_{n+2k}, i) is injective, we conclude that

$$x = \Psi_{-,n+2k}^{n+k}(y) + \Psi_{+,n+2k}^{n+k}(z) \in \text{Im } \Phi_{n+2k}^{n+k}.$$

It remains to show that the desired y_j or z_j exists. We only prove the existence of y_j , and the argument for z_j is similar. Now assume that

$$i \geq \tau^-(\alpha_j) - \frac{(n-1)p - q}{2}.$$

This implies that

$$i - \frac{kp}{2} \geq \tau^-(\alpha_j) - \frac{(n+k-1)p - q}{2}.$$

The hypothesis of the lemma and the definition of $\tau^-(\alpha_j)$ in Definition 6.8, together with Lemma 6.11(3), imply that

$$i \equiv \tau^-(\alpha_j) - \frac{(n-1)p - q}{2} \pmod{p}.$$

As a result, there exists an integer $N \geq 0$ such that

$$i - \frac{kp}{2} = \tau^-(\alpha_j) - \frac{(n+k-1)p - q}{2} + Np.$$

From Definition 6.8, we know that

$$\tau^-(\alpha_j) - \frac{(n+k-1)p - q}{2} = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (n+k - \tau(\alpha_j))p}{2}.$$

As a result, we have

$$i - \frac{kp}{2} = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (n+k - \tau(\alpha_j))p}{2} + Np.$$

The assumption in Case 3 and Lemma 6.12(4) then implies that $N \leq (n+k) - \tau(\alpha_j)$. Hence Lemma 6.12(3) implies the existence of y_j .

(3) If $x \notin \text{Im } \Phi_{n+2k}^{n+k}$, then part (2) means that there exists some j such that

$$\tau^+(\alpha_j) + \frac{(n-1)p-q}{2} < i < \tau^-(\alpha_j) - \frac{(n-1)p-q}{2}.$$

Note that, by Lemma 6.11(3), we must have

$$i \equiv \tau^-(\alpha_j) - \frac{(n-1)p-q}{2} \equiv \tau^+(\alpha_j) + \frac{(n-1)p-q}{2} \pmod{p}.$$

By direct calculation, we have

$$\left(\tau^-(\alpha_j) - \frac{(n-1)p-q}{2}\right) - \left(\tau^+(\alpha_j) + \frac{(n-1)p-q}{2}\right) = (\tau(\alpha_j) - n)p.$$

Then we can choose N with $0 \leq N \leq \tau(\alpha_j) - n - 2$ as desired. □

7.2 The construction of the map

Since Φ_{n+k}^n and Φ_{n+2k}^{n+k} are homogeneous, we can construct Φ_n^{n+2k} for each grading to achieve both the exactness and the commutativity. Given the grading shifts in Lemmas 2.6 and 2.13, the map Φ_n^{n+2k} preserves the gradings. From Lemma 2.5, for any grading i with

$$|i| > \frac{|np-q| - \chi(S)}{2},$$

we have $(\Gamma_n, i) = 0$. From Corollary 2.9, we know either $\Psi_{+,n+2k}^{n+k}$ or $\Psi_{-,n+2k}^{n+k}$ is surjective onto (Γ_{n+2k}, i) for such a grading i . Thus, on such a grading i , the zero map satisfies the exactness for Φ_n^{n+2k} (although we still have to verify the commutativity in Proposition 7.2).

On the other hand, from Lemma 2.19, the restriction of F_{n+2k} on the consecutive p middle gradings is an isomorphism. In particular, when $p = 1$, it is an isomorphism when restricted to each middle grading. Also from Lemma 7.3, it seems that the definition of Φ_{n+k}^{n+2k} on (Γ_{n+2k}, i) should involve $\text{Proj}_n^i \circ G_n$. However, if we simply take

$$\text{Proj}_n^i \circ G_n \circ F_{n+2k}$$

as the definition, the current techniques fall short of demonstrating exactness and commutativity.

We resolve this issue by introducing an isomorphism

$$I: Y \xrightarrow{\cong} Y,$$

and define

$$(7-5) \quad \Phi_n^{n+2k}(x) = \text{Proj}_n^i \circ G_n \circ I \circ F_{n+2k}(x) \quad \text{for } x \in (\Gamma_{n+2k}, i).$$

The construction of I is noncanonical but it helps us to prove the exactness and commutativity.

Remark 7.5 In the first arXiv version of this paper, we deal with the special case $Y = S^3$. In this case $Y \cong \mathbb{C}$ so up to a scalar we have $I = \text{Id}$. In this special case indeed we could prove the exactness and commutativity without explicitly writing down the isomorphism I as follows.

We first define the map I on the basis

$$\mathfrak{B} = \bigcup_{n \in \mathbb{Z}} \mathfrak{B}_n$$

of Y chosen in Section 6.3 that consists of homogeneous elements, and then extend the map on the whole space linearly. We will show it is an isomorphism.

Fix $n_0 \in \mathbb{Z}$ small enough such that Corollary 2.9 and Lemma 2.19 apply. For any $\alpha \in \mathfrak{B}_n$, there exists a grading $i(\alpha) \in (-\frac{1}{2}p, \frac{1}{2}p]$ such that there exists $N(\alpha) \in \mathbb{Z}$ with

$$i(\alpha) = \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} - \frac{(\tau(\alpha) - 2 - n_0)p}{2} + N(\alpha)p.$$

From the above equality, we know that

$$N(\alpha) = \frac{\tau(\alpha) - 2 - n_0}{2} + \frac{2i(\alpha) - \tau^+(\alpha) - \tau^-(\alpha)}{2p}.$$

Note that, except for n_0 , the rest of the terms are bounded, so $N(\alpha) \geq 0$ when n_0 is small enough. Similarly,

$$N(\alpha) + n_0 = \frac{\tau(\alpha) - 2 + n_0}{2} + \frac{2i(\alpha) - \tau^+(\alpha) - \tau^-(\alpha)}{2p},$$

so we have $N(\alpha) + n_0 \leq \tau(\alpha) - 2$ when n_0 is small enough. As a result, by Lemma 6.14(2) and (6) (and the convention after the lemma), we know that for any $\alpha \in \mathfrak{B}_n$,

$$\begin{aligned} \Psi_{-, \tau(\alpha)-2}^{n_0+N(\alpha)}(\eta_{+, n_0+N(\alpha)}^{\tau(\alpha)}(\alpha)) &= \eta_{+, \tau(\alpha)-2}^{\tau(\alpha)}(\alpha) \\ &= \eta_{-, \tau(\alpha)-2}^{\tau(\alpha)}(\alpha) \\ &= \Psi_{+, \tau(\alpha)-2}^{\tau(\alpha)-2-N(\alpha)}(\eta_{-, \tau(\alpha)-2-N(\alpha)}^{\tau(\alpha)}(\alpha)). \end{aligned}$$

Then by Lemma 6.7, there exists $w \in (\Gamma_{n_0}, i(\alpha))$ such that

$$(7-6) \quad \begin{aligned} \Psi_{+, n_0+N(\alpha)}^{n_0}(w) &= \eta_{+, n_0+N(\alpha)}^{\tau(\alpha)}(\alpha), \\ \Psi_{-, \tau(\alpha)-2-N(\alpha)}^{n_0}(w) &= \eta_{-, \tau(\alpha)-2-N(\alpha)}^{\tau(\alpha)}(\alpha). \end{aligned}$$

Let

$$\text{Proj}: \Gamma_{n_0} \rightarrow \bigoplus_{i \in (-\frac{1}{2}p, \frac{1}{2}p]} (\Gamma_{n_0}, i).$$

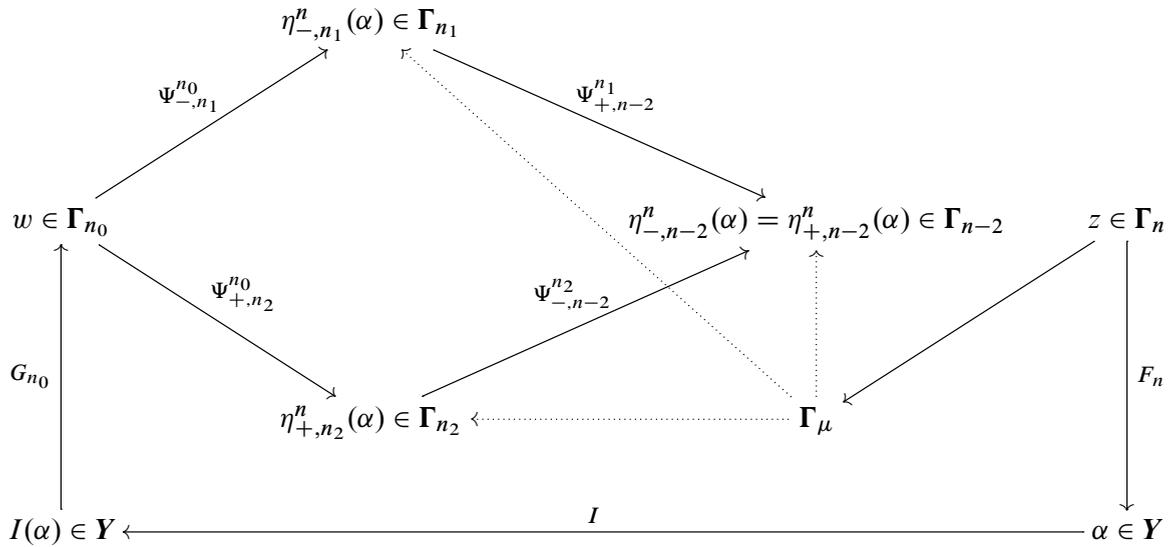
From Lemma 2.19, we know

$$\text{Proj} \circ G_{n_0}: Y \rightarrow \bigoplus_{i \in (-\frac{1}{2}p, \frac{1}{2}p]} (\Gamma_{n_0}, i)$$

is an isomorphism. Hence we define

$$I(\alpha) = (\text{Proj} \circ G_{n_0})^{-1}(w).$$

Writing $n = \tau(\alpha)$, $n_1 = \tau(\alpha) - 2 - N(\alpha)$ and $n_2 = n_0 + N(\alpha)$, the following diagram might be helpful for understanding the construction of I :



Remark 7.6 For a general 3-manifold Y , our construction of I is noncanonical since there are many choices such as the basis \mathfrak{B} and the element w for each $\alpha \in \mathfrak{B}$. However, one could still ask whether we could simply pick $I = \text{Id}$ or not. If we take $I = \text{Id}$, then Proposition 7.2 can finally be reduced to Conjecture 7.7, which we state below. We believe that the following conjecture is true, although currently we do not find a proof for it. Hence, in order to fulfill the main purpose of the paper, we introduce the isomorphism I to bypass this conjecture.

Conjecture 7.7 For any $\alpha \in \mathfrak{B}$, and any integer $n \leq \tau(\alpha) - 2$, we have

$$\eta_{\pm,n}^{\tau(\alpha)}(\alpha) = \text{Proj}_n^{j_{\pm}} \circ G_n(\alpha),$$

where

$$j_{\pm} = \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} \mp \frac{(\tau(\alpha) - 2 - n)p}{2},$$

and

$$\text{Proj}_n^{j_{\pm}} : \Gamma_n \rightarrow (\Gamma_n, j_{\pm})$$

is the projection.

Lemma 7.8 (1) Suppose $\alpha \in \mathfrak{B}$ and $n_0, w, N(\alpha)$ are chosen as above. Suppose n and k are two integers such that $n_0 \leq k \leq n$. Then

(a) $\Psi_{-,n}^k \circ \Psi_{+,k}^{n_0}(w) \neq 0$

if and only if

(b) $k \leq n_0 + N(\alpha)$ and $n - k \leq \tau(\alpha) - 2 - n_0 - N(\alpha)$ (in particular, we have $n \leq \tau(\alpha) - 2$).

- (2) The map $I : Y \rightarrow Y$ is an isomorphism.
- (3) For an element $\alpha \in \mathfrak{B}$, an integer n and a grading i , the following two statements are equivalent.
 - (a) We have $\text{Proj}_n^i \circ G_n \circ I(\alpha) \neq 0$.
 - (b) We have $n \leq \tau(\alpha) - 2$ and there exists $N \in \mathbb{Z}$ such that $N \in [0, \tau(\alpha) - 2 - n]$ and

$$i = \frac{\tau^+(\alpha) + \tau^-(\alpha) - (\tau(\alpha) - 2 - n)p}{2} + Np.$$

- (4) Suppose for an integer n and a grading i we have $\alpha_1, \dots, \alpha_L \in \mathfrak{B}$ such that $\text{Proj}_n^i \circ G_n \circ I(\alpha_j) \neq 0$ for all $1 \leq j \leq L$. Then $\text{Proj}_n^i \circ G_n \circ I(\alpha_1), \dots, \text{Proj}_n^i \circ G_n \circ I(\alpha_L)$ are linearly independent.
- (5) Suppose $\alpha \in \mathfrak{B}$. For any $n \in \mathbb{Z}$ such that $n \leq \tau(\alpha) - 2$, we have

$$\text{Proj}_n^{i_{\pm}} \circ G_n \circ I(\alpha) = \eta_{\pm, n}^{\tau(\alpha)}(\alpha) \quad \text{where } i_{\pm} = \frac{\tau^+(\alpha) + \tau^-(\alpha) \mp (\tau(\alpha) - 2 - n)}{2}.$$

Proof (1) First, when $k > n_0 + N(\alpha)$, from the construction of w , we know that

$$\Psi_{+, k}^{n_0}(w) = \Psi_{+, k}^{n_0 + N(\alpha)} \circ \Psi_{+, n_0 + N(\alpha)}^{n_0}(w) = \Psi_{+, k}^{n_0 + N(\alpha)} \circ \eta_{+, n_0 + N(\alpha)}^{\tau(\alpha)}(\alpha) = 0.$$

The last equality is from Lemma 6.14(6). Similarly, if $n - k > \tau(\alpha) - 2 - n_0 - N(\alpha)$, we know from Lemma 2.11 that

$$\begin{aligned} \Psi_{-, n}^k \circ \Psi_{+, k}^{n_0}(w) &= \Psi_{+, n}^{n + n_0 - k} \circ \Psi_{-, n + n_0 - k}^{n_0}(w) \\ &= \Psi_{+, n}^{n + n_0 - k} \circ \Psi_{-, n + n_0 - k}^{\tau(\alpha) - 2 - N(\alpha)} \circ \Psi_{-, \tau(\alpha) - 2 - N(\alpha)}^{n_0}(w) \\ &= \Psi_{+, n}^{n + n_0 - k} \circ \Psi_{-, n + n_0 - k}^{\tau(\alpha) - 2 - N(\alpha)} \circ \eta_{-, \tau(\alpha) - 2 - N(\alpha)}^{\tau(\alpha)}(\alpha) \quad \text{by definition of } w \\ &= 0 \quad \text{by Lemma 6.14(6).} \end{aligned}$$

Next, we need to show that $\Psi_{-, n}^k \circ \Psi_{+, k}^{n_0}(w) \neq 0$ when $k \leq n_0 + N(\alpha)$ and $n - k \leq \tau(\alpha) - 2 - n_0 - N(\alpha)$. Again, from Lemma 2.11 we have

$$\begin{aligned} \Psi_{+, n + N(\alpha) + n_0 - k}^n \circ \Psi_{-, n}^k \circ \Psi_{+, k}^{n_0}(w) &= \Psi_{-, n_0 + N(\alpha) + n - k}^{n_0 + N(\alpha)} \circ \Psi_{+, n_0 + N(\alpha)}^{n_0}(w) \\ &= \Psi_{-, n_0 + N(\alpha) + n - k}^{n_0 + N(\alpha)} \circ \eta_{+, n_0 + N(\alpha)}^{\tau(\alpha)}(\alpha) \quad \text{by definition of } w \\ &= \eta_{+, n_0 + N(\alpha) + n - k}^{\tau(\alpha)}(\alpha) \quad \text{by Lemma 6.14(6)} \\ &\neq 0 \quad \text{by Lemma 6.14(3).} \end{aligned}$$

- (2) Suppose $\mathfrak{B} = \{\alpha_1, \dots, \alpha_L\}$, where $L = \dim_{\mathbb{C}} Y$. We order the elements α_i such that

$$\tau(\alpha_i) \geq \tau(\alpha_{i+1}).$$

Let $w_i, N_i = N(\alpha_i)$ be the data associated to α_i as above. Since

$$w_i \in \bigoplus_{j \in (-\frac{1}{2}p, \frac{1}{2}p]} (\Gamma_{n_0}, j) \quad \text{for any } i,$$

and by Lemma 2.19, the map

$$\text{Proj} \circ G_{n_0} : Y \rightarrow \bigoplus_{j \in (-\frac{1}{2}p, \frac{1}{2}p]} (\Gamma_{n_0}, j)$$

is an isomorphism, in order to show that I is an isomorphism, it suffices to show that w_1, \dots, w_L are linearly independent.

Now suppose there are complex numbers $\lambda_1, \dots, \lambda_L$ such that

$$\sum_{i=1}^L \lambda_i w_i = 0.$$

Our goal is to show that all λ_i are zero. The idea is to apply various maps $\Psi_{+, \tau(\alpha_1)-2}^{n_0+N_1} \circ \Psi_{-, n_0+N_1}^{n_0}$ to filter out different indices by part (1) of the lemma.

Applying the map $\Psi_{+, \tau(\alpha_1)-2}^{n_0+N_1} \circ \Psi_{-, n_0+N_1}^{n_0}$, from the construction of w_i , the order of α_i and part (1) of the lemma, we know

$$\begin{aligned} 0 &= \Psi_{+, \tau(\alpha_1)-2}^{n_0+N_1} \circ \Psi_{-, n_0+N_1}^{n_0} \left(\sum_{i=1}^L \lambda_i w_i \right) \\ &= \sum_{\alpha_i} \lambda_i \eta_{\pm \tau(\alpha_1)-2}^{\tau(\alpha_i)}(\alpha_i), \end{aligned}$$

where the summation in the second line is over all α_i with

$$\tau(\alpha_i) = \tau(\alpha_1) \quad \text{and} \quad N_i = N_1.$$

From Lemma 6.14(2) and the convention after the lemma, we know $\eta_{+, \tau(\alpha_1)-2}^{\tau(\alpha_1)} = \eta_{-, \tau(\alpha_1)-2}^{\tau(\alpha_1)}$. From Lemma 6.14 again, we know that $\eta_{\pm \tau(\alpha_1)-2}^{\tau(\alpha_i)}(\alpha_i)$ are linearly independent, and as a result all relevant λ_i must be zero. Suppose i_0 is the smallest index in the rest. By our choice of α_i , the element α_{i_0} has the largest τ among the rest of the α_i . Hence we can apply the map $\Psi_{+, \tau(\alpha_{i_0})-2}^{n_0+N_{i_0}} \circ \Psi_{-, n_0+N_{i_0}}^{n_0}$ to filter out α_i with smaller τ . Repeating this argument, we could prove that all λ_i must be zero.

(3) Let $n_0, w, i(\alpha)$ and $N(\alpha)$ be constructed as above. We first prove that (b) \implies (a). Note that, when constructing the isomorphism I , from Corollary 2.9 and Lemma 2.18 we can take $n'_0 = n_0 - 2$ and $w' = (\psi_{+, n_0-1}^{n_0-2})^{-1} \circ (\psi_{+, n_0}^{n_0-1})^{-1}(w)$ that will lead to the same I as n_0 and w . (Note that, by construction, passing from n_0 to $n_0 - 2$ will increase $N(\alpha)$ by 1.)

Remark 7.9 The main goal of the current paper is to derive an integral surgery formula. For n that is sufficiently large, we already know a large surgery as in [Li and Ye 2021]. When n is small enough, we can pass to $-n$ for the mirror of the knot. As a result, instead of changing n_0 for particular n , we could assume a universal bound for all the integers n that we care about and make n_0 universally small.

As a result, we can always assume that n_0 is small enough compared with any given n . Now recall by construction

$$i(\alpha) = \frac{\tau^+(\alpha) + \tau^-(\alpha)}{2} - \frac{(\tau(\alpha) - 2 - n_0)p}{2} + N(\alpha)p,$$

and by the assumption in (b) we have

$$i = \frac{\tau^+(\alpha) + \tau^-(\alpha) - (\tau(\alpha) - 2 - n)p}{2} + Np.$$

We can assume that n_0 is small enough such that $N(\alpha) > N$. Take $k = N(\alpha) - N + n_0$. By construction we have $w \in (\Gamma_{n_0}, i(\alpha))$, so from Lemma 2.6 we know that

$$\Psi_{-,n}^k \circ \Psi_{+,k}^{n_0}(w) \in (\Gamma_n, i(\alpha) - \frac{(k - n_0)p}{2} + \frac{(n - k)p}{2}) = (\Gamma_n, i).$$

As a result, we conclude from the definition of w and Lemma 2.18 that

$$\begin{aligned} \Psi_{-,n}^k \circ \Psi_{+,k}^{n_0}(w) &= \Psi_{-,n}^k \circ \Psi_{+,k}^{n_0} \circ \text{Proj}_{n_0}^{i(\alpha)} \circ G_{n_0} \circ I(\alpha) \\ &= \text{Proj}_n^i \circ \Psi_{-,n}^k \circ \Psi_{+,k}^{n_0} \circ G_{n_0} \circ I(\alpha) \\ &= \text{Proj}_n^i \circ G_n \circ I(\alpha). \end{aligned}$$

Then it is straightforward to verify that $k - n_0 \leq N(\alpha)$ and $n - k \leq \tau(\alpha) - 2 - n_0 - N(\alpha)$. As a result, we conclude from part (1) that

$$\text{Proj}_n^i \circ G_n \circ I(\alpha) \neq 0.$$

Next we show that (a) \implies (b). Again assume that n_0 is small enough compared with the given n . Then there exists $i' \in (-\frac{1}{2}p, \frac{1}{2}p]$ such that there exists $N' \in \mathbb{Z}$ with

$$i' = i - \frac{(n - n_0)p}{2} + N'p.$$

By Lemma 2.18, we know that

$$\text{Proj}_n^i \circ G_n \circ I(\alpha) = \Psi_{-,n}^{n_0+N'} \circ \Psi_{+,n_0+N'}^{n_0} \circ \text{Proj}_{n_0}^{i'} \circ G_{n_0} \circ I(\alpha).$$

From the construction of $I(\alpha)$ and Lemma 2.19 we know $\text{Proj}_n^i \circ G_n \circ I(\alpha) \neq 0$ only if $i' = i(\alpha)$, in which case

$$\text{Proj}_n^i \circ G_n \circ I(\alpha) = \Psi_{-,n}^{n_0+N'} \circ \Psi_{+,n_0+N'}^{n_0} \circ \text{Proj}_{n_0}^{i'} \circ G_{n_0} \circ I(\alpha) = \Psi_{-,n}^{n_0+N'} \circ \Psi_{+,n_0+N'}^{n_0}(w).$$

Hence $\text{Proj}_n^i \circ G_n \circ I(\alpha) \neq 0$ implies that

$$N' \leq N(\alpha) \quad \text{and} \quad n - N' \leq \tau(\alpha) - 2 - N(\alpha)$$

by part (1). Taking $N = N(\alpha) - N'$, it is then straightforward to check that

$$N \in [0, \tau(\alpha) - 2 - n] \quad \text{and} \quad i = \frac{\tau^+(\alpha) + \tau^-(\alpha) - (\tau(\alpha) - 2 - n)p}{2} + Np.$$

(4) The proof is similar to that of (2).

(5) It follows from the proofs of parts (1) and (3). □

7.3 The exact triangle

In this subsection, we prove the exact triangle. Note that we choose the basis \mathfrak{B} of Y as in Section 6.3.

Proof of Proposition 7.1 We will verify the exactness at each space of the triangle.

The exactness at $\Gamma_{n+k} \oplus \Gamma_{n+k}$ This follows from Proposition 5.1.

The exactness at Γ_n From Lemma 7.3 and the construction of Φ_n^{n+2k} in (7-5), we know that $\text{Im } \Phi_n^{n+2k} \subset \ker \Phi_{n+k}^n$. Now pick an arbitrary

$$x \in (\Gamma_n, i) \cap \ker \Phi_{n+k}^n = \text{Im}(\text{Proj}_n^i \circ G_n).$$

Since I is an isomorphism, we can assume that

$$x = \sum_{j=1}^l \text{Proj}_n^i \circ G_n(\lambda_j \cdot I(\alpha_j)),$$

where $\alpha_j \in \mathfrak{B}$ and $\text{Proj}_n^i \circ G_n \circ I(\alpha_j) \neq 0$. From Lemma 7.8(3), we know that this implies that for any $j \in [1, l] \cap \mathbb{Z}$, we have $n \leq \tau(\alpha_j) - 2$ and there exists $N_j \in \mathbb{Z}$ such that $N_j \in [0, \tau(\alpha_j) - 2 - n]$ and

$$i = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (\tau(\alpha_j) - 2 - n)p}{2} + N_j p.$$

Now, for k sufficiently large, we have $n + 2k > \tau(\alpha_j)$. Taking

$$N'_j = n + k + 1 - \tau(\alpha_j) + N_j \in \mathbb{Z},$$

it is straightforward to verify that when k is sufficiently large, we have

$$N'_j \in [0, n + 2k - \tau(\alpha_j)] \quad \text{and} \quad i = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (n + 2k - \tau(\alpha_j))p}{2} + N'_j p.$$

Hence by Lemma 6.12(3), there exists $y_j \in (\Gamma_{n+2k}, i)$ such that $F_{n+2k}(y_j) = \alpha_j$. As a result, it is straightforward to check that

$$x = \Phi_n^{n+2k} \left(\sum_{j=1}^l \lambda_j \cdot y_j \right) \in \text{Im } \Phi_n^{n+2k}.$$

The exactness at Γ_{n+2k} Suppose $x \in (\Gamma_{n+2k}, i)$ and

$$F_{n+2k}(x) = \sum_{j=1}^l \lambda_j \cdot \alpha_j$$

with $\lambda_j \neq 0$ and $\alpha_j \in \mathfrak{B}$.

First, if $x \in \text{Im } \Phi_{n+2k}^{n+k}$, then from Lemma 7.4(2), we know that for any $1 \leq j \leq l$ we have

$$\text{either } i \geq \tau^-(\alpha_j) - \frac{(n-1)p - q}{2} \quad \text{or} \quad i \leq \tau^+(\alpha_j) + \frac{(n-1)p - q}{2}.$$

If we write

$$i = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (\tau(\alpha_j) - 2 - n)p}{2} + N_j p,$$

then for some N_j the inequality

$$i \geq \tau^-(\alpha_j) - \frac{(n-1)p - q}{2}$$

implies that

$$\begin{aligned} N_j &\geq \frac{1}{p} \left(\tau^-(\alpha_j) - \frac{(n-1)p - q}{2} - \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (\tau(\alpha_j) - 2 - n)p}{2} \right) \\ &= \frac{1}{2} \left(1 + \frac{\tau^-(\alpha_j) - \tau^+(\alpha_j) + q}{p} \right) + \frac{\tau(\alpha_j)}{2} - 1 - n \\ &= \tau(\alpha_j) - 1 - n. \end{aligned}$$

Note that the last equality uses the definition of $\tau(\alpha)$ in Definition 6.8. Similarly, we can compute that the inequality

$$i \leq \tau^+(\alpha_j) + \frac{(n-1)p - q}{2}$$

implies that

$$\begin{aligned} N_j &\leq \frac{1}{p} \left(\tau^+(\alpha_j) + \frac{(n-1)p - q}{2} - \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (\tau(\alpha_j) - 2 - n)p}{2} \right) \\ &= \frac{1}{2} \left(-1 + \frac{\tau^+(\alpha_j) - \tau^-(\alpha_j) - q}{p} \right) + \frac{\tau(\alpha_j)}{2} - 1 \\ &= -1. \end{aligned}$$

In summary, $x \in \text{Im } \Phi_{n+2k}^{n+k}$ implies that for all $1 \leq j \leq l$, either $N_j \geq \tau(\alpha_j) - 1 - n$ or $N_j \leq -1$. Hence from Lemma 7.8(3), we know that

$$\text{Proj}_n^i \circ G_n \circ I(\alpha_j) = 0$$

for all $1 \leq j \leq l$ and as a result, $\Phi_n^{n+2k}(x) = 0$.

Second, suppose $x \notin \text{Im } \Phi_{n+2k}^{n+k}$. For any $1 \leq j \leq l$, we can write

$$i = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (\tau(\alpha_j) - 2 - n)p}{2} + N_j p$$

for some N_j . Then from Lemma 7.4(3) we know that there exists j such that $1 \leq j \leq l$, and

$$N_j \in [0, \tau(\alpha_j) - 2 - n] \cap \mathbb{Z}.$$

Hence by Lemma 7.8(3) and (4) we know that

$$\text{Proj}_n^i \circ G_n \circ I(\alpha_j) \neq 0 \Rightarrow \Phi_n^{n+2k}(x) \neq 0.$$

Hence we conclude that

$$\text{Im } \Phi_{n+2k}^{n+k} = \ker \Phi_n^{n+2k}.$$

□

7.4 The commutative diagram

In this subsection, we will prove the commutative diagram presented at the beginning of the section. Note that we choose the basis \mathfrak{B} of Y as in Section 6.3.

Lemma 7.10 *Suppose $n \in \mathbb{Z}$ and i is a grading. Suppose $x \in (\Gamma_n, i)$ such that*

$$F_n(x) = \sum_j^l \lambda_j \alpha_j,$$

with $\lambda_j \neq 0$ and $\alpha_j \in \mathfrak{B}$ for all $1 \leq j \leq l$. Then for any $1 \leq j \leq l$, there exists $N_j \in [0, n + 1 - \tau(\alpha_j)]$ such that

$$i = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (n + 1 - \tau(\alpha_j))p}{2} + N_j p.$$

Proof This is a combination of Lemma 6.11(3), Lemma 6.12(3) and Lemma 7.4(1). The proof is similar to that of Lemma 7.4(2). □

Proof of Proposition 7.2 We only prove the first commutative diagram

$$\begin{array}{ccc} \Gamma_{\frac{1}{2}(2n+2k+1)} & \xrightarrow{\psi_{+,\mu}^{n+k+1} \circ \psi_{-,n+k+1}^{\frac{1}{2}(2n+2k+1)}} & \Gamma_\mu \\ \downarrow \Psi_{+,n+2k}^{n+k+1} \circ \psi_{-,n+k+1}^{\frac{1}{2}(2n+2k+1)} & & \downarrow \psi_{+,n}^\mu \\ \Gamma_{n+2k} & \xrightarrow{\Phi_n^{n+2k}} & \Gamma_n \end{array}$$

The other is similar. Note that at the end of Section 6.3, we introduce new notation of $\eta_{\pm,n-2}^n$ to remove the scalars. Then the second commutative diagram only holds up to a scalar.

First, note that the maps from $\Gamma_{\frac{1}{2}(2n+2k+1)}$ to Γ_μ and Γ_{n+2k} both factor through Γ_{n+k+1} . As a result, we only need to prove the commutative diagram

$$\begin{array}{ccc} \Gamma_{n+k+1} & \xrightarrow{\psi_{+,\mu}^{n+k+1}} & \Gamma_\mu \\ \downarrow \Psi_{+,n+2k}^{n+k+1} & & \downarrow \psi_{+,n}^\mu \\ \Gamma_{n+2k} & \xrightarrow{\Phi_n^{n+2k}} & \Gamma_n \end{array}$$

for sufficiently large k . Now suppose $x \in (\Gamma_{n+k+1}, i)$. Write

$$F_{n+k+1}(x) = \sum_{j=1}^l \lambda_j \cdot \alpha_j$$

with $\lambda_j \neq 0$ and $\alpha_j \in \mathfrak{B}$ for $1 \leq j \leq l$. We want to first establish an identity

$$(7-7) \quad \psi_{+,n}^\mu \circ \psi_{+,\mu}^{n+k+1}(x) = \sum_{\substack{1 \leq j \leq l; \\ N_j = n+k+1-\tau(\alpha_j)}} \lambda_j \cdot \eta_{+,n}^{\tau(\alpha_j)}(\alpha_j),$$

and then show that the other composition has exactly the same expression.

From Lemma 7.10, we know for any $1 \leq j \leq l$, there exists $N_j \in [0, n + k + 1 - \tau(\alpha_j)]$ such that

$$(7-8) \quad i = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (n + k + 1 - \tau(\alpha_j))p}{2} + N_j p.$$

Taking $n'_j = \tau(\alpha_j)$ and $N'_j = 0$, we can apply Lemma 6.12(3) to find an element

$$x_j \in \left(\Gamma_{\tau(\alpha_j)}, \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j)}{2} \right) \quad \text{such that} \quad F_{\tau(\alpha_j)}(x_j) = \alpha_j.$$

It is then straightforward to check that

$$(7-9) \quad y_j = \Psi_{+,n+k+1}^{\tau(\alpha_j)+N_j} \circ \Psi_{-, \tau(\alpha_j)+N_j}^{\tau(\alpha_j)}(x_j) \in (\Gamma_{n+k+1}, i).$$

Write

$$y = x - \sum_{j=1}^l \lambda_j \cdot y_j \in (\Gamma_{n+k+1}, i).$$

From Lemma 2.18 we know that

$$F_{n+k+1}(y) = 0.$$

As a result, by Lemma 6.6,

$$\psi_{+,n}^\mu \circ \psi_{+, \mu}^{n+k+1}(x) = \sum_{j=1}^l \lambda_j \cdot \psi_{+,n}^\mu \circ \psi_{+, \mu}^{n+k+1}(y_j).$$

Note that, unless $N_j = n + k + 1 - \tau(\alpha_j)$, we have

$$\psi_{+, \mu}^{n+k+1} \circ \Psi_{+,n+k+1}^{\tau(\alpha_j)+N_j} = 0$$

by the exactness. As a result,

$$\begin{aligned} \psi_{+,n}^\mu \circ \psi_{+, \mu}^{n+k+1}(x) &= \sum_{\substack{1 \leq j \leq l \\ N_j = n+k+1-\tau(\alpha_j)}} \lambda_j \cdot \psi_{+,n}^\mu \circ \psi_{+, \mu}^{n+k+1} \circ \Psi_{-, \tau(\alpha_j)+N_j}^{\tau(\alpha_j)}(x_j) \\ &= \sum_{\substack{1 \leq j \leq l; n \leq \tau(\alpha_j)-2 \\ N_j = n+k+1-\tau(\alpha_j)}} \lambda_j \cdot \psi_{+,n}^\mu \circ \psi_{+, \mu}^{\tau(\alpha_j)}(x_j) + \sum_{\substack{1 \leq j \leq l; n \geq \tau(\alpha_j)-1 \\ N_j = n+k+1-\tau(\alpha_j)}} \lambda_j \cdot \psi_{+,n}^\mu \circ \psi_{+, \mu}^{\tau(\alpha_j)}(x_j) \\ &= \sum_{\substack{1 \leq j \leq l; n \leq \tau(\alpha_j)-2 \\ N_j = n+k+1-\tau(\alpha_j)}} \lambda_j \cdot \psi_{+,n}^\mu \circ \psi_{+, \mu}^{\tau(\alpha_j)}(x_j) \\ &= \sum_{\substack{1 \leq j \leq l; n \leq \tau(\alpha_j)-2 \\ N_j = n+k+1-\tau(\alpha_j)}} \lambda_j \cdot \eta_{+,n}^{\tau(\alpha_j)}(\alpha_j), \end{aligned}$$

where the second equality is by Lemma 2.12, the third equality is by equation (7-10), and the last equality is by definition of $\eta_{+,n}^{\tau(\alpha_j)}$.

This verifies equation (7-7) if we show that

$$\sum_{\substack{1 \leq j \leq l; n \geq \tau(\alpha_j) - 1 \\ N_j = n + k + 1 - \tau(\alpha_j)}} \lambda_j \cdot \psi_{+,n}^\mu \circ \psi_{+,\mu}^{\tau(\alpha_j)}(x_j) = 0.$$

To verify this last equality, assume that $n \geq \tau(\alpha_j) - 1$. Then from Lemma 2.12 and the exactness of the bypass maps, we have

$$(7-10) \quad \psi_{+,n}^\mu \circ \psi_{+,\mu}^{\tau(\alpha_j)}(x_j) = \psi_{+,n}^\mu \circ \psi_{+,\mu}^{n+1} \circ \Psi_{-,n+1}^{\tau(\alpha_j)}(x) = 0.$$

Now we deal with $\Phi_n^{n+2k} \circ \Psi_{+,n+2k}^{n+k+1}(x)$. Since $F_{n+k+1}(y) = 0$, Lemma 2.19 implies that

$$\Psi_{+,n+2k}^{n+k+1}(y) = 0.$$

Hence

$$\Psi_{+,n+2k}^{n+k+1}(x) = \sum_{j=1}^l \lambda_j \cdot \Psi_{+,n+2k}^{n+k+1}(y_j),$$

where y_j is defined as in (7-9). By definition we have $y_j \in (\Gamma_{n+1+k}, i)$, so from Lemma 2.6, we know

$$\Psi_{+,n+2k}^{n+k+1}(y_j) \in \left(\Gamma_{n+2k}, i - \frac{(k-1)p}{2} \right).$$

By (7-9) and Lemma 2.18, we know that

$$F_{n+2k} \circ \Psi_{+,n+2k}^{n+k+1}(y_j) = F_{\tau(\alpha_j)}(x_j) = \alpha_j.$$

Hence

$$\Phi_n^{n+2k} \circ \Psi_{+,n+2k}^{n+k+1}(x) = \sum_{j=1}^l \lambda_j \cdot \text{Proj}_n^{i - \frac{1}{2}(k-1)p} \circ G_n \circ I(\alpha_j).$$

We write

$$i - \frac{(k-1)p}{2} = \frac{\tau^+(\alpha_j) + \tau^-(\alpha_j) - (\tau(\alpha_j) - 2 - n)p}{2} + N'_j p.$$

Comparing the above formula with (7-8), we know

$$N'_j = N_j + \tau(\alpha_j) - n - k - 1.$$

By construction, $N_j \leq n + k + 1 - \tau(\alpha_j)$, which means $N'_j \leq 0$. Hence from Lemma 7.8 we know

$$\text{Proj}_n^{i - \frac{1}{2}(k-1)p} \circ G_n \circ I(\alpha_j) \neq 0$$

if and only if $N'_j = 0$, ie $N_j = n + k + 1 - \tau(\alpha)$. Also when $N'_j = 0$ from Lemma 7.8(5) we know

$$\text{Proj}_n^{i - \frac{1}{2}(k-1)p} \circ G_n \circ I(\alpha_j) = \eta_{+,n}^{\tau(\alpha_j)}(\alpha_j).$$

We can focus on indices j such that $\text{Proj}_n^{i - \frac{1}{2}(k-1)p} \circ G_n \circ I(\alpha_j) \neq 0$ because if an index j makes $\text{Proj}_n^{i - \frac{1}{2}(k-1)p} \circ G_n \circ I(\alpha_j) = 0$, then on one hand it does not contribute to $\Phi_n^{n+2k} \circ \Psi_{+,n+2k}^{n+k+1}(x)$ since the corresponding summand is 0, and on the other hand we have $N_j \neq n + k + 1 - \tau(\alpha)$, hence per equation (7-7) it does not contribute to $\psi_{+,n}^\mu \circ \psi_{+,\mu}^{n+k+1}(x)$, either. Also, we know from Lemma 7.8(3) that when $\text{Proj}_n^{i - \frac{1}{2}(k-1)p} \circ G_n \circ I(\alpha_j) \neq 0$ we must have $n \leq \tau(\alpha_j) - 2$.

As a result, we know

$$\begin{aligned} \Phi_n^{n+2k} \circ \Psi_{+,n+2k}^{n+k+1}(x) &= \sum_{j=1}^l \lambda_j \cdot \text{Proj}_n^{j-\frac{1}{2}(k-1)p} \circ G_n \circ I(\alpha_j) = \sum_{\substack{1 \leq j \leq l; n \leq \tau(\alpha_j) - 2 \\ N_j = n+k+1-\tau(\alpha_j)}} \lambda_j \cdot \eta_{+,n}^{\tau(\alpha_j)}(\alpha_j) \\ &= \psi_{+,n}^{\mu} \circ \psi_{+,n}^{n+k+1}(x), \end{aligned}$$

where the last inequality holds by equation (7-7). □

References

- [Baldwin and Sivek 2015] **J A Baldwin, S Sivek**, *Naturality in sutured monopole and instanton homology*, J. Differential Geom. 100 (2015) 395–480 MR Zbl
- [Baldwin and Sivek 2016a] **J A Baldwin, S Sivek**, *A contact invariant in sutured monopole homology*, Forum Math. Sigma 4 (2016) art. id. e12 MR Zbl
- [Baldwin and Sivek 2016b] **J A Baldwin, S Sivek**, *Instanton Floer homology and contact structures*, Selecta Math. 22 (2016) 939–978 MR Zbl
- [Baldwin and Sivek 2021] **J A Baldwin, S Sivek**, *Framed instanton homology and concordance*, J. Topol. 14 (2021) 1113–1175 MR Zbl
- [Baldwin and Sivek 2022a] **J A Baldwin, S Sivek**, *Framed instanton homology and concordance, II*, preprint (2022) arXiv 2206.11531
- [Baldwin and Sivek 2022b] **J A Baldwin, S Sivek**, *Khovanov homology detects the trefoils*, Duke Math. J. 171 (2022) 885–956 MR Zbl
- [Baldwin and Sivek 2023] **J A Baldwin, S Sivek**, *Instantons and L-space surgeries*, J. Eur. Math. Soc. 25 (2023) 4033–4122 MR Zbl
- [Baldwin and Sivek 2024] **J A Baldwin, S Sivek**, *Characterizing slopes for 5_2* , J. Lond. Math. Soc. 109 (2024) art. id. e12951 MR Zbl
- [Boardman 1999] **J M Boardman**, *Conditionally convergent spectral sequences*, from “Homotopy invariant algebraic structures” (J-P Meyer, J Morava, W S Wilson, editors), Contemp. Math. 239, Amer. Math. Soc., Providence, RI (1999) 49–84 MR Zbl
- [Eftekhary 2018] **E Eftekhary**, *Bordered Floer homology and existence of incompressible tori in homology spheres*, Compos. Math. 154 (2018) 1222–1268 MR Zbl
- [Floer 1990] **A Floer**, *Instanton homology, surgery, and knots*, from “Geometry of low-dimensional manifolds, I” (S K Donaldson, C B Thomas, editors), Lond. Math. Soc. Lect. Note Ser. 150, Cambridge Univ. Press (1990) 97–114 MR Zbl
- [Ghosh et al. 2024] **S Ghosh, Z Li, C-MM Wong**, *On the tau invariants in instanton and monopole Floer theories*, J. Topol. 17 (2024) art. id. e12346 MR Zbl
- [Hedden and Levine 2024] **M Hedden, A S Levine**, *A surgery formula for knot Floer homology*, Quantum Topol. 15 (2024) 229–336 MR Zbl
- [Honda 2000] **K Honda**, *On the classification of tight contact structures, I*, Geom. Topol. 4 (2000) 309–368 MR Zbl

- [Juhász 2006] **A Juhász**, *Holomorphic discs and sutured manifolds*, *Algebr. Geom. Topol.* 6 (2006) 1429–1457 MR Zbl
- [Kronheimer and Mrowka 2010] **P Kronheimer, T Mrowka**, *Knots, sutures, and excision*, *J. Differential Geom.* 84 (2010) 301–364 MR Zbl
- [Kronheimer and Mrowka 2011] **P B Kronheimer, T S Mrowka**, *Knot homology groups from instantons*, *J. Topol.* 4 (2011) 835–918 MR Zbl
- [Li 2020] **Z Li**, *Contact structures, excisions and sutured monopole Floer homology*, *Algebr. Geom. Topol.* 20 (2020) 2553–2588 MR Zbl
- [Li 2021a] **Z Li**, *Gluing maps and cobordism maps in sutured monopole and instanton Floer theories*, *Algebr. Geom. Topol.* 21 (2021) 3019–3071 MR Zbl
- [Li 2021b] **Z Li**, *Knot homologies in monopole and instanton theories via sutures*, *J. Symplectic Geom.* 19 (2021) 1339–1420 MR Zbl
- [Li and Ye 2021] **Z Li, F Ye**, *SU(2) representations and a large surgery formula*, preprint (2021) arXiv 2107.11005
- [Li and Ye 2022] **Z Li, F Ye**, *Instanton Floer homology, sutures, and Heegaard diagrams*, *J. Topol.* 15 (2022) 39–107 MR Zbl
- [Li and Ye 2024] **Z Li, F Ye**, *An enhanced Euler characteristic of sutured instanton homology*, *Int. Math. Res. Not.* 2024 (2024) 2873–2936 MR Zbl
- [Li and Ye 2025] **Z Li, F Ye**, *Knot surgery formulae for instanton Floer homology, II: Applications*, *Math. Ann.* 391 (2025) 6291–6371 MR Zbl
- [Lidman et al. 2022] **T Lidman, J Pinzón-Caicedo, C Scaduto**, *Framed instanton homology of surgeries on L-space knots*, *Indiana Univ. Math. J.* 71 (2022) 1317–1347 MR Zbl
- [Ozsváth and Szabó 2004] **P Ozsváth, Z Szabó**, *Holomorphic disks and knot invariants*, *Adv. Math.* 186 (2004) 58–116 MR Zbl
- [Ozsváth and Szabó 2008] **P S Ozsváth, Z Szabó**, *Knot Floer homology and integer surgeries*, *Algebr. Geom. Topol.* 8 (2008) 101–153 MR Zbl
- [Ozsváth and Szabó 2011] **P S Ozsváth, Z Szabó**, *Knot Floer homology and rational surgeries*, *Algebr. Geom. Topol.* 11 (2011) 1–68 MR Zbl
- [Sarkar 2015] **S Sarkar**, *Moving basepoints and the induced automorphisms of link Floer homology*, *Algebr. Geom. Topol.* 15 (2015) 2479–2515 MR Zbl
- [Scaduto 2015] **C W Scaduto**, *Instantons and odd Khovanov homology*, *J. Topol.* 8 (2015) 744–810 MR Zbl
- [Weibel 1994] **C A Weibel**, *An introduction to homological algebra*, *Cambridge Stud. Adv. Math.* 38, Cambridge Univ. Press (1994) MR Zbl
- [Zemke 2015] **I Zemke**, *Graph cobordisms and Heegaard Floer homology*, preprint (2015) arXiv 1512.01184

*Academy of Mathematics and Systems Science, Chinese Academy of Sciences
Beijing, China*

*Department of Mathematics, Harvard University
Cambridge, MA, United States*

zhenkun@amss.ac.cn, fanye@math.harvard.edu

Proposed: András I Stipsicz

Seconded: Ciprian Manolescu, Peter Ozsváth

Received: 27 September 2022

Revised: 3 August 2024

Morse actions of discrete groups on symmetric spaces: local-to-global principle

MICHAEL KAPOVICH

BERNHARD LEEB

JOAN PORTI

Our main result is a local-to-global principle for Morse quasigeodesics, maps and actions. As an application of our techniques we show algorithmic recognizability of Morse actions and construct Morse “Schottky subgroups” of higher-rank semisimple Lie groups via arguments not based on Tits ping-pong. Our argument is purely geometric and proceeds by constructing equivariant Morse quasiisometric embeddings of trees into higher-rank symmetric spaces.

22E40, 53C35; 20F65

1. Introduction	2343
2. Preliminaries	2346
3. Morse maps	2350
4. Group-theoretic applications	2361
Appendix A. Further properties of Morse quasigeodesics	2371
Appendix B. Geometry of nonpositively curved symmetric spaces	2379
References	2388

1 Introduction

This is a sequel to our paper [Kapovich et al. 2017] and mostly consists of the material of Section 7 of our earlier paper [Kapovich et al. 2014] (the only additional material appears in Theorem 4.9 and the appendices). We recall that quasigeodesics in Gromov hyperbolic spaces can be recognized locally by looking at sufficiently large finite pieces; see [Coornaert et al. 1990]. In our earlier papers [Kapovich et al. 2017; 2018a; 2018b] as well as [Kapovich and Leeb 2018b; 2018a], for higher-rank symmetric spaces X (of noncompact type) we introduced an analogue of hyperbolic quasigeodesics, which we call *Morse quasigeodesics*. Morse quasigeodesics are defined relatively to a certain face τ_{mod} of the model spherical face σ_{mod} of X (which can be defined as a fundamental domain for the action on the Tits building of X of the identity component $\text{Isom}_0(X)$ of the isometry group of X). In addition to the quasiisometry constants L and A , τ_{mod} -Morse quasigeodesics come equipped with two other parameters, a positive number D

and a Weyl-convex subset Θ of the open star of τ_{mod} in the modal spherical chamber σ_{mod} . In [Kapovich et al. 2014; 2017; 2018b] we also defined τ_{mod} -Morse maps $Y \rightarrow X$ from Gromov-hyperbolic spaces to symmetric spaces. These maps are defined by the property that they send geodesics to *uniformly τ_{mod} -Morse quasigeodesics*, ie τ_{mod} -Morse quasigeodesics with a fixed set of parameters, (Θ, D, L, A) .

The main result of this paper is a *local* characterization of Morse quasigeodesics in X :

Theorem 1.1 (local-to-global principle for Morse quasigeodesics) *For L, A, Θ, Θ', D there exist S, L', A', D' such that every S -local (Θ, D, L, A) -local Morse quasigeodesic in X is a (Θ', D', L', A') -Morse quasigeodesic.*

Here S -locality of a certain property of a map means that this property is satisfied for restrictions of this map to subintervals of length S . We refer to Theorem 3.34 and Theorem 3.34 for the details. Based on this principle, we prove in Section 3.7 a local-to-global principle for Morse maps from hyperbolic metric spaces to symmetric spaces.

Below, G is a semisimple Lie group acting isometrically and transitively on X , and K is a maximal compact subgroup of G , so that X is diffeomorphic to G/K . We will assume that G is commensurable with the isometry group $\text{Isom}(X)$ in the sense that we allow finite kernel and cokernel for the natural map $G \rightarrow \text{Isom}(X)$. However, we require G to act trivially on the model spherical chamber of X ; see Section B for details.

We prove several consequences of the local-to-global principle:

1. (Theorems 4.4 and 4.7) The *structural stability* of Morse subgroups of G , generalizing Sullivan's structural stability theorem in rank one [1985]; see also [Kapovich et al. 2024] for a detailed proof. While structural stability for Anosov subgroups was known earlier (Labourie and Guichard–Wienhard), our method is more general and applies to a wider class of discrete subgroups; see [Kapovich and Leeb \geq 2025].

Theorem 1.2 (openness of the space of Morse actions) *For a word hyperbolic group Γ , the subset of τ_{mod} -Morse actions of Γ on X is open in $\text{Hom}(\Gamma, G)$.*

Theorem 1.3 (structural stability) *Let Γ be word hyperbolic. Then for τ_{mod} -Morse actions $\rho : \Gamma \curvearrowright X$, the boundary embedding $\alpha_\rho : \partial_\infty \Gamma \rightarrow \text{Flag}(\tau_{\text{mod}})$ depends continuously on the action ρ .*

In particular, actions sufficiently close to a faithful Morse action are again discrete and faithful. We supplement this structural stability theorem with a stability theorem on *domains of proper discontinuity*, Theorem 4.9.

2. (Section 4.3) The locality of the Morse property implies that Morse subgroups are algorithmically recognizable:

Theorem 1.4 (semidecidability of Morse property of group actions) *Let Γ be word hyperbolic. Then there exists an algorithm whose inputs are homomorphisms $\rho : \Gamma \rightarrow G$ (defined on generators of Γ) and which terminates if and only if ρ defines a τ_{mod} -Morse action $\Gamma \curvearrowright X$.*

In other words, the property of being a τ_{mod} -Morse subgroup of G is *semidecidable*. If an action of Γ on X is not Morse, the algorithm runs forever. Note that in view of [Kapovich 2016], there are no algorithms (in the sense of BSS computability) which would recognize if a representation $\Gamma \rightarrow \text{Isom}(\mathbb{H}^3)$ is *not geometrically finite*.

3. (Section 4.2) We illustrate our techniques by constructing Morse–Schottky actions of free groups on higher-rank symmetric spaces. Unlike all previously known constructions, our proof does not rely on ping-pong arguments, but is purely geometric and proceeds by constructing equivariant quasiisometric embeddings of trees. The key step is the observation that a certain local *straightness* property for sufficiently spaced sequences of points in the symmetric space implies the global Morse property. This observation is also at the heart of the proof of the local-to-global principle for Morse actions.

Since [Kapovich et al. 2014] was originally posted, several improvements on the material of its Section 7 and, hence, of the present paper were made:

(a) Different forms of combination theorems for Anosov subgroups were proven in [Dey et al. 2019] and [Dey and Kapovich 2023; 2025], written in collaboration with Subhadip Dey. The first one was a generalization of the technique in Section 4.2 of the present paper, but the other two generalizations are based on a form of the ping-pong argument.

(b) Explicit estimates in the local-to-global principle for Morse quasigeodesics and, hence, Morse embeddings, were obtained by Max Riestenberg [2025]. Riestenberg’s estimates are based on replacing certain limiting arguments used in the present paper with differential-geometric and Lie-theoretic arguments.

Organization of the paper The notions of Morse quasigeodesics and actions are discussed in detail in Section 3. In that section, among other things, we establish local-to-global principles for Morse quasigeodesics.

In Section 4 we apply local-to-global principles to discrete subgroups of Lie groups: We show that Morse actions are structurally stable and algorithmically recognizable. We also construct Morse–Schottky actions of free groups on symmetric spaces. In Appendices A and B we prove further properties of Morse quasigeodesics that we found to be useful in our work.

Acknowledgements Kapovich was supported by NSF grants DMS-12-05312 and DMS-16-04241, by KIAS (the Korea Institute for Advanced Study) through the KIAS scholar program, and by a Simons Foundation Fellowship, grant 391602. Porti was supported by Spanish State Research Agency, through grant PID2021-125625NB-100 and Maria de Maeztu Program CEX2020-001084-M.

The three authors are also grateful to the GEAR grant which partially supported the IHP trimester in Winter of 2012 through DMS 1107452, 1107263, 1107367 “*RNMS: Geometric structures and representation varieties*” (the GEAR Network), and to the Max Planck Institute for Mathematics in Bonn, where some of this work was done. We are grateful to the referee for carefully reading the paper and for making various useful suggestions.

2 Preliminaries

2.1 Basic notions of geometry of symmetric spaces

Throughout the paper we will be using definitions, notation and results of our earlier work. For the reader's convenience, we also review these notions in detail in Appendix B.

We refer the reader to our earlier papers for the various notions related to symmetric spaces such as *polyhedral Finsler metrics* on symmetric spaces [Kapovich and Leeb 2018b], the *opposition involution* ι of σ_{mod} , *model faces* τ_{mod} of σ_{mod} and the associated τ_{mod} -flag-manifolds $\text{Flag}(\tau_{\text{mod}})$ [Kapovich et al. 2017, Sections 2.2.2 and 2.2.3], *Weyl-convex* subsets $\Theta \subset \sigma_{\text{mod}}$ [Kapovich et al. 2017, Definition 2.7], *type map* $\theta: \partial_{\infty} X \rightarrow \sigma_{\text{mod}}$, open Schubert cells $C(\tau) \subset \text{Flag}(\tau_{\text{mod}})$ [Kapovich et al. 2017, Section 2.4], Δ -valued distances d_{Δ} on X [Kapovich et al. 2017, Section 2.6], Θ -regular geodesic segments (see [Kapovich et al. 2017, Section 2.5.3]), *parallel sets*, *stars*, *open stars* and Θ -stars, $\text{st}(\tau)$, $\text{ost}(\tau)$, and $\text{st}_{\Theta}(\tau)$, *Weyl sectors* $V(x, \tau)$ [Kapovich et al. 2017, Section 2.4], *Weyl cones* $V(x, \text{st}(\tau))$ and Θ -cones $V(x, \text{st}_{\Theta}(\tau))$, *diamonds* $\diamond_{\tau_{\text{mod}}}(x, y)$ and Θ -diamonds $\diamond_{\Theta}(x, y)$ [Kapovich et al. 2017, Section 2.5], *τ_{mod} -regular sequences and groups* [Kapovich et al. 2017, Section 4.2], τ_{mod} -convergence subgroups, flag-convergence, the Finsler interpretation of flag-convergence (see [Kapovich and Leeb 2018b, Sections 4.5 and 5.2] and [Kapovich et al. 2017]), τ_{mod} -limit sets $\Lambda_{\tau_{\text{mod}}}(\Gamma) \subset \text{Flag}(\tau_{\text{mod}})$ [Kapovich et al. 2017, Section 4.5]), visual limit set [Kapovich et al. 2017, page 4], uniformly τ_{mod} -regular sequences and subgroups [Kapovich et al. 2017, Section 4.6]), Morse subgroups [Kapovich et al. 2017, Section 5.4] and, more generally, Morse quasigeodesics and Morse maps [Kapovich et al. 2018b, Definitions 5.31 and 5.33]), antipodal limit sets [Kapovich et al. 2017, Definition 5.1] and antipodal maps to flag-manifolds [Kapovich et al. 2018b, Definition 6.11].

In the paper we will be frequently using convexity of Θ -cones in X :

Proposition 2.1 [Kapovich et al. 2017, Proposition 2.10] *For every Weyl-convex subset $\Theta \subset \text{st}(\tau_{\text{mod}})$, for every $x \in X$ and $\tau \in \text{Flag}(\tau_{\text{mod}})$, the cone $V(x, \text{st}_{\Theta}(\tau)) \subset X$ is convex.*

2.2 Standing notation and conventions

- We will use the notation X for a symmetric space of *noncompact type*, G for a semisimple Lie group acting isometrically and transitively on X , and K for a maximal compact subgroup of G , so that X is diffeomorphic to G/K . We will assume that G is commensurable with the isometry group $\text{Isom}(X)$ in the sense that we allow finite kernel and cokernel for the natural map $G \rightarrow \text{Isom}(X)$. In particular, the image of G in $\text{Isom}(X)$ contains the identity component $\text{Isom}(X)_o$.
- We let $\tau_{\text{mod}} \subseteq \sigma_{\text{mod}}$ be a fixed ι -invariant face type.
- We will use the notation $x_n \xrightarrow{f} \tau \in \text{Flag}(\tau_{\text{mod}})$ for the *flag-convergence* of a τ_{mod} -regular sequence $x_n \in X$ to a simplex $\tau \in \text{Flag}(\tau_{\text{mod}})$.

- We will be using the notation Θ, Θ' for an ι -invariant, compact, Weyl-convex [Kapovich et al. 2017, Definition 2.7] subset of the open star $\text{ost}(\tau_{\text{mod}}) \subset \sigma_{\text{mod}}$.
- We will always assume that $\Theta < \Theta'$, meaning that $\Theta \subset \text{int}(\Theta')$.
- Constants $L, A, D, \epsilon, \delta, l, a, s, S$ are meant to be always strictly positive and $L \geq 1$.

2.3 ζ -angles

We fix as auxiliary datum an ι -invariant type $\zeta = \zeta_{\text{mod}} \in \text{int}(\tau_{\text{mod}})$. (We will omit the subscript in ζ_{mod} in order to avoid cumbersome notation for ζ -angles.) For a simplex $\tau \subset \partial_\infty X$ of type τ_{mod} , ie $\tau \in \text{Flag}(\tau_{\text{mod}})$, we define $\zeta(\tau) \in \tau$ as the ideal point of type ζ_{mod} . Given two such simplices $\tau_\pm \in \text{Flag}(\tau_{\text{mod}})$ and a point $x \in X$, define the ζ -angles

$$(2.2) \quad \angle_x^\zeta(\tau_-, \tau_+) := \angle_x^\zeta(\tau_-, \xi_+) := \angle_x(\xi_-, \xi_+),$$

where $\xi_\pm = \zeta(\tau_\pm)$.

Similarly, define the ζ -Tits angle

$$(2.3) \quad \angle_{\text{Tits}}^\zeta(\tau_-, \tau_+) = \angle_{\text{Tits}}^\zeta(\tau_-, \xi_+) := \angle_{\text{Tits}}(\xi_-, \xi_+) = \angle_x(\xi_-, \xi_+),$$

where x belongs to a flat $F \subset X$ such that $\tau_-, \tau_+ \subset \partial_{\text{Tits}} F$. Then simplices τ_\pm (of the same type) are antipodal if and only if

$$\angle_{\text{Tits}}^\zeta(\tau_-, \tau_+) = \pi$$

for some, equivalently, every, choice of ζ as above.

Remark 2.4 We observe that the ideal points ζ_\pm are opposite, $\angle_{\text{Tits}}(\zeta_-, \zeta_+) = \pi$, if and only if they can be seen under angle $\simeq \pi$ (ie close to π) from some point in X . More precisely, there exists $\epsilon(\zeta_{\text{mod}})$ such that:

If $\angle_x(\zeta_-, \zeta_+) > \pi - \epsilon(\zeta_{\text{mod}})$ for some point x , then ζ_\pm are opposite.

This follows from the angle comparison $\angle_x(\zeta_-, \zeta_+) \leq \angle_{\text{Tits}}(\zeta_-, \zeta_+)$ and the fact that the Tits distance between ideal points of the fixed type ζ_{mod} takes only finitely many values.

For a τ_{mod} -regular unit tangent vector $v \in T_x X$ we denote by $\tau(v) \subset \partial_\infty X$ the unique simplex of type τ_{mod} such that ray ρ_v with the initial direction v represents an ideal point in $\text{ost}(\tau(v))$. We put $\zeta(v) = \zeta(xy) = \zeta(\tau(v))$, where xy is a geodesic segment in X with the initial direction v . Note that $\zeta(v)$ depends continuously on $v \in TX$.

For a τ_{mod} -regular segment xy in X we let $\tau(xy) = \tau(v)$, where v is the unit vector tangent to xy .

For τ_{mod} -regular segments xy, xz and $\tau \in \text{Flag}(\tau_{\text{mod}})$, we define the ζ -angles

$$(2.5) \quad \angle_x^\zeta(\tau, y) = \angle_x^\zeta(y, \tau) = \angle_x^\zeta(\tau(xy), \tau), \quad \text{and} \quad \angle_x^\zeta(y, z) = \angle_x^\zeta(\tau(xy), \tau(xz)).$$

Note that the ζ -angle $\angle_x^\zeta(y, z)$ depends not on y, z but rather on the simplices $\tau(xy), \tau(xz)$. These ζ -angles will play the role of angles between diamonds $\diamond_{\tau_{\text{mod}}}(x, y)$ and $\diamond_{\tau_{\text{mod}}}(x, z)$, meeting at x . Note that if X has rank 1, then the ζ -angles are just the ordinary Riemannian angles.

We observe that compactness of the set of Θ -regular unit vectors in $T_x X$ and transitivity of the G -action on X imply that the map $v \mapsto \zeta(v)$ from the set of Θ -regular unit vectors to $G\zeta$ is uniformly continuous. This implies uniform continuity of the ζ -angle function $\angle_x^\zeta(\tau, y)$.

2.4 Distances to parallel sets versus angles

In this section we collect some geometric facts regarding parallel sets in symmetric spaces, primarily dealing with estimation of distances from points in X to parallel sets.

Remark 2.6 The constants and functions in this section are not explicit, and their existence is proven by compactness arguments. For explicit computations here and in Theorem 3.18, we refer the reader to the PhD thesis of Max Riestenberg [2025].

We first prove a lemma which strengthens Corollary 2.46 of [Kapovich et al. 2017].

Lemma 2.7 *Suppose that τ_\pm are antipodal simplices in $\partial_{\text{Tits}} X$. Then every geodesic ray γ asymptotic to a point $\xi \in \text{ost}(\tau_+)$ is strongly asymptotic to a geodesic ray in $P(\tau_-, \tau_+)$.*

Proof If ξ belongs to the interior of the simplex τ_+ , then the assertion follows from Corollary 2.46 of [Kapovich et al. 2017]:

Weyl sectors $V(x_1, \tau)$ and $V(x_2, \tau)$ are strongly asymptotic if and only if x_1 and x_2 lie in the same horocycle at τ .

We now consider the general case. Suppose that ξ belongs to an open simplex $\text{int}(\tau')$ such that τ is a face of τ' . Then there exists an apartment $a \subset \partial_{\text{Tits}} X$ containing both ξ (and, hence, τ' as well as τ) and the simplex τ_- . Let $F \subset X$ be the maximal flat with $\partial_\infty F = a$. Then F contains a geodesic asymptotic to points in τ_- and τ_+ . Therefore, F is contained in $P(\tau_-, \tau_+)$. On the other hand, by Corollary 2.46 of [Kapovich et al. 2017], applied to the simplex τ' , we conclude that γ is strongly asymptotic to a geodesic ray in F . \square

The following lemma provides a quantitative strengthening of the conclusion of Lemma 2.7:

Lemma 2.8 *Let Θ be a compact subset of $\text{ost}(\tau_+)$. Then those rays $x\xi$ with $\theta(\xi) \in \Theta$ are uniformly strongly asymptotic to $P(\tau_-, \tau_+)$, ie $d(\cdot, P(\tau_-, \tau_+))$ decays to zero along them uniformly in terms of $d(x, P(\tau_-, \tau_+))$ and Θ .*

Proof Suppose that the assertion of lemma is false, ie there exists $\epsilon > 0$, a sequence $T_i \in \mathbb{R}_+$ diverging to infinity, and a sequence of rays $\rho_i = x_i \xi_i$ with $\xi_i \in \Theta$ and $d(x_i, P(\tau_-, \tau_+)) \leq d$, such that

$$(2.9) \quad d(y, P(\tau_-, \tau_+)) \geq \epsilon \quad \text{for all } y \in \rho([0, T_i]).$$

Using the action of the stabilizer of $P(\tau_-, \tau_+)$, we can assume that the points x_i belong to a certain compact subset of X . Therefore, the sequence of rays $x_i \xi_i$ subconverges to a ray $x \xi$ with $d(x, P(\tau_-, \tau_+)) \leq d$ and $\xi \in \Theta$. The inequality (2.9) then implies that the entire limit ray $x \xi$ is contained outside of the open ϵ -neighborhood of the parallel set $P(\tau_-, \tau_+)$. However, in view of Lemma 2.7, the ray $x \xi$ is strongly asymptotic to a geodesic in $P(\tau_-, \tau_+)$. Contradiction. \square

We next relate distances from points $x \in X$ to parallel sets and the ζ -angles at x . Suppose that the simplices τ_{\pm} , equivalently, the ideal points $\zeta_{\pm} = \zeta(\tau_{\pm})$ (see Section 2.3), are opposite. Then

$$\angle_x^{\zeta}(\tau_-, \tau_+) = \angle_x(\zeta_-, \zeta_+) = \pi$$

if and only if x lies in the parallel set $P(\tau_-, \tau_+)$. Furthermore, $\angle_x^{\zeta}(\tau_-, \tau_+) \simeq \pi$ if and only if x is close to $P(\tau_-, \tau_+)$, and both quantities control each other near the parallel set. More precisely:

- Lemma 2.10** (i) *If $d(x, P(\tau_-, \tau_+)) \leq d$, then $\angle_x^{\zeta}(\tau_-, \tau_+) \geq \pi - \epsilon(d)$ with $\epsilon(d) \rightarrow 0$ as $d \rightarrow 0$.*
 (ii) *For sufficiently small ϵ , $\epsilon \leq \epsilon'(\zeta_{\text{mod}})$, we have: The inequality $\angle_x^{\zeta}(\tau_-, \tau_+) \geq \pi - \epsilon$ implies that $d(x, P(\tau_-, \tau_+)) \leq d(\epsilon)$ for some function $d(\epsilon)$ which converges to 0 as $\epsilon \rightarrow 0$.*

Proof The intersection of parabolic subgroups $P_{\tau_-} \cap P_{\tau_+}$ preserves the parallel set $P(\tau_-, \tau_+)$ and acts transitively on it. Compactness and the continuity of $\angle \cdot (\zeta_-, \zeta_+)$ therefore imply that $\pi - \angle \cdot (\zeta_-, \zeta_+)$ attains on the boundary of the tubular r -neighborhood of $P(\tau_-, \tau_+)$ a strictly positive maximum and minimum, which we denote by $\phi_1(r)$ and $\phi_2(r)$. Furthermore, $\phi_i(r) \rightarrow 0$ as $r \rightarrow 0$. We have the estimate

$$\pi - \phi_1(d(x, P(\tau_-, \tau_+))) \leq \angle_x(\zeta_-, \zeta_+) \leq \pi - \phi_2(d(x, P(\tau_-, \tau_+))).$$

The functions $\phi_i(r)$ are (weakly) monotonically increasing. This follows from the fact that, along rays asymptotic to ζ_- or ζ_+ , the angle $\angle \cdot (\zeta_-, \zeta_+)$ is monotonically increasing and the distance $d(\cdot, P(\tau_-, \tau_+))$ is monotonically decreasing. The estimate implies the assertions. \square

The control of $d(\cdot, P(\tau_-, \tau_+))$ and $\angle \cdot (\zeta_-, \zeta_+)$ “spreads” along the Weyl cone $V(x, \text{st}(\tau_+))$, since the latter is asymptotic to the parallel set $P(\tau_-, \tau_+)$. Moreover, the control improves if one enters the cone far into a τ_{mod} -regular direction. More precisely:

Lemma 2.11 *Let $y \in V(x, \text{st}_{\Theta}(\tau_+))$ be a point with $d(x, y) \geq l$.*

- (i) *If $d(x, P(\tau_-, \tau_+)) \leq d$, then*

$$d(y, P(\tau_-, \tau_+)) \leq D'(d, \Theta, l) \leq d,$$

with $D'(d, \Theta, l) \rightarrow 0$ as $l \rightarrow +\infty$.

- (ii) *For sufficiently small ϵ , $\epsilon \leq \epsilon'(\zeta_{\text{mod}})$, we have: If $\angle_x(\zeta_-, \zeta_+) \geq \pi - \epsilon$, then*

$$\angle_y(\zeta_-, \zeta_+) \geq \pi - \epsilon'(\epsilon, \Theta, l) \geq \pi - \epsilon(d(\epsilon)),$$

with $\epsilon'(\epsilon, \Theta, l) \rightarrow 0$ as $l \rightarrow +\infty$.

Proof The distance from $P(\tau_-, \tau_+)$ takes its maximum at the tip x of the cone $V(x, \text{st}(\tau_+))$, because it is monotonically decreasing along the rays $x\xi$ for $\xi \in \text{st}(\tau_+)$. This yields the right-hand bounds d and, applying Lemma 2.10 twice, $\epsilon(d(\epsilon))$.

Those rays $x\xi$ with uniformly τ_{mod} -regular type $\theta(\xi) \in \Theta$ are uniformly strongly asymptotic to $P(\tau_-, \tau_+)$, ie $d(\cdot, P(\tau_-, \tau_+))$ decays to zero along them uniformly in terms of d and Θ ; see Lemma 2.8. This yields the decay $D'(d, \Theta, l) \rightarrow 0$ as $l \rightarrow +\infty$. The decay of ϵ' follows by applying Lemma 2.10 again. \square

3 Morse maps

In this section we investigate the Morse property of sequences and maps. The main aim of this section is to establish a local criterion for being Morse. To do so we introduce a local notion of *straightness* for sequences of points in X . Morse sequences are in general not straight, but they become straight after suitable modification, namely by sufficiently coarsifying them and then passing to the sequence of successive midpoints. Conversely, the key result is that sufficiently spaced straight sequences are Morse. We conclude that there is a local-to-global characterization of the Morse property.

3.1 Morse quasigeodesics

Definition 3.1 (Morse quasigeodesic) A (Θ, D, L, A) -Morse quasigeodesic in X is an (L, A) -quasigeodesic $p: I \rightarrow X$ (defined on an interval $I \subset \mathbb{R}$) such that for all $t_1, t_2 \in I$, the subpath $p|_{[t_1, t_2]}$ is D -close to a Θ -diamond $\diamond_{\Theta}(x_1, x_2)$ with $d(x_i, p(t_i)) \leq D$.

We will refer to a quadruple (Θ, D, L, A) as a *Morse datum* and abbreviate $M = (\Theta, D, L, A)$. Set $M + D' = (\Theta, D + D', L, A + 2D')$. We say that M contains Θ if M has the form (Θ, D, L, A) for some $D \geq 0, L \geq 1, A \geq 0$.

The following lemma is immediate from the definition of an M -Morse quasigeodesic.

Lemma 3.2 (perturbation lemma) *If p, p' are paths in X such that p is M -Morse and $d(p, p') \leq D'$, then p' is $M + D'$ -Morse.*

A Morse quasigeodesic p is called a *Morse ray* if its domain is a half-line. If $I = \mathbb{R}$, then a Morse quasigeodesic is called a *Morse quasiline*.

Morse quasirays do not in general converge at infinity (in the visual compactification of X), but they τ_{mod} -converge at infinity. This is a consequence of:

Lemma 3.3 (conicality) *Every Morse quasiray $p: [0, \infty) \rightarrow X$ is uniformly Hausdorff close to a subset of a cone $V(p(0), \text{st}_{\Theta}(\tau))$ for a unique simplex τ of type τ_{mod} .*

Proof The subpaths $p|_{[0,t_0]}$ are uniformly Hausdorff close to Θ -diamonds. These subconverge to a cone $V(x, \text{st}_\Theta(\tau))$ uniformly close to $p(0)$ and τ a simplex of type τ_{mod} . This establishes the existence. Since $p(n) \xrightarrow{f} \tau$, the uniqueness of τ follows from the uniqueness of τ_{mod} -limits; see [Kapovich et al. 2017, Lemma 4.23]. \square

Definition 3.4 (end of Morse quasiray) We call the unique simplex given by the previous lemma the *end* of the Morse quasiray $p: [0, \infty) \rightarrow X$, and denote it by

$$p(+\infty) \in \text{Flag}(\tau_{\text{mod}}).$$

Hausdorff close Morse quasirays have the same end by Lemma 3.3. In Section 3.3 we will prove uniform continuity of ends of Morse quasirays with respect to the topology of *coarse convergence* of quasirays.

3.2 Morse maps

We now turn to *Morse maps* with more general domains (than just intervals).

Definition 3.5 Let Y be a Gromov-hyperbolic geodesic metric space. A map $f: Y \rightarrow X$ is called M -Morse if it sends geodesics in Y to M -Morse quasigeodesics.

Thus, every Morse map is a quasiisometric embedding. While this definition makes sense for general metric spaces, in [Kapovich et al. 2018b] we proved that the domain of a Morse map is necessarily hyperbolic.

More generally, one can define Morse maps on *quasigeodesic metric spaces*:

Definition 3.6 (quasigeodesic metric space) A metric space Z is called (l, a) -*quasigeodesic* if all pairs of points in Z can be connected by (l, a) -quasigeodesics. A space is called *quasigeodesic* if it is (l, a) -quasigeodesic for some pair of parameters l, a .

Every quasigeodesic space is quasiisometric to a geodesic metric space. Namely, if Z is a (λ, α) -quasigeodesic space then it is quasiisometric to the 1-skeleton of its $(\lambda + \alpha)$ -Rips complex (one proves this by repeating the proof of [Druţu and Kapovich 2018, Theorem 8.52]). The quasigeodesic spaces considered in this paper are discrete groups equipped with word metrics, for which the claim is clear since we can use the Cayley graph as the geodesic space. The next definition is a slight generalization of the notion of Morse maps defined above.

Definition 3.7 (Morse embedding) Let (Θ, D, L, A) be a Morse datum. An (Θ, D, L, A, l, a) -*Morse embedding* from an (l, a) -quasigeodesic space Z into X is a map $f: Z \rightarrow X$ which sends (l, a) -quasigeodesics in Z to (Θ, D, L, A) -Morse quasigeodesics in X .

Of course, every (l, a) -quasigeodesic metric space is also (l', a') -quasigeodesic space for any $l' \geq l, a' \geq a$. The next lemma shows that this choice of quasigeodesic constants is essentially irrelevant.

Lemma 3.8 *Let $f: Z \rightarrow X$ be a map from a Gromov-hyperbolic (l, a) -quasigeodesic space Z . If f is $M = (\Theta, D, L, A, l, a)$ -Morse then for any (l', a') , it sends (l', a') -quasigeodesics in Z to $M' = (\Theta, D', L', A')$ -Morse quasigeodesics in X . Here the datum M' depends only on M, l', a' and the hyperbolicity constant δ of Z .*

Proof This follows from the definition of Morse quasigeodesics, and the Morse lemma applied to Z . \square

Notice that the parameter Θ in the Morse datum M' is the same as in M . Hence, we arrive at:

Definition 3.9 A map $f: Z \rightarrow X$ of a quasigeodesic hyperbolic space Z is called Θ -Morse if it sends uniform quasigeodesics in Z to Θ -Morse uniform quasigeodesics in X .

This notion depends only on the quasiisometry class of Z , ie the precomposition of a Θ -Morse embedding with a quasiisometry is again Θ -Morse. For this to be true we have to require control on the images of quasigeodesics of arbitrarily bad (but uniform) quality.

Let Γ be a hyperbolic group with a fixed finite generating set S , and let Y be the Cayley graph of Γ with respect to S . For $x \in X$, an isometric action $\Gamma \curvearrowright X$ determines the *orbit map* $o_x: \Gamma \rightarrow \Gamma x \subset X$. Every such map extends to the Cayley graph Y of Γ , sending edges to geodesics in X .

Definition 3.10 An isometric action $\Gamma \curvearrowright X$, or a representation $\rho: \Gamma \rightarrow G$, is called M -Morse (with respect to a basepoint $x \in X$) if the (extended) orbit map $o_x: Y \rightarrow X$ is M -Morse. Similarly, a subgroup $\Gamma < G$ is Morse if the inclusion homomorphism $\Gamma \hookrightarrow G$ is Morse.

The Morse property of an action and the parameter Θ , of course, do not depend on the choice of a generating set of Γ and a basepoint x , but the triple (D, L, A) does. Thus, it makes sense to talk about Θ -Morse and τ_{mod} -Morse actions of hyperbolic groups, where $\Theta \subset \text{ost}(\tau_{\text{mod}})$. In [Kapovich et al. 2017; 2018b] and [Kapovich and Leeb 2018b] we gave many alternative definitions of Morse actions, including the equivalence of this definition to the notion of Anosov subgroups.

3.3 Continuity at infinity

Let X, Y be proper metric spaces. We fix a basepoint $y \in Y$.

Definition 3.11 A sequence of maps $f_n: Y \rightarrow X$ is said to *coarsely converge* to a map $f: Y \rightarrow X$ if there exists $C < \infty$ such that for every R there exists $N = N(C, R)$ for which

$$d(f_n|_B, f|_B) \leq C \quad \text{for all } n \geq N,$$

where $B = B(y, R)$.

Note the difference between this definition and the notion of uniform convergence on compacts: since we are working in the coarse setting, requiring the distance between maps to be less than ϵ close to zero is pointless.

In view of the Arzelà–Ascoli theorem, the space of (L, A) -coarse Lipschitz maps $Y \rightarrow X$ sending y to a fixed bounded subset of X is coarsely sequentially compact: every sequence contains a coarsely converging subsequence.

In the next lemma we assume that Y is a geodesic δ -hyperbolic space and X is a symmetric space of noncompact type. The lemma itself is an immediate consequence of the perturbation lemma, Lemma 3.2.

Lemma 3.12 *Suppose that $p_n: \mathbb{R}_+ \rightarrow X$ is a sequence of M -Morse rays which coarsely converges to a map $p: \mathbb{R}_+ \rightarrow X$. Then p is M' -Morse, where $M' = M + C$ and the constant C is the one appearing in the definition of coarse convergence.*

In particular, a coarse limit of a sequence of (uniformly) Morse quasigeodesics is again Morse.

For the next lemma, we equip the flag-manifold $F = \text{Flag}(\tau_{\text{mod}})$ with some background metric d_F .

Lemma 3.13 *Suppose that $p_n: \mathbb{R}_+ \rightarrow X$ is a sequence of M -Morse rays coarsely converging to a M -Morse ray $p: \mathbb{R}_+ \rightarrow X$. Then the sequence $\tau_n := p_n(\infty)$ of ends of the quasirays p_n converges to $\tau = p(\infty)$. Moreover, the latter convergence is uniform in the following sense. For every $\epsilon > 0$ there exists n_0 depending only on M and C and $N(R, C)$ (appearing in Definition 3.11) such that for all $n \geq n_0$, $d_F(\tau_n, \tau) \leq \epsilon$.*

Proof Suppose that the claim is false. Then in view of coarse compactness of the space of M -Morse maps sending y to a fixed compact subset of X , there exists a sequence (p_n) as in the lemma, coarsely converging to p , such that the sequence $p_n(\infty) = \tau_n$ converges to $\tau' \neq p(\infty) = \tau$. By the coarse convergence $p_n \rightarrow p$, there exists $C < \infty$ and a sequence $t_n \rightarrow \infty$ such that $d(p_n(t_n), p(t_n)) \leq C$. By the definition of Morse quasigeodesics, there exists a sequence of cones $V(x_n, \text{st}(\tau_n))$ (with x_n in a bounded subset $B \subset X$) such that the image of p_n is contained in the D -neighborhood of $V(x_n, \text{st}(\tau_n))$. Thus, the sequence $(p_n(t_n))$ flag-converges to τ' , while $(p(t_n))$ flag-converges to τ . According to [Kapovich et al. 2017, Lemma 4.23], altering a sequence by a uniformly bounded amount does not change the flag-limit. Therefore, the sequence $(p(t_n))$ also flag-converges to τ' . Hence, $\tau = \tau'$. A contradiction. \square

3.4 A Morse lemma for straight sequences

In order to motivate the results of this section we recall the following *sufficient condition* for a piecewise geodesic path in a Hadamard manifold Y of curvature ≤ -1 to be quasigeodesic; see eg [Kapovich and Liu 2019].

Proposition 3.14 *Suppose that c is a piecewise geodesic path in Y whose angles at the vertices are $\geq \alpha > 0$ and whose edges are longer than L , where α and L satisfy*

$$(3.15) \quad \cosh(L/2) \sin(\alpha/2) \geq \nu > 1.$$

Then c is an $(L(\nu), A(\nu))$ -quasigeodesic.

By considering c with vertices on a horocycle in the hyperbolic plane, one sees that the inequality in this proposition is sharp.

Corollary 3.16 *If L is sufficiently large and α is sufficiently close to π , then c is (uniformly) quasi-geodesic.*

In higher rank, we do not have an analogue of the inequality (3.15); instead, we will be generalizing the corollary. However, *angles* in the corollary will be replaced with ζ -*angles*. We will show (in a “string of diamonds” theorem, Theorem 3.30) that if a piecewise geodesic path c in X has sufficiently long edges, and ζ -angles between consecutive segments sufficiently close to π , then c is M -Morse for a suitable Morse datum.

In the following, we consider finite or infinite sequences (x_n) of points in X . For the next definition, we remind the reader of the definition of ζ -angles given in (2.5).

Definition 3.17 (straight and spaced sequence) We call a sequence (x_n) (Θ, ϵ) -*straight* if the segments $x_n x_{n+1}$ are Θ -regular and

$$\angle_{x_n}^{\zeta}(x_{n-1}, x_{n+1}) \geq \pi - \epsilon$$

for all n . We call it l -*spaced* if the segments $x_n x_{n+1}$ have length $\geq l$.

Note that every straight sequence can be extended to a biinfinite straight sequence.

Straightness is a local condition. The goal of this section is to prove the following local-to-global result asserting that sufficiently straight and spaced sequences satisfy a higher-rank version of the Morse lemma (for quasigeodesics in hyperbolic space).

Theorem 3.18 (Morse lemma for straight spaced sequences) *For Θ, Θ', δ there exist l, ϵ such that every (Θ, ϵ) -straight l -spaced sequence (x_n) is δ -close to a parallel set $P(\tau_-, \tau_+)$ with simplices τ_{\pm} of type τ_{mod} , and it moves from τ_- to τ_+ in the sense that its nearest-point projection \bar{x}_n to $P(\tau_-, \tau_+)$ satisfies*

$$(3.19) \quad \bar{x}_{n \pm m} \in V(\bar{x}_n, \text{st}_{\Theta'}(\tau_{\pm})) \quad \text{for all } n \text{ and } m \geq 1.$$

Remark 3.20 (global spacing) (1) As a corollary of this theorem, we will show that straight spaced sequences are quasigeodesic:

$$d(x_n, x_{n+m}) \geq c|m - 2\delta$$

with a constant $c = c(\Theta') > 0$. See Corollary 3.29. In particular, by interpolating the sequence (x_n) via geodesic segments we obtain a Morse quasigeodesic in X .

- (2) Theorem 3.18 is a higher-rank generalization of two familiar facts from geometry of Gromov-hyperbolic geodesic metric spaces: The fact that local quasigeodesics (with suitable parameters) are global quasigeodesics and the Morse lemma stating that quasigeodesics stay uniformly close to geodesics. In the higher rank, quasigeodesics, of course, need not be close to geodesics, but, instead (under the straightness assumption), are close to diamonds/Weyl cones/parallel sets. A different version of the same phenomenon is established in our paper [Kapovich et al. 2018b, Theorem 1.3], where closeness of certain quasigeodesics to diamonds/Weyl cones/parallel sets is proven under the *uniform τ_{mod} -regularity* (rather than the straightness) assumption. As far as we know, neither result directly implies the other, ie neither uniform regularity directly implies straightness, nor vice-versa.
- (3) One can obviously strengthen the Corollary 3.16 by stating that for each $\epsilon < \pi$ there exists $L_0(\epsilon)$ such that if $\alpha \geq \pi - \epsilon$ and $L \geq L_0(\epsilon)$, then c is a uniform quasigeodesic in X . A similar strengthening is false for symmetric spaces of rank ≥ 2 . For instance, when $W \cong S_3$ and $\epsilon = 2\pi/3$, then no matter what Θ, Θ' and l are, the conclusion of Theorem 3.18 fails already for sequences contained in a single flat.

In order to prove the theorem, we start by considering half-infinite sequences and prove that they keep moving away from an ideal simplex of type τ_{mod} if they do so initially.

For the next definition we recall the definition of ζ -angles given in Section 2.3. Given a face $\tau \subset \partial_{\text{Tits}} X$ of type τ_{mod} and distinct points $x, y \in X$, we have the angle

$$\angle_x^\zeta(\tau, y) := \angle_x^\zeta(z, y) = \angle_x(z, \zeta(xy)),$$

where z is a point (distinct from x) on the geodesic ray $x\xi$, where $\xi \in \tau$ is the point of type ζ . (See (2.5).)

Definition 3.21 (moving away from an ideal simplex) We say that a sequence (x_n) moves ϵ -away from a simplex τ of type τ_{mod} if

$$\angle_{x_n}^\zeta(\tau, x_{n+1}) \geq \pi - \epsilon \quad \text{for all } n.$$

Lemma 3.22 (moving away from ideal simplices) For small ϵ and large $l, \epsilon \leq \epsilon_0$ and $l \geq l(\epsilon, \Theta)$, the following holds:

If the sequence $(x_n)_{n \geq 0}$ is (Θ, ϵ) -straight l -spaced and if

$$\angle_{x_0}^\zeta(\tau, x_1) \geq \pi - 2\epsilon,$$

then (x_n) moves ϵ -away from τ .

Proof By Lemma 2.11(ii), the unit-speed geodesic segment $c: [0, t_1] \rightarrow X$ from $p(0)$ to $p(1)$ moves $\epsilon(d(2\epsilon))$ -away from τ at all times, and $\epsilon'(2\epsilon, \Theta, l)$ -away at times $\geq l$, which includes the final time t_1 . For $l(\epsilon, \Theta)$ sufficiently large, we have $\epsilon'(2\epsilon, \Theta, l) \leq \epsilon$. Then c moves ϵ -away from τ at time t_1 , which means that $\angle_{x_1}^\zeta(\tau, x_0) \leq \epsilon$. Straightness at x_1 and the triangle inequality yield that again $\angle_{x_1}^\zeta(\tau, x_2) \geq \pi - 2\epsilon$. One proceeds by induction. □

Note that there do exist simplices τ satisfying the hypothesis of the previous lemma. For instance, one can extend the initial segment x_0x_1 backwards to infinity and choose $\tau = \tau(x_1x_0)$.

Now we look at *biinfinite* sequences.

We assume in the following that $(x_n)_{n \in \mathbb{Z}}$ is (Θ, ϵ) -straight l -spaced for small ϵ and large l . As a first step, we study the asymptotics of such sequences and use the argument for Lemma 3.22 to find a pair of opposite ideal simplices τ_{\pm} such that (x_n) moves from τ_- towards τ_+ .

Lemma 3.23 (moving towards ideal simplices) *For small ϵ and large l , $\epsilon \leq \epsilon_0$ and $l \geq l(\epsilon, \Theta)$, the following holds:*

There exists a pair of opposite simplices τ_{\pm} of type τ_{mod} such that the inequality

$$(3.24) \quad \angle_{x_n}^{\zeta}(\tau_{\mp}, x_{n \mp 1}) \geq \pi - 2\epsilon$$

holds for all n .

Proof 1. For every n define a compact set $C_n^{\mp} \subset \text{Flag}(\tau_{\text{mod}})$

$$C_n^{\pm} = \{\tau_{\pm} : \angle_{x_n}^{\zeta}(\tau_{\pm}, x_{n \mp 1}) \geq \pi - 2\epsilon\}.$$

As in the proof of Lemma 3.22, straightness at x_{n+1} implies that $C_n^- \subset C_{n+1}^-$. Hence the family $\{C_n^-\}_{n \in \mathbb{Z}}$ form a nested sequence of nonempty compact subsets and therefore have nonempty intersection containing a simplex τ_- . Analogously, there exists a simplex τ_+ which belongs to C_n^+ for all n .

2. It remains to show that the simplices τ_-, τ_+ are antipodal. Using straightness and the triangle inequality, we see that

$$(\star) \quad \angle_{x_n}^{\zeta}(\tau_-, \tau_+) \geq \pi - 5\epsilon$$

for all n . Hence, if $5\epsilon < \epsilon(\zeta)$, then the simplices τ_-, τ_+ are antipodal in view of Remark 2.4. □

The pair of opposite simplices (τ_-, τ_+) which we found determines a parallel set in X . The second step is to show that (x_n) is uniformly close to it.

Lemma 3.25 (close to parallel set) *For small ϵ and large l , $\epsilon \leq \epsilon(\delta)$ and $l \geq l(\Theta, \delta)$, the sequence (x_n) is δ -close to $P(\tau_-, \tau_+)$.*

Proof The statement follows from the combination of the inequality (\star) in the second part of the proof of Lemma 3.23, and Lemma 2.10. □

The third and final step of the proof is to show that the nearest point projection (\bar{x}_n) of (x_n) to $P(\tau_-, \tau_+)$ moves from τ_- towards τ_+ .

Lemma 3.26 (projection moves towards ideal simplices) *For small ϵ and large l , with $\epsilon \leq \epsilon_0$ and $l \geq l(\epsilon, \Theta, \Theta')$, the segments $\bar{x}_n\bar{x}_{n+1}$ are Θ' -regular and*

$$\angle_{\bar{x}_n}^{\zeta}(\tau_-, \bar{x}_{n+1}) = \pi \quad \text{for all } n.$$

Proof By the previous lemma, (x_n) is δ_0 -close to $P(\tau_-, \tau_+)$ if ϵ_0 is sufficiently small and l is sufficiently large. Since $x_n x_{n+1}$ is Θ -regular, the triangle inequality for Δ -lengths yields that the segment $\bar{x}_n \bar{x}_{n+1}$ is Θ' -regular, again if l is sufficiently large.

Let ξ_+ denote the ideal endpoint of the ray extending this segment, ie $\bar{x}_{n+1} \in \bar{x}_n \xi_+$. Then x_{n+1} is $2\delta_0$ -close to the ray $x_n \xi_+$. We obtain that (in view of the uniform continuity of ζ -angles observed in Section 2.3)

$$\angle_{\text{Tits}}^\zeta(\tau_-, \xi_+) \geq \angle_{x_n}^\zeta(\tau_-, \xi_+) \simeq \angle_{x_n}^\zeta(\tau_-, x_{n+1}) \simeq \pi,$$

where the last step follows from inequality (3.24). The discreteness of Tits distances between ideal points of fixed type ζ implies that in fact

$$\angle_{\text{Tits}}^\zeta(\tau_-, \xi_+) = \pi,$$

ie the ideal points $\zeta(\tau_-)$ and $\zeta(\xi_+)$ are antipodal. But the only simplex opposite to τ_- in $\partial_\infty P(\tau_-, \tau_+)$ is τ_+ , so $\tau(\xi_+) = \tau_+$ and

$$\angle_{\bar{x}_n}^\zeta(\tau_-, \bar{x}_{n+1}) = \angle_{\bar{x}_n}^\zeta(\tau_-, \xi_+) = \pi,$$

as claimed. □

Proof of Theorem 3.18 It suffices to consider biinfinite sequences.

The conclusion of Lemma 3.26 is equivalent to $\bar{x}_{n+1} \in V(\bar{x}_n, \text{st}_{\Theta'}(\tau_+))$. Combining Lemmas 3.25 and 3.26, we thus obtain the theorem for $m = 1$.

The convexity of Θ' -cones, cf Proposition 2.1, implies that

$$V(\bar{x}_{n+1}, \text{st}_{\Theta'}(\tau_+)) \subset V(\bar{x}_n, \text{st}_{\Theta'}(\tau_+)),$$

and the assertion follows for all $m \geq 1$ by induction. □

Remark 3.27 The conclusion of the theorem implies flag-convergence $x_{\pm n} \rightarrow \tau_{\pm}$ as $n \rightarrow +\infty$. However, the sequences $(x_n)_{n \in \pm\mathbb{N}}$ do not in general converge at infinity, but accumulate at compact subsets of $\text{st}_{\Theta'}(\tau_{\pm})$.

3.5 Lipschitz retractions to straight paths

Consider a (possibly infinite) closed interval J in \mathbb{R} ; we will assume that J has integer or infinite bounds. Suppose that $p: J \cap \mathbb{Z} \rightarrow P = P(\tau_-, \tau_+) \subset X$ is an l -separated, λ -Lipschitz, $(\Theta, 0)$ -straight coarse sequence pointing away from τ_- and towards τ_+ . We extend p to a piecewise geodesic map $p: J \rightarrow P$ by sending intervals $[n, n + 1]$ to geodesic segments $p(n)p(n + 1)$ via affine maps. We retain the name p for the extension.

Lemma 3.28 *There exists $L = L(l, \lambda, \Theta)$ and an L -Lipschitz retraction of X to p , ie an L -Lipschitz map $r: X \rightarrow J$ such that $r \circ p = \text{Id}$. In particular, $p: J \cap \mathbb{Z} \rightarrow X$ is an (\bar{L}, \bar{A}) -quasigeodesic, where \bar{L} and \bar{A} depend only on l, λ and Θ .*

Proof It suffices to prove existence of a retraction. Since P is convex in X , it suffices to construct a map $P \rightarrow J$. Pick a generic point $\xi = \xi_+ \in \tau_+$ and let $b_\xi: P \rightarrow \mathbb{R}$ denote the Busemann function normalized so that $b_\xi(p(z)) = 0$ for some $z \in J \cap \mathbb{Z}$. Then the Θ -regularity assumption on p implies that the slope of the piecewise linear function $b_\xi \circ p: J \rightarrow \mathbb{R}$ is strictly positive, bounded away from 0. The assumption that p is l -separated λ -Lipschitz implies that

$$l \leq |p'(t)| \leq \lambda$$

for each t (where the derivative exists). The straightness assumption on p implies that the function $h := b_\xi \circ p: J \rightarrow \mathbb{R}$ is strictly increasing. By combining these observations, we conclude that h is an L -bi-Lipschitz homeomorphism to $h(J)$ for some $L = L(l, \lambda, \Theta)$. Let $\rho: \mathbb{R} \rightarrow J$ denote the nearest-point projection to the interval J ; it is a 1-Lipschitz map. Lastly, we define

$$r: P \rightarrow J \quad \text{by} \quad r = h^{-1} \circ \rho \circ b_\xi.$$

Since b_ξ is 1-Lipschitz, the map r is L -Lipschitz. By the construction, $r \circ p = \text{Id}$. □

Corollary 3.29 Fix $\delta > 0$ and Θ, Θ' such that $\Theta \subset \text{int}(\Theta')$. Let $l = l(l, \Theta, \Theta', \delta)$ and $\epsilon = \epsilon(l, \Theta, \Theta', \delta)$ be the constants as in Theorem 3.18. Suppose that $p: J \cap \mathbb{Z} \rightarrow X$ is an l -spaced, λ -Lipschitz, (Θ, ϵ) -straight sequence. Then for $L = L(l - 2\delta, \lambda + 2\delta, \Theta')$ we have:

- (1) There exists an $(L, 2\delta)$ -coarse Lipschitz retraction $X \rightarrow J$.
- (2) The map p is a (Θ', D', L', A') -quasigeodesic with D', L', A' depending only on $l, \lambda, \Theta, \Theta', \epsilon$.

Proof The statement immediately follows the above lemma combined with Theorem 3.18. □

Reformulating in terms of piecewise geodesic paths, we obtain:

Theorem 3.30 (string of diamonds theorem) For any pair of Weyl convex subsets $\Theta < \Theta'$ and a number $D \geq 0$, there exist positive numbers ϵ, S, L, A depending on the datum (Θ, Θ', D) such that the following holds.

Suppose that c is an arc-length parametrized piecewise geodesic path (finite or infinite) in X obtained by concatenating geodesic segments $x_i x_{i+1}$ such that, for all i ,

- (1) each segment $x_i x_{i+1}$ is Θ -regular and has length $\geq S$, and
- (2) $\angle_{x_i}^\xi(x_{i-1}, x_{i+1}) \geq \pi - \epsilon$.

Then the path c is (Θ', D, L, A) -Morse.

3.6 Local Morse quasigeodesics

According to Theorem 3.30, sufficiently straight and spaced straight piecewise geodesic paths are Morse. In this section we will now prove that, conversely, the Morse property implies straightness in a suitable sense, namely that for sufficiently spaced quadruples the associated midpoint triples are arbitrarily straight. (For the quadruples themselves this is in general not true.)

Definition 3.31 (quadruple condition) For points $x, y \in X$, we let $\text{mid}(x, y)$ denote the midpoint of the geodesic segment xy . A map $p: I \rightarrow X$ satisfies the (Θ, ϵ, l, s) -quadruple condition if for all $t_1, t_2, t_3, t_4 \in I$ with $t_2 - t_1, t_3 - t_2, t_4 - t_3 \geq s$, the triple of midpoints

$$(\text{mid}(t_1, t_2), \text{mid}(t_2, t_3), \text{mid}(t_3, t_4))$$

is (Θ, ϵ) -straight and l -spaced.

Proposition 3.32 (Morse implies quadruple condition) For $L, A, \Theta, \Theta', D, \epsilon, l$ there exists a scale $s = s(L, A, \Theta, \Theta', D, \epsilon, l)$ such that every (Θ, D, L, A) -Morse quasigeodesic satisfies the $(\Theta', \epsilon, l, s')$ -quadruple condition for every $s' \geq s$.

Proof Let $p: I \rightarrow X$ be an (L, A, Θ, D) -Morse quasigeodesic, and let $t_1, \dots, t_4 \in I$ be such that $t_2 - t_1, t_3 - t_2, t_4 - t_3 \geq s$. We abbreviate $p_i := p(t_i)$ and $m_i = \text{mid}(p_i, p_{i+1})$.

Regarding straightness, it suffices to show that the segment m_2m_1 is Θ' -regular and that $\angle_{m_2}^\xi(p_2, m_1) \leq \frac{1}{2}\epsilon$ provided s is sufficiently large in terms of the given data.

By the Morse property, there exists a diamond $\diamond_\Theta(x_1, x_3)$ such that

$$d(x_1, p_1), d(x_3, p_3) \leq D \quad \text{and} \quad p_2 \in N_D(\diamond_\Theta(x_1, x_3)).$$

The diamond spans a unique parallel set $P(\tau_-, \tau_+)$. (Necessarily, we have $x_3 \in V(x_1, \text{st}_\Theta(\tau_+))$ and $x_1 \in V(x_3, \text{st}_\Theta(\tau_-))$.)

We denote by \bar{p}_i and \bar{m}_i the projections of p_i and m_i to the parallel set.

We first observe that m_2 (and m_3) is arbitrarily close to the parallel set if s is large enough. If this were not true, a limiting argument would produce a geodesic line at strictly positive finite Hausdorff distance $\in (0, D]$ from $P(\tau_-, \tau_+)$ and asymptotic to ideal points in $\text{st}_\Theta(\tau_\pm)$. However, all lines asymptotic to ideal points in $\text{st}_\Theta(\tau_\pm)$ are contained in $P(\tau_-, \tau_+)$.

Next, we look at the directions of the segments $\bar{m}_2\bar{m}_1$ and $\bar{m}_2\bar{p}_2$ and show that they have the same τ -direction. Since \bar{p}_2 is $2D$ -close to $V(\bar{p}_1, \text{st}_\Theta(\tau_+))$, the point \bar{p}_1 is $2D$ -close to $V(\bar{p}_2, \text{st}_\Theta(\tau_-))$, and hence also \bar{m}_1 is $2D$ -close to $V(\bar{p}_2, \text{st}_\Theta(\tau_-))$. Therefore, $\bar{p}_1, \bar{m}_1 \in V(\bar{p}_2, \text{st}_\Theta(\tau_-))$ if s is large enough. Similarly, $\bar{m}_2 \in V(\bar{p}_2, \text{st}_{\Theta'}(\tau_+))$ and hence $\bar{p}_2 \in V(\bar{m}_2, \text{st}_{\Theta'}(\tau_-))$. The convexity of Θ' -cones, see Proposition 2.1, implies that also $\bar{m}_1 \in V(\bar{m}_2, \text{st}_{\Theta'}(\tau_-))$. In particular, $\angle_{\bar{m}_2}^\xi(\bar{p}_2, \bar{m}_1) = 0$ if s is sufficiently large.

Since m_2 is arbitrarily close to the parallel set if s is sufficiently large, it follows by another limiting argument that $\angle_{m_2}^\xi(p_2, m_1) \leq \epsilon/2$ if s is sufficiently large.

Regarding the spacing, we use that $\bar{m}_1 \in V(\bar{p}_2, \text{st}_{\Theta'}(\tau_-))$ and $\bar{m}_2 \in V(\bar{p}_2, \text{st}_{\Theta'}(\tau_+))$. It follows that

$$d(\bar{m}_1, \bar{m}_2) \geq c \cdot (d(\bar{m}_1, \bar{p}_2) + d(\bar{p}_2, \bar{m}_2))$$

with a constant $c = c(\Theta') > 0$, and hence that $d(m_1, m_2) \geq l$ if s is sufficiently large. □

Theorem 3.18 and Proposition 3.32 say that the Morse property for quasigeodesics is equivalent to straightness (of associated spaced sequences of points). Since straightness is a local condition, this leads to a local-to-global result for Morse quasigeodesics, namely that the Morse property holds globally if it holds locally up to a sufficiently large scale.

Definition 3.33 (local Morse quasigeodesic) An S -local (Θ, D, L, A) -Morse quasigeodesic in X is a map $p: I \rightarrow X$ such that for all t_0 , the subpath $p|_{[t_0, t_0+S]}$ is a (Θ, D, L, A) -Morse quasigeodesic.

Note that local Morse quasigeodesics are uniformly coarse Lipschitz.

Theorem 3.34 (local-to-global principle for Morse quasigeodesics) For L, A, Θ, Θ', D there exist S, L', A', D' such that every S -local (Θ, D, L, A) -local Morse quasigeodesic in X is an (Θ', D', L', A') -Morse quasigeodesic.

Proof We choose an auxiliary Weyl convex subset Θ'' such that $\Theta < \Theta'' < \Theta'$.

Let $p: I \rightarrow X$ be an S -local (Θ, D, L, A) -local Morse quasigeodesic. We consider its coarsification on a (large) scale s and the associated midpoint sequence, ie we put $p_n^s = p(ns)$ and $m_n^s = \text{mid}(p_n^s, p_{n+1}^s)$.

Whereas the coarsification itself does not in general become arbitrarily straight as the scale s increases, this is true for its midpoint sequence due to Proposition 3.32. We want it to be sufficiently straight and spaced so that we can apply to it the Morse lemma from Theorem 3.18. Therefore we first fix an auxiliary constant δ , and further auxiliary constants l, ϵ as determined by Theorem 3.18 in terms of Θ', Θ'' and δ . Then Proposition 3.32 applied to the (Θ, D, L, A) -Morse quasigeodesics $p|_{[t_0, t_0+S]}$ yields that (m_n^s) is (Θ'', ϵ) -straight and l -spaced if $S \geq 3s$ and the scale s is large enough depending on $L, A, \Theta, \Theta'', D, \epsilon, l$.

Now we can apply Theorem 3.18 to (m_n^s) . It yields a nearby sequence (\bar{m}_n^s) with $d(\bar{m}_n^s, m_n^s) \leq \delta$, which has the following property: For all $n_1 < n_2 < n_3$, the segments $\bar{m}_{n_1}^s \bar{m}_{n_3}^s$ are uniformly regular and the points $m_{n_2}^s$ are δ -close to the diamonds $\diamond_{\Theta'}(\bar{m}_{n_1}^s, \bar{m}_{n_3}^s)$.

Since the subpaths $p|_{[ns, (n+1)s]}$ filling in (p_n^s) are (L, A) -quasigeodesics (because $S \geq s$), and it follows that for all $t_1, t_2 \in I$ the subpaths $p|_{[t_1, t_2]}$ are D' -close to Θ' -diamonds with D' depending on L, A, s .

The conclusion of Theorem 3.18 also implies a global spacing for the sequence (m_n^s) , compare Remark 3.20, ie $d(m_n^s, m_{n'}^s) \geq c \cdot |n - n'|$ with a positive constant c depending on Θ', l . Hence p is a global (L', A') -quasigeodesic with L', A' depending on L, A, s, c .

Combining this information, we obtain that p is an (Θ', D', L', A') -Morse quasigeodesic for certain constants L', A' and D' depending on L, A, Θ, Θ' and D , provided that the scale S is sufficiently large in terms of the same data. \square

3.7 Local-to-global principle for Morse maps

We now deduce from our local-to-global result for Morse quasigeodesics, Theorem 3.34, a local-to-global result for Morse embeddings.

We restrict to the setting of maps of Gromov-hyperbolic (l, a) -quasigeodesic metric spaces Z to symmetric spaces X .

Definition 3.35 (local Morse embedding) We call a map $f: Z \rightarrow X$ an S -local (Θ, D, L, A) -Morse embedding if for any (l, a) -quasigeodesic $q: I \rightarrow Z$ defined on an interval I of length $\leq S$, the image path $f \circ q$ is a (Θ, D, L, A) -Morse quasigeodesic in X .

Theorem 3.36 (local-to-global principle for Morse embeddings of Gromov hyperbolic spaces) For $l, a, L, A, \Theta, \Theta', D$ there exists a scale S and a datum (D', L', A') such that every S -local (Θ, D, L, A) -Morse embedding from an (l, a) -quasigeodesic Gromov hyperbolic space into X is a (Θ', D', L', A') -Morse embedding.

Proof Let $f: Z \rightarrow X$ denote the local Morse embedding. It sends every (l, a) -quasigeodesic $q: I \rightarrow Z$ to an S -local (Θ, D, L, A) -Morse quasigeodesic $p = f \circ q$ in X . By Theorem 3.34, p is (L', A', Θ', D') -Morse if $S \geq S(l, a, L, A, \Theta, \Theta', D)$, where L', A', D' depend on the given data. \square

Below is a reformulation of this theorem in the case of geodesic Gromov-hyperbolic spaces.

Let Z be a δ -hyperbolic geodesic space. An R -ball $B(z, R)$ in Z need not be convex, but it is δ -quasiconvex. In particular, the restriction of the metric from Z to $B(z, R)$ results in a $(1, \delta)$ -quasigeodesic metric space.

Theorem 3.37 (local-to-global principle for Morse embeddings of geodesic spaces) For L, A, Θ, Θ', D and δ there exists a scale R and a datum (D', L', A') such that if Z is a δ -hyperbolic geodesic metric space and the restriction of f to any R -ball is $(\Theta, D, L, A, 1, \delta)$ -Morse, then $f: Z \rightarrow X$ is (Θ', D', L', A') -Morse.

4 Group-theoretic applications

As a consequence of the local-to-global criterion for Morse maps, in this section we establish that the Morse property for isometric group actions is an open condition. Furthermore, for two nearby Morse actions, the actions on their τ_{mod} -limit sets are also close, ie conjugate by an equivariant homeomorphism close to identity. In view of the equivalence of Morse property with the asymptotic properties discussed earlier, this implies structural stability for asymptotically embedded groups. Another corollary of the local-to-global result is the algorithmic recognizability of Morse actions.

We conclude the section by illustrating our technique by constructing Morse–Schottky actions of free groups on higher-rank symmetric spaces.

4.1 Stability of Morse actions

We consider isometric actions $\Gamma \curvearrowright X$ of finitely generated groups.

Definition 4.1 (Morse action) We call an action $\Gamma \curvearrowright X$ Θ -Morse if one (any) orbit map $\Gamma \rightarrow \Gamma x \subset X$ is a Θ -Morse embedding with respect to a(ny) word metric on Γ . We call an action $\Gamma \curvearrowright X$ τ_{mod} -Morse if it is Θ -Morse for some τ_{mod} -Weyl convex compact subset $\Theta \subset \text{ost}(\tau_{\text{mod}})$.

Remark 4.2 (Morse actions are τ_{mod} -regular and undistorted) (i) It follows immediately from the definition of Morse quasigeodesics that Θ -Morse actions are τ_{mod} -regular for the simplex type τ_{mod} determined by Θ .

(ii) Morse subgroups of G are *undistorted* in the sense that the orbit maps are quasiisometric embeddings. In [Kapovich and Leeb 2018b] we prove that Morse subgroups of G satisfy a stronger property: they are *coarse Lipschitz retracts* of G . This retraction property is stronger than nondistortion: every finitely generated subgroup which is a coarse retract of G is undistorted in G , but there are examples of undistorted subgroups which are not coarse retracts. For instance, the group $\Phi := F_2 \times F_2$ admits an undistorted embedding in the isometry group of $X = \mathbb{H}^2 \times \mathbb{H}^2$. On the other hand, pick an epimorphism $\phi: F_2 \rightarrow \mathbb{Z}$ and define the subgroup $\Gamma < \Phi$ as the kernel of the homomorphism

$$(\gamma_1, \gamma_2) \mapsto \phi(\gamma_1) - \phi(\gamma_2).$$

Then Γ is a finitely generated undistorted subgroup of Φ (see eg [Ol'shanskii and Sapir 2001, Theorem 2]), but is not finitely presented (see eg [Baumslag and Roseblade 1984]). Hence, $\Gamma < G = \text{Isom}(\mathbb{H}^2) \times \text{Isom}(\mathbb{H}^2)$ is undistorted but is not a coarse Lipschitz retract.

We denote by $\text{Hom}_{\tau_{\text{mod}}}(\Gamma, G) \subset \text{Hom}(\Gamma, G)$ the subset of τ_{mod} -Morse actions $\Gamma \curvearrowright X$.

By analogy with *local Morse quasigeodesics*, we define *local Morse group actions* $\rho: \Gamma \curvearrowright X$ of a hyperbolic group (with a fixed finite generating set):

Definition 4.3 An action ρ is called S -locally (Θ, D, L, A) -locally Morse, or (Θ, D, L, A) -locally Morse on the scale S , with respect to a basepoint $x \in X$, if the orbit map $\Gamma \rightarrow \Gamma \cdot x \subset X$ induces an S -local (Θ, D, L, A) -local Morse embedding of the Cayley graph of Γ .

According to our local-to-global result for Morse embeddings (see Theorem 3.37), an action of a word hyperbolic group is Morse if and only if it is local Morse on a sufficiently large scale. Since this is a finite condition, it follows that the Morse property is stable under perturbation of the action:

Theorem 4.4 (Morse is open for word hyperbolic groups) *For any word hyperbolic group Γ the subset $\text{Hom}_{\tau_{\text{mod}}}(\Gamma, G)$ is open in $\text{Hom}(\Gamma, G)$. More precisely, if $\rho \in \text{Hom}_{\tau_{\text{mod}}}(\Gamma, G)$ is M -Morse with respect to a basepoint $x \in X$ then there exists a neighborhood of ρ in $\text{Hom}(\Gamma, G)$ consisting entirely of M' -Morse representations with respect to x , where M' depends only on M .*

Proof Let $\rho: \Gamma \curvearrowright X$ be a Morse action. We fix a word metric on Γ and a basepoint $x \in X$. Then there exist data $M = (L, A, \Theta, D)$ such that the orbit map $\Gamma \rightarrow \Gamma x \subset X$ extends to a (Θ, D, L, A) -Morse map of the Cayley graph Y on Γ .

We relax the Morse parameters slightly, namely we consider (L, A, Θ, D) -Morse quasigeodesics as $(L, A + 1, \Theta, D + 1)$ -Morse quasigeodesics satisfying strict inequalities. For every scale S , the orbit map $\Gamma \rightarrow \Gamma x \subset X$ defines an $(L, A + 1, \Theta, D + 1, S)$ -local Morse embedding $Y \rightarrow X$. Due to Γ -equivariance, this is a finite condition in the sense that it is equivalent to a condition involving only finitely many orbit points. Since we relaxed the Morse parameters, the same condition is satisfied by all actions sufficiently close to ρ .

Theorem 3.37 provides a scale S such that all S -local $(\Theta, D + 1, L, A + 1)$ -Morse embeddings $Y \rightarrow X$ are M' -Morse for some Morse datum M' depending only on $(L, A + 1, \Theta, D + 1, S)$. It follows that actions sufficiently close to ρ are τ_{mod} -Morse. □

Lemma 4.5 *Let Φ be a finite group and G a semisimple Lie group with finitely many connected components and finite center. Then $\text{Hom}(\Phi, G)/G$ is finite, where $g \in G$ act on representations $\rho: \Phi \rightarrow G$ by postcompositions with the inner automorphism Inn_g of G defined by g ,*

$$\rho \mapsto \text{Inn}_g \circ \rho.$$

Proof As before, we fix a maximal compact subgroup $K < G$. Since every homomorphism $\rho: \Phi \rightarrow G$ has compact image, its image is conjugate to a subgroup of K . Thus, it suffices to prove finiteness of $\text{Hom}(\Phi, K)/K$. The group K has the structure of a real-algebraic group; hence, $\text{Hom}(\Phi, K)$ also has a natural structure of a compact (with respect to the Lie group topology of K) algebraic subset of $K \times K \times \dots \times K$ (n times, where n is the number of generators of Φ). For $\rho \in \text{Hom}(\Phi, K)$ the Zariski tangent space $T_\rho(\text{Hom}(\Phi, K)/K)$ is isomorphic to the group of 1-cocycles

$$Z^1(\Phi, \mathfrak{k}_{\text{Ad} \circ \rho}),$$

where we regard the Lie algebra \mathfrak{k} as a $\mathbb{R}\Phi$ -module via the composition of ρ and the adjoint representation of K . For every finite group Φ and an $\mathbb{R}\Phi$ -module M , we have $H^i(\Phi, M) = 0$ for all $i > 0$; see eg [Brown 1982, Corollary 10.2]. Applying this to $M = \mathfrak{k}_{\text{Ad} \circ \rho}$, we see that every point in $\text{Hom}(\Phi, K)/K$ is isolated; see eg [Raghunathan 1972, Theorem 6.7]. Since $\text{Hom}(\Phi, K)/K$ is compact, it follows that it is finite. □

Corollary 4.6 *For every hyperbolic group Γ the space of **faithful** Morse representations*

$$\text{Hom}_{\text{inj}, \tau_{\text{mod}}}(\Gamma, G)$$

is open in $\text{Hom}_{\tau_{\text{mod}}}(\Gamma, G)$.

Proof Every hyperbolic group Γ has the unique maximal finite normal subgroup $\Phi \triangleleft \Gamma$ (if Γ is not elementary then Φ is the kernel of the action of Γ on $\partial_\infty \Gamma$). Since Morse actions are properly discontinuous, the kernel of every Morse representation $\Gamma \rightarrow G$ is contained in Φ . Since $\text{Hom}(\Phi, G)/G$ is finite (see Lemma 4.5), it follows that the set of faithful Morse representations is open in $\text{Hom}_{\tau_{\text{mod}}}(\Gamma, G)$. □

The result on the openness of the Morse condition for actions of word hyperbolic groups (see Theorem 4.4) can be strengthened in the sense that the asymptotics of Morse actions vary continuously:

Theorem 4.7 (Morse actions are structurally stable) *The boundary map at infinity of a Morse action depends continuously on the action.*

Proof According to Theorem 4.4 nearby actions are uniformly Morse. The assertion therefore follows from the fact that the ends of Morse quasirays vary uniformly continuously; cf Lemma 3.13. \square

Remark 4.8 (i) Since the boundary maps at infinity are embeddings, the Γ -actions on the τ_{mod} -limit sets are topologically conjugate to each other and, for nearby actions, by a homeomorphism close to the identity.

(ii) In rank one, our argument yields a different proof for Sullivan's structural stability theorem [1985] for convex cocompact group actions on rank-one symmetric spaces. Other proofs can be found in [Labourie 2006; Guichard and Wienhard 2012] (for Anosov subgroups in higher rank), [Corlette 1990; Izeki 2000; Bowditch 1998] for rank-one symmetric spaces.

Our next goal is to extend the topological conjugation from the limit set to the domains of proper discontinuity. Recall that in [Kapovich et al. 2018a] we constructed domains of proper discontinuity and cocompactness for τ_{mod} -Morse group actions on flag-manifolds $\text{Flag}(\nu_{\text{mod}}) = G/P_{\nu_{\text{mod}}}$. Such domains depend on a certain auxiliary datum, a *balanced thickening* $\text{Th} \subset W$, which is a $W_{\tau_{\text{mod}}}$ -left invariant subset satisfying certain conditions; see [Kapovich et al. 2018a, Section 3.4]. Let $\nu_{\text{mod}} \subset \sigma_{\text{mod}}$ be an ι -invariant face such that Th is invariant under the action of $W_{\nu_{\text{mod}}}$ via the *right* multiplication (this is automatic if $\nu_{\text{mod}} = \sigma_{\text{mod}}$ since $W_{\sigma_{\text{mod}}} = \{e\}$). The thickening $\text{Th} \subset W$ defines a thickening $\text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma)) \subset \text{Flag}(\nu_{\text{mod}})$. One of the main results of [Kapovich et al. 2018a] (Theorem 1.7) is that each τ_{mod} -Morse subgroup $\Gamma < G$ acts properly discontinuously and cocompactly on

$$\Omega_{\text{Th}}(\Gamma) := \text{Flag}(\nu_{\text{mod}}) - \text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma)).$$

Theorem 4.9 (stability of Morse quotient spaces) *Suppose that $\rho_n: \Gamma \rightarrow \rho_n(\Gamma) = \Gamma_n < G$ is a sequence of faithful τ_{mod} -Morse representations converging to a τ_{mod} -Morse embedding $\rho: \Gamma \hookrightarrow G$. Then:*

- (1) *The sequence of thickenings $\text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma_n))$ Hausdorff-converges to $\text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma))$.*
- (2) *If $\gamma_n \in \Gamma$ is a divergent sequence, then, after extraction, the sequence $(\rho_n(\gamma_n))$ flag-converges to a point in $\Lambda_{\tau_{\text{mod}}}(\Gamma)$.*
- (3) *For all sufficiently large n , there exists an equivariant diffeomorphism*

$$h_n: \Omega_{\text{Th}}(\Gamma) \rightarrow \Omega_{\text{Th}}(\Gamma_n) \subset \text{Flag}(\tau_{\text{mod}})$$

such that the sequence of maps (h_n) converges to the inclusion map $\Omega_{\text{Th}}(\Gamma) \rightarrow \text{Flag}(\tau_{\text{mod}})$ uniformly on compacts.

- (4) *In particular, the quotient orbifolds $\Omega_{\text{Th}}(\Gamma_n)/\Gamma_n$ are diffeomorphic to $\Omega_{\text{Th}}(\Gamma)/\Gamma$ for all sufficiently large n .*

Proof (1) First of all, suppose that a sequence $\tau_n \in \text{Flag}(\tau_{\text{mod}})$ converges to $\tau \in \text{Flag}(\tau_{\text{mod}})$. Then, since $\text{Flag}(\nu_{\text{mod}}) = G/P_{\nu_{\text{mod}}}$, there is a sequence $g_n \in G$, $g_n \rightarrow e$, such that $g_n(\tau) = \tau_n$. Since

$$g_n(\text{Th}(\tau)) = \text{Th}(g_n\tau) = \text{Th}(\tau_n),$$

it follows that we have Hausdorff-convergence of subsets $\text{Th}(\tau_n) \rightarrow \text{Th}(\tau)$. Moreover, this convergence of subsets is uniform: There exists $n_0 = n(\delta)$ such that if $d(\tau_n, \tau) < \delta$ for all $n \geq n_0$, then $d(\text{Th}(\tau_n), \text{Th}(\tau)) < \epsilon = \epsilon(\delta)$ for all $n \geq n_0$. Here $\epsilon \rightarrow 0$ as $\delta \rightarrow 0$. Since the sequence of limit sets $\Lambda_{\tau_{\text{mod}}}(\Gamma_n)$ Hausdorff-converges to $\Lambda_{\tau_{\text{mod}}}(\Gamma)$, it follows that the sequence of thickenings $\text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma_n))$ Hausdorff-converges to $\text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma))$. This proves (1).

(2) Consider a sequence of geodesic rays $e\xi_n$ in the Cayley graph Y of Γ such that γ_n lies in an R -neighborhood of $e\xi_n$ for all n . Then, in view of the uniform M' -Morse property for the representations ρ_n , each point $\rho_n(\gamma_n)(x)$ belongs to the D' -neighborhood of the Weyl cone $V(x, \text{st}(\tau_n))$, where $\tau_n = \alpha_n(\xi_n)$, and $\alpha_n: \partial_\infty \Gamma \rightarrow \Lambda_{\tau_{\text{mod}}}(\Gamma_n)$ is the asymptotic embedding. Thus, by the definition of flag-convergence, the sequences $(\rho_n(\gamma_n))$ and (τ_n) have the same flag-limit in $\text{Flag}(\tau_{\text{mod}})$. By part (1), the sequence (τ_n) subconverges to a point in $\Lambda_{\tau_{\text{mod}}}(\Gamma)$. Hence, the same holds for $(\rho_n(\gamma_n))$.

(3) The proof of this part is mostly standard; see [Izeki 2000] in the case when X is a hyperbolic space. The quotient orbifold $O = \Omega_{\text{Th}}(\Gamma)/\Gamma$ has a natural (F, G) -structure where $F = \text{Flag}(\nu_{\text{mod}})$. The orbifold O has finitely many components, let Z be one of them and let $\hat{Z} \subset \Omega_{\text{Th}}(\Gamma)$ be a component projecting to Z . It suffices to construct maps h_n on each component \hat{Z} and then extend these maps to maps h_n of $\Omega_{\text{Th}}(\Gamma)$ by ρ_n -equivariance.

The covering map $\hat{Z} \rightarrow Z$ induces an epimorphism $\phi: \pi_1(Z) \rightarrow \Gamma_Z$, where Γ_Z is the Γ -stabilizer of \hat{Z} . Let $\text{dev}: \tilde{Z} \rightarrow \hat{Z} \subset \Omega_{\text{Th}}(\Gamma)$ be the developing map, where $\tilde{Z} \rightarrow Z$ is the universal covering. By the Ehresmann–Thurston holonomy theorem (see [Lok 1984], [Canary et al. 1987], [Goldman 1988] or [Kapovich 2001, Section 7.1]), for all sufficiently large n , the homomorphism $\phi_n := \rho_n \circ \phi$ is the holonomy of an (F, G) -structure on Z . Moreover, the developing maps $\text{dev}_n: \tilde{Z} \rightarrow F$ converge to dev uniformly on compacts in the C^∞ -topology. Since $\pi_1(\hat{Z})$ is contained in the kernel of ϕ , it is also in the kernel of ϕ_n . Hence, the maps dev_n descend to maps $\widehat{\text{dev}}_n: \hat{Z} \rightarrow F$. The sequence $\widehat{\text{dev}}_n$ still converges to the identity embedding $\hat{Z} \hookrightarrow F$ uniformly on compacts. Pick a compact fundamental set $C \subset \hat{Z}$ for the Γ_Z -action, ie a compact subset whose Γ -orbit equals \hat{Z} . In view of part (1) of the theorem, $\widehat{\text{dev}}_n(C) \subset \Omega_{\text{Th}}(\Gamma_n)$ for all sufficiently large n . Therefore, we can assume that $\widehat{\text{dev}}_n(\hat{Z})$ is contained in a component \hat{Z}_n of $\Omega_{\text{Th}}(\Gamma_n)$. By the compactness of the quotient orbifolds, $\widehat{\text{dev}}_n$ projects to a finite-to-one (smooth) orbi-covering map $c_n: Z \rightarrow Z_n := \hat{Z}_n/\rho_n(\Gamma_Z)$. Hence, $\widehat{\text{dev}}_n: \hat{Z} \rightarrow \hat{Z}_n$ is a covering map as well. If \hat{Z}_n were simply connected, it would follow that $\widehat{\text{dev}}_n$ is a diffeomorphism as required (and this is how Izeki concludes his proof in [Izeki 2000]). We will prove that $\widehat{\text{dev}}_n$ is a diffeomorphism by a direct argument.

Suppose that each $\widehat{\text{dev}}_n$ is not injective. Then, by the equivariance of these maps, after extraction, there exist convergent sequences $z_n \rightarrow z$ and $z'_n \rightarrow z'$ in \hat{Z} and a sequence $\gamma_n \in \Gamma$ such that

$$\rho_n(\gamma_n) \widehat{\text{dev}}_n(z_n) = \widehat{\text{dev}}_n(z'_n), \quad \gamma_n(z_n) \neq z'_n.$$

If the sequence (γ_n) were contained in a finite subset of Γ we would obtain a contradiction with the uniform convergence on compacts $\widehat{\text{dev}}_n \rightarrow \text{id}$ on \widehat{Z} . Hence, after extraction, we may assume that (γ_n) is a divergent sequence. We therefore obtain a dynamical relation between the points z, z' via the sequence $(\rho_n(\gamma_n))$. According to part (2), the sequence $(\rho_n(\gamma_n))$ flag-accumulates to $\Lambda_{\tau_{\text{mod}}}(\Gamma)$. The dynamical relation then contradicts fatness of the balanced thickening Th ; see [Kapovich et al. 2018a, Section 5.2] and the proof of Theorem 6.8 in [Kapovich et al. 2018a].

We conclude that the maps

$$\widehat{\text{dev}}_n: \widehat{Z} \rightarrow \widehat{Z}_n$$

are diffeomorphisms for all sufficiently large n . Since $\rho_n: \Gamma \rightarrow \Gamma_n$ are isomorphisms, equivariance of the developing maps implies that the maps $h_n: \Omega_{\text{Th}}(\Gamma) \rightarrow \Omega_{\text{Th}}(\Gamma_n)$ are diffeomorphisms for sufficiently large n .

(4) This part is an immediate corollary of part (3). \square

Remark 4.10 (i) When X is a hyperbolic space, the equivariant diffeomorphism $h_n: \Omega(\Gamma) \rightarrow \Omega(\Gamma_n)$ combined with the equivariant homeomorphism of the limit sets $\Lambda(\Gamma) \rightarrow \Lambda(\Gamma_n)$ yield an equivariant homeomorphism $\partial_\infty X \rightarrow \partial_\infty X$; see [Tukia 1985; Izeki 2000]. Such an extension does not exist in higher rank since, in general, there is no equivariant homeomorphism of thickened limit sets $\text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma)) \rightarrow \text{Th}(\Lambda_{\tau_{\text{mod}}}(\Gamma_n))$. This can be already seen for group actions on products of hyperbolic planes.

(ii) An analogue of Theorem 4.9 holds when we replace the group actions on flag-manifolds with actions on Finsler compactifications of the symmetric space and replace flag-manifold thickenings $\text{Th}(\Lambda_{\tau_{\text{mod}}})$ with Finsler thickenings $\text{Th}_{F\ddot{u}}(\Lambda_{\tau_{\text{mod}}}) \subset \partial_{F\ddot{u}} X$. Proving this requires extending Ehresmann–Thurston holonomy theorem to the category of smooth manifolds with corners and we will not pursue it here.

4.2 Schottky actions

In this section we apply our local-to-global result for straight sequences (Theorem 3.18) to construct Morse actions of free groups, generalizing and sharpening¹ Tits' ping-pong construction.

We consider two oriented τ_{mod} -regular geodesic lines a, b in X . Let $\tau_{\pm a}, \tau_{\pm b} \in \text{Flag}(\tau_{\text{mod}})$ denote the simplices which they are τ -asymptotic to, and let $\theta_{\pm a}, \theta_{\pm b} \in \sigma_{\text{mod}}$ denote the types of their forward/backward ideal endpoints in $\partial_\infty X$. (Note that $\theta_{-a} = \iota(\theta_a)$ and $\theta_{-b} = \iota(\theta_b)$.) Let Θ be a compact convex subset of $\text{ost}(\tau_{\text{mod}}) \subset \sigma_{\text{mod}}$, which is invariant under ι .

Definition 4.11 (generic pair of geodesics) We call the pair of geodesics (a, b) *generic* if the four simplices $\tau_{\pm a}, \tau_{\pm b}$ are pairwise opposite.

¹In the sense that we obtain free subgroups which are not only embedded, but also asymptotically embedded in G .

Let $\alpha, \beta \in G$ be axial isometries with axes a and b respectively and translating in the positive direction along these geodesics. Then $\tau_{\pm a}$ and $\tau_{\pm b}$ are the attractive/repulsive fixed points of α and β on $\text{Flag}(\tau_{\text{mod}})$.

For every pair of numbers $m, n \in \mathbb{N}$ we consider the representation of the free group in two generators

$$\rho_{m,n}: F_2 = \langle A, B \rangle \rightarrow G$$

sending the generator A to α^m and B to β^n . We regard it as an isometric action $\rho_{m,n}: F_2 \curvearrowright X$.

Definition 4.12 (Schottky subgroup) A τ_{mod} -Schottky subgroup of G is a free τ_{mod} -asymptotically embedded subgroup of G .

If G has rank one, this definition amounts to the requirement that Γ is convex cocompact and free. Equivalently, this is a discrete finitely generated subgroup of G which contains no nontrivial elliptic and parabolic elements and has totally disconnected limit set; see [Kapovich 2008]. We note that this definition essentially agrees with the standard definition of Schottky groups in rank-one Lie groups, provided one allows fundamental domains at infinity for such groups to be bounded by pairwise disjoint compact submanifolds which need not be topological spheres; see [Kapovich 2008] for the detailed discussion.

Theorem 4.13 (Morse–Schottky actions) *If the pair of geodesics (a, b) is generic and if $\theta_{\pm a}, \theta_{\pm b} \in \text{int}(\Theta)$, then the action $\rho_{m,n}$ is Θ -Morse for sufficiently large m, n . Thus, such $\rho_{m,n}$ is injective and its image is a τ_{mod} -Schottky subgroup of G .*

Remark 4.14 In particular, these actions are faithful and undistorted; compare Remark 4.2.

Proof Let $S = \{A^{\pm 1}, B^{\pm 1}\}$ be the standard generating set. We consider the sequences (γ_k) in F_2 with the property that $\gamma_k^{-1}\gamma_{k+1} \in S$ and $\gamma_{k+1} \neq \gamma_{k-1}$ for all k . They correspond to the geodesic segments in the Cayley tree of F_2 associated to S which connect vertices.

Let $x \in X$ be a basepoint. In view of Lemma 3.8 we must show that the corresponding sequences $(\gamma_k x)$ in the orbit $F_2 \cdot x$ are uniformly Θ -Morse. (Meaning for instance that the maps $\mathbb{R} \rightarrow X$ sending the intervals $[k, k + 1)$ to the points $\gamma_k x$ are uniform Θ -Morse quasigeodesics.) As in the proof of Theorem 3.34 we will obtain this by applying our local-to-global result for straight spaced sequences (Theorem 3.18) to the associated midpoint sequences. Note that the sequences $(\gamma_k x)$ themselves cannot be expected to be straight.

Taking into account the Γ -action, the uniform straightness of all midpoint sequences depends on the geometry of a finite configuration in the orbit. It is a consequence of the following fact. Consider the midpoints $y_{\pm m}$ of the segments $x\alpha^{\pm m}(x)$ and $z_{\pm n}$ of the segments $x\beta^{\pm n}(x)$.

Lemma 4.15 *For sufficiently large m and n , the quadruple $\{y_{\pm m}, z_{\pm n}\}$ is arbitrarily separated and Θ -regular. Moreover, for any of the four points, the segments connecting it to the other three points have arbitrarily small ζ -angles with the segment connecting it to x .*

Proof The four points are arbitrarily separated from each other and from x because the axes a and b diverge from each other due to our genericity assumption.

By symmetry, it suffices to verify the rest of the assertion for the point y_m , i.e. we show that the segments $y_m y_{-m}$ and $y_m z_n$ are Θ -regular for large m, n and that $\lim_{m \rightarrow \infty} \angle_{y_m}^\xi(x, y_{-m}) = 0$ and $\lim_{n, m \rightarrow \infty} \angle_{y_m}^\xi(x, z_n) = 0$.

The orbit points $\alpha^{\pm m}x$ and the midpoints $y_{\pm m}$ are contained in a tubular neighborhood of the axis a . Therefore, the segments $y_m x$ and $y_m y_{-m}$ are Θ -regular for large m and $\angle_{y_m}(x, y_{-m}) \rightarrow 0$. This implies that also $\angle_{y_m}^\xi(x, y_{-m}) \rightarrow 0$.

To verify the assertion for (y_m, z_n) we use that, due to genericity, the simplices τ_a and τ_b are opposite and we consider the parallel set $P = P(\tau_a, \tau_b)$. Since the geodesics a and b are forward asymptotic to P , it follows that the points x, y_m, z_n have uniformly bounded distance from P . We denote their projections to P by $\bar{x}, \bar{y}_m, \bar{z}_n$.

Let $\Theta'' \subset \text{int}(\Theta)$ be an auxiliary Weyl convex subset such that $\theta_{\pm a}, \theta_{\pm b} \in \text{int}(\Theta'')$. We have that $\bar{y}_m \in V(\bar{x}, \text{st}_{\Theta''}(\tau_a))$ for large m because the points y_m lie in a tubular neighborhood of the ray with initial point \bar{x} and asymptotic to a . Similarly, $\bar{z}_n \in V(\bar{x}, \text{st}_{\Theta''}(\tau_b))$ for large n . It follows that $\bar{x} \in V(\bar{y}_m, \text{st}_{\Theta''}(\tau_b))$ and, using the convexity of Θ -cones (Proposition 2.1), that $\bar{z}_n \in V(\bar{y}_m, \text{st}_{\Theta''}(\tau_b))$.

The cone $V(y_m, \text{st}_{\Theta''}(\tau_b))$ is uniformly Hausdorff close to the cone $V(\bar{y}_m, \text{st}_{\Theta''}(\tau_b))$ because the Hausdorff distance of the cones is bounded by the distance $d(y_m, \bar{y}_m)$ of their tips. Hence there exist points $x', z'_n \in V(y_m, \text{st}_{\Theta''}(\tau_b))$ uniformly close to x, z_n . Since $d(y_m, x'), d(y_m, z'_n) \rightarrow \infty$ as $m, n \rightarrow \infty$, it follows that the segments $y_m x$ and $y_m z_n$ are Θ -regular for large m, n . Furthermore, since $\angle_{y_m}^\xi(x', z'_n) = 0$ and both $\angle_{y_m}(x, x') \rightarrow 0$ and $\angle_{y_m}(z_n, z'_n) \rightarrow 0$, it follows that $\angle_{y_m}^\xi(x, z_n) \rightarrow 0$. \square

To conclude the proof of Theorem 4.13, the lemma implies that for any given l, ϵ the midpoint triples of the four point sequences $(\gamma_k x)$ are (Θ, ϵ) -straight and l -spaced if m, n are sufficiently large; compare the quadruple condition (Definition 3.31). This means that the midpoint sequences of all sequences $(\gamma_k x)$ are (Θ, ϵ) -straight and l -spaced for large m, n . Theorem 3.18 then implies that the sequences $(\gamma_k x)$ are uniformly Θ -Morse. \square

Remark 4.16 (1) Generalizing the above argument to free groups with finitely many generators, one can construct Morse Schottky subgroups for which the set $\theta(\Lambda) \subset \sigma_{\text{mod}}$ of types of limit points is arbitrarily Hausdorff close to a given ι -invariant Weyl convex subset Θ . This provides an alternative approach to the second main theorem in [Benoist 1997] using coarse geometric arguments.

(2) Theorem 4.13 was generalized (by arguments similar to the proof of Theorem 4.13) in [Dey et al. 2019] to free products of Morse subgroups of G .

4.3 Algorithmic recognition of Morse actions

In this section, we describe an algorithm which has an isometric action $\rho: \Gamma \curvearrowright X$ and a point $x \in X$ as its input and terminates if and only if the action ρ is Morse (otherwise, the algorithm runs forever).

We begin by describing briefly *Riley's algorithm* [1983] accomplishing a similar task, namely, detecting geometrically finite actions on $X = \mathbb{H}^3$. Suppose that we are given a finite (symmetric) set of generators $g_1 = 1, \dots, g_m$ of a subgroup $\Gamma \subset \text{PO}(3, 1)$ and a basepoint $x \in X = \mathbb{H}^3$. The idea of the algorithm is to construct a finite sided Dirichlet fundamental domain D for Γ (with center at x): every geometrically finite subgroup of $\text{PO}(3, 1)$ admits such a domain. (The latter is false for geometrically finite subgroups of $\text{PO}(n, 1)$ with $n \geq 4$ but is, nevertheless, true for convex cocompact subgroups.) Given a finite sided convex fundamental domain, one concludes that Γ is geometrically finite. Here is how the algorithm works: for each k define the subset $S_k \subset \Gamma$ represented by words of length $\leq k$ in the letters g_1, \dots, g_m . For each $g \in S_k$ consider the half-space $\text{Bis}(x, g(x)) \subset X$ bounded by the bisector of the segment $xg(x)$ and containing the point x . Then compute the intersection

$$D_k = \bigcap_{g \in S_k} \text{Bis}(x, g(x)).$$

Check if D_k satisfies the conditions of *Poincaré's fundamental domain theorem*. If it does, then $D = D_k$ is a finite sided fundamental domain of Γ . If not, increase k by 1 and repeat the process. Clearly, this process terminates if and only if Γ is geometrically finite.

One can enhance the algorithm in order to detect if a geometrically finite group is convex cocompact. Namely, after a Dirichlet domain D is constructed, one checks for whether:

- (1) The ideal boundary of a Dirichlet domain D has isolated ideal points (they would correspond to rank-two cusps which are not allowed in convex cocompact groups).
- (2) The ideal boundary of D contains tangent circular arcs with points of tangency fixed by parabolic elements (coming from the "ideal vertex cycles"). Such points correspond to rank-one cusps, which again are not allowed in convex cocompact groups.

Checking (1) and (2) is a finite process; after its completion, one concludes that Γ is convex cocompact.

We refer the reader to [Gilman 1995; 1997], [Gilman and Maskit 1991], [Kapovich 2016] and [Kapovich and Leeb 2018a, Section 1.8] for more details concerning discreteness algorithms for groups acting on hyperbolic planes and hyperbolic 3-spaces.

We now consider group actions on general symmetric spaces. Let Γ be a hyperbolic group with a fixed finite (symmetric) generating set; we equip the group Γ with the word metric determined by this generating set.

For each n , let \mathcal{L}_n denote the set of maps $q: [0, 3n] \cap \mathbb{Z} \rightarrow \Gamma$ which are restrictions of geodesics $\tilde{q}: \mathbb{Z} \rightarrow \Gamma$, such that $q(0) = 1 \in \Gamma$. In view of the geodesic automatic structure on Γ (see eg [Epstein et al. 1992, Theorem 3.4.5]), the set \mathcal{L}_n can be described via a finite state automaton.

Suppose that $\rho: \Gamma \curvearrowright X$ is an isometric action on a symmetric space X ; we fix a basepoint $x \in X$ and the corresponding orbit map $f: \Gamma \rightarrow \Gamma x \subset X$. We also fix an ι -invariant face τ_{mod} of the model spherical simplex σ_{mod} of X . The algorithm that we are about to describe will detect that the action ρ is τ_{mod} -Morse.

Remark 4.17 If the face τ_{mod} is not fixed in advance, we would run algorithms for each face τ_{mod} in parallel.

For the algorithm we will be using a special (countable) increasing family of Weyl convex compact subsets $\Theta = \Theta_i \subset \text{ost}(\tau_{\text{mod}}) \subset \sigma_{\text{mod}}$ which exhausts $\text{ost}(\tau_{\text{mod}})$; in particular, every compact ι -invariant convex subset of $\text{ost}(\tau_{\text{mod}}) \subset \sigma_{\text{mod}}$ is contained in some Θ_i :

$$(4.18) \quad \Theta_i := \left\{ v \in \sigma : \min_{\alpha \in \Phi_{\tau_{\text{mod}}}} \alpha(v) \geq \frac{1}{i} \right\},$$

where $\Phi_{\tau_{\text{mod}}}$ is the subset of the set of simple roots Φ (with respect to σ_{mod}) which vanish on the face τ_{mod} . Clearly, the sets Θ_i satisfy the required properties. Furthermore, we consider only those L and D which are natural numbers.

Next, consider the sequence

$$(L_i, \Theta_i, D_i) = (i, \Theta_i, D_i) \quad \text{for } i \in \mathbb{N}.$$

In order to detect τ_{mod} -Morse actions we will use the local characterization of Morse quasigeodesics given by Theorem 3.18 and Proposition 3.32. Due to the discrete nature of quasigeodesics that we will be considering, it suffices to assume that the additive quasiisometry constant A is zero.

Consider the functions

$$l(\Theta, \Theta', \delta), \epsilon(\Theta, \Theta', \delta)$$

as in Theorem 3.18. Using these functions, for the sets $\Theta = \Theta_i, \Theta' = \Theta_{i+1}$ and the constant $\delta = 1$ we define the numbers

$$l_i = l(\Theta, \Theta', \delta), \epsilon_i = \epsilon(\Theta, \Theta', \delta).$$

Next, for the numbers $L = L_i, D = D_i$ and the sets $\Theta = \Theta_i, \Theta' = \Theta_{i+1}$, consider the numbers

$$s_i = s(L_i, 0, \Theta_i, \Theta_{i+1}, D_i, \epsilon_{i+1}, l_{i+1})$$

as in Proposition 3.32. According to this proposition, every $(L_i, 0, \Theta_i, D_i)$ -Morse quasigeodesic satisfies the $(\Theta_{i+1}, \epsilon_{i+1}, l_{i+1}, s)$ -quadruple condition for all $s \geq s_i$. We note that, a priori, the sequence s_i need not be increasing. We set $S_1 = s_1$ and define a monotonic sequence S_i recursively by

$$S_{i+1} = \max(S_i, s_{i+1}).$$

Then every $(\Theta_i, D_i, L_i, 0)$ -Morse quasigeodesic also satisfies the $(\Theta_{i+1}, \epsilon_{i+1}, l_{i+1}, S_{i+1})$ -quadruple condition.

We are now ready to describe the algorithm. For each $i \in \mathbb{N}$ we compute the numbers l_i, ϵ_i and, then, S_i , as above. We then consider finite discrete paths in Γ , $q \in \mathcal{L}_{S_i}$, and the corresponding discrete paths in X , $p(t) = q(t)x$ with $t \in [0, 3S_i] \cap \mathbb{Z}$. The number of paths q (and, hence, p) for each i is finite, bounded by the growth function of the group Γ .

For each discrete path p we check the $(\Theta_i, \epsilon_i, l_i, S_i)$ -quadruple condition. If for some $i = i_*$, all paths p satisfy this condition, the algorithm terminates: it follows from Theorem 3.18 that the map f sends all normalized discrete biinfinite geodesics in Γ to Morse quasigeodesics in X . Hence, the action $\Gamma \curvearrowright X$ is Morse in this case. Conversely, suppose that the action of Γ is $(\Theta, D, L, 0)$ -Morse. Then f sends all isomeric embeddings $\tilde{q}: \mathbb{Z} \rightarrow \Gamma$ to $(\Theta, D, L, 0)$ -Morse quasigeodesics \tilde{p} in X . In view of the properties of the sequence

$$(L_i, \Theta_i, D_i),$$

it follows that for some i ,

$$(L, \Theta, D) \leq (L_i, \Theta_i, D_i),$$

ie $L \leq L_i$, $\Theta \subset \Theta_i$ and $D \leq D_i$; hence, all the biinfinite discrete paths \tilde{p} are $(\Theta_i, D_i, L_i, 0)$ -Morse quasigeodesic. By the definition of the numbers l_i, ϵ_i, S_i , it then follows that all the discrete paths $p = f \circ q, q \in \mathcal{L}_{S_i}$ satisfy the $(\Theta_{i+1}, \epsilon_{i+1}, l_{i+1}, S_{i+1})$ -quadruple condition. Thus, the algorithm will terminate at the step $i + 1$ in this case.

Therefore, the algorithm terminates if and only if the action is Morse (for some parameters). If the action is not Morse, the algorithm will run forever. \square

Remark 4.19 Applied to a rank-one symmetric space X and a hyperbolic group Γ without a nontrivial normal finite subgroup, the above algorithm verifies if the given representation $\rho: \Gamma \rightarrow \text{Isom}(X)$ is faithful with convex-cocompact image. We could not find this result in the existing literature; cf however [Gilman and Keen 2016].

Appendix A Further properties of Morse quasigeodesics

This is the only part of the paper not contained in [Kapovich et al. 2014]. Here we collect various properties of Morse quasigeodesics that we found to be useful elsewhere in our work.

A.1 Finsler geometry of symmetric spaces

In [Kapovich and Leeb 2018b] (see also [Kapovich et al. 2017]), we considered a certain class of G -invariant “polyhedral” Finsler metrics on X . Their geometric and asymptotic properties turned out to be well adapted to the study of geometric and dynamical properties of regular subgroups. They provide a Finsler geodesic *combing* of X which is, in many ways, more suitable for analyzing the asymptotic

geometry of X than the geodesic combing given by the standard Riemannian metric on X . These Finsler metrics also play a basic role in the present paper. We briefly recall their definition and some basic properties, and refer to [Kapovich and Leeb 2018b, Section 5.1] for more details.

Let $\bar{\theta} \in \text{int}(\tau_{\text{mod}})$ be a type spanning the face type τ_{mod} . The $\bar{\theta}$ -Finsler distance $d^{\bar{\theta}}$ on X is the G -invariant pseudometric defined by

$$d^{\bar{\theta}}(x, y) := \max_{\theta(\xi)=\bar{\theta}} (b_{\xi}(x) - b_{\xi}(y))$$

for $x, y \in X$, where the maximum is taken over all ideal points $\xi \in \partial_{\infty} X$ with type $\theta(\xi) = \bar{\theta}$. It is positive, ie a (nonsymmetric) metric, if and only if the radius of σ_{mod} with respect to $\bar{\theta}$ is $< \pi/2$. This is in turn equivalent to $\bar{\theta}$ not being contained in a factor of a nontrivial spherical join decomposition of σ_{mod} , and is always satisfied if for instance X is irreducible.

If $d^{\bar{\theta}}$ is positive, it is equivalent to the Riemannian metric. In general, if it is only a pseudometric, it is still equivalent to the Riemannian metric d on uniformly regular pairs of points. More precisely, if the pair of points x, y is Θ -regular, then

$$L^{-1}d(x, y) \leq d^{\bar{\theta}}(x, y) \leq Ld(x, y)$$

with a constant $L = L(\Theta) \geq 1$.

Regarding symmetry of the Finsler distance, one has the identity

$$d^{i\bar{\theta}}(y, x) = d^{\bar{\theta}}(x, y),$$

and hence $d^{\bar{\theta}}$ is symmetric if and only if $i\bar{\theta} = \bar{\theta}$. We refer to $d^{\bar{\theta}}$ as a Finsler metric of type τ_{mod} .

The $d^{\bar{\theta}}$ -balls in X are convex but not strictly convex. (Their intersections with flats through their centers are polyhedra.) Accordingly, $d^{\bar{\theta}}$ -geodesics connecting two given points x, y are not unique. To simplify notation, xy will stand for *some* $d^{\bar{\theta}}$ -geodesic connecting x and y . The union of all $d^{\bar{\theta}}$ -geodesic xy equals the τ_{mod} -diamond $\diamond_{\tau_{\text{mod}}}(x, y)$, that is, a point lies on a $d^{\bar{\theta}}$ -geodesic xy if and only if it is contained in $\diamond_{\tau_{\text{mod}}}(x, y)$; see [Kapovich et al. 2017]. Finsler geometry thus provides an alternative description of diamonds. Note that with this description, the diamond $\diamond_{\tau_{\text{mod}}}(x, y)$ is also defined when the segment xy is not τ_{mod} -regular. Such a *degenerate* τ_{mod} -diamond is contained in a smaller totally geodesic subspace, namely in the intersection of all τ_{mod} -parallel sets containing the points x, y . The description of geodesics and diamonds also implies that the unparametrized $d^{\bar{\theta}}$ -geodesics depend only on the face type τ_{mod} , and not on $\bar{\theta}$. We will refer to $d^{\bar{\theta}}$ -geodesics as τ_{mod} -Finsler geodesics. Note that Riemannian geodesics are Finsler geodesics.

We will call a Θ -regular τ_{mod} -Finsler geodesic a Θ -Finsler geodesic. If xy is a Θ -regular (Riemannian) segment, then the union of Θ -Finsler geodesics xy equals the Θ -diamond $\diamond_{\Theta}(x, y)$.

Every τ_{mod} -Finsler ray in X is contained in a τ_{mod} -Weyl cone, and we will use the notation $x\tau$ for a τ_{mod} -Finsler ray contained $V(x, \text{st}(\tau))$. Similarly, every τ_{mod} -Finsler line is contained in a τ_{mod} -parallel set, and we denote by $\tau_{-}\tau_{+}$ an oriented τ_{mod} -Finsler line forward/backward asymptotic to two antipodal simplices $\tau_{\pm} \in \text{Flag}(\tau_{\text{mod}})$ and contained in $P(\tau_{-}, \tau_{+})$.

Examples of Θ -regular Finsler geodesics can be obtained as follows. Let (x_i) be a (finite or infinite) sequence contained in a parallel set $P(\tau_-, \tau_+)$ such that each Riemannian segment $x_i x_{i+1}$ is τ_+ -longitudinal (see Definition B.2) and Θ' -regular. Then the concatenation of these geodesic segments is a Θ -regular Finsler geodesic; see Remark B.3.

Conversely, every Θ -regular Finsler geodesic $c: I \rightarrow X$ can be approximated by a piecewise Riemannian Finsler geodesic c' : pick a number $s > 0$ and consider a maximal s -separated subset $J \subset I$. Then take c' to be the concatenation of Riemannian geodesic segments $c(i)c(j)$ for consecutive pairs $i, j \in J$. In view of this approximation procedure, the “string of diamonds” theorem (Theorem 3.30) holds if instead of Riemannian geodesic segments $x_i x_{i+1}$, we allow Θ -regular Finsler segments.

A.2 Stability of diamonds

Diamonds can be regarded as Finsler-geometric replacements of *geodesic segments* in nonpositively curved symmetric spaces of higher rank.

Riemannian geodesic segments in Hadamard manifolds (and, more generally, CAT(0) metric spaces) depend *uniformly continuously* on their tips: by convexity of the distance function we have,

$$d_{\text{Haus}}(xy, x'y') \leq \max(d(x, x'), d(y, y')).$$

In [Kapovich et al. 2018b, Proposition 3.70] we proved that diamonds $\diamond_{\tau_{\text{mod}}}$ depend *continuously* on their tips.

Below we establish uniform control on how much sufficiently large Θ -diamonds vary with their tips.

Lemma A.1 *For $d' > d > 0$ there exists $C = C(\Theta, \Theta', d, d')$ such that the following holds:*

If a segment $x_-x_+ \subset X$ is Θ -regular with length $\geq C$ and $y_{\pm} \in B(x_{\pm}, d)$, then

- (i) *the segment y_-y_+ is Θ' -regular, and*
- (ii) $\diamond_{\Theta}(x_-, x_+) \subset N_{d'}(\diamond_{\Theta'}(y_-, y_+))$.

Proof (i) The Θ' -regularity of y_-y_+ for sufficiently large C follows from the Δ -triangle inequality.

(ii) Suppose that there exists no constant C for which (ii) holds. Then there are sequences of points x_n^{\pm} with $d(x_n^-, x_n^+) \rightarrow +\infty$, y_n^{\pm} with $d(x_n^{\pm}, y_n^{\pm}) \leq d$, $x_n \in \diamond_{\Theta}(x_n^-, x_n^+)$ and $y_n \in \diamond_{\Theta'}(y_n^-, y_n^+)$ with $d(x_n, \diamond_{\Theta'}(y_n^-, y_n^+)) = d(x_n, y_n) = d'$. We may assume convergence $x_n \rightarrow x_{\infty}$ and $y_n \rightarrow y_{\infty}$ in X .

After extraction, at least one of the sequences (x_n^{\pm}) diverges. There are two cases to consider.

Suppose first that both sequences (x_n^{\pm}) diverge. Then they are uniformly τ_{mod} -regular and, after extraction, we have τ_{mod} -flag convergence $x_n^{\pm}, y_n^{\pm} \rightarrow \tau_{\pm} \in \text{Flag}(\tau_{\text{mod}})$. The limit simplices τ_{\pm} are antipodal (because $x_n \rightarrow x_{\infty}$). We observe that

$$d(x_n, \partial \diamond_{\Theta'}(x_n^-, x_n^+)), d(y_n, \partial \diamond_{\Theta'}(y_n^-, y_n^+)) \rightarrow +\infty.$$

It follows that the sequences of diamonds $\diamond_{\Theta'}(x_n^-, x_n^+)$ and $\diamond_{\Theta'}(y_n^-, y_n^+)$ both Hausdorff converge to the τ_{mod} -parallel set $P = P(\tau_-, \tau_+)$. It holds that $x_\infty \in P$ because $x_n \in \diamond_{\Theta'}(x_n^-, x_n^+)$. On the other hand, $d(x_\infty, P) = d'$ because $d(x_n, \diamond_{\Theta'}(y_n^-, y_n^+)) = d'$, a contradiction.

Second, suppose that only one of the sequences (x_n^\pm) diverges, say, after extraction, $x_n^- \rightarrow x_\infty^-$ and $y_n^- \rightarrow y_\infty^-$ in X to limit points with $d(x_\infty^-, y_\infty^-) \leq d$, and $x_n^+ \rightarrow \tau_+ \in \text{Flag}(\tau_{\text{mod}})$. Now the distance of x_n from the boundary of the Θ' -Weyl cone with tip x_n^+ and containing x_n goes to infinity and it follows that $\diamond_{\Theta'}(x_n^-, x_n^+) \rightarrow V(x_\infty^-, \text{st}_{\Theta'}(\tau_+))$ and, similarly, $\diamond_{\Theta'}(y_n^-, y_n^+) \rightarrow V(y_\infty^-, \text{st}_{\Theta'}(\tau_+))$. The asymptotic limit Weyl cones have Hausdorff distance $d(x_\infty^-, y_\infty^-)$. On the other hand, $x_\infty \in V(x_\infty^-, \text{st}_{\Theta'}(\tau_+))$ and $d(x_\infty, V(y_\infty^-, \text{st}_{\Theta'}(\tau_+))) = d'$, again a contradiction.

This shows that (ii) holds for sufficiently large C . □

We reformulate this result in terms of Finsler geodesics:

Lemma A.2 *There exists $C = C(\Theta, \Theta', d, d')$ such that the following holds: if x_-x_+ is a Θ -Finsler geodesic in X with $d(x_-, x_+) \geq C$, and y_\pm are points with $d(y_\pm, x_\pm) \leq d$, then every point x on x_-x_+ lies within distance d' of a point y on a Θ' -Finsler geodesic y_-y_+ .*

We do not claim here that one can take the same Finsler geodesic y_-y_+ for all points x on x_-x_+ .

We now apply this stability result to Morse quasigeodesics. One, somewhat annoying, feature of the definition of Θ -Morse quasigeodesics $p: I \rightarrow X$ is that $p([t_1, t_2])$ is not required to be uniformly close to a Θ -diamond spanned by $p(t_1), p(t_2)$. (One reason is because the segment $p(t_1)p(t_2)$ need not be Θ -regular.) Nevertheless, Lemma A.1 implies:

Lemma A.3 *For every Morse datum $M = (\Theta, B, L, A)$ and $\Theta' > \Theta$, there exists $C = C(M, \Theta')$ and D' such that the following holds for all (Θ, B, L, A) -Morse quasigeodesics $p: I \rightarrow X$:*

Whenever $d(x_1, x_2) \geq C$, the segment $x_1x_2 = p(t_1)p(t_2)$ is Θ' -regular and $p([t_1, t_2])$ lies in the D' -neighborhood of the Θ' -diamond $\diamond_{\Theta'}(x_1, x_2)$.

A.3 Finsler approximation of Morse quasigeodesics

The next theorem establishes that every (sufficiently long) Morse quasigeodesic is uniformly close to a Finsler geodesic with the same endpoints. In this theorem, for convenience of notation, we will be allowing Morse quasigeodesics p to be defined on closed intervals I in the extended real line; this is just a shorthand for a map $p: I' = I \cap \mathbb{R} \rightarrow X$ such that, as $t \rightarrow \pm\infty$ in I' , $p(t) \rightarrow p(\pm\infty) \in \text{Flag}(\tau_{\text{mod}})$. When we say that such maps p, c are within distance D' from each other, this simply means that their restrictions to I' are within distance $\leq D'$.

Theorem A.4 (Finsler approximation theorem) *For every Morse datum $M = (\Theta, D, L, A)$, $\Theta' > \Theta$, and positive number S , there exist $C = C(M, \Theta', S)$ and $D' = D'(M, \Theta', S)$ satisfying the following.*

Let $p: I = [t_-, t_+] \rightarrow X \cup \text{Flag}(\tau_{\text{mod}})$ be an M -Morse quasigeodesic between the points $x_{\pm} = p(t_{\pm})$ in $X \cup \text{Flag}(\tau_{\text{mod}})$ such that $d(x_-, x_+) \geq C$. Then there exists a Θ' -Finsler geodesic x_-x_+ equipped with a monotonic parametrization $c: I \rightarrow x_-x_+$ such that

- (a) *the maps $p, c: I \rightarrow X$ are within distance $\leq D'$ from each other, and*
- (b) *x_-x_+ is an S -spaced piecewise Riemannian geodesic, ie the Riemannian length of each Riemannian segments of x_-x_+ is $\geq S$.*

Proof We will prove this in the case when both x_{\pm} are in X , since the proofs when one or both points x_{\pm} are in $\text{Flag}(\tau_{\text{mod}})$ are similar: one replaces diamonds with Weyl cones or parallel sets.

By the definition of an M -Morse quasigeodesic, for all subintervals $[s_-, s_+] \subset [t_-, t_+]$ there exists a Θ -diamond

$$\diamond_{\Theta}(y'_-, y'_+)$$

whose D -neighborhood contains $p([s_-, s_+])$, and for $y_{\pm} = p(s_{\pm})$ we have

$$d(y_{\pm}, y'_{\pm}) \leq D.$$

Therefore, applying Lemma A.1(i), we conclude that the Riemannian segment y_-y_+ is Θ' -regular provided that

$$(A.5) \quad d(y_-, y_+) \geq C_1 = C_1(M, \Theta').$$

In view of the quasigeodesic property of p , there exists $s = s(M, \Theta')$ such that the separation condition

$$s_+ - s_- \geq s = s(M, \Theta')$$

implies the inequality (A.5). This, of course, also applies to $[s_-, s_+] = [t_-, t_+]$ and hence, using Lemma A.1(ii), we obtain

$$p(I) \subset N_D(\diamond_{\Theta}(x'_-, x'_+)) \subset N_{D+D_1}(\diamond_{\Theta'}(x_-, x_+)),$$

where $D_1 = D_1(M, \Theta')$. We let

$$\bar{y}_{\pm} \in \diamond' := \diamond_{\Theta'}(x_-, x_+) = V(x_-, \text{st}_{\Theta'}(\tau_+)) \cap V(x_+, \text{st}_{\Theta'}(\tau_-))$$

denote the nearest-point projections of $y_{\pm} = p(s_{\pm})$. There exists $s' = s'(M, \Theta')$ such that as long as $s_+ - s_- \geq s'$, the Riemannian segments $\bar{y}_-\bar{y}_+$ are also Θ' -regular and have length $\geq S$. Furthermore, as in the proof of Proposition 3.32, we can assume that $s' = s'(M, \Theta')$ is chosen so that, whenever $s_+ - s_- \geq s'$, each segment $\bar{y}_-\bar{y}_+$ is τ_+ -longitudinal. We fix this $s' = s'(M, \Theta')$ from now on.

We also assume, in what follows, that $t_+ - t_- \geq s'(M, \Theta')$; this is achieved by assuming that

$$L^{-1}(d(x_-, x_+) - A) \geq s' = s'(M, \Theta').$$

Take a maximal s' -separated subset $J \subset I$ containing t_{\pm} . For each $j \in J$ define the point

$$z_j := \overline{p(j)} \in \diamond'.$$

Then for all consecutive $i, j \in J$, with $s' \leq |j - i| \leq 2s'$, we have

(A.6)
$$L^{-1}s' - (A + 2D + 2D_1) \leq d(z_i, z_j) \leq 2Ls' + (A + 2D + 2D_1).$$

We then let c denote the concatenation of Riemannian segments $z_i z_j$ for consecutive $i, j \in J$, where we use the affine parametrization of $[i, j] \rightarrow z_i z_j$. Thus, c is a Θ' -Finsler geodesic. We now take the smallest $s'' \geq s'(M, \Theta')$ satisfying

$$S \leq L^{-1}s'' - (A + 2D + 2D_1).$$

The inequalities (A.6) imply that c satisfies both requirements of the approximation theorem with

$$D' = 2Ls'' + (A + 2D + 2D_1) + (D + D_1) + (2Ls'' + A). \quad \square$$

Remark A.7 In the case when the domain of p is unbounded, one can prove a slightly sharper result; namely, one can take $\Theta' = \Theta$. Compare [Kapovich and Leeb 2023, Section 6].

A.4 Altering Morse quasigeodesics

Below we consider certain modifications of M -Morse quasigeodesics p in X represented as concatenations $p = p_- \star p_0 \star p_+$, where x_{\pm} are the endpoints of p_0 , and y_{\pm}, x_{\pm} are the endpoints of p_{\pm} . (As in the previous section, we will be allowing y_{\pm} to be in $X \cup \text{Flag}(\tau_{\text{mod}})$.) These modifications will have the form $p' = p'_- \star p'_0 \star p'_+$, where p'_{\pm} and p'_0 are all Morse. We will see that, under certain assumptions, the entire p' is again Morse (for suitable Morse datum M').

We begin by analyzing extensions of p to biinfinite paths.

Lemma A.8 (extension lemma) *Suppose that $p = p_- \star p_0 \star p_+$ is an M -Morse quasigeodesic as above. Assume, furthermore, that*

$$p_{\pm} \subset V_{\pm} = V(x_{\pm}, \text{st}(\tau_{\pm})).$$

Whenever y_{\pm} is in X , we let c_{\pm} be Θ -regular Finsler rays contained in V_{\pm} and connecting y_{\pm} to τ_{\pm} . Then, for every $\Theta' > \Theta$, there exists a Morse datum M' containing Θ' (and depending only on M and Θ') such that the concatenation

$$\widehat{p} = c_- \star p \star c_+$$

is M' -Morse, provided that $d(x_{\pm}, y_{\pm}) \geq C = C(M, \Theta')$.

Proof We fix an auxiliary subset Θ_1 satisfying $\Theta < \Theta_1 < \Theta'$. We let $S = S(\Theta_1, \Theta', 1), \epsilon = \epsilon(\Theta_1, \Theta', 1)$ be constants as in the string of diamonds theorem (Theorem 3.30).

According to Theorem A.4, there exists a Θ' -regular Finsler geodesic

$$\bar{c} = y_- \bar{x}_- \star \bar{x}_- \bar{x}_+ \star \bar{x}_+ y_+$$

within distance $D_1 = D_1(M, \Theta', S)$ from the path p , such that \bar{c} is the concatenation of segments of length $\geq S$ and $d(x_{\pm}, \bar{x}_{\pm}) \leq D_1$. We let $z_{\pm}y_{\pm}$ denote the subsegments of $\bar{x}_{\pm}y_{\pm}$ containing y_{\pm} .

Since $d(x_{\pm}, \bar{x}_{\pm}) \leq D_1$, for each $\epsilon > 0$ and a sufficiently large $C_1 = C_1(D_1, \Theta')$, the inequality $d(x_{\pm}, y_{\pm}) \geq C_1$ implies

$$\angle_{y_{\pm}}^{\xi}(x_{\pm}, \bar{x}_{\pm}) \leq \epsilon.$$

Therefore,

$$\angle_{y_{\pm}}^{\xi}(z_{\pm}, \tau_{\pm}) \geq \pi - 2\epsilon$$

and, hence, the piecewise geodesic path

$$\hat{c} = c_- \star \bar{c} \star c_+$$

is $(\Theta_1, 2\epsilon)$ -straight and S -spaced. By Theorem 3.30, the concatenation \hat{c} is M_1 -Morse, where $M_1 = (\Theta', 1, L, A)$. Since the path \hat{p} is within distance D_1 from \hat{c} , it is M' -Morse, where $M' = M_1 + D_1$. \square

The next lemma was proven in [Dey et al. 2019, Thm. 4.11] in the case when p, p' are finite paths. The proof in the case of (bi)infinite paths is the same and we omit it.

Lemma A.9 (replacement lemma) *Suppose that $p' = p'_- \star p'_0 \star p'_+$ is an M -Morse quasigeodesic in X , which is a concatenation of M -Morse quasigeodesics p'_-, p'_0, p'_+ . Let p_0 be an M -Morse quasigeodesic connecting the initial and the terminal point of p'_0 . Then for every $\Theta' > \Theta$ there exists a Morse datum M' containing Θ' (again, depending only on M and Θ') such that the path*

$$p' = p'_- \star p_0 \star p'_+$$

is M' -Morse.

In the following three lemmata we will modify an M -Morse quasigeodesic $p = p_- \star p_0 \star p_+$ by altering p_{\pm} and keeping p_0 unchanged (“wiggling the head and the tail of p ”).

Lemma A.10 (wiggle lemma, I) *Suppose p is as above and the paths p_{\pm} are both infinite and p_0 connects the points x_+, x_- . We let p'_{\pm} be M -Morse quasigeodesic rays with finite terminal points x_{\pm} and set $p' := p'_- \star p_0 \star p'_+$. Then, given $\Theta' > \Theta$ there exists $\epsilon = \epsilon(M, \Theta') > 0$ and a Morse datum M' (depending only on M and Θ') containing Θ' such that if*

$$\mu := \max(\angle_{x_{\pm}}^{\xi}(p'_{\pm}(\pm\infty), p_{\pm}(\pm\infty))) < \epsilon,$$

then p' is M' -Morse.

Proof We fix an auxiliary compact Weyl-convex subset $\Theta_1 \subset \text{ost}(\tau_{\text{mod}})$ such that $\Theta < \Theta_1 < \Theta'$. Set $\tau_{\pm} = p_{\pm}(\pm\infty), \tau'_{\pm} = p'_{\pm}(\pm\infty)$.

According to Lemma A.9, there exists a Morse datum M_1 containing Θ_1 such that for any Θ_1 -regular Finsler geodesic rays $c_{\pm} := x_{\pm}\tau_{\pm}$, the concatenation $c_- \star p_0 \star c_+$ is M_1 -Morse.

Let $M_2 > M_1 + 1$ be a Morse datum containing Θ' and let $S > 0$ be such that if a path q in X is S -locally $M_1 + 1$ -Morse then q is M_2 -Morse; see Theorem 3.34. Let ϵ be such that for $x \in X$ and $\tau, \tau' \in \text{Flag}(\tau_{\text{mod}})$, if $\angle_x^\xi(\tau, \tau') < \epsilon$ then each Θ_1 -regular Finsler segment of length $\leq S$ in $V(x, \text{st}(\tau'))$ is within unit distance from a Θ_1 -regular Finsler segment of length $\leq S$ in $V(x, \text{st}(\tau))$. We assume now that $\mu < \epsilon$.

Since p'_\pm are M -Morse rays, they are within distance $D_1 = D_1(M, \Theta_1)$ from Θ_1 -regular Finsler rays $c'_\pm = x_\pm \tau'_\pm$ connecting x_\pm and τ'_\pm . Define a new path $c' := c'_- \star p_0 \star c'_+$.

By our choice of ϵ , the Θ_1 -regular Finsler subsegment $s'_\pm = x_\pm y'_\pm$ of c'_\pm of length S is within unit distance from a Θ_1 -regular Finsler subsegment $s_\pm = x_\pm y_\pm$ of c_\pm of length S , where $c_\pm = x_\pm \tau_\pm$ is a Θ_1 -Finsler geodesic connecting x_\pm to τ_\pm .

The concatenation

$$s_- \star p_0 \star s_+$$

is M_1 -Morse and, since c'_\pm are Θ_1 -Finsler geodesic, the path c' is S -locally $M_1 + 1$ -Morse. By our choice of S , the path c' is M_2 -Morse. Since c' is within distance D_1 from p' , the path p' is $M_2 + D_1$ -Morse. Lastly, we set $M' := M_2 + D_1$. □

We generalize this lemma by allowing finite Morse quasigeodesics. We continue with the setting and notation of Lemma A.10; as before, M' in the next two lemmata will depend only on M and Θ' . In Lemma A.11 we now allow paths p_\pm and p'_\pm to be finite, connecting y_\pm, x_\pm and y'_\pm, x_\pm , respectively. (Some of y_\pm, y'_\pm might be in $\text{Flag}(\tau_{\text{mod}})$.) However, we will now assume that the distances $d(x_\pm, y_\pm), d(x'_\pm, y_\pm)$ are sufficiently large, namely $\geq C$.

Lemma A.11 (wiggle lemma, II) *Given $\Theta' > \Theta$ there exist $C \geq 0, \epsilon > 0$ and a Morse datum M' containing Θ' such that if*

$$\mu := \max(\angle_{x_\pm}^\xi(y'_\pm, y_\pm)) < \epsilon \quad \text{and} \quad \nu := \min(d(x_\pm, y_\pm), d(x_\pm, y'_\pm)) \geq C,$$

then p' is M' -Morse.

Proof Pick an auxiliary compact Weyl-convex subset Θ_2 , with $\Theta < \Theta_2 < \Theta'$.

We define biinfinite geodesic extensions \hat{p}, \hat{p}' as in Lemma A.8, by extending (if necessary) the paths p_\pm, p'_\pm via Θ -Finsler geodesics $y_\pm \tau_\pm$ and $y'_\pm \tau'_\pm$. According to Lemma A.8, there exists $C > 0$ and a Morse datum M_2 (containing Θ_2), both depending on M and Θ_2 , such that the path \hat{p} is M_2 -Morse. The same lemma applied to the paths \hat{p}'_\pm implies that they are also M_2 -Morse.

By the construction,

$$\mu := \angle_{x_\pm}^\xi(y'_\pm, y_\pm) = \angle_{x_\pm}^\xi(\tau'_\pm, \tau_\pm).$$

Now, the claim follows from Lemma A.10. □

Lastly, we prove a general wiggle lemma where we allow perturbing of the entire path p . We consider concatenations

$$p = p_- \star p_0 \star p_+ \quad \text{and} \quad p' = p'_- \star p'_0 \star p'_+$$

of M -Morse quasigeodesics, where we assume that p_0 and p'_0 are within distance D_0 from each other. The paths p_{\pm} connect y_{\pm}, x_{\pm} , and the paths p'_{\pm} connect y'_{\pm}, x'_{\pm} .

In the next lemma, we continue with the setting and notation of Lemma A.10.

Lemma A.12 (wiggle lemma, III) *Given $\Theta' > \Theta$ there exist $C \geq 0, \epsilon > 0$ and a Morse datum M' containing Θ' such that if*

$$\mu := \max(\angle_{x_{\pm}}^{\xi}(y'_{\pm}, y_{\pm})) < \epsilon \quad \text{and} \quad \nu := \min(d(x_{\pm}, y_{\pm}), d(x'_{\pm}, y'_{\pm})) \geq C,$$

then p' is M' -Morse.

Proof As before, we fix an auxiliary compact Weyl-convex subset $\Theta_3, \Theta < \Theta_3 < \Theta'$. Then p'_{\pm} are within distance $D_3 = D_3(M, \Theta_3)$ from Θ_3 -regular Finsler geodesics $c_{\pm} := y'_{\pm}x_{\pm}$. We apply Lemma A.11 to the pair of paths

$$p, p'' := c_- \star p_0 \star c_+.$$

It follows that p'' is M_3 -Morse for some Morse datum M_3 containing Θ' provided that

$$\mu \leq \epsilon = \epsilon(M, \Theta_3, \Theta') \quad \text{and} \quad \nu \geq C = C(M, \Theta_3, \Theta').$$

Since the paths p'' and p' are within distance $D' := \max(D_0, D_3)$ from each other, the path p' is $M' := (M_3 + D')$ -Morse. □

Appendix B Geometry of nonpositively curved symmetric spaces and their ideal boundaries

In this appendix we collect various definitions and facts about symmetric spaces of noncompact type, their ideal boundaries and their isometry groups. The material of this section is taken from [Kapovich et al. 2017; 2018b] and [Kapovich and Leeb 2018b].

Let X be a symmetric space of noncompact type. Throughout the paper we will be using the *visual* compactification of X , namely $\bar{X} = X \cup \partial_{\infty} X$ (see [Ballmann et al. 1985]), where $\partial_{\infty} X$ is defined as the set of *asymptotic* equivalence classes of geodesic rays in X : two rays are equivalent if they are at finite Hausdorff distance from each other. We refer to [Ballmann et al. 1985] for the detailed definition of the topology on \bar{X} . We will also occasionally use the notion of *strongly asymptotic* geodesic rays: these are rays $\rho_i: \mathbb{R}_+ \rightarrow X$ for $i = 1, 2$ such that

$$\lim_{t \rightarrow \infty} d(\rho_1(t), \rho_2(t)) = 0.$$

The space of classes of strongly asymptotic rays is a topological space, obtained by taking the quotient space of the space of all geodesic rays in X . The latter is topologized by the topology of uniform convergence on compacts. Given a point $\xi \in \partial_\infty X$, one defines X_ξ , the *space of strong asymptote classes at ξ* as follows. Consider the set $\text{Ray}(\xi)$ consisting of all geodesic rays in X asymptotic to ξ (and, as before, equipped with the topology of uniform convergence on compacts). Then take the quotient of $\text{Ray}(\xi)$ by the equivalence relation where two rays are equivalent if they are strongly asymptotic. One can identify the resulting space X_ξ as follows. Pick a point $\hat{\xi} \in \partial_\infty X$ opposite to ξ and consider the parallel set $P(\xi, \hat{\xi})$, which is the union of all geodesic rays in X which are forward/backward asymptotic to $\xi, \hat{\xi}$. Each point $x \in P(\xi, \hat{\xi})$ defines the ray $x\xi$ asymptotic to ξ . The projection of the subset of such rays to X_ξ is a homeomorphism. Thus, we can identify X_ξ with the parallel set $P(\xi, \hat{\xi})$. One metrizes X_ξ by

$$d(\rho_1, \rho_2) = \lim_{t \rightarrow \infty} d(\rho_1(t), \rho_2(t)).$$

Then the projection $P(\xi, \hat{\xi}) \rightarrow X_\xi$ is an isometry. We refer to [Kapovich et al. 2017, Section 2.8] for details. Given a subset $A \subset X$ we let $\partial_\infty A$ denote the accumulation set of A in $\partial_\infty X$.

Building notions The visual boundary $\partial_\infty X$ of a symmetric space X admits a structure as a thick spherical building (the Tits building of X), which is a certain *spherical* simplicial complex; see [Eberlein 1996; Ballmann et al. 1985]. This complex is either connected (if X has rank ≥ 2) or discrete (if X has rank one, or equivalently, X is negatively curved). In the connected case, this building is equipped with the path-metric \angle_{Tits} induced by the spherical metrics on simplices. This metric space has diameter 1. In the case when the building is discrete, the distance between distinct points is π .

The connected component $\text{Isom}_0(X)$ of the isometry group of X acts isometrically on this spherical building, and transitively on facets (top-dimensional simplices). The quotient $\partial_\infty X / \text{Isom}_0(X)$ is a single spherical simplex, denoted by σ_{mod} , the *model spherical chamber* of the building $\partial_\infty X$. One can identify σ_{mod} with a facet of the building $\partial_\infty X$, a *fundamental domain* for the action $\text{Isom}_0(X)$ on the building. We will use the notation $\theta: \partial_\infty X \rightarrow \sigma_{\text{mod}}$ for the *type projection*, the quotient map $\partial_\infty X \rightarrow \partial_\infty X / \text{Isom}_0(X)$. The full isometry group $\text{Isom}(X)$ acts on both $\partial_\infty X$ and σ_{mod} and the map θ is equivariant with respect to these actions. Note that the action on σ_{mod} , in general, is nontrivial. Throughout the paper, G denotes a Lie group with finitely many components, which acts isometrically on X with finite kernel, so that the action of G on σ_{mod} is trivial and the image of G in $\text{Isom}(X)$ has finite index. In particular, we can identify X with G/K , where K is a maximal compact subgroup of G . The reader can think of G as being equal to the kernel of the action of $\text{Isom}(X)$ on σ_{mod} . However, this would exclude such natural examples as $G = \text{SL}(2, \mathbb{R})$, which acts on the associated symmetric space (the hyperbolic plane) with finite nontrivial kernel (equal to the center of G).

For an algebraically inclined reader, the spherical building above is defined as a simplicial complex whose vertices are conjugacy classes of maximal parabolic subgroups of G . If W is the Weyl group of X (equivalently, the relative Weyl group of G) and r is the rank of X (equivalently, the split real rank of G),

then W is a reflection group acting isometrically on the unit sphere a_{mod} of dimension $r - 1$ and σ_{mod} is a fundamental chamber of this action. From the building viewpoint, a_{mod} is the *model spherical apartment* of the Tits building $\partial_\infty X$.

We return to the geometric discussion of $\partial_\infty X$. We let $\iota: \sigma_{\text{mod}} \rightarrow \sigma_{\text{mod}}$ denote the *opposition involution*: it is the projection to σ_{mod} of *Cartan involutions* of X . We will use the notation $\bar{\zeta} = \theta(\zeta)$ for elements of σ_{mod} . Similarly (cf [Kapovich et al. 2017, Section 2.2.2]), we let $\tau_{\text{mod}} \subseteq \sigma_{\text{mod}}$ denote faces of σ_{mod} : these are the *model simplices* in the Tits building; we will use the notation τ for simplices in $\partial_\infty X$ of type τ_{mod} , so that $\theta(\tau) = \tau_{\text{mod}}$. Given a face τ_{mod} of σ_{mod} , we let $\partial_{\tau_{\text{mod}}}\sigma_{\text{mod}}$ denote the union of faces of σ_{mod} disjoint from τ_{mod} ; we use the notation $\text{ost}(\tau_{\text{mod}})$ for the complement $\sigma_{\text{mod}} \setminus \partial_{\tau_{\text{mod}}}\sigma_{\text{mod}}$, the *open star* of τ_{mod} in σ_{mod} .

Let $W_{\tau_{\text{mod}}}$ denote the stabilizer of the face τ_{mod} in W . A (convex) subset $\Theta \subset \sigma_{\text{mod}}$ is said to be *Weyl-convex* (more precisely, Weyl-convex with respect to a face τ_{mod}) if the $W_{\tau_{\text{mod}}}$ -orbit of Θ is a convex subset in the sphere a_{mod} ; see [Kapovich et al. 2017, Definition 2.7]. We will *always* use the notation Θ for compact ι -invariant Weyl-convex subsets of the open star $\text{ost}(\tau_{\text{mod}})$.

Note that, when X has rank one, the data τ_{mod} and Θ are obsolete since σ_{mod} is a singleton. In this case, we also have that $\partial\sigma_{\text{mod}} = \emptyset$ and $\Theta = \text{int}(\sigma_{\text{mod}}) = \sigma_{\text{mod}}$ is clopen.

Two points in $\partial_\infty X$ are called *antipodal* if their distance in the Tits building $\partial_\infty X$ equals π . Equivalently, these points are connected by a geodesic in X . Equivalently, antipodal points are swapped by a Cartan involution of X . Similarly, two simplices in $\partial_\infty X$ are said to be antipodal if they are swapped by a Cartan involution. We will use the notation $\tau, \hat{\tau}$ for pairs of antipodal simplices. Their types in σ_{mod} are swapped by the opposition involution ι . Two simplices τ, τ' in $\partial_\infty X$ are antipodal if and only if $\iota(\theta(\tau')) = \theta(\tau)$ and the open simplices corresponding to τ, τ' contain antipodal points.

Identify σ_{mod} with a simplex in $\partial_\infty X$. Then G -stabilizers of faces τ_{mod} of σ_{mod} are *standard parabolic subgroups* of G , and are denoted by $P_{\tau_{\text{mod}}}$; these are closed subgroups of G . The set of simplices of type τ_{mod} in $\partial_\infty X$ is identified with the quotient $G/P_{\tau_{\text{mod}}}$, which is a smooth compact manifold called the *flag-manifold* $\text{Flag}(\tau_{\text{mod}})$ of type τ_{mod} ; see [Kapovich et al. 2017, Sections 2.2.2 and 2.2.3]. Given $\tau \in \text{Flag}(\tau_{\text{mod}})$, one defines the *open Schubert cell* $C(\tau) \subset \text{Flag}(\tau_{\text{mod}})$, which is an open subset of $\text{Flag}(\tau_{\text{mod}})$ consisting of elements opposite to τ ; see [Kapovich et al. 2017, Section 2.4]. We will use the notation $\text{int} \tau$ for the *open simplex* obtained by removing from τ all its proper faces. The notation $\text{st}(\tau)$ is used to denote the *star* of τ in $\partial_\infty X$, the union of all faces of $\partial_\infty X$ containing τ . Similarly, $\text{ost}(\tau)$, the *open star* of τ in $\partial_\infty X$, is obtained from $\text{st}(\tau)$ by removing all faces disjoint from τ . Accordingly, $\partial_{\tau_{\text{mod}}}\sigma_{\text{mod}}$ is defined as $\sigma_{\text{mod}} \setminus \text{ost}(\tau_{\text{mod}})$.

Given a Weyl-convex compact subset $\Theta \subset \text{ost} \tau_{\text{mod}} \subset \sigma_{\text{mod}}$, we will define the Θ -*star* of a simplex τ in $\partial_\infty X$ of type τ_{mod} as the preimage of Θ under the restriction $\theta: \text{st}(\tau) \subset \partial_\infty X \rightarrow \sigma_{\text{mod}}$.

Symmetric space notions An isometry of X is called a *transvection* if it preserves a geodesic in X and acts trivially on its normal bundle. *Transvections* in X are precisely the compositions of pairs of Cartan involutions.

A *flat* in X is an isometrically embedded totally geodesic Euclidean subspace of X . The dimension of a *maximal flat* in X is the *rank* of X . We will use the notation F for maximal flats in X . The *parallel set* $P(l)$ of a geodesic l in X is the union of all maximal flats in X containing l . We let τ_{\pm} denote the smallest simplices in $\partial_{\infty}X$ containing the elements $\xi_{\pm} \in \partial_{\infty}l$; these simplices are antipodal. We will use the notation $P(\tau_{-}, \tau_{+})$ for the parallel set $P(l)$, since $P(l)$ can be described as the union of geodesics in X that are forward/backward asymptotic to points in the open simplices $\text{int } \tau_{\pm}$.

Given a maximal flat $F \subset X$, the stabilizer G_F of F in G acts transitively on F . For each $x \in F$ the intersection $G_x \cap G_F$ acts on F as a finite reflection group, isomorphic to the Weyl group W of X . One frequently fixes a basepoint $x = o \in X$ and the *model maximal flat* $F_{\text{mod}} \subset X$ containing o ; the stabilizer G_x is then denoted by K , it is a maximal compact subgroup of G . A fundamental domain for the W -action on F_{mod} is the *model Euclidean Weyl chamber*, denoted by Δ . The ideal boundary $\partial_{\infty}\Delta$ is then identified with σ_{mod} , the model chamber in $\partial_{\infty}X$.

The Euclidean Weyl chamber Δ is the Euclidean cone over the simplex σ_{mod} . Thus, we can also “cone-off” various objects from σ_{mod} . In particular, we define the τ_{mod} -*boundary* $\partial_{\tau_{\text{mod}}}\Delta$ of Δ as the union of rays $o\zeta$ with $\zeta \in \partial_{\tau_{\text{mod}}}\sigma_{\text{mod}}$ (see [Kapovich et al. 2017, Section 2.5.2]), and define the Θ -cone Δ_{Θ} as the union of rays $o\zeta$ with $\zeta \in \Theta$, where $\Theta \subset \text{ost } \tau_{\text{mod}} \subset \sigma_{\text{mod}}$.

The group of transvections along geodesics in F is usually denoted by A ; then (in the case $G < \text{Isom}(X)$) $G = KAK$ is the *Cartan decomposition* of G . The more refined form of this decomposition is $G = KA_+K$, where $A_+ \subset A$ is the subsemigroup consisting of transvections mapping Δ into itself. Geometrically speaking, the Cartan decomposition states that each K -orbit $Ky \subset X$ intersects Δ in exactly one point. The projection $c: y \mapsto Ky \cap \Delta$ is 1-Lipschitz since each orbit Ky meets F_{mod} orthogonally and transversally. This projection leads to the notion of Δ -distance on X (see [Kapovich et al. 2017, Section 2.6] and [Kapovich et al. 2009]): Given a pair of points $x, y \in X$, find $g \in G$ such that $g(x) = o$ and $z = g(y) \in \Delta$. Then

$$\vec{o}z := d_{\Delta}(x, y).$$

The vector $d_{\Delta}(x, y)$ is the complete G -congruence invariant of pairs $(x, y) \in X^2$, and

$$d(x, y) = \|d_{\Delta}(x, y)\|,$$

where $\|\cdot\|$ is the Euclidean norm on F_{mod} (regarded as a Euclidean vector space with o serving as zero). Since c is 1-Lipschitz, the Δ -distance satisfies the triangle inequality

$$\|d_{\Delta}(x, y) - d_{\Delta}(x, z)\| \leq \|d_{\Delta}(y, z)\| = d(y, z).$$

We refer the reader to [Kapovich et al. 2009] for in-depth discussion of *generalized triangle inequalities* satisfied by the Δ -valued distance function on X .

We say that a nondegenerate segment $xy \subset X$ is τ_{mod} -regular if

$$d_{\Delta}(x, y) \notin \partial_{\tau_{\text{mod}}} \Delta,$$

and that it is Θ -regular if

$$d_{\Delta}(x, y) \in \Delta_{\Theta}.$$

See [Kapovich et al. 2017, Section 2.5.3]. In what follows, we will always assume that τ_{mod} is ι -invariant and $\Theta \subset \text{ost}(\tau_{\text{mod}})$ is an ι -invariant Weyl-convex compact subset of $\text{ost}(\tau_{\text{mod}}) \subset \sigma_{\text{mod}}$.

An injective map f from an interval I in \mathbb{R} to X is called τ_{mod} -regular if for all $s < t$ in I the geodesic segment $f(s)f(t)$ is τ_{mod} -regular.

Weyl cones Fix a simplex τ in $\partial_{\infty} X$ and a point $x \in X$. We define the *Weyl cone* $V(x, \text{st}(\tau))$ with the tip x over the star $\text{st}(\tau) \subset \partial_{\infty} X$ as the union of geodesic rays emanating from x and asymptotic to points of $\text{st}(\tau)$. Similarly, assuming that τ has the type τ_{mod} and $\Theta \subset \text{ost}(\tau_{\text{mod}}) \subset \sigma_{\text{mod}}$ is a Weyl-convex compact subset, we define the Θ -cone $V(x, \text{st}_{\Theta}(\tau))$ with the tip x over the Θ -star $\text{st}_{\Theta}(\tau) \subset \partial_{\infty} X$ as the union of geodesic rays emanating from x and asymptotic to points of $\text{st}_{\Theta}(\tau)$. It was proven in [Kapovich et al. 2017, Section 2.5] that Weyl cones $V(x, \text{st}(\tau))$ and Θ -cones $V(x, \text{st}_{\Theta}(\tau))$ are convex in X .

In particular, these cones satisfy the *nested cones property*:

- (1) If $y \in V(x, \text{st}(\tau))$, then $V(y, \text{st}(\tau)) \subset V(x, \text{st}(\tau))$.
- (2) If $y \in V(x, \text{st}_{\Theta}(\tau))$, then $V(y, \text{st}_{\Theta}(\tau)) \subset V(x, \text{st}_{\Theta}(\tau))$.

Finsler geodesics in X These are geodesics with respect to certain G -invariant Finsler metrics on X defined in [Kapovich and Leeb 2018b]. The precise description of all Finsler geodesics is given in [Kapovich and Leeb 2018b, Section 5.1.3]. We merely use the following description of Finsler geodesics in lieu of the definition:

Definition B.1 (Finsler geodesics) Fix an ι -invariant face τ_{mod} of σ_{mod} . Let $I \subset \mathbb{R}$ be an interval. A (continuous) path $c: I \rightarrow X$ is called a *Finsler geodesic* (more precisely, a τ_{mod} -Finsler geodesic) if there exists a pair of antipodal flags $\tau_{\pm} \in \text{Flag}(\tau_{\text{mod}})$ such that $c(I) \subset P(\tau_+, \tau_-)$ and

$$c(t_2) \in V(c(t_1), \text{st}(\tau_+)) \quad \text{for all } t_1 < t_2.$$

Moreover, given an ι -invariant compact Weyl-convex subset $\Theta \subset \text{ost}(\tau_{\text{mod}})$, a Finsler geodesic $c: I \rightarrow X$ is called a Θ -Finsler geodesic if, in addition to the above, it satisfies the following stronger condition:

$$c(t_2) \in V(c(t_1), \text{st}_{\Theta}(\tau_+)) \quad \text{for all } t_1 < t_2,$$

ie c is Θ -regular.

Note that each τ_{mod} -Finsler geodesic is automatically a τ_{mod} -regular path in X .

Diamonds Intersecting cones, we define *diamonds* in X ; see [Kapovich et al. 2017, Section 2.5]. Take two antipodal simplices τ_+, τ_- of the type $\tau_{\text{mod}} = \iota\tau_{\text{mod}}$ and points $x_{\pm} \in X$ such that

$$x_+ \in V(x_-, \text{st}(\tau_+)), x_- \in V(x_+, \text{st}(\tau_-)).$$

Then the intersection

$$\diamond_{\tau_{\text{mod}}}(x_-, x_+) = V(x_-, \text{st}(\tau_+)) \cap V(x_+, \text{st}(\tau_-))$$

is the τ_{mod} -diamond with tips x_{\pm} . Similarly, suppose that $\Theta \subset \text{ost}(\tau_{\text{mod}}) \subset \sigma_{\text{mod}}$ is a Weyl-convex ι -invariant compact subset and

$$x_+ \in V(x_-, \text{st}_{\Theta}(\tau_+)), \quad x_- \in V(x_+, \text{st}_{\Theta}(\tau_-)).$$

Then the Θ -diamond is defined as the intersection

$$\diamond_{\Theta}(x_-, x_+) = V(x_-, \text{st}_{\Theta}(\tau_+)) \cap V(x_+, \text{st}_{\Theta}(\tau_-)).$$

Thus, diamonds are also convex in X .

Diamonds have a nice interpretation in terms of Finsler geometry of X : it is proven in [Kapovich and Leeb 2018b] that $\diamond_{\tau_{\text{mod}}}(x_-, x_+)$ is the union of all Finsler geodesics in X connecting the points x_{\pm} . Similarly, $\diamond_{\Theta}(x_-, x_+)$ is the union of all Θ -regular Finsler geodesics in X connecting the points x_{\pm} .

Longitudinal notions The following definition is taken from [Kapovich et al. 2018b, Section 3.2]. Let $P = P(\tau_-, \tau_+) \subset X$ be a parallel set, where τ_{\pm} are antipodal simplices of type τ_{mod} in $\partial_{\text{Tits}} X$.

Definition B.2 (longitudinal directions and segments in parallel sets) A nondegenerate geodesic segment xy in P is said to be *longitudinal* or, more precisely, τ_+ -longitudinal, if it is contained in a geodesic ray $x\xi \subset P$, where $\xi \in \text{ost}(\tau_+)$.

Remark B.3 (1) Longitudinal directions and segments are, in particular, τ_{mod} -regular.

- (2) A nondegenerate segment xy is longitudinal if and only if all nondegenerate subsegments xy are longitudinal.
- (3) If xy and yz are τ_+ -longitudinal Θ -regular segments in P , then so is the segment xz . This follows, for instance, from the fact that

$$V(y, \text{st}_{\Theta}(\tau_+)) \subset V(x, \text{st}_{\Theta}(\tau_+)).$$

Regular sequences and groups The reader should think of *regularity* conditions for subgroups $\Gamma < G$ as a way of strengthening the discreteness assumption: The discreteness condition means that the sequence of distances $d(x, \gamma_n(x))$ diverges to infinity for every sequence of distinct elements $g_n \in \Gamma$. Regularity conditions on Γ require certain forms of divergence to infinity of vector-valued distances $d_{\Delta}(x, g_n x)$.

We first consider sequences in the euclidean model Weyl chamber Δ . Recall that

$$\partial_{\tau_{\text{mod}}}\Delta = V(0, \partial_{\tau_{\text{mod}}}\sigma_{\text{mod}}) \subset \Delta$$

is the union of faces of Δ which do not contain the sector $V(0, \tau_{\text{mod}})$. Note that

$$\partial_{\tau_{\text{mod}}}\Delta \cap V(0, \tau_{\text{mod}}) = \partial V(0, \tau_{\text{mod}}) = V(0, \partial\tau_{\text{mod}}).$$

The following definitions are taken from [Kapovich et al. 2017, Section 4.2].

Definition B.4 (i) A sequence (δ_n) in Δ is called τ_{mod} -regular if it drifts away from $\partial_{\tau_{\text{mod}}}\Delta$, ie

$$d(\delta_n, \partial_{\tau_{\text{mod}}}\Delta) \rightarrow +\infty.$$

- (ii) A sequence (x_n) in X is τ_{mod} -regular if for some (any) basepoint $o \in X$ the sequence of Δ -distances $d_{\Delta}(o, x_n)$ in Δ has this property.
- (iii) A sequence (g_n) in G is τ_{mod} -regular, if for some (any) point $x \in X$ the orbit sequence $(g_n x)$ in X has this property.
- (iv) A subgroup $\Gamma < G$ is τ_{mod} -regular if all sequences of distinct elements in Γ have this property.

Next, we describe a stronger form of regularity following [Kapovich et al. 2017, Section 4.6].

Definition B.5 (i) A sequence $\delta_n \rightarrow \infty$ in Δ is *uniformly* τ_{mod} -regular if it drifts away from $\partial_{\tau_{\text{mod}}}\Delta$ at a linear rate with respect to its norm,

$$\liminf_{n \rightarrow +\infty} \frac{d(\delta_n, \partial_{\tau_{\text{mod}}}\Delta)}{\|\delta_n\|} > 0.$$

- (ii) A sequence (x_n) in X is *uniformly* τ_{mod} -regular if for some (any) basepoint $o \in X$ the sequence of Δ -distances $d_{\Delta}(o, x_n)$ in Δ has this property.
- (iii) A sequence (g_n) in G is *uniformly* τ_{mod} -regular if for some (any) point $x \in X$ the orbit sequence $(g_n x)$ in X has this property.
- (iv) A subgroup $\Gamma < G$ is *uniformly* τ_{mod} -regular if all sequences of distinct elements in Γ have this property.

Note that (uniform) regularity of a sequence in X is independent of the basepoint and stable under bounded perturbation of the sequence (due to the triangle inequality for Δ -distances). A sequence (x_n) is uniformly τ_{mod} -regular if and only if there exists a compact $\Theta \subset \text{ost}(\tau_{\text{mod}})$ such that for each $x \in X$ all but finitely many vectors $d_{\Delta}(x, x_n)$ belong to Δ_{Θ} .

Remark B.6 The definition of regularity of sequences in G has the following dynamical interpretation, in terms of dynamics on the flag-manifold $\text{Flag}(\tau_{\text{mod}})$, which generalizes the familiar convergence property for sequences of isometries of a Gromov-hyperbolic space Y acting on the visual boundary of Y .

A sequence (g_n) in G is said to be τ_{mod} -contracting if there exists a pair of elements $\tau_{\pm} \in \text{Flag}(\tau_{\text{mod}})$ such that the sequence (g_n) converges to τ_+ uniformly on compacts in $C(\tau_-) \subset \text{Flag}(\tau_{\text{mod}})$. Here, as elsewhere, $C(\tau_-)$ is the open Schubert cell in $\text{Flag}(\tau_{\text{mod}})$ consisting of simplices antipodal to τ_- . In this situation, the simplex τ_+ is the τ_{mod} -limit of (g_n) . A sequence (g_n) is τ_{mod} -regular if and only if there exists a pair of bounded sequences $a_n, b_n \in G$ such that the sequence of compositions $c_n g_n b_n$ is τ_{mod} -contracting. Equivalently, a sequence (g_n) is τ_{mod} -regular if and only if every subsequence in (g_n) contains a further subsequence which is τ_{mod} -contracting. We refer the reader to [Kapovich et al. 2017] for details.

For a subgroup $\Gamma < G$, uniform τ_{mod} -regularity is equivalent to the *visual limit set* $\Lambda \subset \partial_{\infty} X$ being contained in the union of the open τ_{mod} -stars. Here Λ is the accumulation set in $\partial_{\infty} X$ of some (any) Γ -orbit in X .

We next discuss τ_{mod} -limit sets. In a nutshell, the τ_{mod} -limit set of a τ_{mod} -regular subgroup $\Gamma < G$ is the accumulation set in the flag-manifold $\text{Flag}(\tau_{\text{mod}})$ of one (equivalently, every) Γ -orbit in X . However, the way the accumulation is defined is more subtle than in the rank-one case. If G (equivalently, X) has rank one, one can define convergence of a divergent sequence $x_n \in X$ to a point $\lambda \in \partial_{\infty} X$ by taking a sequence of geodesic rays $o\xi_n$ through x_n (with the fixed initial point o). Then $x_n \rightarrow \lambda$ if and only if $\xi_n \rightarrow \lambda$ in $\text{Flag}(\tau_{\text{mod}}) = \partial_{\infty} X$. In higher rank, a geodesic $o\xi_n$ through x_n need not even terminate in a face τ_n of $\partial_{\text{Tits}} X$ of type τ_{mod} . But, if it does, then $x_n \rightarrow \tau \in \text{Flag}(\tau_{\text{mod}})$ if and only if $\tau_n \rightarrow \tau$ in $\text{Flag}(\tau_{\text{mod}})$. In general, due to the τ_{mod} -regularity assumption, one finds (for all sufficiently large n) a unique face τ_n of type τ_{mod} in $\text{Flag}(\tau_{\text{mod}})$ such that ξ_n belongs to the open star $\text{ost}(\tau_n)$ of τ_n . Then, by the definition, $x_n \rightarrow \tau$ if and only if $\tau_n \rightarrow \tau$ in $\text{Flag}(\tau_{\text{mod}})$. Lastly, if $\Gamma < G$ is a τ_{mod} -regular subgroup, then the τ_{mod} -limit set $\Lambda_{\tau_{\text{mod}}}(\Gamma) \subset \text{Flag}(\tau_{\text{mod}})$ is the subset consisting of accumulation points in $\text{Flag}(\tau_{\text{mod}})$ of one (equivalently, every) Γ -orbit in X , where convergence is understood as above.

A subset Λ of $\text{Flag}(\tau_{\text{mod}})$ is said to be *antipodal* if any two distinct elements of Λ are antipodal to each other. A map $\beta: Z \rightarrow \text{Flag}(\tau_{\text{mod}})$ is said to be *antipodal* if it sends distinct elements of Z to antipodal elements of $\text{Flag}(\tau_{\text{mod}})$. We will apply these notions to the limit sets of τ_{mod} -regular subgroups of G .

Morse quasigeodesics, maps and subgroups Lastly, we come to the heart of the matter, the notion of *Morse quasigeodesics* in X (Definition 3.1), *Morse embeddings* (or *Morse maps*) to X (Definition 3.7) and *Morse subgroups* of G and *Morse actions* on X (Definition 4.1).

Below we recall an alternative characterization of Morse quasigeodesics given in [Kapovich et al. 2018b]. A natural way to *coarsify* the notion of regularity for geodesic segments in X is as follows.

Let $B \geq 0$. We say that a pair (x, y) of (not necessarily distinct) points is (Θ, B) -regular if it is oriented B -close to some Θ -regular pair of points (x', y') , ie $d(x, x') \leq B$ and $d(y, y') \leq B$. This is equivalent to the property that the segment xy is oriented B -Hausdorff close to the Θ -regular segment $x'y'$, and we say also that the *segment* xy is (Θ, B) -regular.

We say that a (not necessarily continuous) *path* $p: I \rightarrow X$ is (Θ, B) -regular if for every subinterval $[a', b'] \subset I$, the segment $p(a')p(b')$ is (Θ, B) -regular.

If τ_{mod} and Θ are ι -invariant, then we say that a *subset* of X is (Θ, B) -regular if every pair of points in the subset has this property, and more generally, that a *map* into X is (Θ, B) -regular if it sends any pair of points to a (Θ, B) -regular pair of points in X . Note that the images of (Θ, B) -regular maps are (Θ, B) -regular subsets. We say that the subset or map is (*coarsely*) Θ -regular if it is (Θ, B) -regular for some constant B .

In [Kapovich et al. 2018b] we prove:

Theorem B.7 (regular implies Morse for quasigeodesics) *Uniformly coarsely τ_{mod} -regular quasigeodesics in X are uniform τ_{mod} -Morse quasigeodesics and vice-versa.*

Note that τ_{mod} -Morse embeddings are coarsely uniformly τ_{mod} -regular quasiisometric embeddings. We also have the converse, proven in [Kapovich et al. 2018b]:

Theorem B.8 (regular implies Morse for quasiisometric embeddings) *Coarsely uniformly τ_{mod} -regular quasiisometric embeddings from quasigeodesic metric spaces into model spaces are uniform τ_{mod} -Morse embeddings.*

Morse actions on X are *undistorted* in the sense that the orbit maps are quasiisometric embeddings. In particular, they are properly discontinuous. Furthermore, Θ -Morse actions are (coarsely) Θ -regular. In [Kapovich et al. 2018b] we prove a converse:

Theorem B.9 (URU implies Morse) *Uniformly τ_{mod} -regular undistorted isometric actions by finitely generated groups on X are uniformly τ_{mod} -Morse.*

Below we collect some of the equivalent characterizations of τ_{mod} -Morse (equivalently, τ_{mod} -Anosov) subgroups of G given in [Kapovich et al. 2017; 2018b] and [Kapovich and Leeb 2018b]:

Theorem B.10 *The following are equivalent for a subgroup $\Gamma < G$:*

- (1) Γ is τ_{mod} -Morse.
- (2) Γ is τ_{mod} -URU, ie it is τ_{mod} -uniformly regular, finitely generated and one (equivalently, every) orbit map $\Gamma \rightarrow X$ is a quasiisometric embedding.
- (3) Γ is τ_{mod} -asymptotically embedded, ie Γ is τ_{mod} -regular, Gromov-hyperbolic with the Gromov-compactification $\bar{\Gamma} = \Gamma \cup \partial_{\infty} \Gamma$, and there exists an antipodal homeomorphism $\beta: \partial_{\infty} \Gamma \rightarrow \Lambda_{\tau_{\text{mod}}}(\Gamma)$ which, combined with one (equivalently, every) orbit map $\Gamma \rightarrow X$ defines a continuous map $\bar{\Gamma} \rightarrow X \cup \text{Flag}(\tau_{\text{mod}})$, where the latter is topologized using the natural topologies of X and $\text{Flag}(\tau_{\text{mod}})$, and the topology of τ_{mod} -flag convergence for sequences in X to elements of $\text{Flag}(\tau_{\text{mod}})$.
- (4) Γ is τ_{mod} -RCA (regular, conical and antipodal), ie Γ is τ_{mod} -regular, its τ_{mod} -limit set is antipodal and every element τ of $\Lambda_{\tau_{\text{mod}}}(\Gamma)$ is conical. The latter means that for some (equivalently, every) $x \in X$ there exists $C < \infty$ and an infinite sequence γ_n in Γ such that the sequence $(\gamma_n(x))$ flag-converges to τ in the C -neighborhood of the Weyl cone $V(x, \text{st}(\tau))$.

The map β in (3) is called the *boundary map* of Γ . It is unique as long as Γ is a nonelementary hyperbolic group.

References

- [Ballmann et al. 1985] **W Ballmann, M Gromov, V Schroeder**, *Manifolds of nonpositive curvature*, Progr. Math. 61, Birkhäuser, Boston, MA (1985) MR
- [Baumslag and Roseblade 1984] **G Baumslag, J E Roseblade**, *Subgroups of direct products of free groups*, J. London Math. Soc. 30 (1984) 44–52 MR
- [Benoist 1997] **Y Benoist**, *Propriétés asymptotiques des groupes linéaires*, Geom. Funct. Anal. 7 (1997) 1–47 MR
- [Bowditch 1998] **B H Bowditch**, *Spaces of geometrically finite representations*, Ann. Acad. Sci. Fenn. Math. 23 (1998) 389–414 MR
- [Brown 1982] **K S Brown**, *Cohomology of groups*, Graduate Texts in Math. 87, Springer (1982) MR
- [Canary et al. 1987] **R D Canary, D B A Epstein, P Green**, *Notes on notes of Thurston*, from “Analytical and geometric aspects of hyperbolic space” (D B A Epstein, editor), London Math. Soc. Lecture Note Ser. 111, Cambridge Univ. Press (1987) 3–92 MR
- [Coornaert et al. 1990] **M Coornaert, T Delzant, A Papadopoulos**, *Géométrie et théorie des groupes: les groupes hyperboliques de Gromov*, Lecture Notes in Math. 1441, Springer (1990) MR
- [Corlette 1990] **K Corlette**, *Hausdorff dimensions of limit sets, I*, Invent. Math. 102 (1990) 521–541 MR
- [Dey and Kapovich 2023] **S Dey, M Kapovich**, *Klein–Maskit combination theorem for Anosov subgroups: free products*, Math. Z. 305 (2023) art. id. 35 MR
- [Dey and Kapovich 2025] **S Dey, M Kapovich**, *Klein–Maskit combination theorem for Anosov subgroups: amalgams*, J. Reine Angew. Math. 819 (2025) 1–43 MR
- [Dey et al. 2019] **S Dey, M Kapovich, B Leeb**, *A combination theorem for Anosov subgroups*, Math. Z. 293 (2019) 551–578 MR
- [Druţu and Kapovich 2018] **C Druţu, M Kapovich**, *Geometric group theory*, American Mathematical Society Colloquium Publications 63, Amer. Math. Soc., Providence, RI (2018) MR
- [Eberlein 1996] **P B Eberlein**, *Geometry of nonpositively curved manifolds*, Univ. Chicago Press (1996) MR
- [Epstein et al. 1992] **D B A Epstein, J W Cannon, D F Holt, S V F Levy, M S Paterson, W P Thurston**, *Word processing in groups*, Jones and Bartlett, Boston, MA (1992) MR
- [Gilman 1995] **J Gilman**, *Two-generator discrete subgroups of $\mathrm{PSL}(2, \mathbb{R})$* , Mem. Amer. Math. Soc. 561, Amer. Math. Soc., Providence, RI (1995) MR
- [Gilman 1997] **J Gilman**, *Algorithms, complexity and discreteness criteria in $\mathrm{PSL}(2, \mathbb{C})$* , J. Anal. Math. 73 (1997) 91–114 MR
- [Gilman and Keen 2016] **J Gilman, L Keen**, *Canonical hexagons and the $\mathrm{PSL}(2, \mathbb{C})$ discreteness problem*, preprint (2016) arXiv 1508.00257v3
- [Gilman and Maskit 1991] **J Gilman, B Maskit**, *An algorithm for 2-generator Fuchsian groups*, Michigan Math. J. 38 (1991) 13–32 MR
- [Goldman 1988] **W M Goldman**, *Geometric structures on manifolds and varieties of representations*, from “Geometry of group representations” (W M Goldman, A R Magid, editors), Contemp. Math. 74, Amer. Math. Soc., Providence, RI (1988) 169–198 MR
- [Guichard and Wienhard 2012] **O Guichard, A Wienhard**, *Anosov representations: domains of discontinuity and applications*, Invent. Math. 190 (2012) 357–438 MR

- [Izeki 2000] **H Izeki**, *Quasiconformal stability of Kleinian groups and an embedding of a space of flat conformal structures*, *Conform. Geom. Dyn.* 4 (2000) 108–119 MR
- [Kapovich 2001] **M Kapovich**, *Hyperbolic manifolds and discrete groups*, *Progr. Math.* 183, Birkhäuser, Boston, MA (2001) MR
- [Kapovich 2008] **M Kapovich**, *Kleinian groups in higher dimensions*, from “Geometry and dynamics of groups and spaces” (M Kapranov, S Kolyada, Y I Manin, P Moree, L Potyagailo, editors), *Progr. Math.* 265, Birkhäuser, Boston, MA (2008) 487–564 MR
- [Kapovich 2016] **M Kapovich**, *Discreteness is undecidable*, *Internat. J. Algebra Comput.* 26 (2016) 467–472 MR
- [Kapovich and Leeb 2018a] **M Kapovich, B Leeb**, *Discrete isometry groups of symmetric spaces*, from “Handbook of group actions, IV” (L Ji, A Papadopoulos, S-T Yau, editors), *Adv. Lect. Math.* 41, International, Somerville, MA (2018) 191–290 MR
- [Kapovich and Leeb 2018b] **M Kapovich, B Leeb**, *Finsler bordifications of symmetric and certain locally symmetric spaces*, *Geom. Topol.* 22 (2018) 2533–2646 MR
- [Kapovich and Leeb 2023] **M Kapovich, B Leeb**, *Relativizing characterizations of Anosov subgroups, I*, *Groups Geom. Dyn.* 17 (2023) 1005–1071 MR
- [Kapovich and Leeb \geq 2025] **M Kapovich, B Leeb**, *Relativizing characterizations of Anosov subgroups, II*, in preparation
- [Kapovich and Liu 2019] **M Kapovich, B Liu**, *Geometric finiteness in negatively pinched Hadamard manifolds*, *Ann. Acad. Sci. Fenn. Math.* 44 (2019) 841–875 MR
- [Kapovich et al. 2009] **M Kapovich, B Leeb, J Millson**, *Convex functions on symmetric spaces, side lengths of polygons and the stability inequalities for weighted configurations at infinity*, *J. Differential Geom.* 81 (2009) 297–354 MR
- [Kapovich et al. 2014] **M Kapovich, B Leeb, J Porti**, *Morse actions of discrete groups on symmetric space*, preprint (2014) arXiv 1403.7671
- [Kapovich et al. 2017] **M Kapovich, B Leeb, J Porti**, *Anosov subgroups: dynamical and geometric characterizations*, *Eur. J. Math.* 3 (2017) 808–898 MR
- [Kapovich et al. 2018a] **M Kapovich, B Leeb, J Porti**, *Dynamics on flag manifolds: domains of proper discontinuity and cocompactness*, *Geom. Topol.* 22 (2018) 157–234 MR
- [Kapovich et al. 2018b] **M Kapovich, B Leeb, J Porti**, *A Morse lemma for quasigeodesics in symmetric spaces and Euclidean buildings*, *Geom. Topol.* 22 (2018) 3827–3923 MR
- [Kapovich et al. 2024] **M Kapovich, S Kim, J Lee**, *Structural stability of meandering-hyperbolic group actions*, *J. Inst. Math. Jussieu* 23 (2024) 753–810 MR
- [Labourie 2006] **F Labourie**, *Anosov flows, surface groups and curves in projective space*, *Invent. Math.* 165 (2006) 51–114 MR
- [Lok 1984] **W L Lok**, *Deformations of locally homogeneous spaces and Kleinian groups*, PhD thesis, Columbia Univ. (1984) Available at <https://www.proquest.com/docview/303285563>
- [Ol’shanskii and Sapir 2001] **A Y Ol’shanskii, M V Sapir**, *Length and area functions on groups and quasi-isometric Higman embeddings*, *Internat. J. Algebra Comput.* 11 (2001) 137–170 MR
- [Raghunathan 1972] **M S Raghunathan**, *Discrete subgroups of Lie groups*, *Ergebnisse der Math.* 68, Springer (1972) MR

- [Riesterberg 2025] **J M Riesterberg**, *A quantified local-to-global principle for Morse quasigeodesics*, *Groups Geom. Dyn.* 19 (2025) 37–107 MR
- [Riley 1983] **R Riley**, *Applications of a computer implementation of Poincaré’s theorem on fundamental polyhedra*, *Math. Comp.* 40 (1983) 607–632 MR
- [Sullivan 1985] **D Sullivan**, *Quasiconformal homeomorphisms and dynamics, II: Structural stability implies hyperbolicity for Kleinian groups*, *Acta Math.* 155 (1985) 243–260 MR
- [Tukia 1985] **P Tukia**, *On isomorphisms of geometrically finite Möbius groups*, *Inst. Hautes Études Sci. Publ. Math.* 61 (1985) 171–214 MR

*Department of Mathematics, University of California, Davis
Davis, CA, United States*

*Mathematisches Institut, Universität München
München, Germany*

*Departament de Matemàtiques and Centre de Recerca Matemàtica, Universitat Autònoma de Barcelona
Barcelona, Spain*

kapovich@math.ucdavis.edu, b.l@lmu.de, joan.porti@uab.cat

Proposed: Bruce Kleiner

Seconded: Tobias H Colding, David Fisher

Received: 9 January 2023

Revised: 26 May 2024

Nearly geodesic immersions and domains of discontinuity

COLIN DAVALO

We study nearly geodesic immersions in higher-rank symmetric spaces of noncompact type, which we define as immersions that satisfy a bound on their fundamental form, generalizing the notion of immersions in hyperbolic space with principal curvature in $(-1, 1)$. This notion depends on the choice of a flag manifold embedded in the visual boundary, and immersions satisfying this bound admit a natural domain in this flag manifold that comes with a fibration. As an application we give an explicit fibration of some domains of discontinuity for some Anosov representations. Our method can be applied in particular to some Θ -positive representations for each notion of Θ -positivity.

53C35; 22E40

1. Introduction	2391
2. Symmetric spaces of noncompact type	2397
3. Representations of hyperbolic groups	2406
4. Busemann functions on symmetric spaces	2409
5. Nearly geodesic immersions	2414
6. Pencils of tangent vectors	2426
7. Fibered domains in flag manifolds	2430
8. Applications	2439
Appendix A. Maximal representations of surface groups into $\mathrm{SO}_o(2, n)$	2450
Appendix B. Projective structures with (quasi-)Hitchin holonomy	2456
References	2459

1 Introduction

Let Γ_g be the fundamental group of a closed orientable surface S_g of genus $g \geq 2$. An interesting open subset of the space of representations $\rho: \Gamma_g \rightarrow \mathrm{PSL}(2, \mathbb{C})$ is the space of *quasi-Fuchsian* representations, which are representations whose limit set is a Jordan curve, or equivalently representations that are quasi-isometric embeddings.

An interesting open subset of this space is the set of *nearly Fuchsian* representations, which is the set of representations that admit a ρ -equivariant immersion from the universal cover of S_g to the hyperbolic space \mathbb{H}^3 whose image has principal curvature in $(-1, 1)$. Epstein [16] studied immersions satisfying this bound. He proved in particular that these immersions must be proper embeddings and that their Gauss map fibers a domain in $\partial\mathbb{H}^3$. Nearly Fuchsian representations must therefore be discrete, and one can even conclude that they are quasi-isometric embeddings. A nearly Fuchsian representation is called *almost Fuchsian* if the immersion u can be chosen to be a minimal immersion. These representations have been studied, among others, by Uhlenbeck [36].

A generalization of quasi-Fuchsian representations for higher-rank semisimple Lie groups are *Anosov representations*, introduced by Labourie [32] for representations of surface groups and by Guichard and Wienhard [24] for representations of any Gromov hyperbolic group. Given a subset Θ of the set of simple restricted roots of the Lie group G , a Θ -Anosov representation is a representation that satisfies some strong contraction property in the associated flag manifold, and it is in particular discrete. The set of Θ -Anosov representations $\rho: \Gamma \rightarrow G$ is moreover open.

We consider a generalization of Epstein's bound on the second fundamental form of an immersion in a symmetric space of noncompact type \mathbb{X} associated to a semisimple Lie group G . Given a unit vector τ in the model Weyl chamber, or equivalently a flag manifold \mathcal{F}_τ embedded in the visual boundary of \mathbb{X} , we define the notion of τ -*nearly geodesic* and *uniformly τ -nearly geodesic* immersions using Busemann functions. We prove that a *uniformly τ -nearly geodesic* immersion is an embedding, a quasi-isometric embedding, and admits a domain in the corresponding flag manifold \mathcal{F}_τ , which admits a fibration whose fibers are the bases of some *pencils of tangent vectors* in \mathbb{X} , which is a notion generalizing pencils of quadrics that we introduce.

Given the fundamental group Γ of a compact manifold N we also study τ -*nearly Fuchsian* representations, namely representations $\rho: \Gamma \rightarrow G$ that admit an equivariant and τ -nearly geodesic immersion $u: \tilde{N} \rightarrow \mathbb{X}$. We prove that they form an open subset of the space Anosov representations.

As an application we study some domains of discontinuity associated to Anosov representations in flag manifolds, constructed by Guichard and Wienhard [24] and Kapovich, Leeb and Porti [30]. Dumas and Sanders conjectured in [14] that, in the context of complex Lie groups, the quotient of the domain of discontinuity associated to a Tits–Bruhat ideal fibers equivariantly over the universal cover of the surface S_g for every representation of a surface group Γ_g that admits a totally geodesic equivariant immersion. We show that for a nearly Fuchsian representation ρ of the fundamental group of N , the domain associated with the immersion u coincides with a domain of discontinuity for ρ , and fibers equivariantly over the universal cover of N . This technique can be applied to representations of surface groups, and in particular to some Θ -positive representations for every notion of Θ -positivity. One can also apply these techniques to fundamental groups of hyperbolic manifolds of higher dimension: we consider two examples in the end of Section 8.5.

Independently and with different techniques, Alessandrini, Maloni, Tholozan and Wienhard proved that any domain of discontinuity associated to any representation that factors through a cocompact representation into a rank-1 simple Lie group admits an equivariant fibration [3]. However the fibers are not easy to describe in general.

It is not true that all the quotients of domains of discontinuity for Anosov representations of surface groups always fiber over the surface. Gromov, Lawson and Thurston showed that there exist convex-cocompact, and hence Anosov, representations of surface groups into $SO(1, 4)$ such that the quotient of its cocompact domain of discontinuity in the unique associated flag manifold is hyperbolic and does not fiber over the surface; see [21].

Although the only method we provide here to construct nearly geodesic immersions equivariant with respect to a representation involves deforming locally a totally geodesic immersion, there may be examples that cannot be deformed into generalized Fuchsian representations. Bronstein has constructed generalized almost-Fuchsian representations in the isometry group of \mathbb{H}^4 that cannot be deformed into Fuchsian representations [9].

1.1 Nearly geodesic immersions

We introduce and study a generalization of the condition of having principal curvature in $(-1, 1)$ in the setting of higher-rank semisimple Lie groups of noncompact type G . More specifically we choose a unit vector τ in the associated model Weyl chamber \mathfrak{a}^+ . This choice is equivalent to the choice of a flag manifold \mathcal{F}_τ embedded in the visual boundary $\partial_{\text{vis}}\mathbb{X}$ of the symmetric space \mathbb{X} associated to G . Let $\iota: \mathfrak{S}\mathfrak{a}^+ \rightarrow \mathfrak{S}\mathfrak{a}^+$ be the antipodal involution; to a point $a \in \mathcal{F}_\tau \cup \mathcal{F}_{\iota(\tau)} \subset \partial_{\text{vis}}\mathbb{X}$ and a basepoint $o \in \mathbb{X}$ one can associate a *Busemann function* $b_{a,o}$ which can be interpreted as the distance to a in \mathbb{X} , relative to o .

Definition 1.1 (Definition 5.1) An immersion $u: M \rightarrow \mathbb{X}$ is called τ -nearly geodesic if for all $a \in \mathcal{F}_\tau \cup \mathcal{F}_{\iota(\tau)}$ the function $b_{a,o} \circ u: M \rightarrow \mathbb{R}$ has positive Hessian in any critical direction $v \in TM$, for the metric on M induced by the immersion.

This property is satisfied for totally geodesic immersions whose tangent vectors are τ -regular, namely whose Cartan projection is not orthogonal to $w \cdot \tau$ for any w in the Weyl group (see Definition 5.6). It is equivalent to an open bound on the second fundamental form that depends on the Cartan projection of the image of the differential of the immersion (see Proposition 5.2). When $G = \text{PSL}(2, \mathbb{C})$, a τ -nearly geodesic immersion for the only $\tau \in \mathfrak{S}\mathfrak{a}^+$ is exactly an immersion with sectional curvature in $(-1, 1)$ in $\mathbb{X} = \mathbb{H}^3$ (see Proposition 5.5).

If the immersion is complete and *uniformly* τ -nearly geodesic, namely if the Hessians of Busemann functions in critical directions are uniformly bounded from below, we show moreover that it is a τ -regular embedding, a quasi-isometric embedding (see Proposition 5.16) and that the nearest-point projection $\pi_u^\tau \circ u: \mathbb{X} \rightarrow u(M)$ is well defined for some explicit appropriate Finsler pseudodistance on \mathbb{X} depending on τ (see Proposition 5.17).

If a τ -nearly geodesic immersion is equivariant with respect to a representation of a group acting cocompactly and properly on M , it provides information about the representation:

Theorem 1.2 (Theorem 5.22) *A τ -nearly Fuchsian representation $\rho: \Gamma \rightarrow G$ is Θ -Anosov for some nonempty set of roots Θ that depends on τ and ρ .*

This theorem can be stated in a simpler form if we consider a vector $\tau \in \mathbb{S}\mathfrak{a}^+$ colinear to a coroot. Given a simple Lie group, the set Δ of simple roots can be partitioned into one or two *Weyl orbits of simple roots*; see Figure 4. To a Weyl orbit of simple roots Θ one can associate a unique normalized coroot $\tau_\Theta \in \mathbb{S}\mathfrak{a}^+$ in the model Weyl chamber. The previous theorem can be restated in this case:

Theorem 1.3 (Theorem 5.20) *A τ_Θ -nearly Fuchsian representation for some Weyl orbit of simple roots $\Theta \subset \Delta$ is Θ -Anosov.*

When $\tau = \tau_\Theta$ we also prove a sufficient condition for a surface to be τ -nearly geodesic:

Theorem 1.4 (Theorem 5.23) *Let Θ be a Weyl orbit of simple roots. Let $u: S \rightarrow \mathbb{X}$ be an immersion that satisfies, for all $v \in TS$ and $\alpha \in \Theta$,*

$$(1) \quad \|\Pi_u(v, v)\| < c_\Theta \alpha(\mu(du(v)))^2.$$

Then u is a τ_Θ -nearly geodesic immersion.

Here $\mu: T\mathbb{X} \rightarrow \mathfrak{a}^+$ denotes the *Cartan projection*. The constant c_Θ depends on the scaling of the metric chosen on \mathbb{X} , and on Θ .

1.2 An associated domain in a flag manifold

In Section 6 we introduce and study *pencils of tangent vectors*, or d -pencils, which are vector subspaces $\mathcal{P} \subset T_x\mathbb{X}$ of dimension d for some $x \in \mathbb{X}$. When $G = \mathrm{PSL}(n, \mathbb{R})$ these can be thought of as *pencils of quadrics* with zero trace with respect to some scalar product (see Proposition 6.5). To a pencil we associate a subset of the flag manifold \mathcal{F}_τ that we call its *base*, which is a smooth submanifold if the pencil is τ -regular, namely if all nonzero vectors $v \in \mathcal{P}$ are τ -regular; see Lemma 6.7. When $G = \mathrm{PSL}(n, \mathbb{R})$ and $\mathcal{F}_\tau \simeq \mathbb{R}\mathbb{P}^{n-1}$ the base of the pencil corresponds to the intersection of all the quadric hypersurfaces defined by the elements of the corresponding pencil of quadrics.

To a complete and uniformly τ -nearly geodesic immersion $u: M \rightarrow \mathbb{X}$, we associate an open domain $\Omega_u^\tau \subset \mathcal{F}_\tau$ consisting of points $a \in \mathcal{F}_\tau$ for which $b_{a,o} \circ u$ is proper and bounded from below for one, and hence any basepoint $o \in \mathbb{X}$. We define a projection $\pi_u: \Omega_u^\tau \rightarrow M$ that associates to $a \in \Omega_u^\tau$ the point in M at which $b_{a,o} \circ u$ is minimal. This point will be unique because $b_{a,o} \circ u$ is convex and strictly convex in critical directions.

Theorem 1.5 (Theorem 7.3) *Let $u: M \rightarrow \mathbb{X}$ be a complete and uniformly τ -nearly geodesic immersion. The map $\pi_u: \Omega_u^\tau \rightarrow M$ is a fibration. The fiber $\pi_u^{-1}(x)$ at a point $x \in M$ is the base $\mathcal{B}_\tau(\mathcal{P}_x)$ of the τ -regular pencil of tangent vectors $\mathcal{P}_x = du(T_x M)$.*

A domain of discontinuity Ω for a representation $\rho: \Gamma \rightarrow G$ in a G -homogeneous space is an open $\rho(\Gamma)$ -invariant set on which $\rho(\Gamma)$ acts properly discontinuously (some authors call these domains of *proper* discontinuity). A *cocompact* domain of discontinuity is a domain on which the action of $\rho(\Gamma)$ is cocompact.

If the representation $\rho: \Gamma \rightarrow G$ admits an equivariant τ -nearly geodesic immersion $u: \tilde{N} \rightarrow \mathbb{X}$, the domain $\Omega_\rho^\tau := \Omega_u^\tau$ does not depend on u , and is a cocompact domain of discontinuity for ρ . The fibration π_u from Theorem 1.5 is ρ -equivariant so the quotient of Ω_ρ^τ fibers over the surface. We show in Theorem 7.11 that the domain Ω_ρ^τ coincides with some domains of discontinuity associated to *Tits–Bruhat ideals* for Anosov representations, constructed by Kapovich, Leeb and Porti [30]. The Tits–Bruhat ideals needed to describe the domains Ω_ρ^τ can always be constructed as a *metric thickening*.

We prove in Section 7.4 that the diffeomorphism type of the domains of discontinuity obtained by Tits–Bruhat ideals is invariant under deformation in the space of Anosov representations. We therefore understand the topology of some domains of discontinuity for all representations in some connected component of a Θ -Anosov representation.

1.3 Applications

A union of connected components of representations of a surface group into a higher-rank simple Lie group G consisting only of discrete representations is called a *higher-rank Teichmüller space*; see Wienhard [37]. The Hitchin component is a higher-rank Teichmüller space for any split real simple Lie group. When G is of Hermitian type, the space of *maximal* representations, namely representations with maximal Toledo invariant, is also a higher-rank Teichmüller space. More generally, Guichard and Wienhard defined a notion of Θ -positive representations for some simple Lie groups and some subsets of simple roots Θ , of which Hitchin and maximal representations are a particular instance. Beyer and Pozzetti proved that the spaces of Θ -positive representations in $SO(p, q)$ for $3 \leq p < q$ are higher-rank Teichmüller spaces [5]. Bradlow, Collier, García-Prada, Gothen and Oliveira [7] and Guichard, Labourie and Wienhard [22] proved that for each notion of Θ -positivity there exist higher-rank Teichmüller spaces consisting of Θ -positive representations.

The classical Teichmüller space can be defined as a connected component of the space of conjugacy classes of discrete and faithful representations of a surface group into $PSL(2, \mathbb{R})$. It can also be interpreted as the space of marked hyperbolic structures on the corresponding surface, a particular instance of (G, X) -structures in the sense of Klein. To a manifold M equipped with a (G, X) -structure one can associate its *holonomy representation* $\rho: \pi_1(M) \rightarrow G$, defined up to conjugation by elements of G .

Hitchin underlined similarities between the Teichmüller space and the Hitchin component, and asked about the geometric significance of elements in the Hitchin component. In the spirit of this question it is natural to ask the following:

Question 1.6 Can Θ -positive representations $\rho: \Gamma_g \rightarrow G$ be characterized as the holonomies on some (G, X) -structures on a compact manifold M ?

In higher rank, Guichard and Wienhard [24] constructed geometric structures associated to Anosov representations by constructing cocompact domains of discontinuity in the corresponding flag manifolds. Such domains have been constructed in more generality by Kapovich, Leeb and Porti [30]. Since Θ -positive representations are Θ -Anosov [22], the construction of Kapovich, Leeb and Porti provides geometric structures associated to any Θ -positive representation.

The topology on the quotient of these domains, which are the manifolds on which we put a geometric structure, is not completely clear. Alessandrini, Li and the author [2] studied the topology of this manifold for Hitchin representations in $\mathrm{PSL}(2n, \mathbb{R})$. Here we study more generally the domains of discontinuity from [30] that coincide with the domains Ω_ρ^τ .

Suppose that G is center-free. Let \mathfrak{h} be an \mathfrak{sl}_2 Lie subalgebra in the Lie algebra \mathfrak{g} of G , namely a Lie subalgebra isomorphic to $\mathfrak{sl}_2(\mathbb{R})$. One can associate to \mathfrak{h} a representation $\iota_{\mathfrak{h}}: \mathrm{SL}(2, \mathbb{R}) \rightarrow G$ and a $\iota_{\mathfrak{h}}$ -equivariant and totally geodesic embedding $u_{\mathfrak{h}}: \mathbb{H}^2 \rightarrow \mathbb{X}$. Choose $\tau \in \mathbb{S}\mathfrak{a}^+$ such that $u_{\mathfrak{h}}$ is τ -regular, so that $u_{\mathfrak{h}}$ is τ -nearly geodesic. A representation preserving and acting properly and cocompactly on $u_{\mathfrak{h}}(\mathbb{H}^2)$ will be called \mathfrak{h} -generalized Fuchsian.

Theorem 1.7 (Theorem 8.7) *Let $\rho: \Gamma_g \rightarrow G$ be an \mathfrak{h} -generalized Fuchsian representation. Suppose that the associated domain of discontinuity Ω_ρ^τ is nonempty. Let \mathcal{C} be the connected component of the space of Θ -Anosov representations that contains ρ , for some nonempty Θ depending on \mathfrak{h} . Every representation in \mathcal{C} is the restricted holonomy of a (G, \mathcal{F}_τ) -structure **on a fiber bundle** over S_g .*

The holonomy of the structure along the fibers is trivial, so one can consider the restricted holonomy even if the fiber has nontrivial fundamental group. The fiber bundle is induced as the reduction of an $(\mathrm{SO}(2, \mathbb{R}) \times \mathcal{C}_K(\mathfrak{h}))$ -principal bundle over S_g via the action of $\mathrm{SO}(2, \mathbb{R}) \times \mathcal{C}_K(\mathfrak{h})$ on a codimension-2 submanifold of the flag manifold \mathcal{F}_τ associated to \mathfrak{h} . This codimension-2 submanifold, which is the fiber of the fiber bundle, can be described as the union of connected components of the τ -base of a pencil of tangent vectors in \mathbb{X} associated to \mathfrak{h} .

This theorem applies to connected components of Θ -positive representations that contain an \mathfrak{h}_Θ -generalized Fuchsian representation, where \mathfrak{h}_Θ is a Θ -principal \mathfrak{sl}_2 Lie subalgebra. These components are known to form higher-rank Teichmüller spaces of Θ -positive representations; see Guichard, Labourie and Wienhard [22]. Suppose that G admits a notion of Θ -positivity and is not isomorphic to $\mathrm{PSL}(2, \mathbb{R})$.

Corollary 1.8 (Corollary 8.11) *Every Θ -positive representation $\rho: \Gamma_g \rightarrow G$ in a component containing an \mathfrak{h}_Θ -generalized Fuchsian representation is the restricted holonomy of a $(G, \mathcal{F}_{\tau_{\Theta'}})$ -structure on a fiber bundle over S_g , for every Weyl orbit of simple roots $\Theta' \subset \Theta$.*

In this theorem $\mathcal{F}_{\tau_{\Theta}}$ is a flag manifold associated to a set of simple roots $\Theta(\tau_{\Theta}) \subset \Delta$. In Figure 4 we give the list of all Weyl orbit of simple roots, together with the corresponding sets of simple roots $\Theta(\tau_{\Theta})$. Corollary 1.8 describes exactly one geometric structure for each notion of Θ -positivity, except for Hitchin representations in nonsimply laced split Lie groups, for which it describes two geometric structures.

Any maximal representation $\rho: \Gamma_g \rightarrow \mathrm{Sp}(2n, \mathbb{R})$ for $n \geq 3$ is the restricted holonomy of a projective structure on a fiber bundle over S_g with fiber a Stiefel manifold. In the case $n = 2$, such structures are already well described by Collier, Tholozan, Toulisse in [13], and it is not clear if our method can apply to the *Gothen components*. In Appendix A we compare their fibration of the corresponding domain of discontinuity with the one that we construct for nearly Fuchsian representations.

Every Hitchin representation $\rho: \Gamma_g \rightarrow \mathrm{PSL}(n, \mathbb{R})$ is the restricted holonomy of a structure modeled on $\mathcal{F}_{1,n-1}$ on a fiber bundle over S_g , where $\mathcal{F}_{1,n-1}$ is the space of partial flags in \mathbb{R}^n of the form (ℓ, H) where $\ell \subset H \subset \mathbb{R}^n$, $\dim(\ell) = 1$ and $\dim(H) = n - 1$.

Similarly every Θ -positive representation $\rho: \Gamma_g \rightarrow \mathrm{SO}(p, q)$ for $3 \leq p$ and $p + 1 < q$ is the restricted holonomy of a structure modeled on \mathcal{F}_2 on a fiber bundle over S_g , where \mathcal{F}_2 is the space of isotropic planes in $\mathbb{R}^{p,q}$. If $q = p + 1$, there are some exceptional components that are conjectured to consist only of Zariski-dense representations, so it is not clear if our method can be applied to these.

Acknowledgments

I would like to thank Max Riestenberg and Gabriele Viaggi for very precious discussions on the geometry of symmetric spaces. I would like to thank Thi Dang and Johannes Horn for raising interesting questions about nearly geodesic immersions. Finally I would like to thank Beatrice Pozzetti for her comments and support throughout this work and the preparation of the paper. The author was funded through the DFG Emmy Noether project 427903332 of Pozzetti.

2 Symmetric spaces of noncompact type

In this section we recall the general theory of symmetric spaces of noncompact type and fix some notation. References for the results mentioned can be found in [15; 26]. We then illustrate some of these notions for some families of Lie groups. Finally we introduce the notion of the Weyl orbit of simple roots.

2.1 Symmetric space associated to a semisimple Lie group

Let G be a connected semisimple Lie group with finite center and no compact factors, namely of noncompact type.

Let \mathfrak{g} be the Lie algebra of G , and let B be the *Killing form* on \mathfrak{g} . Since \mathfrak{g} is semisimple it admits a *Cartan involution*, namely an involutive automorphism $\theta: \mathfrak{g} \rightarrow \mathfrak{g}$ such that $(v, w) \mapsto -B(v, \theta(w))$ is a scalar product on \mathfrak{g} . Any two Cartan involutions are conjugated by Ad_g for some $g \in G$.

Let \mathbb{X} be the space of Cartan involutions of \mathfrak{g} . For any $x \in X$ we will write the corresponding Cartan involution $\theta_x: \mathfrak{g} \rightarrow \mathfrak{g}$. This involution determines a B -orthogonal decomposition $\mathfrak{g} = \mathfrak{t}_x \oplus \mathfrak{p}_x$, where \mathfrak{t}_x is the $+1$ eigenspace of θ , and \mathfrak{p}_x the -1 eigenspace.

For $x \in X$, define K_x to be the group of elements $k \in G$ such that Ad_k commutes with θ_x . This subgroup is a maximal compact subgroup of G . Given any $x \in \mathbb{X}$, one can identify \mathbb{X} with the homogeneous space G/K_x . The Lie subalgebra \mathfrak{t}_x is the Lie algebra of the compact K_x , and thus the space \mathfrak{p}_x is naturally identified with $T_x\mathbb{X}$.

Let $\langle \cdot, \cdot \rangle_x$ be the scalar product defined for $v, w \in \mathfrak{g}$ as

$$(2) \quad \langle v, w \rangle_x = B(v, \theta_x(w)).$$

This scalar product restricted to $\mathfrak{p}_x \simeq T_x\mathbb{X}$ defines a Riemannian metric $g_{\mathbb{X}}$ on \mathbb{X} . We will denote by $d_{\mathbb{X}}$ the induced Riemannian distance on \mathbb{X} . With this metric, the space \mathbb{X} is a symmetric space in the sense that for all $x \in \mathbb{X}$ there is an isometry σ_x of X such that $d_x\sigma = -\text{Id}$.

The symmetric space \mathbb{X} is of *noncompact type*. It is simply connected and has nonpositive sectional curvature. In particular it is a *Hadamard manifold*.

Remark 2.1 We only consider symmetric spaces \mathbb{X} associated to semisimple Lie groups G , having their Riemannian metric defined via the Killing form.

2.2 Reduced root systems

Fix a basepoint $o \in \mathbb{X}$. Let \mathfrak{a} be a choice of a *maximal abelian subalgebra* of \mathfrak{p}_o . These maximal abelian subalgebras are all conjugated by elements of K_x . The dimension $\text{rank}(\mathbb{X})$ of \mathfrak{a} will be called the *rank* of \mathbb{X} .

Remark 2.2 In general $\text{rank}(\mathbb{X}) \leq \text{rank}(G)$, where $\text{rank}(G)$ is the dimension of any Cartan subalgebra in \mathfrak{g} .

Let $\alpha \in \mathfrak{a}^*$ be a linear form. Let \mathfrak{g}_α be the set of elements $v \in \mathfrak{g}$ such that for all $\tau \in \mathfrak{a}$

$$\text{ad}_\tau(v) = \alpha(\tau)v.$$

The *reduced root system* Σ is the set of linear forms $\alpha \in \mathfrak{a}^*$ such that $\mathfrak{g}_\alpha \neq \{0\}$. An element $\tau \in \mathfrak{a}$ is *regular* if for all $\alpha \in \Sigma \setminus \{0\}$, $\alpha(\tau) \neq 0$.

Let us choose a regular element $\tau_0 \in \mathfrak{a}$. Let Σ^+ be the associated set of *positive roots*, namely the set of $\alpha \in \Sigma$ such that $\alpha(\tau_0) > 0$. There exists a unique set Δ of linearly independent roots in Σ^+ such that any root in Σ^+ can be written as a linear combination of roots in Δ . The roots in Δ are called *simple roots*.

Let the *Weyl group* W be quotient of the subgroup of elements in K_x whose adjoint action stabilizes \mathfrak{a} by the subgroup of elements that fix \mathfrak{a} pointwise.

For any root $\alpha \in \Sigma \setminus \{0\}$ there is an element $\sigma_\alpha \in W$ whose action on \mathfrak{a} is the orthogonal symmetry with respect to $\text{Ker}(\alpha)$ in \mathfrak{a} . The Weyl group acts linearly on \mathfrak{a} , and it is generated by the elements $(\sigma_\alpha)_{\alpha \in \Delta}$. The *model Weyl chamber* \mathfrak{a}^+ is the cone $\{\tau \in \mathfrak{a} \mid \forall \alpha \in \Delta, \alpha(\tau) \geq 0\}$. For any $\tau_0 \in \mathfrak{a}$ there is a unique $\tau \in \mathfrak{a}^+$ such that for some $w \in W$, $w \cdot \tau_0 = \tau$. An element $\tau \in \mathfrak{a}^+$ is Θ -regular for $\Theta \subset \Delta$ if for all $\alpha \in \Theta$, $\alpha(\tau) \neq 0$.

We denote by $\mathbb{S}\mathfrak{a}$ and $\mathbb{S}\mathfrak{a}^+$ the unit sphere in \mathfrak{a} and the unit sphere intersected with the model Weyl chamber \mathfrak{a}^+ , respectively, for the metric (2).

Let $w_0 \in W$ be the only element such that $w_0 \cdot \mathfrak{a}^+ = -\mathfrak{a}^+$. This element is called *the longest element* of the Weyl group. Let $\iota: \mathfrak{a}^+ \rightarrow \mathfrak{a}^+$ be the *opposition involution*, namely the involution such that for $\tau \in \mathbb{S}\mathfrak{a}^+$, $\iota(\tau) = -w_0 \cdot \tau$. An element $\tau \in \mathbb{S}\mathfrak{a}^+$ is called *symmetric* if $\iota(\tau) = \tau$.

Given a set of simple roots $\Theta \subset \Delta$, we say that the *model Θ -facet* is the set of elements $\tau \in \mathbb{S}\mathfrak{a}^+$ such that for all $\alpha \in \Delta \setminus \Theta$, $\alpha(\tau) = 0$. The *open model Θ -facet* is the set of elements $\tau \in \mathbb{S}\mathfrak{a}^+$ such that for all $\alpha \in \Delta \setminus \Theta$, $\alpha(\tau) = 0$, and for all $\alpha \in \Theta$, $\alpha(\tau) > 0$. For an element $\tau \in \mathbb{S}\mathfrak{a}^+$ we will write $\Theta(\tau)$ for the unique set of simple roots such that α lies in the open model $\Theta(\tau)$ -facet.

2.3 Maximal flats, visual boundary and parabolic subgroups

A *flat* in \mathbb{X} is a complete totally geodesic subspace of \mathbb{X} on which the sectional curvature completely vanishes. A flat F is maximal if $\dim(F) = \text{rank}(\mathbb{X})$. Flats passing through a point $x \in \mathbb{X}$ are in one-to-one correspondence with abelian subalgebras of \mathfrak{p}_x , and maximal flats correspond to maximal subalgebras. As a consequence the action of G on the space of maximal flats is transitive. The maximal flat corresponding to \mathfrak{a} will be called the *model flat*. Moreover for any $x \in X$ and $v \in T_x\mathbb{X}$, there is a maximal flat F such that $x \in F$ and $v \in T_x F$.

We say that two geodesic rays parametrized with unit length $\eta_1, \eta_2: \mathbb{R}_{\geq 0} \rightarrow \mathbb{X}$ are *asymptotic* if there exists a positive constant C such that for all $t > 0$, $d_{\mathbb{X}}(\eta_1(t), \eta_2(t)) \leq C$. This defines an equivalence relation on the space of rays.

The *visual boundary* $\partial_{\text{vis}}\mathbb{X}$ of the symmetric space \mathbb{X} is the space of classes of asymptotic geodesic rays parametrized with unit speed. The group G acts by isometries on \mathbb{X} , and hence it acts on $\partial_{\text{vis}}\mathbb{X}$.

A unit vector $v \in T_x\mathbb{X}$ at a point $x \in \mathbb{X}$ *points towards* $a \in \partial_{\text{vis}}\mathbb{X}$ if the geodesic ray γ such that $\gamma(0) = x$ and $\gamma'(0) = v$ is in the class corresponding to a . Since \mathbb{X} is a Hadamard manifold, for any $x \in \mathbb{X}$ and $a \in \partial_{\text{vis}}\mathbb{X}$ there is a unique unit vector that points towards a . We will denote this vector by $v_{a,x}$ throughout the paper. There exists a unique topology on $\partial_{\text{vis}}\mathbb{X}$ such that for any $x \in X$ the map $\phi_x: a \mapsto v_{a,x}$ is a homeomorphism between $\partial_{\text{vis}}\mathbb{X}$ and $T_x^1\mathbb{X}$.

The visual boundary of the model flat can be identified with $\mathbb{S}\mathfrak{a}$, and is included in the visual boundary of \mathbb{X} . The G -orbit of the point in the visual boundary of \mathbb{X} associated to $\tau \in \mathbb{S}\mathfrak{a}^+$ will be denoted by \mathcal{F}_τ .

A Θ -facet is a subset of $\partial_{\text{vis}}\mathbb{X}$ that is the image of the model Θ -facet by the action of an element of G . We define similarly the notion of *open Θ -facet*. The stabilizer of the open model Θ -facet will be denoted by P_Θ , and we will denote by \mathcal{F}_Θ the associated *flag manifolds*, namely the quotient G/P_Θ . Since P_Θ is also the stabilizer of any point in the open Θ -facet, there exists a natural G -equivariant diffeomorphism between the G -homogeneous spaces \mathcal{F}_Θ and \mathcal{F}_τ for any τ in the open model Θ -facet.

Example 2.3 When $G = \text{PSL}(n, \mathbb{R})$, a parabolic subgroup is the stabilizer of a partial flag f . Any point in $\partial_{\text{vis}}\mathbb{X}$ belongs to a unique open facet, which corresponds to a partial flag. The type of the partial flag, namely the dimensions of the subspaces that form the flag, determine a set of roots Θ_f . The points in $\partial_{\text{vis}}\mathbb{X}$ are in one-to-one correspondence with partial flags decorated with a point in the open Θ_f -model facet. This decoration can be interpreted as a collection of weights associated to the subspaces of the partial flag.

Given any two points $a, a' \in \partial_{\text{vis}}\mathbb{X}$, one can define their *Tits angle* $\angle_{\text{Tits}}(a, a')$ as the minimum of $\angle(v_{a,x}, v_{a',x})$ for $x \in \mathbb{X}$. This minimum is obtained when $x \in \mathbb{X}$ lies in a common flat with a and a' .

2.4 Cartan and Iwasawa decomposition

The *Cartan projection* $\mu: T\mathbb{X} \rightarrow \mathfrak{a}^+$ is the function that maps any vector $w \in T_x\mathbb{X}$ to the unique element $\mu(w) \in \mathfrak{a}^+$ of the model Weyl chamber such that for some $g \in G$, $g \cdot w = \mu(w)$.

The *generalized distance* $d_a(x, y)$ based at $x \in \mathbb{X}$ of a point $y \in \mathbb{X}$ is the Cartan projection $\mu(w)$ of the unique vector $w \in T_x\mathbb{X} \simeq \mathfrak{p}_x$ such that $\exp(w) \cdot x = y$. This generalized distance is 1-Lipschitz in the following sense:

Lemma 2.4 *Let $x, y, z \in \mathbb{X}$. Then*

$$|d_a(x, z) - d_a(x, y)| \leq d_{\mathbb{X}}(y, z).$$

A proof of this lemma can be found for instance in [35, Corollary 3.8]. Here $|\cdot|$ means the norm induced by the metric (2).

We say that a vector $v \in T\mathbb{X}$ is Θ -regular for a set of simple root Θ if its Cartan projection is Θ -regular, namely it avoids the walls of the Weyl chamber associated with elements of Θ . We will later introduce a similar notion of a τ -regular vector in Definition 5.6.

Let $T_{a,x}: P_a \rightarrow G$ be the map that associates to $g \in G_a$ the limit

$$\lim_{t \rightarrow +\infty} \exp(-tv_{a,x})g \exp(tv_{a,x}).$$

It is a well-defined continuous morphism. Let $N_{a,x}$ be the kernel of $T_{a,x}$, and $\mathfrak{n}_{a,x}$ its Lie algebra. The generalized Iwasawa decomposition is useful to compute Busemann functions.

Theorem 2.5 (generalized Iwasawa decomposition) *Let $x \in \mathbb{X}$ and $a \in \partial_{\text{vis}}\mathbb{X}$. Then the map*

$$N_{a,x} \times \exp(\mathfrak{a}_{a,x}) \times K_x \rightarrow G, \quad (n, \exp(v), k) \mapsto n \exp(v)k,$$

is a diffeomorphism. In particular for every $x \in \mathbb{X}$ and $a \in \partial_{\text{vis}}\mathbb{X}$ there is a splitting

$$\mathfrak{g} = \mathfrak{n}_{a,x} \oplus \mathfrak{a}_{a,x} \oplus \mathfrak{k}_x.$$

The sum $\mathfrak{n}_{a,x} \oplus \mathfrak{a}_{a,x}$ is orthogonal with respect to $\langle \cdot, \cdot \rangle_x$.

In this theorem, $\mathfrak{a}_{a,x} \subset \mathfrak{p}_x$ is the centralizer of $v_{a,x}$.

2.5 Examples

In this subsection we consider the case when the semisimple Lie group G is equal to $\text{PSL}(n, \mathbb{R})$, $\text{PSL}(n, \mathbb{C})$, $\text{PSp}(2n, \mathbb{R})$ or $\text{PSO}(p, q)$. The notation introduced here will be used in the examples throughout the paper.

Let $G = \text{PSL}(n, \mathbb{R})$ and let us fix a volume form on \mathbb{R}^n . Let \mathcal{S}_n for $n \geq 2$ be the space of all scalar products on \mathbb{R}^n having volume 1. The group $\text{PSL}(n, \mathbb{R})$ acts transitively on \mathcal{S}_n by changing the basis, ie for $g \in \text{PSL}(n, \mathbb{R})$, $q \in X$ and $v, w \in \mathbb{R}^n$ we have that $g \cdot q(v, w) = q(g^{-1}(v), g^{-1}(w))$. For any $q \in \mathcal{S}_n$ the space \mathcal{S}_n can be identified with the quotient $\text{PSL}(n, \mathbb{R}) / \text{PSO}(q) \cong \text{PSL}(n, \mathbb{R}) / \text{PSO}(n, \mathbb{R})$.

Let θ_q at a point $q \in X$ be the involutive automorphism of $\mathfrak{sl}(n, \mathbb{R})$ defined by $u \mapsto -u^T$, where u^T is the transpose of u with respect to the scalar product q . This is a Cartan involution. The space \mathcal{S}_n is the symmetric space of noncompact type associated to $G = \text{PSL}(n, \mathbb{R})$.

The space \mathfrak{p}_q is the space of symmetric endomorphisms with respect to q , and \mathfrak{k}_q is the space of antisymmetric endomorphisms with respect to q . The scalar product $\langle \cdot, \cdot \rangle_q$ at a point $q \in \mathcal{S}_n$ is equal to $\langle u, v \rangle_q = 2n \text{Tr}(u^T v)$ for $u, v \in \mathfrak{sl}(n, \mathbb{R})$.

We choose the standard scalar product $q \in \mathcal{S}_n$ on \mathbb{R}^n to be our basepoint of \mathcal{S}_n . A maximal abelian subalgebra $\mathfrak{a} \subset \mathfrak{p}_q \subset \mathfrak{sl}(n, \mathbb{R})$ is equal to the algebra of diagonal matrices:

$$\mathfrak{a} = \left\{ \text{Diag}(\sigma_1, \dots, \sigma_n) \mid \sigma_1, \dots, \sigma_n \in \mathbb{R}^n, \sum_{i=1}^n \sigma_i = 0 \right\}.$$

A choice of simple root is $\Delta = \{\alpha_1, \dots, \alpha_{n-1}\}$ where for any $1 \leq i \leq n-1$ and any $\tau = \text{Diag}(\sigma_1, \dots, \sigma_n) \in \mathfrak{a}$, $\alpha_i(\tau) = \sigma_i - \sigma_{i+1}$.

The Weyl chamber associated to this choice is

$$\mathfrak{a}^+ = \left\{ \text{Diag}(\sigma_1, \dots, \sigma_n) \mid \sigma_1 \geq \dots \geq \sigma_n \in \mathbb{R}^n, \sum_{i=1}^n \sigma_i = 0 \right\}.$$

The Weyl group W is isomorphic to \mathfrak{S}_n . It acts on \mathfrak{a} by permuting the entries.

Let $G = \text{PSL}(n, \mathbb{C})$. The space \mathcal{H}_n of all positive-definite Hermitian bilinear forms of \mathbb{C}^n having volume 1 can be identified with $\text{PSL}(n, \mathbb{C}) / \text{PSU}(n, \mathbb{C})$. It can be given in a similar way a Riemannian metric that makes it a symmetric space of noncompact type associated to $G = \text{PSL}(n, \mathbb{C})$.

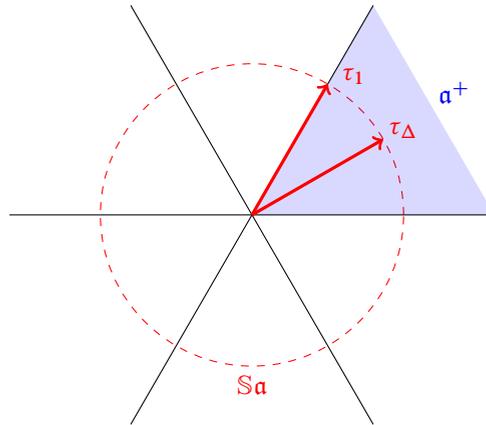


Figure 1: The model restricted Cartan algebra \mathfrak{a} and its Weyl chamber \mathfrak{a}^+ for $\text{PSL}(3, \mathbb{R})$.

The subalgebra $\mathfrak{a} \subset \mathfrak{sl}(n, \mathbb{R}) \subset \mathfrak{sl}(n, \mathbb{C})$ defined previously is still a maximal abelian subalgebra of \mathfrak{p}_q . One has $\text{rank}(\mathcal{S}_n) = \text{rank}(\mathcal{H}_n) = \text{rank}(\text{PSL}(n, \mathbb{R})) = n - 1$, but $\text{rank}(\text{PSL}(n, \mathbb{C})) = 2n - 2$.

Let $G = \text{PSp}(2n, \mathbb{R})$. Let ω be a symplectic form on \mathbb{R}^{2n} . Let \mathcal{X}_n be the space of endomorphisms J on \mathbb{R}^{2n} such that $J^2 = -\text{Id}$ and $(v, w) \mapsto \omega(v, J(w))$ is a scalar product on \mathbb{R}^{2n} . The semisimple Lie group $\text{PSp}(2n, \mathbb{R})$ acts on \mathcal{X}_n by conjugation. The space \mathcal{X}_n can be identified with $\text{PSp}(2n, \mathbb{R}) / \text{PSU}(n, \mathbb{R})$. This is one of the models for the *Siegel space*; see for instance [12].

For $J \in \mathcal{X}_n$, let us write $\theta_J = \text{Ad}_J$. This is the Cartan involution of $\mathfrak{sp}(2n, \mathbb{R})$. The Siegel space is the symmetric space of noncompact type associated with $G = \text{PSp}(2n, \mathbb{R})$.

Let ω and $J \in \mathcal{X}_n$ be such that for $x = (x_1, \dots, x_{2n})$ and $y = (y_1, \dots, y_{2n})$,

$$\omega(x, y) = \sum_{i=1}^n x_i y_{2n-i} - \sum_{i=n+1}^{2n} x_i y_{2n-i}, \quad J(x) = (-x_{2n}, \dots, -x_{n+1}, x_n, \dots, x_1).$$

We chose this special symplectic form so that the following set of diagonal matrices is a maximal abelian subalgebra $\mathfrak{a} \subset \mathfrak{p}_J \subset \mathfrak{sp}(2n, \mathbb{R})$:

$$\mathfrak{a} = \{\text{Diag}(\sigma_1, \dots, \sigma_n, -\sigma_n, \dots, -\sigma_1) \mid \sigma_1, \dots, \sigma_n \in \mathbb{R}^n\}.$$

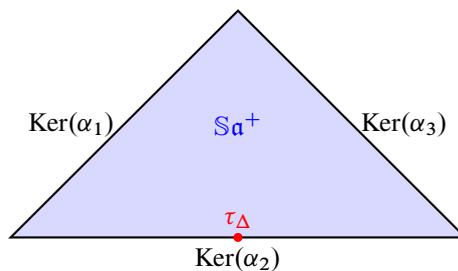


Figure 2: The projectivization of the Weyl chamber $\mathfrak{S}\mathfrak{a}^+$ for $G = \text{PSL}(4, \mathbb{R})$ in an affine chart.

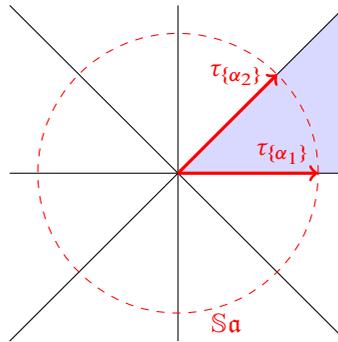


Figure 3: The model restricted Cartan algebra \mathfrak{a} and its Weyl chamber \mathfrak{a}^+ for $\mathrm{PSp}(4, \mathbb{R})$.

A choice of simple roots is $\Delta = \{\alpha_1, \dots, \alpha_n\}$ where $\alpha_i(\tau) = \sigma_i - \sigma_{i+1}$ for $1 \leq i \leq n-1$ and $\alpha_n(\tau) = 2\sigma_n$, with $\tau = \mathrm{Diag}(\sigma_1, \dots, \sigma_n, -\sigma_n, \dots, -\sigma_1)$.

The Weyl chamber associated to this choice is

$$\mathfrak{a}^+ = \{\mathrm{Diag}(\sigma_1, \dots, \sigma_n, -\sigma_n, \dots, -\sigma_1) \mid \sigma_1 \geq \dots \geq \sigma_n \geq 0 \in \mathbb{R}^n\}.$$

The Weyl group W is isomorphic to the subgroup of elements in \mathfrak{S}_{2n} that commutes with the involution $\iota: i \mapsto 2n + 1 - i$. It acts on \mathfrak{a} by permuting the entries.

Let $G = \mathrm{SO}(p, q)$ with $p < q$. Let $\mathbb{R}^{p,q}$ be the vector space \mathbb{R}^{p+q} equipped with a symmetric bilinear form $\langle \cdot, \cdot \rangle$ of signature (p, q) defined in the standard basis by

$$\langle x, y \rangle = \sum_{i=1}^p (x_i y_{p+q-i} + x_{p+q-i} y_i) - \sum_{i=1}^{p-q} x_{p+i} y_{p+i}.$$

A model for the associated symmetric space \mathbb{X} is the space of spacelike subspaces $U \subset \mathbb{R}^{p,q}$, namely subspaces on which $\langle \cdot, \cdot \rangle$ is positive definite.

A maximal abelian subalgebra $\mathfrak{a} \subset \mathfrak{p}_U \subset \mathfrak{so}(p, q)$ is the algebra

$$\mathfrak{a} = \{\mathrm{Diag}(\sigma_1, \dots, \sigma_p, 0, \dots, 0, -\sigma_p, \dots, -\sigma_1) \mid \sigma_1, \dots, \sigma_p\}.$$

A choice of simple roots is $\Delta = \{\alpha_1, \dots, \alpha_p\}$ where $\alpha_i(\tau) = \sigma_i - \sigma_{i+1}$ for $1 \leq i \leq p-1$, and $\alpha_p(\tau) = \sigma_p$.

The Weyl chamber associated to this choice is

$$\mathfrak{a}^+ = \{\mathrm{Diag}(\sigma_1, \dots, \sigma_p, 0, \dots, 0, -\sigma_p, \dots, -\sigma_1) \mid \sigma_1 \geq \dots \geq \sigma_p \geq 0\}.$$

The Weyl group W is isomorphic to the subgroup of elements in \mathfrak{S}_{2p} that commutes with the involution $\iota: i \mapsto 2p + 1 - i$. It acts on \mathfrak{a} by permuting the first and last p entries.

2.6 Weyl orbits of simple roots

In this subsection we introduce *Weyl orbits of simple roots*, which are special sets of simple roots. To a Weyl orbit of simple roots Θ one can associate a unit vector in the Weyl chamber $\tau_\Theta \in \mathfrak{S}\mathfrak{a}^+$ which is colinear to a coroot.

We consider the restricted root system Σ associated with the semisimple Lie group G , with a choice of a set of positive roots Σ^+ and of simple roots Δ . Two simple roots α and β are *conjugates* if there is an element w in the Weyl group W such that $\alpha = w \cdot \beta = \beta \circ w^{-1}$.

Definition 2.6 A set of simple roots $\Theta \subset \Delta$ is called a *Weyl orbit of simple roots* if it is an equivalence class for the conjugation relation on the set of simple roots Δ .

Proposition 2.7 Let Θ be a Weyl orbit of simple roots. There exists a unique unit vector $\tau_\Theta \in \mathbb{S}\mathfrak{a}^+$ such that for any $\alpha \in \Theta$ there is some $w \in W$ such that $w \cdot \tau_\Theta$ is orthogonal to $\ker(\alpha)$. The vector $\tau_\Theta \in \mathbb{S}\mathfrak{a}^+$ will be called the **normalized coroot** associated to Θ .

The normalized coroot associated to Θ is colinear to a coroot which is itself conjugate via the Weyl group to the coroot associated to any $\alpha \in \Theta$.

Proof Let $\alpha \in \Theta$. Let $\tau_0 \in \mathbb{S}\mathfrak{a}$ be a unit vector orthogonal to $\ker(\alpha)$. Since every orbit for the action of the Weyl group on $\mathbb{S}\mathfrak{a}^+$ intersects exactly once the model Weyl chamber, there exists a unique vector $\tau_\Theta \in \mathbb{S}\mathfrak{a}^+$ such that $\tau_\Theta = w \cdot \tau_0$ for some $w \in W$.

This definition does not depend of the choice of $\alpha \in \Theta$, because if $\beta \in \Theta$ then for some $w_0 \in W$, $w_0 \cdot \beta = \alpha$ and hence any vector τ'_0 orthogonal to $\ker(\beta)$ can be written $\tau'_0 = w_0 \cdot \tau_0$ or $\tau'_0 = (w_0 \sigma_\beta) \cdot \tau_0$. So $W \cdot \tau_0 = W \cdot \tau'_0$. Therefore $W \cdot \tau_0 \cap \mathbb{S}\mathfrak{a}^+ = W \cdot \tau'_0 \cap \mathbb{S}\mathfrak{a}^+ = \{\tau_\Theta\}$. \square

Note that in particular τ_Θ is symmetric.

Remark 2.8 If Θ is a Weyl orbit of simple roots, $\mathcal{F}_{\tau_\Theta}$ is not in general the same flag manifold as $\mathcal{F}_\Theta = G/P_\Theta$.

The *Dynkin diagram* associated to the restricted root system Σ is the graph with vertex set Δ such that for all $\alpha, \beta \in \Delta$ distinct roots there is a link between α and β of multiplicity depending on the order k of $\sigma_\alpha \sigma_\beta$ or $\sigma_\beta \sigma_\alpha$, where $\sigma_\alpha, \sigma_\beta \in W$ are the symmetries associated with the roots α and β . If $k = 2$ we consider that there is no link, there is a simple link if $k = 3$, a double link if $k = 4$ and a triple link if $k = 6$. These are the only cases that occur for spherical Dynkin diagrams. If two roots have different norms, we orient the edge towards the root with largest norm.

Proposition 2.9 Consider the Dynkin diagram associated with the reduced root system Σ , and remove all the double or triple edges. A set $\Theta \subset \Delta$ is a Weyl orbit of simple roots if and only if it is a connected component of this graph.

Proof Let α and β be two simple roots such that $\sigma_\alpha = w \sigma_\beta w^{-1}$ for some $w \in W$. Then the simple root α and $w \cdot \beta$ are proportional, and hence $\alpha = w \cdot \beta$ or $\alpha = -w \cdot \beta$. In any case α and β are conjugated. Reciprocally if α and β are conjugates, then σ_α and σ_β are conjugated.

The system $(W, (\sigma_\alpha)_{\alpha \in \Delta})$ is a Coxeter system. Generators of a Coxeter system are conjugated to one another if and only if there is a path of single edges between the corresponding vertices in the Dynkin diagram [19, Proposition 2.1]. Hence Weyl orbits of simple roots correspond exactly to connected components for the modified Dynkin diagram. \square

We can now describe the Weyl orbits of simple roots in Δ for the restricted system of roots associated to a simple group G . For this we use the classification of the Dynkin diagrams that occur as the reduced root system for a symmetric space \mathbb{X} associated to G .

Corollary 2.10 *If the restricted root system Σ is of type A_n, D_n for $n \geq 2, E_6, E_7$ or E_8 , then the only Weyl orbit of simple roots in Δ is Δ .*

If the root system Σ is of type B_n, C_n for $n \geq 2, F_4$ or G_2 , then Δ can be partitioned into its only two Weyl orbits of simple roots.

Example 2.11 We keep notation from Section 2.5. If $G = \text{PSL}(n, \mathbb{R})$, with previous notation Δ is the only Weyl orbit of simple roots and

$$\tau_\Delta = \frac{1}{2\sqrt{n}} \text{Diag}(1, 0, \dots, 0, -1).$$

The flag manifold \mathcal{F}_{p_Δ} can be identified with

$$\mathcal{F}_{1,n-1} = \{(\ell, H) \mid \ell \subset H \subset \mathbb{R}^n, \dim(\ell) = 1, \dim(H) = n - 1\}.$$

If $G = \text{Sp}(2n, \mathbb{R})$, with previous notation $\Theta = \{\alpha_n\}$ and $\Theta' = \{\alpha_1, \alpha_2, \dots, \alpha_{n-1}\}$ are the two Weyl orbits of simple roots of Δ . One has

$$\tau_\Theta = \frac{1}{2\sqrt{n}} \text{Diag}(1, 0, \dots, 0, -1), \quad \tau_{\Theta'} = \frac{1}{2\sqrt{n}} \text{Diag}(1, 1, 0, \dots, 0, -1, -1).$$

The flag manifold $\mathcal{F}_{\tau_\Theta}$ can be identified with $\mathbb{R}P^{2n-1}$ and $\mathcal{F}_{\tau_{\Theta'}}$ can be identified with the Grassmannian of planes P in \mathbb{R}^{2n} that are isotropic for ω , namely such that $\omega|_P = 0$.

In general, the Weyl orbits of simple roots for any root system are summarized in Figure 4. The table also includes an illustration of the set of roots $\Theta(\tau_\Theta)$ such that $\mathcal{F}_{\Theta(\tau_\Theta)} \simeq \mathcal{F}_{\tau_\Theta}$. The sets of roots are illustrated in the diagram as the set of filled vertices. Using notation from [33, Table 1, page 293], the basis (e_i) is an orthonormal basis such that $e_i^\vee = \epsilon_i$ for B_n, C_n, D_n and F_4 , and $e_i^\vee - (1/(n + 1)) \sum_{k=1}^{n+1} e_k^\vee = \epsilon_i$ for A_n, E_7, E_8 and G_2 . For E_6 , we write $e = \epsilon^\vee$.

The table can be checked as follows: for each Weyl orbit of simple root one can check that the vector τ_Θ is orthogonal to the kernel of a root conjugate to a root in Θ , and lies in the model Weyl chamber. Then one can check that the simple roots that do not vanish on τ_Θ are the ones in $\Theta(\tau_\Theta)$, as depicted in Figure 4.

Δ	Θ	τ_Θ	$\Theta(\tau_\Theta)$
A_n		$(1/\sqrt{2})(e_1 - e_{n+1})$	
BC_2		e_1 $(1/\sqrt{2})(e_1 + e_2)$	
$B_n, n \geq 3$		e_1 $(1/\sqrt{2})(e_1 + e_2)$	
$C_n, n \geq 3$		e_1 $(1/\sqrt{2})(e_1 + e_2)$	
D_4		$(1/\sqrt{2})(e_1 + e_2)$	
$D_n, n \geq 5$		$(1/\sqrt{2})(e_1 + e_2)$	
E_6		$\sqrt{2}e$	
E_7		$(1/\sqrt{2})(e_8 - e_7)$	
E_8		$(1/\sqrt{2})(e_1 - e_9)$	
F_4		$(1/\sqrt{2})(e_1 + e_2)$ e_1	
G_2		$(1/\sqrt{2})(e_1 - e_3)$ $(1/\sqrt{6})(2e_1 - e_2 - e_3)$	

Figure 4: Weyl orbits of simple roots and their associated normalized coroots.

3 Representations of hyperbolic groups

3.1 Gromov hyperbolic groups

Let Γ be a finitely generated group. Let F be any finite generating system for Γ that is symmetric, namely such that $s^{-1} \in F$ for all $s \in F$. We can define the norm of an element $\gamma \in \Gamma$ as

$$|\gamma|_F = \min\{n \mid n = s_1 s_2 \cdots s_n, s_i \in S\}.$$

This norm defines the word distance on Γ by taking $d_F(\gamma_1, \gamma_2) = |\gamma_1^{-1} \gamma_2|_F$ for $\gamma_1, \gamma_2 \in \Gamma$.

A map $f: Y \rightarrow X$ between two metric spaces X and Y is called a *quasi-isometric embedding* if there exist C and D such that for all $x_1, x_2 \in X$,

$$\frac{1}{C} d_Y(x_1, x_2) - D \leq d_X(f(x_1), f(x_2)) \leq C d_Y(x_1, x_2) + D.$$

By extension, we say that a representation ρ is a *quasi-isometric embedding* if some, and hence any, ρ -equivariant map $u_0: \Gamma \rightarrow \mathbb{X}$ is a quasi-isometry, where Γ acts on itself by left multiplication.

This notion does not depend on the choice of F ; indeed if F' is another finite generating system, the identity map $(\Gamma, d_F) \rightarrow (\Gamma, d_{F'})$ is a quasi-isometric embedding.

The group Γ is called *hyperbolic* if, as a metric space, it is hyperbolic in the sense of Gromov. We denote by $\partial\Gamma$ the Gromov boundary of a hyperbolic group Γ , which we equip with the usual topology [20].

Given a discrete representation, we will need to consider the limit cone of the Cartan projections of elements of the group.

Definition 3.1 The *limit cone* of a discrete representation $\rho: \Gamma \rightarrow G$ is the closed subset

$$C_\rho = \bigcap_{n \in \mathbb{N}} \overline{\{[d_{\mathfrak{a}}(o, \rho(\gamma) \cdot o)], |\gamma|_w \geq n\}} = \bigcap_{n \in \mathbb{N}} \overline{\left\{ \frac{d_{\mathfrak{a}}(o, \rho(\gamma) \cdot o)}{d_{\mathbb{X}}(o, \rho(\gamma) \cdot o)}, |\gamma|_w \geq n \right\}} \subset \mathbb{S}\mathfrak{a}^+.$$

Recall that the generalized distance $d_{\mathfrak{a}}$ was defined in Section 2. This definition does not depend on the choice of the basepoint $o \in \mathbb{X}$.

Lemma 3.2 *The limit cone C_ρ of a discrete representation $\rho: \Gamma \rightarrow G$ of a nonelementary hyperbolic group is connected.*

A hyperbolic group is nonelementary if it is neither finite nor virtually cyclic.

Remark 3.3 If ρ is a Zariski dense representation, Benoist proved that C_ρ is convex [4]. We give a proof of the connectedness of C_ρ to avoid considerations of Zariski density.

Proof We define the distance d on $\mathbb{S}\mathfrak{a}$, as $d([x], [y]) = |x/|x| - y/|y||$ for $x, y \in \mathfrak{a}$.

Assume that there exists a partition $A \cup B$ of C_ρ into two open and closed sets. These sets are compact and hence are at uniform distance $\epsilon > 0$. Let F be a finite symmetric generating set for Γ . Let M be the maximum for $\gamma \in F$ of $d_{\mathbb{X}}(o, \rho(\gamma) \cdot o)$. Let $E \subset \Gamma$ be the subset of elements γ such that either $d_{\mathbb{X}}(o, \rho(\gamma) \cdot o) \leq 12M/\epsilon$ or

$$d([d_{\mathfrak{a}}(o, \rho(\gamma) \cdot o)], C_\rho) \geq \frac{1}{3}\epsilon.$$

Since ρ is discrete and by the definition of C_ρ , E is finite. It is included in the ball of radius R centered at the identity of Γ for $|\cdot|_F$ and some $R > 0$.

We say that two elements γ' and γ'' in $\Gamma \setminus E$ are *E-connected* if one can construct a sequence (γ_n) of elements of $\Gamma \setminus E$ such that for all $1 \leq i < N$, $\gamma_{i+1} = a_i \gamma_i b_i$ for some $a_i, b_i \in F$, with $\gamma' = \gamma_0$ and $\gamma'' = \gamma_N$.

Using only right translations, namely $a_i = \text{Id}$, one can see that any point in $\partial\Gamma$ has a neighborhood whose intersection with Γ is *E-connected*. Indeed given $D > 0$ for any $\gamma \in \Gamma$ with $|\gamma|_F > R + D$, the set of elements γ' such that the geodesic from Id to γ' passes at distance at most D of γ is *E-connected*. Any point in the visual boundary admits neighborhood whose intersection with Γ has this form.

Using only left translations, namely $b_i = \text{Id}$, one can see that the intersection of a neighborhood of $\partial\Gamma$ and Γ is *E-connected*. Indeed, the action of Γ on $\partial\Gamma$ is minimal because Γ is nonelementary. Therefore there exist $b_1 \cdots b_n = \gamma \in \Gamma$ with $b_i \in F$ such that $\gamma \cdot x$ lies in the interior of V . Hence any $\gamma' \in V \cap \Gamma$

close enough to $\gamma \cdot x$ is E -connected to $U \cap \Gamma$. Therefore given $x, y \in \partial\Gamma$ with respective neighborhoods U and V whose intersection with Γ is E -connected, the intersection $(U \cup V) \cap \Gamma$ is also E -connected. Since $\partial\Gamma$ is compact one can find a cofinite E -connected subset of Γ .

Since $A, B \neq \emptyset$, the definition of the limit cone allows us to pick two large enough elements $\gamma', \gamma'' \in \Gamma \setminus E$ such that

$$d([\mathbf{d}_\alpha(o, \rho(\gamma') \cdot o)], A) \leq \frac{1}{3}\epsilon, \quad d([\mathbf{d}_\alpha(o, \rho(\gamma'') \cdot o)], B) \leq \frac{1}{3}\epsilon.$$

One can assume that γ' and γ'' are E -connected by taking them large enough. Therefore there exist $\gamma \in \Gamma \setminus E$ and $a, b \in F$ such that

$$d([\mathbf{d}_\alpha(o, \rho(\gamma) \cdot o)], A) \leq \frac{1}{3}\epsilon, \quad d([\mathbf{d}_\alpha(o, \rho(a\gamma b) \cdot o)], B) \leq \frac{1}{3}\epsilon.$$

Using Lemma 2.4 one gets that

$$\begin{aligned} |\mathbf{d}_\alpha(o, \rho(\gamma) \cdot o) - \mathbf{d}_\alpha(o, \rho(a\gamma b) \cdot o)| &\leq |\mathbf{d}_\alpha(o, \rho(\gamma) \cdot o) - \mathbf{d}_\alpha(o, \rho(a\gamma) \cdot o)| + d_{\mathbb{X}}(o, \rho(b) \cdot o), \\ |\mathbf{d}_\alpha(\rho(\gamma)^{-1} \cdot o, o) - \mathbf{d}_\alpha(\rho(\gamma)^{-1} \rho(a)^{-1} \cdot o, o)| &\leq d_{\mathbb{X}}(\rho(a)^{-1} \cdot o, o). \end{aligned}$$

And therefore

$$\left| \frac{\mathbf{d}_\alpha(o, \rho(\gamma) \cdot o)}{d_{\mathbb{X}}(o, \rho(\gamma) \cdot o)} - \frac{\mathbf{d}_\alpha(o, \rho(a\gamma b) \cdot o)}{d_{\mathbb{X}}(o, \rho(a\gamma b) \cdot o)} \right| < \frac{1}{3}\epsilon.$$

This contradicts the fact that A and B are at distance ϵ , so \mathcal{C}_ρ is connected. □

The assumption that Γ is nonelementary is necessary; see the end of Section 5.4.

3.2 Anosov representations

The Anosov properties are more restrictive for a representation than the property of being a quasi-isometric embedding. These notions are interesting in high rank because the Anosov properties hold for an open set of representations, whereas the property of being a quasi-isometric embedding is not necessarily open in $\text{Hom}(\Gamma, G)$ when the rank of \mathbb{X} is at least 2.

Definition 3.4 [6, Section 4] Let $\Theta \subset \Delta$ be a nonempty set of simple roots. A representation $\rho: \Gamma \rightarrow G$ is Θ -Anosov if for every root $\alpha \in \Theta$ there exists some constants $b, c > 0$ such that for every $\gamma \in \Gamma$

$$\alpha(\mathbf{d}_\alpha(o, \rho(\gamma) \cdot o)) \geq b|\gamma|_F - c.$$

This definition does not depend on the choice of the generating set F and the basepoint o .

A Δ -Anosov representation in the case when G is a split real simple Lie group is called a *Borel–Anosov* representation.

Remark 3.5 A representation is P -Anosov for a parabolic subgroup P if it is Θ -Anosov for the corresponding set of simple roots $\Theta \subset \Delta$.

Let $\alpha \in \Delta$ and $\tau_0 \in \text{Sa}$ be orthogonal to $\text{Ker}(\alpha)$. The evaluation of α to $\mathbf{d}_\alpha(x, y)$ satisfies

$$\alpha(\mathbf{d}_\alpha(x, y)) = \alpha(\tau_0) d_{\mathbb{X}}(x, y) \cos(\angle(\mathbf{d}_\alpha(x, y), \tau_0)).$$

Anosov representations are necessarily quasi-isometric embeddings. Reciprocally a quasi-isometric embedding is $\{\alpha\}$ -Anosov if and only if the angle $\langle \mathbf{d}_a(o, \rho(\gamma) \cdot o), \tau_0 \rangle$ is not too small in absolute value for $\gamma \in \Gamma$ large enough.

In particular we have the following characterization of Anosov representations:

Theorem 3.6 [29] *A representation $\rho: \Gamma \rightarrow G$ is Θ -Anosov for $\Theta \subset \Delta$ if and only if it is a quasi-isometric embedding and if $\text{Ker}(\alpha) \cap \mathcal{C}_\rho = \emptyset$ for all $\alpha \in \Theta$.*

Representations that are Θ -Anosov admit a natural continuous and ρ -equivariant map $\xi_\rho^\Theta: \partial\Gamma \rightarrow \mathcal{F}_\Theta = G/\tau_\Theta$, where $\partial\Gamma$ is the Gromov boundary of Γ .

In the proof of Theorem 7.11 we will use the following results about the boundary maps of Anosov representations. For two points $o, x \in \mathbb{X}$ let $\ell(o, x) \in \partial_{\text{vis}}\mathbb{X}$ be the class of the unique geodesic ray with unit speed starting from o and passing through x .

Theorem 3.7 [6, Section 4] *Let $\rho: \Gamma \rightarrow G$ be a Θ -Anosov representation for a nonempty set $\Theta \subset \Delta$. There exists a unique ρ -equivariant continuous and dynamic preserving map $\xi_\rho^\Theta: \partial\Gamma \rightarrow \mathcal{F}_\Theta$. This map is such that for any $o \in \mathbb{X}$ and any sequence $(\gamma_n)_{n \in \mathbb{N}}$ of elements of Γ converging to $\zeta \in \partial\Gamma$, the Δ -facet containing any limit point of the sequence $(\ell(o, \rho(\gamma_n) \cdot o))_{n \in \mathbb{N}}$ also contains the Θ -facet $\xi_\rho^\Theta(\zeta)$.*

For instance, when $G = \text{PSL}(n, \mathbb{R})$ and if $\Theta = \{\alpha_k\}$, one can associate a partial flag to any point in $\partial_{\text{vis}}\mathbb{X}$. If the representation ρ is $\{\alpha_k\}$ -Anosov, the partial flag associated to any limit point of $(\ell(o, \rho(\gamma_n) \cdot o))_{n \in \mathbb{N}}$ contains the same k -dimensional plane, which will be denoted by $\xi_\rho^k(\zeta)$.

Kapovich, Leeb and Porti also proved a generalization of the Morse lemma. Here is a version of this result. Let us fix any metric on Γ quasi-isometric to a word metric.

Theorem 3.8 [31, Theorem 1.3] *Let $\rho: \Gamma \rightarrow G$ be a Θ -Anosov representation. Let $o \in \mathbb{X}$ be a basepoint. There exists a constant $D > 0$ such that for every $\gamma \in \Gamma$, there exists a geodesic ray $\eta: \mathbb{R}_{>0} \rightarrow \mathbb{X}$ at distance at most D from $\rho(\gamma) \cdot o$ with $\eta(0) = o$, whose class $[\eta] \in \partial_{\text{vis}}\mathbb{X}$ lies in a common Δ -facet with $\xi_\rho^\Theta(\zeta_\gamma)$. Here $\zeta_\gamma \in \partial\Gamma$ is the endpoint of any geodesic ray in Γ starting at the identity and going through γ .*

4 Busemann functions on symmetric spaces

Busemann functions are natural functions on Hadamard manifolds associated to points in the visual boundary. These functions will play a key role in the definition of τ -nearly geodesic immersions, and in the fibration of domains of discontinuity. In this section we prove the main properties of Busemann functions and compute their Hessian.

4.1 Main properties of Busemann functions

Busemann functions can be interpreted as the distance of a point $x \in \mathbb{X}$ to a point a in the visual boundary relative to a basepoint $o \in \mathbb{X}$.

Definition 4.1 The Busemann function associated to $a \in \partial_{\text{vis}} \mathbb{X}$ and based at $o \in \mathbb{X}$ is the map $b_{a,o} : X \rightarrow \mathbb{R}$ that associates to $x \in \mathbb{X}$ the limit

$$\lim_{t \rightarrow +\infty} d_{\mathbb{X}}(x, \gamma(t)) - d_{\mathbb{X}}(o, \gamma(t)),$$

for any geodesic ray $\gamma : \mathbb{R}^+ \rightarrow \mathbb{X}$ in the class of a .

This definition makes sense because \mathbb{X} is a Hadamard manifold [15]. The definition implies that for any $x, o, o' \in X$ and $a \in \partial_{\text{vis}} \mathbb{X}$, the Busemann cocycle holds:

$$(3) \quad b_{a,o'}(x) = b_{a,o}(x) + b_{a,o'}(o).$$

For symmetric spaces, this function can be computed using the generalized Iwasawa decomposition. First we prove that unipotent elements preserve the level sets of Busemann functions:

Lemma 4.2 Let $x, o \in \mathbb{X}$ and $a \in \partial_{\text{vis}} \mathbb{X}$ be two points. Let n be an element of the unipotent subgroup $N_{a,o}$ of G . Then

$$b_{a,o}(n \cdot x) = b_{a,o}(x).$$

Proof The Busemann cocycle implies that $b_{a,o}(n \cdot x) - b_{a,o}(x) = b_{a,x}(n \cdot x)$. This is by definition the limit when $t \rightarrow \infty$ of the difference

$$\begin{aligned} d_{\mathbb{X}}(n \cdot x, \exp(tv_{a,x}) \cdot x) - d_{\mathbb{X}}(x, \exp(tv_{a,x}) \cdot x) \\ = d_{\mathbb{X}}(x, n^{-1} \exp(tv_{a,x}) \cdot x) - d_{\mathbb{X}}(x, \exp(tv_{a,x}) \cdot x) \\ \leq d_{\mathbb{X}}(n^{-1} \exp(tv_{a,x}) \cdot x, \exp(tv_{a,x}) \cdot x). \end{aligned}$$

But since $n \in N_{a,x}$, $\exp(-tv_{a,x})n \exp(tv_{a,x})$ converges to the identity when $t \rightarrow +\infty$, so this distance converges to 0. Hence $b_{a,x}(n \cdot x) = 0$. □

Recall that $v_{a,o}$ is the unit vector in $T_o \mathbb{X}$ pointing towards $a \in \partial_{\text{vis}} \mathbb{X}$. To compute a Busemann function one needs to understand it on maximal flats. Let $x = \exp(w) \cdot o$ for $w \in \mathfrak{p}_o$. Suppose that a, o and x lie in the same flat subspace, namely $[w, v_{a,o}] = 0$. The Busemann function on this Euclidean space is equal to

$$b_{a,o}(x) = -d_{\mathbb{X}}(x, o) \cos(\angle_o(a, x)) = \langle -v_{a,x}, w \rangle_x.$$

Using these facts we can write Busemann functions in the symmetric space \mathbb{X} explicitly. Let $o, x \in \mathbb{X}$ be a basepoint and $a \in \partial_{\text{vis}} \mathbb{X}$.

Corollary 4.3 Let $o, x \in X$ and $a \in \partial_{\text{vis}} \mathbb{X}$. The Busemann function can be computed as

$$b_{a,o}(x) = \langle -v_{a,o}, w \rangle_o,$$

where $w \in \mathfrak{a}_{a,o}$ is given by the generalized Iwasawa decomposition, namely is the unique element such that one can write $x = n \exp(w)k \cdot o$ with $n \in N_{a,o}$ and $k \in K_o$.

Since G acts by isometries on \mathbb{X} , Busemann functions are G -equivariant in the following sense:

Corollary 4.4 *Let $o \in \mathbb{X}$ and $a \in \partial_{\text{vis}}\mathbb{X}$. For any $g \in G$ and any $x \in \mathbb{X}$, $b_{g \cdot a, g \cdot o}(g \cdot x) = b_{a, o}(x)$.*

The gradient of Busemann functions is characterized as follows:

Proposition 4.5 *The gradient of the Busemann function based at any point $o \in \mathbb{X}$ associated to $a \in \partial_{\text{vis}}\mathbb{X}$ is the vector field $(-v_{a, x})_{x \in \mathbb{X}}$ of unit vectors pointing towards a .*

Proof The differential $d_x b_{a, o}$ of $b_{a, x}$ at x associates to an element $w \in \mathfrak{p}_x$ the value $\langle w', -v_{a, x} \rangle_x$, where w' is the projection of w to $\mathfrak{a}_{a, x}$ with respect to the decomposition $\mathfrak{g} = \mathfrak{n}_{a, x} \oplus \mathfrak{a}_{a, x} \oplus \mathfrak{k}_x$. Note that $\mathfrak{n}_{a, x}$ and \mathfrak{k} are orthogonal to $v_{a, o} \in \mathfrak{a}_{a, x}$ with respect to $\langle \cdot, \cdot \rangle_x$. Hence $w = w'$, so the gradient of $b_{a, o}$ at x is $-v_{a, x}$. □

Busemann functions vary smoothly when the base flag varies in a flag manifold.

Lemma 4.6 *For any $o \in \mathbb{X}$ and $\tau \in \text{Sa}^+$ the map $\mathcal{F}_\tau \times \mathbb{X} \rightarrow \mathbb{R}$ given by $(a, x) \mapsto b_{a, o}(x)$ is smooth.*

Proof Let P be the stabilizer of an element $a \in \mathcal{F}_\tau$. By Corollary 4.4, for $g \in G$ and $x, y \in \mathbb{X}$, $b_{g \cdot a_0, o}(y) = b_{a_0, o}(g^{-1} \cdot y) - b_{a_0, o}(g^{-1} \cdot o)$.

Hence the map $G \times \mathbb{X} \rightarrow \mathbb{R}$, $(g, x) \mapsto b_{g \cdot a_0, o}(x)$ is smooth, and defines a smooth map from the quotient $G/P \times X \simeq \mathcal{F}_\tau \times \mathbb{X}$. □

Example 4.7 Let $\mathbb{X} = S_n$ or \mathcal{H}_n , the symmetric space associated with $\text{PSL}_n(\mathbb{K})$ for $\mathbb{K} = \mathbb{R}$ or \mathbb{C} , as in Section 2.5. Let (e_1, \dots, e_n) be a basis of \mathbb{K}^n . The projective space $\mathbb{P}(\mathbb{K}^n)$ can be identified with the G -orbit \mathcal{F}_{τ_1} of the point $a \in \partial_{\text{vis}}\mathbb{X}$ corresponding to the limit point where t goes to $+\infty$ of the geodesic ray:

$$t \mapsto \begin{pmatrix} e^{-t(n-1)} & 0 & \dots & 0 \\ 0 & e^t & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & e^t \end{pmatrix}.$$

The point $a \in \mathcal{F}_\tau \simeq \mathbb{P}(\mathbb{K}^n)$ is identified with the first basis vector since the stabilizer of both points by the respective actions of $\text{PSL}_n(\mathbb{K})$ are equal.

The Busemann function $b_{[v], q_0}$ where $q_0 \in \mathbb{X}$ and $[v] \in \mathbb{P}(\mathbb{K}^n)$ associates the value $\sqrt{(n-1)/n} \log(q(v, v))$ to $q \in \mathbb{X}$, where v is a representative of $[v]$ such that $q_0(v, v) = 1$.

The asymptotic behavior of Busemann functions along geodesic rays is determined by the Tits angle between the endpoints.

Lemma 4.8 *Let $a \in \partial_{\text{vis}}\mathbb{X}$ and $x \in \mathbb{X}$. Let η be a geodesic ray converging to $b \in \partial_{\text{vis}}\mathbb{X}$. Then there exists a constant $C > 0$ such that for all $t \in \mathbb{R}$,*

$$|b_{a,x_0}(\eta(t)) + t \cos(\angle_{\text{Tits}}(a, b))| \leq C.$$

Proof There exists some element $g \in G$ such that $g \cdot a$ and $g \cdot b$ belong to $\partial_{\text{vis}}F$ with F the model flat in \mathbb{X} . Moreover there exists a geodesic ray η' at bounded distance from $g \cdot \eta$ that belongs to the flat F . On the flat subspace F , the Busemann function can be computed:

$$b_{a,\eta'(0)}(\eta'(t)) = -t \cos(\angle_{\text{Tits}}(g \cdot a, g \cdot b)). \quad \square$$

4.2 Computation of the Hessian

We compute here the Hessian of Busemann functions in the symmetric space \mathbb{X} . This computation will be used in the proof of Theorem 5.23.

Lemma 4.9 *Let $a \in \partial_{\text{vis}}\mathbb{X}$, and $x, o \in \mathbb{X}$. The Hessian of the Busemann function $b_{a,o}$ at a point $x \in \mathbb{X}$ is given by the following quadratic form on $T_x\mathbb{X}$:*

$$(4) \quad v \mapsto \langle \sqrt{\text{ad}_{v_{a,x}}^2}(v), v \rangle_x.$$

Here $\sqrt{\text{ad}_{v_{a,x}}^2}$ is the only root of the endomorphism $\text{ad}_{v_{a,x}} \circ \text{ad}_{v_{a,x}}|_{\mathfrak{p}_x} : \mathfrak{p}_x \rightarrow \mathfrak{p}_x$ that is symmetric and semipositive for the scalar product $\langle \cdot, \cdot \rangle_x$.

This quadratic form is semipositive, and vanishes exactly on $\mathfrak{z}(v_{a,x}) \cap \mathfrak{p}_x$. For $v \in (\mathfrak{z}(v_{a,x}) \cap \mathfrak{p}_x)^\perp$, it satisfies

$$(5) \quad \text{Hess}_x(v, v) \geq \|v\|^2 \min_{\alpha \in \Sigma, \alpha(\mu(a)) \neq 0} |\alpha(\mu(v_{a,x}))|.$$

Recall that $\mathfrak{z}(v)$ for $v \in \mathfrak{g}$ is the centralizer in \mathfrak{g} of v .

Remark 4.10 The Hessian of a Busemann function is related to the sectional curvature of the symmetric space. When measured along a tangent plane spanned by two orthogonal unit vectors $v, w \in T_o\mathbb{X} \simeq \mathfrak{p}_o \subset \mathfrak{g}$, the sectional curvature of \mathbb{X} is equal to:

$$\kappa_{v,w} = -\langle [v, [v, w]], w \rangle_o = -\langle \text{ad}_v^2(w), w \rangle_o.$$

This lemma implies that Busemann functions are strictly convex except on flats. This is a more general fact about Hadamard manifolds; see [15].

Proof The Busemann functions with respect to two different basepoints differ only by a constant. Hence we can assume here without any loss of generality that $x = o$.

Let $v \in T_o\mathbb{X}$ be a vector. The generalized Iwasawa decomposition, and the fact that the exponential map is a local diffeomorphism, implies that there exists a neighborhood I of 0 in \mathbb{R} such that for all $t \in I$,

$$(6) \quad \exp(tv) = \exp(n_t) \exp(w_t) \exp(k_t)$$

for $n_t \in \mathfrak{n}_{a,o}$, $w_t \in \mathfrak{a}_{a,x}$ and $k_t \in \mathfrak{t}_x$, and such that the map $t \mapsto (n_t, w_t, k_t)$ is smooth. Let us denote by $(\dot{n}, \dot{w}, \dot{k})$ and $(\ddot{n}, \ddot{w}, \ddot{k})$ the first and second derivative of this map at $t = 0$.

The limited development of order 2 at $t = 0$ of (6) yields

$$\exp(tv) = \exp(\dot{n}t + \frac{1}{2}\ddot{n}t^2) \exp(\dot{w}t + \frac{1}{2}\ddot{w}t^2) \exp(\dot{k}t + \frac{1}{2}\ddot{k}t^2) + o(t^2).$$

But the Baker–Campbell–Hausdorff formula [26] implies that the right side of this equality is equal to

$$\exp(\dot{n}t + \dot{w}t + \dot{k}t + \frac{1}{2}\ddot{n}t^2 + \frac{1}{2}\ddot{w}t^2 + \frac{1}{2}\ddot{k}t^2 + \frac{1}{2}([\dot{n}, \dot{w}] + [\dot{n}, \dot{k}] + [\dot{w}, \dot{k}])t^2 + o(t^2)).$$

Hence we get the following two equalities:

$$v = \dot{n} + \dot{w} + \dot{k}, \quad 0 = \frac{1}{2}\ddot{n} + \frac{1}{2}\ddot{w} + \frac{1}{2}\ddot{k} + \frac{1}{2}([\dot{n}, \dot{w}] + [\dot{n}, \dot{k}] + [\dot{w}, \dot{k}]).$$

However since $v, \dot{w} \in \mathfrak{p}_x$, $\theta_x(\dot{n} + \dot{k}) = -\dot{n} - \dot{k}$. Hence $\dot{k} = -\frac{1}{2}(\dot{n} + \theta_x(\dot{n}))$. This lets us simplify the last part of the previous equation:

$$[\dot{n}, \dot{w}] + [\dot{n}, \dot{k}] + [\dot{w}, \dot{k}] = [\dot{n}, \dot{w}] - \frac{1}{2}[\dot{n}, \theta_x(\dot{n})] - \frac{1}{2}[\dot{w}, \dot{n} + \theta_x(\dot{n})].$$

The metric on \mathbb{X} can be written $\langle \cdot, \cdot \rangle_x = B(\cdot, \theta_x(\cdot))$ on \mathfrak{p}_x with B the Killing form, defined on \mathfrak{g} .

Since $v_{a,x}$ is orthogonal to $\mathfrak{n}_{a,o}$ and \mathfrak{t}_x , $B(v_{a,x}, \ddot{n}) = B(v_{a,x}, \ddot{k}) = 0$. Moreover $[\dot{n}, \dot{w}] \in \mathfrak{n}_{a,o}$ so

$$B(v_{a,x}, [\dot{w}, \dot{n} + \theta_x(\dot{n})]) = B(v_{a,x}, [\dot{n}, \dot{w}]) = 0.$$

In particular one gets

$$\text{Hess}_x(b_{a,o})(v, v) = \langle -v_{a,x}, \ddot{w} \rangle_x = \frac{1}{2}B(-v_{a,x}, [\dot{n}, \theta_x(\dot{n})]) = \frac{1}{2}B([-v_{a,x}, \dot{n}], \theta_x(\dot{n})).$$

Let $\Sigma_a \subset \Sigma$ be the set of roots α such that $\alpha(\mu(a)) \neq 0$. The Lie algebra decomposes into root spaces:

$$\mathfrak{g} = \mathfrak{z}(v_{a,x}) \oplus \bigoplus_{\alpha \in \Sigma_a} \mathfrak{g}_{a,x}^\alpha.$$

Here $\text{Ad}_g(\mathfrak{g}_{a,x}^\alpha)$ for any element $g \in G$ such that $g \cdot v_{a,x} = \mu(v_{a,x})$ is the model root space \mathfrak{g}^α .

The restriction of $\text{ad}_{v_{a,x}}$ on $\mathfrak{g}_{a,x}^\alpha$ is a homothety of ratio $\alpha(\mu(v_{a,x}))$. The vector v can be decomposed in this direct sum:

$$v = v^0 + \sum_{\alpha \in \Sigma_a} v^\alpha.$$

The endomorphism $\sqrt{\text{ad}_{v_{a,x}}^2}$ associates to v the vector

$$\sum_{\alpha \in \Sigma_a} |\alpha(\mu(v_{a,x}))| v^\alpha \in \mathfrak{p}_x.$$

Let Σ_a^+ be the set of roots α such that $\alpha(\mu(v_{a,x})) > 0$. The vector \dot{n} can be expressed as

$$\dot{n} = 2 \sum_{\alpha \in \Sigma_a^+} v^\alpha.$$

Hence, as desired,

$$\begin{aligned} (7) \quad \frac{1}{2} B([-v_{a,x}, \dot{n}], \theta_x(\dot{n})) &= -2 \sum_{\alpha \in \Sigma_a^+} \alpha(\mu(v_{a,x})) B(v^\alpha, \theta_x(v^\alpha)) \\ &= \sum_{\alpha \in \Sigma_a^+} |\alpha(\mu(v_{a,x}))| \langle v^\alpha, v^\alpha \rangle_x = \langle \sqrt{\text{ad}_{v_{a,x}}^2}(v), v \rangle_x. \end{aligned}$$

This is equal to zero if and only if $v = v^0$. □

5 Nearly geodesic immersions

In this section we introduce a local condition for an immersion into the symmetric space of noncompact type \mathbb{X} that generalizes the notion of an immersion with principal curvature in $(-1, 1)$ inside \mathbb{H}^n .

5.1 Curvature bound and Busemann functions

We introduce the key definition of a nearly geodesic immersion, which relies on Busemann functions (see Section 4). Let M be a smooth connected manifold, $u: M \rightarrow \mathbb{X}$ be an immersion, $o \in \mathbb{X}$ a basepoint and let $\tau \in \mathbb{S}\mathfrak{a}^+$ be a unit vector in the model Weyl chamber.

Definition 5.1 An immersion $u: M \rightarrow \mathbb{X}$ is called τ -nearly geodesic if for all $a \in \mathcal{F}_\tau \cup \mathcal{F}_{l(\tau)}$ and $v \in TM$ such that $d(b_{a,o} \circ u)(v) = 0$, the function $b_{a,o} \circ u$ has positive Hessian in the direction v .

The Hessian considered in this definition is computed with the induced metric $u^*g_{\mathbb{X}}$ on M . Recall that $\mathcal{F}_{l(\tau)}$ is the opposite flag manifold to \mathcal{F}_τ for $\tau \in \mathbb{S}\mathfrak{a}^+$.

We will first show that the nearly geodesic condition can be written as a bound on the fundamental form \mathbb{I}_u , depending on the Cartan projection of the surface tangent vectors.

Since the Hessian of a Busemann function $b_{a,o}$ on \mathbb{X} does not depend on o , we will denote it by Hess_{b_a} . Recall that $v_{a,o}$ is the unit vector in $T_o\mathbb{X}$ pointing towards $a \in \partial_{\text{vis}}\mathbb{X}$. The second fundamental form \mathbb{I}_u for $x \in M$ of the immersion u is the difference $u^*\nabla^{\mathbb{X}} - \nabla^M$, where $\nabla^{\mathbb{X}}$ is the Levi-Civita connection on $T\mathbb{X}$ associated to $g_{\mathbb{X}}$ and ∇^M is the Levi-Civita connection on $TM \subset u^*T\mathbb{X}$ associated to the metric $u^*g_{\mathbb{X}}$. The second fundamental form is a symmetric 2-tensor with values in u^*N , where $N \subset T\mathbb{X}$ is the normal tangent bundle to $u(M)$.

Proposition 5.2 An immersion $u: M \rightarrow \mathbb{X}$ is τ -nearly geodesic if and only if for all $y \in M$, for all $a \in \mathcal{F}_\tau \cup \mathcal{F}_{l(\tau)}$ and $v \in T_yM$ such that $\langle du(v), v_{a,u(y)} \rangle_{u(y)} = 0$:

$$(8) \quad \text{Hess}_{b_a}(du(v), du(v)) + \langle \mathbb{I}_u(v, v), v_{a,u(y)} \rangle_{u(y)} > 0.$$

The first term of this expression is always nonnegative, and the second one can be made negative up to changing the sign of $v_{a,u(y)}$, so the first term needs to be positive for the inequality to hold. In particular $\langle du(v), v_{a,u(y)} \rangle_{u(y)} = 0$ implies that $du(v)$ cannot lie in the same flat as $v_{a,u(y)}$.

We will prove a sufficient condition that has a simpler form in Theorem 5.23 when $\tau = \tau_\Theta$ for a Weyl orbit of simple roots Θ .

Proof Let $y \in M$ and $a \in \mathcal{F}_\tau \cup \mathcal{F}_{l(\tau)}$. The function $b_{a,o} \circ u$ is critical at y in the direction $v \in T_y M$ if and only if $\langle du(v), v_{a,u(y)} \rangle_{u(y)} = 0$.

Let $\gamma: \mathbb{R} \rightarrow M$ be a geodesic for the metric $u^*g_{\mathbb{X}}$ on M such that $\gamma(0) = y$ and $\gamma'(0) = v$. The Hessian of $b_{a,o} \circ u$ on M is equal to the derivative at $t = 0$ of the differential of the Busemann function, namely

$$t \mapsto \langle v_{a,u(\gamma(t))}, du(\gamma'(t)) \rangle_{u(\gamma(t))}.$$

The first term, $\langle \nabla_{du(v)}^{\mathbb{X}} v_{a,u(\gamma(t))}, du(\gamma'(t)) \rangle_{u(\gamma(t))}$, is equal to $\text{Hess}_{b_a}(du(v), du(v))$. The second term can be written

$$\nabla_{du(v)}^{\mathbb{X}} du(\gamma') = u_* \nabla_v^M du(\gamma') + \Pi_u(v, v).$$

However, γ is a geodesic, so $\nabla_v^M du(\gamma') = 0$; therefore the Hessian of $b_{a,o} \circ u$ on M in the direction v is equal to

$$\text{Hess}_{b_a}(du(v), du(v)) + \langle \Pi_u(v, v), v_{a,u(y)} \rangle_{u(y)} > 0. \quad \square$$

A consequence of Proposition 5.2 is that the property of being τ -nearly geodesic is locally an open property for the \mathcal{C}^2 -topology, which is the topology associated with the uniform convergence over any compact set of the first two differentials.

Corollary 5.3 *Let $u_0: M \rightarrow \mathbb{X}$ be a τ -nearly geodesic map for some $\tau \in \mathbb{S}\mathfrak{a}^+$. For all compact $K \subset M$, there exists a neighborhood U of u_0 for the \mathcal{C}^2 -topology in the space of \mathcal{C}^2 maps from M to \mathbb{X} and a neighborhood V of τ in $\mathbb{S}\mathfrak{a}^+$ such that for all $\tau' \in V$ and $u \in U$, u satisfies the τ' -nearly geodesic immersion condition on K .*

Let G be the isometry group of the n -dimensional hyperbolic space \mathbb{H}^n for some $n \in \mathbb{N}$ with its usual metric with sectional curvature equal to -1 . We prove that the notion of τ -nearly geodesic immersion generalizes the notion of immersion with principal curvatures in $(-1, 1)$ in \mathbb{H}^n . Principal curvatures are only defined for hypersurfaces, but the following definition allows us to generalize the notion of having bounded principal curvature:

Definition 5.4 An immersion $u: M \rightarrow \mathbb{H}^n$ has *principal curvature in $(-1, 1)$* if and only if for all $v \in TM$, $\|du(v)\|^2 > \|\Pi_u(v, v)\|$.

Since \mathbb{H}^n is a rank-1 symmetric space, $\mathbb{S}\mathfrak{a}^+$ contains a single element.

Proposition 5.5 *An immersion $u: M \rightarrow \mathbb{H}^n$ is nearly geodesic for the only element $\tau \in \mathbb{S}\mathfrak{a}^+$ if and only if u has principal curvature in $(-1, 1)$.*

Proof Let $x \in \mathbb{X} = \mathbb{H}^n$ and $a \in \mathcal{F}_\tau = \mathcal{F}_{\iota(\tau)} = \mathbb{C}\mathbb{P}^1 = \partial\mathbb{H}^n$. For any $w \in T_x\mathbb{X}$, $\text{Hess}_{b_a}(w, w) = \lambda \|w^\perp\|^2$, where w^\perp is the orthogonal projection of w onto the orthogonal in $T_x\mathbb{X}$ of $v_{a,x}$ by Lemma 4.9, with some constant λ which is equal to 1 for the metric of sectional curvature equal to -1 on \mathbb{H}^n (see Remark 4.10).

If u is τ -nearly geodesic, then it is an immersion and for every $y \in M$ and $v \in T_yM$ there exists $a \in \partial\mathbb{H}^n$ such that $v_{a,u(y)}$ is positively colinear with $-\Pi_u(du(v), du(v))$. By Proposition 5.2, and since $v \perp v_{a,u(y)}$, one has

$$\|du(v)\|^2 - \|\Pi_u(v, v)\| > 0.$$

Therefore the principal curvature of u is in $(-1, 1)$.

Conversely if u is an immersion with principal curvatures in $(-1, 1)$, let $a \in \partial\mathbb{H}^n$, $y \in M$ and $v \in T_yM$ be such that $b_{a,o} \circ u$ is critical in the direction v . Hence $v_{a,u(y)}$ is perpendicular to $du(v)$ so $\text{Hess}_{b_a}(du(v), du(v)) = \|du(v)\|^2$. Therefore the fact that u has principal curvature in $(-1, 1)$ implies that the hypothesis of Proposition 5.2 holds, so u is τ -nearly geodesic. \square

In general, the property of being τ -nearly geodesic implies that the surface is regular in the following sense:

Definition 5.6 A tangent vector $v \in T\mathbb{X}$ is called τ -regular if its Cartan projection $\mu(v)$ does not belong to $\bigcup_{w \in W}(w \cdot \tau)^\perp$.

We say that an immersion $u: M \rightarrow \mathbb{X}$ is τ regular if for all $v \in TM$, $du(v)$ is τ -regular.

Being regular, namely having the Cartan projection in the interior of α^+ , and being τ -regular are in general unrelated. However when $\tau = \tau_\Theta$ for a Weyl orbit of simple roots Θ , a τ -regular vector $v \in T\mathbb{X}$ is exactly a Θ -regular vector, namely for all $\alpha \in \Theta$, $\alpha(\mu(v)) \neq 0$.

Proposition 5.7 Let $\tau \in \mathbb{S}\alpha^+$. If u is a τ -nearly geodesic immersion, the tangent vectors $du(v)$ for $v \in TM$ are τ -regular.

Proof Let $v \in T_yM$ for some $y \in M$. Assume that $du(v)$ is not τ -regular, so its Cartan projection is orthogonal to $w \cdot \tau$ for some $w \in W$. Therefore there is a unit vector which lies in a common maximal flat with $du(v)$, and whose Cartan projection is equal to τ . This vector is equal to $v_{a,u(y)}$ for some $a \in \mathcal{F}_\tau \cup \mathcal{F}_{\iota(\tau)}$.

Since $v_{a,u(y)}$ and $du(v)$ are in a common flat, $\text{Hess}_{b_a}(du(v), du(v)) = 0$. One can assume that $\langle \Pi_u(v, v), v_{a,u(y)} \rangle_{u(y)} \leq 0$ up to exchanging a with its symmetric with respect to $u(y)$ which is still in $\mathcal{F}_\tau \cup \mathcal{F}_{\iota(\tau)}$. Moreover since $\langle v_{a,u(y)}, du(v) \rangle_{u(y)} = 0$, this is a contradiction with the criterion from Proposition 5.2, so the immersion u cannot be τ -nearly geodesic. \square

The property of being τ -nearly geodesic is not necessarily satisfied for totally geodesic immersions, but it is satisfied for τ -regular totally geodesic immersions.

Proposition 5.8 A totally geodesic immersion is τ -nearly geodesic if and only if it is τ -regular.

Proof An immersion u is totally geodesic if and only if $\mathbb{I}_u = 0$. If u is a τ -nearly geodesic immersion that is totally geodesic, for every $y \in M$, $v \in T_y M$ and every $a \in \mathcal{F}_\tau$, $y \in M$ and $v \in T_y M$ such that $\langle du(v), v_{a,u(y)} \rangle_{u(y)} = 0$, Proposition 5.2 implies that

$$\text{Hess}_{b_a}(du(v), du(v)) > 0.$$

This implies that for no $a \in \mathcal{F}_\tau$ such that $\langle du(v), v_{a,u(y)} \rangle_{u(y)} = 0$ the vector $v_{a,u(y)}$ lies in a common flat with $du(v)$ by Lemma 4.9. Hence the Cartan projection of $du(v)$ is not orthogonal to $w \in \tau$ for any $w \in W$.

Conversely, if the totally geodesic immersion is τ -regular, $\text{Hess}_{b_a}(du(v), du(v))$ is never equal to 0 for any $y \in M$, $v \in T_y M$ and $a \in \mathcal{F}_\tau$ such that $\langle du(v), v_{a,u(y)} \rangle_{u(y)} = 0$. Since Hess_{b_a} is nonnegative, Proposition 5.2 implies that u is τ -nearly geodesic. \square

5.2 Uniformly nearly geodesic immersions

If the nearly geodesic condition for an immersion is satisfied uniformly, one can prove that the exponential of some multiple of Busemann functions are strictly convex on the image of the immersion.

Definition 5.9 Let $\tau \in \mathbb{S}\mathfrak{a}^+$. An immersion $u: M \rightarrow \mathbb{X}$ is *uniformly* τ -nearly geodesic if there exists $\epsilon > 0$ such that for all $v \in TM$ such that $\|du(v)\| = 1$, one has for all $a \in \mathcal{F}_\tau$ satisfying $v_{a,o} \perp v$:

$$\text{Hess}_{b_a}(du(v), du(v)) + \langle \mathbb{I}_u(v, v), v_{a,o} \rangle_o \geq \epsilon.$$

Remark 5.10 When $\mathbb{X} = \mathbb{H}^n$, being uniformly nearly geodesic for a hypersurface is equivalent to having principal curvature in $(-\lambda, \lambda)$ for some $\lambda < 1$.

Suppose that $M = \tilde{N}$ is the universal cover of a compact smooth manifold N . Let Γ be the fundamental group of N . A ρ -equivariant immersion $u: M \rightarrow \mathbb{X}$ for some representation $\rho: \Gamma \rightarrow G$ which is τ -nearly geodesic is necessarily uniformly τ -nearly geodesic since T^1N is compact.

If we consider a uniformly τ -nearly geodesic immersion u , not only are Busemann functions convex in critical directions, but for some $\lambda > 0$, $e^{\lambda b_{a,o} \circ u}$ is strictly convex on M .

Lemma 5.11 Let $\tau \in \mathbb{S}\mathfrak{a}^+$. Let $u: M \rightarrow \mathbb{X}$ be a uniformly τ -nearly geodesic immersion. For some $\lambda > 0$, for all $a \in \mathcal{F}_\tau$ the function $\exp(\lambda b_{a,o} \circ u)$ has positive Hessian for the metric $u^*(g_{\mathbb{X}})$. Moreover there exists some $\epsilon > 0$ such that for any $a \in \mathcal{F}_\tau$ and any geodesic $\eta: \mathbb{R} \rightarrow M$ the functions $f_\eta = \exp(\lambda b_{a,o} \circ u \circ \eta)$ satisfy $f'' \geq \epsilon f$.

Recall that the metric on M that we consider to define geodesics is the induced metric $u^*(g_{\mathbb{X}})$.

Proof Let $o \in \mathbb{X}$ and let U_τ^ϵ be the compact set of pairs $(v, \mathbb{I}) \in T^1\mathbb{X}_o \times T_o\mathbb{X}$ such that for all $a \in \mathcal{F}_\tau$ satisfying $v_{a,o} \perp v$,

$$\text{Hess}_{b_a}(v, v) + \langle \mathbb{I}, v_{a,o} \rangle_o \geq \epsilon.$$

Let us consider

$$C = \inf_{a \in \mathcal{F}_\tau, (v, \mathbb{I}) \in U_\tau^\xi} \frac{\text{Hess}_{b_a}(v, v) + \langle \mathbb{I}, v_{a,x} \rangle_o}{\langle v_{a,o}, v \rangle_o^2}.$$

This infimum is the infimum of a continuous function taking values in $\mathbb{R} \cup \{+\infty\}$ on a compact set. Indeed the numerator must be strictly positive whenever the denominator vanishes, and the denominator is always positive. Hence $C \in \mathbb{R} \cup \{+\infty\}$.

Let λ be any real number greater than $\max(1 - C, 0)$. Let η be any geodesic in M . Let us write $g = b_{a,o} \circ u \circ \eta$. Note that $g'' \geq C(g')^2$ by definition of C . Therefore

$$(e^{\lambda g})''/e^{\lambda g} = \lambda g'' + \lambda^2(g')^2 \geq (C\lambda + \lambda^2)(g')^2 + (\lambda - C)g'' \geq \lambda(g')^2 \geq 0.$$

Note also that $(e^{\lambda g})''/e^{\lambda g} \geq \lambda g''$. Consider the following quantity:

$$M = \inf_{a \in \mathcal{F}_\tau, (v, \mathbb{I}) \in U_\tau^\xi} \max(\text{Hess}_{b_a}(v, v) + \langle \mathbb{I}, v_{a,x} \rangle_o, \langle v_{a,o}, v \rangle_o^2).$$

Note that $M \leq \max(g'', (g')^2)$. Since $K \subset U_\tau$, this quantity is strictly positive as it is an infimum taken on a compact set of a positive function. Hence the function $f = e^{\lambda g}$ is strictly convex and satisfies $f'' > \lambda M f$. □

5.3 Convexity of a Finsler distance

When $\mathbb{X} = \mathbb{H}^n$, and given $y \in \mathbb{H}^n$, for any nearly geodesic immersion $u: M \rightarrow \mathbb{H}^n$ the function $x \mapsto \exp(d_{\mathbb{H}^n}(u(x), y))$ is strictly convex. However for a general symmetric space of higher rank, the τ -nearly geodesic condition doesn't imply the convexity for the Riemannian metric at critical points.

This leads us to consider a Finsler pseudodistance $d_{\mathbb{X}}^\tau$ on \mathbb{X} associated to an element $\tau \in \mathbb{S}\mathfrak{a}^+$. We show in this section that this pseudodistance satisfies a similar convexity property for any τ -nearly geodesic immersion. This pseudodistance is symmetric when τ is symmetric (namely τ is fixed by the opposition involution ι on $\mathbb{S}\mathfrak{a}^+$) and it is equal to the Riemannian distance when $\text{rank}(\mathbb{X}) = 1$. The convexity of this distance allows us to prove the injectivity and properness of complete τ -nearly geodesic immersions. This Finsler pseudodistance is studied in [28, Section 5].

Let us define, for $\tau_0 \in \mathfrak{a}$,

$$|\tau_0|_\tau = \max_{w \in W} \langle w \cdot \tau, \tau_0 \rangle.$$

The map $\tau_0 \mapsto |\tau_0|_\tau$ is nonnegative, homogeneous and subadditive, thus we call it in general a pseudonorm.

This pseudonorm is not necessarily symmetric: $|\tau_0|_\tau = |-\tau_0|_{\iota(\tau)}$. In particular it is symmetric if and only if τ is symmetric. Figure 5 illustrates the unit ball of this norm in \mathfrak{a} for two examples of semisimple Lie groups whose associated symmetric space has rank 2: on the left $G = \text{SL}(3, \mathbb{R})$ and $\tau = \tau_\Delta$, in the middle $G = \text{SL}(3, \mathbb{R})$ and $\tau = \tau_1$ (such that $\mathcal{F}_{\tau_1} \simeq \mathbb{R}\mathbb{P}^2$) and on the right $G = \text{Sp}(4, \mathbb{R})$ and $\tau = \tau_{\{\alpha_n\}}$ with the notation from Section 2.5.

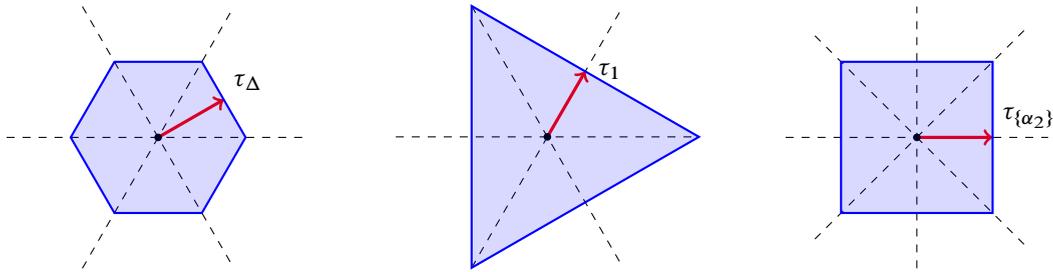


Figure 5: The unit ball in \mathfrak{a} of $|\cdot|_{\tau_\Delta}$ and $|\cdot|_{\tau_1}$ for $G = \text{SL}(3, \mathbb{R})$ and $|\cdot|_{\tau_{\{\alpha_2\}}}$ for $G = \text{Sp}(4, \mathbb{R})$.

This pseudonorm isn't necessarily positive on nonzero vectors. However if a nonzero vector $v \in \mathfrak{a}$ has zero norm, the Weyl group does not act irreducibly on \mathfrak{a} since the W -orbit of τ is orthogonal to v , which means that the underlying Lie group G is not simple.

Example 5.12 Let $n \geq 2$ be an integer, $G = \text{PSL}(2, \mathbb{R})^n$ and $\mathbb{X} = (\mathbb{H}^2)^n$. A model flat in \mathbb{X} is the product of a geodesic in each of the n copies of \mathbb{H}^2 . Let τ_k be the tangent vector to the geodesic on the k^{th} copy of \mathbb{H}^2 . The pseudodistance on \mathbb{X} defined by the pseudonorm $|\cdot|_{\tau_k}$ is the distance in \mathbb{H}^2 of the k^{th} components, which is not a distance on \mathbb{X} .

However if G is simple and τ is symmetric $|\cdot|_\tau$ is a norm. This norm is W -invariant and hence it defines a G -invariant not necessarily symmetric Finsler metric on \mathbb{X} such that for $v \in T\mathbb{X}$, $\|v\|_\tau = |\mu(v)|_\tau$, where μ is the Cartan projection [34, Theorem 6.2.1].

For any semisimple Lie group G , we denote by $d_{\mathbb{X}}^\tau: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_{\geq 0}$ the corresponding pseudodistance on \mathbb{X} , namely $d_{\mathbb{X}}^\tau(x, y)$ is the infimum for all piecewise \mathcal{C}^1 -paths η from x to y of

$$\int \|\eta'\|_\tau = \int |\mu(\eta')|_\tau.$$

This distance can be characterized in terms of Busemann functions.

Proposition 5.13 Let $x, y \in X$ be two points. The pseudodistance $d_{\mathbb{X}}^\tau$ between these two points satisfies

$$d_{\mathbb{X}}^\tau(x, y) = \max_{a \in \mathcal{F}_\tau} b_{a,x}(y).$$

Proof Let $o \in \mathbb{X}$, $v \in T_o\mathbb{X}$ and $a \in \mathcal{F}_\tau$. As usual $v_{a,o}$ is the unit vector based at o pointing towards a . The maximum for $a \in \mathcal{F}_\tau$ of $\langle v, v_{a,o} \rangle$ is reached when v and $v_{a,o}$ are in a common flat [15, Proposition 24].

If we assume that v and $v_{a,o}$ are in a common flat, the maximum is equal to $|\mu(v)|_\tau$. Given two points $x, y \in \mathbb{X}$, any piecewise \mathcal{C}^1 curve η such that $\eta(0) = x$ and $\eta(1) = y$ satisfies, for all $a \in \mathcal{F}_\tau$,

$$(b_a \circ \eta)' = \langle v_{a,\eta(t)}, \eta'(t) \rangle_{\eta(t)} \leq \|\eta'\|_\tau.$$

Hence

$$b_{a,x}(y) \leq \int \|\eta'\|_\tau.$$

Moreover equality is reached for the Riemannian geodesic such that $\eta(0) = x$ and $\eta(1) = y$. Indeed there is a point $a \in \mathcal{F}_\tau$ that lies in a common flat with x and y such that $|\eta'(t)|_\tau = \langle \eta'(t), v_{a,\eta(t)} \rangle_{\eta(t)}$ for all $t \in [0, 1]$. Hence $b_{a,x}(y) = d_{\mathbb{X}}^\tau(x, y)$. Note that the curve reaching this minimum is not unique in general. □

This pseudodistance satisfies the desired convexity condition:

Proposition 5.14 *Let $u: M \rightarrow \mathbb{X}$ be a uniformly τ -nearly geodesic immersion. There exists $\lambda > 0$ such that for all $x \in \mathbb{X}$ the following function is strictly convex for the metric $u^*g_{\mathbb{X}}$:*

$$f: y \in M \mapsto \exp(\lambda d_{\mathbb{X}}^\tau(x, u(y))).$$

A strictly convex continuous function on the Riemannian manifold $(M, u^*(g_{\mathbb{X}}))$ is a function that is strictly convex on any geodesic.

Proof By Lemma 5.11 there exists $\lambda > 0$ such that for any $a \in \mathcal{F}_\tau$, the function $\exp(\lambda b_{a,o} \circ u)$ is strictly convex on M . One can then write f as

$$f(y) = \exp(\lambda d_{\mathbb{X}}^\tau(x, u(y))) = \sup_{a \in \mathcal{F}_\tau} \exp(b_{a,x} \circ u(y)).$$

Hence f is the supremum of a family of convex functions, so it is convex. Moreover the supremum is taken over a compact family of strictly convex functions, so it is strictly convex. □

A consequence of the convexity of this Finsler distance is that the immersion u is injective, which is an interesting property of τ -nearly geodesic surfaces. We say that u is *complete* if M is complete for the induced metric $u^*(g_{\mathbb{X}})$.

Proposition 5.15 *Let $u: M \rightarrow \mathbb{X}$ be a complete uniformly τ -nearly geodesic immersion. Then u is an embedding.*

Proof Consider $y_0 \in M$. The function $y \in M \mapsto \exp(\lambda d_{\mathbb{X}}^\tau(u(y), u(y_0)))$ is strictly convex for some $\lambda > 0$ and admits a minimum at $y = y_0$. The completeness of the metric $u^*(g_{\mathbb{X}})$ implies that there is a geodesic joining any two points. Hence the minimum of any strictly convex function is unique, so u is injective. Therefore it is an embedding. □

Moreover the immersion u cannot be too distorted: the metric induced by u is quasi-isometric to the ambient metric on \mathbb{X} . The notion of quasi-isometric embedding was recalled in Section 3.

Proposition 5.16 *Let $u: M \rightarrow \mathbb{X}$ be a complete uniformly τ -nearly geodesic immersion. Then u is a quasi-isometric embedding for the induced metric $u^*g_{\mathbb{X}}$ on M . In particular u is proper.*

Proof Let $y_0 \in M$ and let $o = u(y_0)$. Let $\epsilon > 0$ and $\lambda > 0$ be the constants provided by Lemma 5.11. Let $\gamma: \mathbb{R}_{\geq 0} \rightarrow M$ be a geodesic ray parametrized with unit speed in M for the metric $u^*g_{\mathbb{X}}$ with $\gamma(0) = y_0$.

Let $a \in \mathcal{F}_\tau$ be such that $b_{a,o}(u \circ \gamma(1)) = d_{\mathbb{X}}^\tau(o, u \circ \gamma(1))$. Consider the function

$$f : t \in \mathbb{R}_{\geq 0} \mapsto \exp(\lambda b_{a,u \circ \gamma(t)}(u \circ \gamma(t))).$$

It is strictly convex and satisfies $f(1) \geq f(0)$ so $f'(1) \geq 0$. Moreover $f'' > \epsilon f$, so $f(t) \geq \cosh(\epsilon(t-1)) > \frac{1}{2}e^{\epsilon(t-1)}$. In particular,

$$d_{\mathbb{X}}^\tau(o, u \circ \gamma(t)) \geq b_{a,o}(u \circ \gamma(t)) \geq \frac{\epsilon}{\lambda}(t-1) - \frac{\log(2)}{\lambda}.$$

For all $y \in M$ there exists a geodesic ray γ passing through y . If d_u is the Riemannian distance on M induced by $u^*g_{\mathbb{X}}$,

$$d_{\mathbb{X}}^\tau(u(x_0), u(x)) \geq \frac{\epsilon}{\lambda}(d_u(x, y) - 1) - \frac{\log(2)}{\lambda}.$$

This Finsler metric is equivalent to the Riemannian metric $g_{\mathbb{X}}$ if G is simple, and in general it is dominated by the Riemannian metric. Moreover u is 1-Lipschitz with respect to the induced metric, so u is a quasi-isometric embedding. □

Using the convexity of this Finsler pseudodistance one can define a continuous projection from the whole symmetric \mathbb{X} to M . This projection will not be used in what follows, but the fibration of the domains in \mathcal{F}_τ constructed in Section 7 is an extension of it.

Proposition 5.17 *Let $u : M \rightarrow \mathbb{X}$ be a complete uniformly τ -nearly geodesic immersion. For every $x \in \mathbb{X}$, there exists a unique point $\pi_u^\tau(x) \in M$ that minimizes*

$$y \in M \mapsto d_{\mathbb{X}}^\tau(x, u(y)).$$

The function $\pi_u^\tau : \mathbb{X} \rightarrow M$ is continuous, and $\pi_u^\tau(u(y)) = y$ for $y \in M$.

Proof Let λ, ϵ be the two constants provided by the Lemma 5.11 and let $x \in \mathbb{X}$. The following function is strictly convex on M :

$$y \mapsto \exp(\lambda d_{\mathbb{X}}^\tau(x, u(y))).$$

It is moreover proper since u is proper by Proposition 5.16. Hence it has a unique minimum, so π_u^τ is well defined.

If we consider a sequence $(x_n) \in \mathbb{X}$ of points that converge to $x \in \mathbb{X}$, then the sequence $(\pi_u^\tau(x_n))$ is bounded since ρ is discrete. Moreover any of its limit points is a minimum of $d_{\mathbb{X}}^\tau(x, u(y))$, so the sequence converges to $\pi_u^\tau(x)$. The function π_u^τ is hence continuous. □

5.4 Anosov property for nearly Fuchsian representations

Let N be a compact manifold with fundamental group Γ . We call a representation $\rho : \Gamma \rightarrow G$ that admits a τ -nearly geodesic equivariant immersion $u : \tilde{N} \rightarrow \mathbb{X}$ a τ -nearly Fuchsian representation.

Proposition 5.18 *The set of τ -nearly Fuchsian representations is open in the space of representations $\rho : \Gamma \rightarrow G$, for the compact-open topology.*

Proof One can continuously deform any ρ -equivariant immersion $u: \tilde{N} \rightarrow \mathbb{X}$ to a ρ' -equivariant smooth map $u': \tilde{N} \rightarrow \mathbb{X}$ for ρ' close to ρ . Indeed fix a Riemannian metric on N , and let $\eta: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a smooth function that is positive on $[0, R]$ for R large enough and vanishes on $[R', +\infty)$ for some $R' > R$. One can define $u'(y)$ for $y \in \tilde{N}$ as the barycenter of the points $x_\gamma^y = \rho'(\gamma^{-1}) \cdot u(\gamma \cdot y)$ with weight $\lambda_\gamma^y = \eta(d(y, \gamma \cdot y))$ for $\gamma \in \Gamma$. Concretely this means that we consider the unique local minimum of the convex function

$$D: x \in \mathbb{X} \mapsto \sum_{\gamma \in \Gamma} \lambda_\gamma^y d(x, x_\gamma^y)^2.$$

Note that $\rho'(\gamma_0) \cdot x_\gamma^y = x_{\gamma\gamma_0^{-1}}^y$ and $\lambda_\gamma^y = \lambda_{\gamma\gamma_0^{-1}}^y$ for $\gamma_0 \in \Gamma$. Therefore u' is ρ' -equivariant. Since \mathbb{X} is a Hadamard manifold D is strictly convex, so the barycenter map is well defined and smooth. Therefore for ρ' close enough to ρ , u' is an immersion which is close to u for the C^2 -topology on any compact fundamental domain of the action of Γ on \tilde{N} . In particular u' is a τ -nearly geodesic immersion for ρ close enough to ρ' . □

The condition that u is τ -nearly geodesic is local, but it will imply some coarse property on u and therefore on ρ . Recall that the limit cone \mathcal{C}_ρ was defined in Section 3 (Definition 3.1). Due to flats, Busemann functions are not strictly convex in critical directions on \mathbb{X} . However Busemann functions are strictly convex in critical directions on $u(\tilde{N})$. We deduce that τ -nearly geodesic surfaces must coarsely avoid these flats, which in turn can be interpreted as a property of the limit cone \mathcal{C}_ρ ; see Definition 3.1.

Proposition 5.19 *Let $\rho: \Gamma \rightarrow G$ be a τ -nearly Fuchsian representation. Then*

$$(9) \quad \mathcal{C}_\rho \cap \bigcup_{w \in W} (w \cdot \tau)^\perp = \emptyset.$$

Recall that W is the Weyl group associated to G .

Proof Let $x_0, x \in \tilde{N}$ and $o = u(x_0)$. Let $w \in W$. Then there exist two points $a \in \mathcal{F}_\tau$ and $a' \in \mathcal{F}_{\iota(\tau)}$ that are opposite from o , namely $v_{a,o} = -v_{a',o}$, and such that

$$b_{a,o}(u(x)) = \langle \mathbf{d}_a(o, u(x)), w \cdot \tau \rangle, \quad b_{a',o}(u(x)) = \langle \mathbf{d}_a(o, u(x)), -w \cdot \tau \rangle.$$

This holds for a and a' that lie in a maximal flat containing o and $u(x)$.

Let η be a geodesic parametrized with unit length in \tilde{N} for $u^*(g_{\mathbb{X}})$ such that

$$\gamma(0) = x_0 \quad \text{and} \quad \gamma(d_u(x_0, x)) = x.$$

Let $\lambda, \epsilon > 0$ be the constants given by Lemma 5.11. Consider the function

$$f: t \mapsto \exp(\lambda b_{a,o} \circ u \circ \gamma(t)).$$

Since $v_{a,o} = -v_{a',o}$, up to exchanging a and a' we can assume that $f'(0) \geq 0$. By Lemma 5.11, $f'' > \epsilon f$. Together with the fact that $f(0) = 1$, this implies that for all $t \in [0, d_u(x_0, x)]$,

$$f(t) \geq \cosh(\epsilon t).$$

Hence $\langle \mathbf{d}_a(o, u(x)), w \cdot p \rangle \geq (\epsilon/\lambda)d_u(x_0, x) - \log(2)/\lambda$. Since u is a quasi-isometric embedding, there exist $c, D > 0$ such that for all $x \in \tilde{N}$ the distance for the induced metric $u^*(g_{\mathbb{X}})$ between x and x_0 is at least

$$cd_{\mathbb{X}}(o, u(x)) - D.$$

In conclusion

$$\left\langle \frac{\mathbf{d}_a(o, u(x))}{d_{\mathbb{X}}(o, u(x))}, w \cdot p \right\rangle \geq \frac{c\epsilon}{\lambda} - \frac{\log(2) + \epsilon D}{\lambda d_{\mathbb{X}}(o, u(x))}.$$

Any element of \mathcal{C}_ρ therefore has a scalar product at least $c\epsilon/\lambda > 0$ with $w \cdot \tau$, for any $w \in W$. This implies that the limit cone cannot intersect $\bigcup_{w \in W} (w \cdot \tau)^\perp$. \square

The set $\mathbb{S}\mathfrak{a}^+ \setminus \bigcup_{w \in W} (w \cdot \tau)^\perp$ contains a single connected component if and only if $(w \cdot \tau)^\perp$ is always a wall of the Weyl chamber decomposition of \mathfrak{a} , namely when $\tau = \tau_\Theta$ for a Weyl orbit of simple roots $\Theta \subset \Delta$ (this notion was defined Section 2.6). In this case

$$\mathbb{S}\mathfrak{a}^+ \setminus \bigcup_{w \in W} (w \cdot \tau_\Theta)^\perp = \mathbb{S}\mathfrak{a}^+ \setminus \bigcup_{\alpha \in \Theta} \text{Ker}(\alpha).$$

Hence we get the following:

Theorem 5.20 *Let $\Theta \subset \Delta$ be a Weyl orbit of simple roots. A τ_Θ -nearly Fuchsian representation $\rho: \Gamma \rightarrow G$ is Θ -Anosov.*

In particular, only hyperbolic groups admit τ_Θ -nearly Fuchsian representations; see [6, Theorem 3.2].

If $\tau \in \mathbb{S}\mathfrak{a}^+$ does not correspond to a Weyl orbit of simple roots, let us assume that Γ is a nonelementary Gromov hyperbolic group, so that the limit cone of Γ is connected; see Lemma 3.2.

To a τ -nearly Fuchsian representation $\rho: \Gamma \rightarrow G$ one can associate the connected component σ_ρ^τ which \mathcal{C}_ρ lies inside:

$$\mathbb{S}\mathfrak{a}^+ \setminus \bigcup_{w \in W} (w \cdot \tau)^\perp.$$

To a connected component of this space one can associate a nonempty set $\Theta(\sigma_\rho^\tau)$ of simple roots. Recall that for $\tau_0 \in \mathbb{S}\mathfrak{a}^+$, $\Theta(\tau_0) \subset \Delta$ is the set of simple roots α such that $\alpha(\tau_0) \neq 0$.

Lemma 5.21 *Let $\tau \in \mathbb{S}\mathfrak{a}^+$ and let $\sigma \subset \mathbb{S}\mathfrak{a}^+$ be a connected component of*

$$(10) \quad \mathbb{S}\mathfrak{a}^+ \setminus \bigcup_{w \in W} (w \cdot \tau)^\perp.$$

Let $\Theta(\sigma) \subset \Delta$ be the set of simple roots α such that $\sigma \cap \text{Ker}(\alpha) = \emptyset$. This set is nonempty, and there exists some $\tau_0 \in \sigma$ such that $\Theta(\tau_0) = \Theta(\sigma)$.

In other words there is $\tau_0 \in \sigma$ such that for any simple root $\alpha \in \Delta$, $\alpha(\tau_0) \neq 0$ if and only if for all $\tau_0' \in \sigma$, $\alpha(\tau_0') \neq 0$. Figure 6 illustrates the lines in $\mathbb{S}\mathfrak{a}^+$ corresponding to $\bigcup_{w \in W} (w \cdot \tau)^\perp$ for some $\tau \in \mathbb{S}\mathfrak{a}^+$, as well as some connected component of the complement σ . In this example $\Theta(\sigma)$ contains only one root.

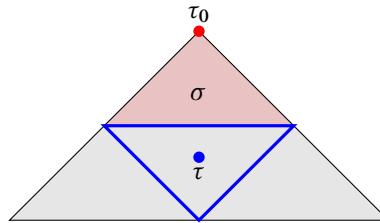


Figure 6: Illustration for $G = \text{PSL}(4, \mathbb{R})$ of a connected component σ of $\mathbb{S}\mathfrak{a}^+ \setminus \bigcup_{w \in W} (w \cdot \tau)^\perp$ in an affine chart.

Proof Let $W_0 \subset W$ be the subgroup of the Weyl group generated by symmetries associated to $\alpha \in \Delta \setminus \Theta(\sigma)$. Let $\hat{\sigma} \subset \mathbb{S}\mathfrak{a}$ be the connected component of σ in

$$\mathbb{S}\mathfrak{a} \setminus \bigcup_{w \in W} (w \cdot \tau)^\perp.$$

This connected component $\hat{\sigma}$ is stabilized by W_0 . Indeed let α be in $\Delta \setminus \Theta(\sigma)$. By definition there is some $v \in \sigma$ such that $\alpha(v) = 0$, and hence that is fixed by the symmetry associated to α . Thus the connected component of v in $\mathbb{S}\mathfrak{a}$ is stabilized by this symmetry, and hence by the group W_0 .

Let $\tau_0 \in \sigma$ be any element and let $\tau_0' \in \mathbb{S}\mathfrak{a}^+$ be the element that, up to the action of W , is positively colinear to

$$\sum_{w \in W_0} w \cdot \tau_0.$$

This sum does not vanish since $\langle \tau, w \cdot \tau_0 \rangle$ has constant sign for $w \in W_0$. This element τ_0' is W_0 -invariant, and hence for all $\alpha \in \Delta \setminus \Theta(\sigma)$, $\alpha(\tau_0') = 0$.

Moreover, since the Lie group G is semisimple, the action of W has no global fixed point on $\mathbb{S}\mathfrak{a}$, and hence $W_0 \neq W$, which proves that $\Theta(\sigma) \neq \emptyset$. □

Theorem 5.22 *A τ -nearly Fuchsian representation $\rho: \Gamma \rightarrow G$ from a nonelementary hyperbolic group Γ is $\Theta(\sigma_\rho^\tau)$ -Anosov.*

Note that $\Theta(\sigma_\rho^\tau) \neq \emptyset$, because of Lemma 5.21.

Proof We use the characterization of Anosov representations from Theorem 3.6. We already proved that τ -nearly Fuchsian representations are quasi-isometric embeddings in Proposition 5.16.

The limit cone \mathcal{C}_ρ lies inside σ_ρ^τ , which avoids $\text{Ker}(\alpha)$ for $\alpha \in \Theta(\sigma_\rho^\tau)$. Hence ρ is $\Theta(\sigma_\rho^\tau)$ -Anosov. □

The nonelementary assumption is necessary; indeed the following representation $\rho: \mathbb{Z} \rightarrow \text{SL}(3, \mathbb{R})$ is not Anosov for any set of roots:

$$n \mapsto \begin{pmatrix} 4^n & 0 & 0 \\ 0 & 2^{-n} & 0 \\ 0 & 0 & 2^{-n} \end{pmatrix}.$$

However, this representation preserves a geodesic which is τ -regular for almost every $\tau \in \mathbb{S}\mathfrak{a}^+$.

5.5 A sufficient condition for an immersion to be nearly geodesic

Let Θ be a Weyl orbit of simple roots as in Section 2.6. Let $\alpha \in \Theta$ be any root. We define the following constant:

$$(11) \quad c_\Theta = \min_{\beta \in \Sigma, \beta(\tau_\Theta) \neq 0} \frac{|\beta(\tau_\Theta)|}{\|\alpha\|^2}.$$

Here $\|\alpha\|$ for $\alpha \in \Theta$ denotes the maximum of $|\alpha(\tau)|$ for $\tau \in \mathbb{S}a$ a unit vector. This quantity is the same for any $\alpha \in \Theta$, since Θ is a Weyl orbit of simple roots.

A sufficient condition for the immersion u to be a τ_Θ -nearly geodesic surface is the following:

Theorem 5.23 *Let $u: S \rightarrow \mathbb{X}$ be an immersion that satisfies, for all $v \in TS$ and $\alpha \in \Theta$,*

$$(12) \quad \|\Pi_u(v, v)\|_{\tau_\Theta} < c_\Theta \alpha(\mu(du(v)))^2.$$

Then u is a τ_Θ -nearly geodesic immersion.

Note that $\|\cdot\|_{\tau_\Theta} < \|\cdot\|$, so having (12) with the Riemannian metric in the left-hand side instead of the Finsler pseudodistance from Section 5.5 is also a sufficient condition.

This property is a generalization of the property of having principal curvature in $(-1, 1)$, where the norm of the tangent vector is replaced by the evaluation of roots of the Cartan projection.

Proof Let us show that (12) implies the condition of Proposition 5.2:

$$\text{Hess}_{b_a}(du(v), du(v)) + \langle \Pi_u(v, v), v_{a,u(y)} \rangle_{u(y)} > 0.$$

Let $x = u(y)$. Let us write $du(v) = w_0 + w^\perp$, where $w_0 \in \mathfrak{z}(v_{a,x}) \cap \mathfrak{p}_x$ and $w^\perp \in (\mathfrak{z}(v_{a,x}) \cap \mathfrak{p}_x)^\perp$. Because of Lemma 4.9, one has

$$(13) \quad \text{Hess}_{b_a}(du(v), du(v)) \geq \|w^\perp\|^2 \min_{\beta \in \Sigma, \beta(\tau_\Theta) \neq 0} |\beta(\tau_\Theta)|.$$

Lemma 2.4 implies that for any $\alpha \in \Sigma$,

$$\alpha(\mu(w_0)) + \|\alpha\| \times \|w^\perp\| \geq \alpha(\mu(du(v))).$$

We assumed that $\langle v_{a,x}, du(v) \rangle_x = 0$. Since $v_{a,x} \in \mathfrak{z}(v_{a,x}) \cap \mathfrak{p}_x$, one has therefore $\langle v_{a,x}, w^\perp \rangle_x = 0$ and hence $\langle v_{a,x}, w_0 \rangle_x = 0$. Moreover $v_{a,x}$ and w_0 are in a common flat. Since $\mu(v_{a,x}) = \tau_\Theta$, this implies that $\alpha(\mu(w_0)) = 0$ for some root $\alpha \in \Theta$. Therefore for this root α ,

$$(14) \quad \|w^\perp\| \geq \|\alpha\|^{-1} \alpha(\mu(du(v))).$$

Recall that for any $w \in T_x \mathbb{X}$ and $a \in \mathcal{F}_{\tau_\Theta}$, $\langle w, v_{a,x} \rangle_x \leq \|w\|_{\tau_\Theta}$, as a consequence of [15, Proposition 24].

Equations (13) and (14) together imply the following inequality, with c_Θ defined in (11):

$$\text{Hess}_{b_a}(du(v), du(v)) + \langle \Pi_u(v, v), v_{a,u(y)} \rangle_{u(y)} \geq c_\Theta \alpha(\mu(du(v)))^2 - \|\Pi_u(v, v)\|_{\tau_\Theta}.$$

The rightmost term is strictly positive because of (12). □

Example 5.24 Let $G = \mathrm{PSL}(n, \mathbb{R})$. We choose the standard metric on \mathbb{X} that comes from the Killing form. In particular the Euclidean metric on \mathfrak{a} is given by

$$\langle \mathrm{Diag}(\lambda_1, \dots, \lambda_n), \mathrm{Diag}(\mu_1, \dots, \mu_n) \rangle = 2n \sum_{i=1}^n \lambda_i \mu_i.$$

In this case $\tau_\Delta = \mathrm{Diag}(1/(2\sqrt{n}), 0, \dots, 0, -1/(2\sqrt{n}))$. The minimum nonzero value of $\beta(\tau_\Theta)$ for $\beta \in \Sigma$ is reached for the root $\alpha_1: \mathrm{Diag}(\lambda_1, \dots, \lambda_n) \mapsto \lambda_1 - \lambda_2$, and is equal to $1/(2\sqrt{n})$ if $n \geq 3$ and $1/\sqrt{n}$ if $n = 2$.

The norm of any root α is equal to the norm of α_1 . But $|\alpha_1(\tau)| \leq |\lambda_1| + |\lambda_2| \leq \sqrt{2} \sqrt{\lambda_1^2 + \lambda_2^2} \leq (1/\sqrt{n}) \|\tau\|$ with equality for some $\tau \in \mathfrak{S}\mathfrak{a}$. Hence $\|\alpha_1\| = 1/\sqrt{n}$, so if $n \geq 3$,

$$c_\Delta = 2\sqrt{n},$$

and $c_\Delta = \sqrt{2}$ if $n = 2$. Note that if we rescale the metric on $\mathbb{X} = \mathbb{H}^2$ so that the sectional curvature is equal to -1 , (12) is exactly the condition of having principal curvature in $(-1, 1)$.

6 Pencils of tangent vectors

In this section we recall the classical notion of a pencil of quadrics, then we generalize it to the notion of a pencil of tangent vectors in a symmetric space of noncompact type and its base in a flag manifold. Bases of pencils appear as the fibers of the fibration that will be constructed in Section 7.

6.1 Pencils of quadrics

Some references for the notion of pencils of quadrics can be found in [17]. Let V be a finite-dimensional vector space over $\mathbb{K} = \mathbb{R}$ or \mathbb{C} .

Definition 6.1 A *pencil of quadrics*, or more precisely a *d-pencil of quadrics*,¹ on V is a linear subspace \mathcal{P} of dimension d in the space $S(V)$ of symmetric bilinear forms on V if $\mathbb{K} = \mathbb{R}$, or in the space $\mathcal{H}(V)$ of Hermitian forms on V if $\mathbb{K} = \mathbb{C}$.

The *base* $b(\mathcal{P})$ of a d -pencil \mathcal{P} is the set of points $[v] \in \mathbb{P}(V)$ such that for all $q \in \mathcal{P}$, $q(v, v) = 0$.

The following is a criterion for a pencil of quadrics to have a smooth base:

Lemma 6.2 Let \mathcal{P} be a pencil of quadrics such that all nonzero $q \in \mathcal{P}$ are nondegenerate bilinear forms. The map $p: V \rightarrow \mathcal{P}^*$ given by $v \mapsto (q \mapsto q(v, v))$ is a submersion at every $v \in V \setminus \{0\}$ such that $[v] \in b(\mathcal{P})$. In particular $b(\mathcal{P})$ is a smooth manifold of codimension d .

Proof Let (q_1, \dots, q_d) be a basis of \mathcal{P} . Let us consider some $v \in V \setminus \{0\}$ such that $q_1(v, v) = \dots = q_d(v, v) = 0$. The kernel of the differential of p is the intersection of the orthogonal spaces $[v]^\perp q_i$ with respect to q_i of the line generated by v for $1 \leq i \leq d$. Since the forms q_i are nondegenerate, these are hyperplanes.

¹In the literature, for instance in [17], a *pencil of quadrics* is often just a 2-pencil of quadrics.

Suppose that their intersection does not have codimension d . In particular the linear forms $q_i(v, \cdot)$ for $1 \leq i \leq d$ are not linearly independent, so there exists a linear combination of the bilinear forms that is degenerate, but is a nonzero element of \mathcal{P} , contradicting our assumption.

Hence the kernel of p has codimension d , so p is a submersion at v . □

The base of a pencil of quadratics is smooth and has codimension d around each of its points which are nonsingular, meaning that they are not degenerate points for any quadric in the pencil. We generalize this notion of singular points in the next section.

6.2 Pencils of tangent vectors in symmetric spaces

In this subsection we consider pencils of tangent vectors in a symmetric space \mathbb{X} of noncompact type, which are related to pencils of quadrics when $G = \text{PSL}(n, \mathbb{R})$.

Definition 6.3 A pencil of tangent vectors at $x \in \mathbb{X}$, or more precisely a d -pencil, is a vector subspace $\mathcal{P} \subset T_x \mathbb{X}$ of dimension d for some point $x \in \mathbb{X}$.

To a pencil one can associate some subsets of any G -orbit in the visual boundary. Recall that for $a \in \partial_{\text{vis}} \mathbb{X}$ and $x \in \mathbb{X}$, the unit vector $v_{a,x} \in T_x \mathbb{X}$ is the unit vector pointing towards a . Let $\tau \in \mathbb{S}\mathfrak{a}^+$.

Definition 6.4 The τ -base of the pencil \mathcal{P} , whose basepoint is $x \in \mathbb{X}$, is the set $\mathcal{B}_\tau(\mathcal{P})$ of elements $a \in \mathcal{F}_\tau$ such that $v_{a,x}$ is orthogonal to \mathcal{P} .

When $G = \text{PSL}(n, \mathbb{R})$ and $\mathbb{X} = \mathcal{S}_n$, a pencil at $q \in \mathcal{S}_n$ corresponds to a subspace \mathcal{P}' of symmetric bilinear forms on \mathbb{R}^n , namely a pencil of quadrics, that is compatible with q in the sense that the trace of the associated q -symmetric matrices vanishes.

Proposition 6.5 Let $\tau \in \mathbb{S}\mathfrak{a}^+$ be such that $\mathcal{F}_\tau \simeq \mathbb{R}\mathbb{P}^{n-1}$. The τ -base of the pencil \mathcal{P} is identified via this identification with the base of the pencil of quadrics \mathcal{P}' .

Proof The τ -base of \mathcal{P} is the space of lines $[C]$ where $C \in \mathbb{R}^n$ is a column vector with $\text{Tr}(CC^{\perp q}M) = 0$ for all $M \in \mathcal{P}'$, since $CC^{\perp q}$ is colinear to $v_{[C],q}$. Hence the τ -base of the pencil is also the set of lines $[C]$ such that $C^{\perp q}MC = 0$, namely the base of the pencil of quadrics \mathcal{P}' . □

We now generalize Lemma 6.2 to general pencils of tangent vectors.

Definition 6.6 A point $a \in \mathcal{F}_\tau$ in the base $\mathcal{B}_\tau(\mathcal{P})$ of a pencil \mathcal{P} at $x \in X$ is called *singular* if for some $w \in \mathcal{P}$ one has $[w, v_{a,x}] = 0$.

We denote by $\mathcal{B}_\tau^*(\mathcal{P}) \subset \mathcal{B}_\tau(\mathcal{P})$ the set of nonsingular points, which we will also call the *regular base*.

Lemma 6.7 Let \mathcal{P} be a pencil of tangent vectors at x in \mathbb{X} . The function which associates to $a \in \mathcal{F}_\tau$ the linear form $v \mapsto \langle v_{a,x}, v \rangle_x$ on \mathcal{P} is a submersion at $a \in \mathcal{B}_\tau(\mathcal{P})$ if and only if $a \in \mathcal{B}_\tau^*(\mathcal{P})$. In particular $\mathcal{B}_\tau^*(\mathcal{P})$ is always a smooth codimension d submanifold of \mathcal{F}_τ .

Proof Let $\phi: \mathcal{F}_\tau \rightarrow \mathcal{P}^*$ be the map that associates to $a \in \mathcal{F}_\tau$ the linear form $v \mapsto \langle v_{a,x}, v \rangle_x$.

Suppose that $a \in \mathcal{B}_\tau(\mathcal{P}) \setminus \mathcal{B}_\tau^*(\mathcal{P})$. Then there exists some $w \in \mathcal{P}$ such that $[w, v_{a,x}] = 0$. The map $\psi: k \in K_x \mapsto k \cdot a \in \mathcal{F}_\tau$ is a submersion, so for every tangent vector in $T_a\mathcal{F}_\tau$ the differential of $a \mapsto v_{a,x}$ in this direction is $\text{ad}_k(v_{a,x})$ for some $k \in \mathfrak{k}_x$. The differential of $a \mapsto \langle v_{a,x}, w \rangle_x$ in this direction is equal to $\langle \text{ad}_k(v_{a,x}), w \rangle_x = -B(\text{ad}_k(v_{a,x}), w) = -B(k, [v_{a,x}, w]) = 0$. Hence the image of the differential of ϕ is not surjective: it is not a submersion.

Suppose that $a \in \mathcal{B}_\tau^*(\mathcal{P})$. Let $v \in \mathcal{P}$ be any nonzero vector and consider $[v_{a,x}, v] = k \in \mathfrak{k}_x$. The differential of $a \mapsto \langle v_{a,x}, v \rangle_x$ in the corresponding tangent direction is equal to $\langle \text{ad}_k(v_{a,x}), v \rangle_x = -B(\text{ad}_k(v_{a,x}), v) = -B([v_{a,x}, v], [v_{a,x}, v]) \neq 0$. Since for all $v \in \mathcal{P}$ there is a direction in which the differential of $a \mapsto \langle v_{a,x}, v \rangle_x$ does not vanish, the map ϕ is therefore a submersion at a . □

A pencil of tangent vectors \mathcal{P} at $x \in \mathbb{X}$ is called τ -regular if all its nonzero vectors are τ -regular as in Definition 5.6. In particular a τ -regular pencil satisfies $\mathcal{B}_\tau^*(\mathcal{P}) = \mathcal{B}_\tau(\mathcal{P})$, so the τ -base of a τ -regular vector is a smooth codimension- d submanifold of \mathcal{F}_τ .

Because of Lemma 6.7, the topology of the base of a regular pencil does not vary if the pencil is deformed continuously.

Corollary 6.8 *Let \mathcal{P}_0 and \mathcal{P}_1 be two pencils at $x \in \mathbb{X}$ in the same connected component of the space of τ -regular pencils at x . Then $\mathcal{B}_\tau(\mathcal{P}_1)$ and $\mathcal{B}_\tau(\mathcal{P}_2)$ are diffeomorphic.*

Proof Since the space of regular pencils is an open subset of the Grassmannian of planes in $T_x\mathbb{X}$, there exists a smooth path $(\mathcal{P}_t)_{t \in [0,1]}$ of regular pencils between \mathcal{P}_0 and \mathcal{P}_1 . Because of Lemma 6.7 the set $\{(a, t) \mid a \in \mathcal{F}_\tau, t \in [0, 1]\}$ is a submanifold with boundary $\mathcal{F}_\tau \times [0, 1]$ that comes with a natural submersion $(a, t) \mapsto t$. Since this manifold is compact, all the fibers are diffeomorphic by the Ehresmann fibration theorem. □

Example 6.9 Let $G = \text{PSL}(3, \mathbb{R})$ and $\mathbb{X} = \mathcal{S}_3$. We identify the tangent space $T_{q_0}\mathcal{S}_3$ at the point q_0 corresponding to the standard scalar product on \mathbb{R}^3 with the space of three-by-three symmetric matrices with real coefficients and zero trace. Consider the following two pencils:

$$\mathcal{P}_{\text{irr}} = \left\langle \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix} \right\rangle, \quad \mathcal{P}_{\text{red}} = \left\langle \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right\rangle.$$

Let $\tau_1 \in \text{Sa}^+$ be such that \mathcal{F}_{p_1} is diffeomorphic in a $\text{PSL}(3, \mathbb{R})$ -equivariant way to \mathbb{RP}^2 , and let τ_Δ be the normalized coroot associated to the Weyl orbit of simple roots Δ . It satisfies $\mathcal{F}_{\tau_\Delta} \simeq \mathcal{F}_{1,2}$, the space of complete flags in \mathbb{R}^3 .

The pencils \mathcal{P}_{irr} and \mathcal{P}_{red} are not τ_1 -regular: $\mathcal{B}_{\tau_1}(\mathcal{P}_{\text{irr}})$ is the disjoint union of a point and a line where $\mathcal{B}_{\tau_1}^*(\mathcal{P}_{\text{irr}})$ contains only the point. In this particular case the regular base is a connected component of the

base, so it is a smooth compact codimension-2 submanifold. The set $\mathcal{B}_{\tau_1}(\mathcal{P}_{\text{red}})$ is a single point that is singular for the pencil. Here we see that a singular point can still be a point around which the base is a smooth codimension-2 submanifold.

Both pencils are τ_Δ -regular, but their τ_Δ -bases are different.

A flag $(\ell, H) = ([x], [y]^\perp)$ with $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ nonzero vectors such that $x_1^2 + x_2^2 + x_3^2 = y_1^2 + y_2^2 + y_3^2$ and $x_1y_1 + x_2y_2 + x_3y_3 = 0$ belongs to $\mathcal{B}_{\tau_\Delta}(\mathcal{P}_{\text{red}})$ if and only if

$$x_1^2 - x_3^2 = y_1^2 - y_3^2, \quad 2x_1x_3 = 2y_1y_3.$$

Up to replacing y by $-y$, these equations are equivalent to $x_1 = y_1, x_2 = -y_2, x_3 = y_3$ and $x_1^2 + x_3^2 = x_2^2$. The corresponding flags (ℓ, H) in the affine chart $(x, y) \mapsto [x, 1, y]$ of $\mathbb{R}\mathbb{P}^2$ are the tangent point and tangent lines to the circle of radius 1 centered at the origin.

A flag $(\ell, H) = ([x], [y]^\perp)$ with $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ nonzero vectors such that $x_1^2 + x_2^2 + x_3^2 = y_1^2 + y_2^2 + y_3^2$ and $x_1y_1 + x_2y_2 + x_3y_3 = 0$ belongs to $\mathcal{B}_{\tau_\Delta}(\mathcal{P}_{\text{irr}})$ if and only if

$$2x_1^2 - 2x_3^2 = 2y_1^2 - 2y_3^2, \quad 2x_1x_2 + 2x_2x_3 = 2y_1y_2 + 2y_2y_3.$$

Let $\ell_0 = \langle(1, 0, 1)\rangle$ and $H_0 = \langle(1, 0, -1), (0, 1, 0)\rangle$. The corresponding flags (ℓ, H) belong to one of the three circles in $\mathcal{F}_{1,2}$ defined by

- $\ell = \ell_0$ and H is any plane through ℓ ,
- $H = H_0$ and ℓ is any line in H ,
- $\ell \subset H_0$ and $\ell_0 \subset H$.

Indeed one can check that these flags satisfy the equations. In order to check that these are the only solutions, one can see that these are the fibers of a fibration over the surface with three connected components; see Section 8.3.2.

The τ_Δ -base $\mathcal{B}_{\tau_\Delta}(\mathcal{P}_{\text{red}})$ is a circle, whereas $\mathcal{B}_{\tau_\Delta}(\mathcal{P}_{\text{irr}})$ is the union of three circles. Hence Corollary 6.8 implies that they must lie in different connected components of the space of τ_Δ -regular pencils.

Since the pencils will be the fibers of the domains of discontinuity that we will construct, proving that the domain is nonempty will be equivalent to having nonempty pencils. We present here a topological argument to prove that some pencils are nonempty:

Proposition 6.10 *Let $\tau \in \mathbb{S}\mathfrak{a}^+$ and \mathcal{P} be a τ -regular pencil of tangent vectors based at $x \in \mathbb{X}$ of dimension d . If the τ -base of \mathcal{P} is empty, then \mathcal{F}_τ fibers over the sphere S^{d-1} .*

If moreover $d = 2$, the fundamental group of \mathcal{F}_τ is infinite.

Proof To $a \in \mathcal{F}_\tau$ we associate $\pi_0(a) \in \mathcal{P}$ the orthogonal projection of $v_{a,x} \in T_x\mathbb{X}$ onto $\mathcal{P} \subset T_x\mathbb{X}$. Since the τ -base of \mathcal{P} is empty, one can define a map $\pi: \mathcal{F}_\tau \rightarrow \mathbb{S}\mathcal{P}$ into the unit sphere of \mathcal{P} where $\pi(a) = \pi_0(a)/\|\pi_0(a)\|$. This map is a submersion. Indeed let $a \in \mathcal{F}_\tau$, and let $\mathcal{P}_0 \subset \mathcal{P}$ be the orthogonal to $\pi(a)$ in \mathcal{P} . Lemma 6.7 applied to \mathcal{P}_0 implies that π is a submersion at a .

This submersion is proper since \mathcal{F}_τ is compact, and hence it is a fibration. This fibration induces a long exact sequence, where F is the fiber:

$$\cdots \rightarrow \pi_1(\mathcal{F}_\tau) \rightarrow \pi_1(\mathcal{SP}) \rightarrow \pi_0(F) \rightarrow \cdots.$$

Since F is compact, $\pi_0(F)$ is finite and if $d = 2$, $\pi_1(\mathcal{SP}) \simeq \mathbb{Z}$, so $\pi_1(\mathcal{F}_\tau)$ is infinite. □

We conclude this section by the following remark, which says that regular pencils cannot be tangent to flats:

Proposition 6.11 *If a 2-pencil is tangent to a flat, then it is not τ -regular for any $\tau \in \mathbb{S}\mathfrak{a}^+$.*

Proof Up to the action of G one can identify \mathcal{P} with a plane in \mathfrak{a} . But for any $\tau \in \mathbb{S}\mathfrak{a}^+$, the orthogonal of τ intersects this plane. Hence there is an element of \mathcal{P} whose Cartan projection is orthogonal to $w \cdot \tau$ for some w in the Weyl group. □

7 Fibered domains in flag manifolds

In this section we associate an open domain $\Omega_u^\tau \subset \mathcal{F}_\tau$ to any complete uniformly τ -nearly geodesic immersion $u: M \rightarrow \mathbb{X}$ with $\tau \in \mathbb{S}\mathfrak{a}^+$, and show that this domain is a smooth fiber bundle over M where the fibers are τ -bases of the pencils that are the tangent planes to $u(M)$. This construction is the analog of the *Gauss map* for hypersurfaces in \mathbb{H}^n . We also mention what happens with our construction for totally geodesic immersions that are not τ -regular.

If $M = \tilde{N}$ for some compact manifold N with torsion-free fundamental group Γ , and if u is equivariant with respect to a representation ρ , we show that the domain Ω_u^τ is a cocompact domain of discontinuity for the action of ρ and its quotient fibers over N . This domain always coincides with some domain of discontinuity associated to the Tits–Bruhat ideals constructed by Kapovich, Leeb and Porti [30]. Finally we prove the invariance of the topology of the quotients of these domains of discontinuity.

7.1 A domain associated to a nearly geodesic immersion

Let $\tau \in \mathbb{S}\mathfrak{a}^+$ be any unit vector and $u: M \rightarrow \mathbb{X}$ be a complete uniformly τ -nearly geodesic immersion. We consider a particular domain of the flag manifold \mathcal{F}_τ , defined for any nearly geodesic immersion $u: M \rightarrow \mathbb{X}$ using Busemann functions. For this we fix a basepoint $o \in \mathbb{X}$, but the definition will not depend on this choice.

Definition 7.1 Let Ω_u^τ be the set of elements $a \in \mathcal{F}_\tau$ such that the function $b_{a,o} \circ u$ is proper and bounded from below.

We have additional properties if u is a complete uniformly τ -nearly geodesic immersion.

Lemma 7.2 *Let $a \in \mathcal{F}_\tau$. There exists a critical point $x \in M$ for the function $b_{a,o} \circ u$ if and only if $a \in \Omega_u^\tau$. In this case this point is unique, and the Hessian of $b_{a,o} \circ u$ at this point is positive. The domain Ω_u^τ is open.*

Proof Let $a \in \mathcal{F}_\tau$. Suppose that $b_{a,o} \circ u$ is critical at $y \in M$. Since u is τ -nearly geodesic, the Hessian of $b_{a,o} \circ u$ at y is positive. Moreover, due to Lemma 5.11 there exists $\lambda > 0$ such that $\exp(\lambda b_{a,o} \circ u)$ has positive Hessian everywhere on M .

A convex function with positive Hessian on a complete connected Riemannian manifold has a unique minimum, and is proper. The function $\exp(\lambda b_{a,o} \circ u)$, and hence the function $b_{a,o} \circ u$, are therefore proper and have a unique minimum. In particular $a \in \Omega_\rho^\tau$.

Conversely if $a \in \Omega_\rho^\tau$, $b_{a,o} \circ u$ is proper it admits a global minimum, which is a critical point.

If a function has a critical point with positive Hessian, every small deformation of the function for the \mathcal{C}^0 -topology still admits a critical point. Indeed if x is a critical point with positive Hessian of a convex function f , for any small enough closed disk D around x one has $f > f(x)$ on ∂D . If we fix such a disk D , for g close enough in the \mathcal{C}^0 -topology to f , one still has $g > g(x)$ on the compact ∂D . Therefore g restricted to D must admit a global minimum that is reached at some point y in the interior of D . This point y is a local minimum and hence a critical point of g .

In conclusion Ω_ρ^τ is open. □

We thus can define the projection $\pi_u: \Omega_u^\tau \rightarrow M$ associated to u as the map that associates to $a \in \Omega_u^\tau$ the unique critical point $\pi_u(a) \in M$ of $b_{a,o} \circ u$. This is an extension at infinity of the nearest-point projection from Proposition 5.17.

Theorem 7.3 *Let $u: M \rightarrow \mathbb{X}$ be a complete and uniformly τ -nearly geodesic immersion. The map $\pi_u: \Omega_u^\tau \rightarrow M$ is a fibration. The fiber $\pi_u^{-1}(x)$ at a point $x \in M$ is the base $\mathcal{B}_\tau(\mathcal{P}_x)$ of the τ -regular pencil $\mathcal{P}_x = du(T_x M)$.*

Figure 7 illustrates this construction in the rank-1 case $G = \text{PSL}(2, \mathbb{C})$, for a totally geodesic immersion u . The associated symmetric space \mathbb{H}^3 is depicted with Poincaré’s ball model. Since \mathbb{H}^3 has rank 1, its

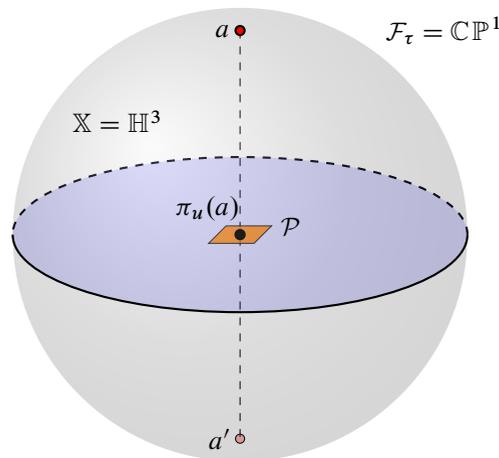


Figure 7: Fibration of the domain Ω_u^τ in the rank-1 case, $G = \text{PSL}(2, \mathbb{C})$.

visual boundary contains a single orbit $\mathcal{F}_\tau \simeq \mathbb{C}\mathbb{P}^1$. The image of u is the disk bounded by the equator. The pencil \mathcal{P} is depicted as a parallelogram. Its τ -base is a fiber of the fibration, and is the codimension-2 submanifold $\mathcal{B}_\tau(\mathcal{P}) = \{a, a'\}$.

Remark 7.4 If some element $g \in G$ preserves $u(M)$, then the map $\pi_u \circ u$ commutes with the action of g . In particular if $M = \tilde{N}$ for a compact manifold N with fundamental group Γ and if u is ρ -equivariant for some $\rho: \Gamma \rightarrow G$, then π_u is ρ -equivariant, and hence defines a fibration $\bar{\pi}_u: \Omega_u^\tau / \rho(\Gamma) \rightarrow N$.

The two important steps in the proof of Theorem 7.3 are to check that the fibers are distinct and far enough from one another using Lemma 7.2, and that these fibers are smooth manifolds using Lemma 6.7.

Proof of Theorem 7.3 Consider the set

$$E = \{(a, x) \in \Omega_u^\tau \times M \mid d_x(b_{a,o} \circ u) = 0\}.$$

Because of Lemma 7.2 the Hessian of $b_{a,o} \circ u$ is nondegenerate at critical points, and hence E is locally the zero set of a submersion, so it is a codimension-2 submanifold of $\Omega_u^\tau \times M$.

Let $\pi_1: \Omega_u^\tau \times M \rightarrow \Omega_u^\tau$ and $\pi_2: \Omega_u^\tau \times M \rightarrow M$ be the projections onto the first and second factors, respectively.

Lemma 7.2 implies that π_1 restricted to E is a bijection. Moreover, again because of the nondegeneracy of the Hessian of $b_{a,o} \circ u$, the tangent space $T_{(a,x)}E$ at $(a, x) \in E$ intersects trivially $T_x M \subset T_{(a,x)}(\Omega_u^\tau \times M)$. Hence π_1 restricted to E is a local diffeomorphism, and therefore a diffeomorphism.

Let $(a, x) \in E$. By definition $d_a(b_{a,o} \circ u): v \mapsto \langle du(v), v_{a,u(x)} \rangle_{u(x)}$ vanishes, so $v_{a,u(x)} \perp du(T_x M) = \mathcal{P}_x$. Hence a belongs to the τ -base of \mathcal{P}_x . Because of Proposition 5.7, this pencil is τ -regular and hence its τ -base contains no singular points. Lemma 6.7 implies that the tangent space $T_{(a,x)}E$ at $(a, x) \in E$ intersects trivially $T_a \Omega_u^\tau \subset T_{(a,x)}(\Omega_u^\tau \times M)$. The map π_2 restricted to E is therefore a submersion at (a, x) .

As a conclusion, $\pi_u = \pi_2 \circ \pi_1^{-1}$ is a smooth submersion. The τ -base of the pencil \mathcal{P}_x is compact in \mathcal{F}_τ , and it is included in Ω_u^τ because of Lemma 7.2. Hence π_u is a proper submersion over a connected manifold; by the Ehresmann fibration theorem it is a fibration. \square

7.2 Totally geodesic immersions that are not nearly geodesic

In this subsection let $u: M \rightarrow \mathbb{X}$ be a complete totally geodesic immersion. Let $\tau \in \mathbb{S}\mathfrak{a}^+$. We don't assume in this subsection that u is τ -regular, and hence τ -nearly geodesic.

One can still define Ω_u^τ as the set of $a \in \mathcal{F}_\tau$ such that $b_{a,o} \circ u$ is proper and bounded from below, but we can't always expect the domain to have compact fibers in this case. Lemma 7.2 can be adapted as follows:

Lemma 7.5 *A point $a \in \mathcal{F}_\tau$ belongs to Ω_u^τ if and only if the function $b_{a,o} \circ u$ admits a critical point $\pi_u(a) \in M$ at which the Hessian is positive. In this case the critical point is unique and is a global minimum of $b_{a,o} \circ u$. The domain Ω_u^τ is open.*

Proof Note that for $a \in \mathcal{F}_\tau$ the function $b_{a,o} \circ u$ is convex, since u is totally geodesic and $b_{a,o}$ is convex, but not necessarily strictly convex. Let us show that if $a \in \Omega_\rho^\tau$, $b_{a,o} \circ u$ is strictly convex at critical points.

Let $a \in \Omega_\rho^\tau$, and let $y \in M$ be a critical point of $b_{a,o} \circ u$, namely a global minimum since $b_{a,o} \circ u$ is convex. Assume for contradiction that the Hessian of $b_{a,o}$ in the direction $du(v)$ vanishes for some $v \in T_y M$.

There must exist a flat that contains a and $du(v)$ by Lemma 4.9. Let η be the geodesic ray starting at $du(v)$ in \mathbb{X} . The function $b_{a,o}$ is linear on η since a and η belong to a common flat. However the derivative of $b_{a,o}$ along η vanishes at $u(y)$, so $b_{a,o}$ is constant along η . Moreover u is totally geodesic and the whole geodesic ray starting at $du(v)$ in \mathbb{X} belongs to the image of u , so $b_{a,o} \circ u$ is not proper, contradicting the assumption that $a \in \Omega_\rho^\tau$.

For all $a \in \Omega_\rho^\tau$ we showed that the functions $b_{a,o} \circ u$ are convex and strictly convex at any critical point. Such a critical point is therefore unique since $b_{a,o} \circ u$ is convex. The rest of the proof goes as in Lemma 7.2. □

We define a map $\pi_u: \Omega_u^\tau \rightarrow M$, using Lemma 7.5. We show that this map is a fibration. Recall that the regular base $\mathcal{B}_\tau^*(\mathcal{P})$ defined in Section 6 is a subset of the base $\mathcal{B}_\tau(\mathcal{P})$ that is always a smooth codimension- d submanifold.

Theorem 7.6 *To every $a \in \Omega_u^\tau$, we associate the unique critical point $\pi_u(a) \in M$ of $b_{a,o} \circ u$ provided by Lemma 7.5. The map $\pi_u: \Omega_u^\tau \rightarrow M$ is a smooth fibration, and the fiber of this map at $y \in M$ is the regular base $\mathcal{B}_\tau^*(du(T_y M))$.*

Proof By the same argument as for Theorem 7.3, π_u is a smooth submersion.

However we need to proceed differently to prove that this map is a fibration, since the fiber is not necessarily compact. Let $g \in G$ be an element that stabilizes $u(M) \subset X$. The map π_u is equivariant with respect to g , namely for all $a \in \Omega_\rho^\tau$,

$$\pi_u(g \cdot a) = g \cdot \pi_u(a).$$

Let $y \in M$. Recall that the exponential map for the Lie group G defines a map $\exp: T_{u(y)}\mathbb{X} \simeq \mathfrak{p}_{u(y)} \subset \mathfrak{g} \rightarrow G$. Moreover since $u(M)$ is totally geodesic, any element of $\exp(du(T_y M))$ is a transvection on this totally geodesic subspace; hence it stabilizes $u(M)$. We consider the map

$$\phi: T_y M \times \mathcal{B}_\tau^*(du(T_y M)) \rightarrow \Omega_u^\tau, \quad (v, a) \mapsto \exp(du(v)) \cdot a,$$

which is an immersion between spaces of equal dimension. Moreover it is a bijection, and hence it is a diffeomorphism. Through the identification $\exp: T_y M \rightarrow M$, this gives Ω_u^τ the structure of a fibration with projection π_u . □

Since the regular base is open, it is compact if and only if the regular points form a union of connected component of $\mathcal{B}_\tau(\mathcal{P})$. This is for instance the case if $G = \text{PSL}(3, \mathbb{R})$ and $\mathcal{P} = \mathcal{P}_{\text{irr}}$ as in Example 6.9.

Example 7.7 Let $G = \text{SL}(3, \mathbb{R})$, and let ρ be a representation of the form $\rho = \iota_{\text{irr}} \circ \rho_0$ for some Fuchsian representation $\rho_0: \Gamma_g \rightarrow \text{SO}(1, 2) \simeq \text{PSL}(2, \mathbb{R})$ of a surface group and the natural inclusion $\iota_{\text{irr}}: \text{SO}(1, 2) \rightarrow \text{SL}(3, \mathbb{R})$. This representation admits a ρ -equivariant totally geodesic map $u: \tilde{S}_g \rightarrow \mathbb{X}$ (see Section 8.1 for more details). The pencil $\mathcal{P} = du(T_y \tilde{S}_g)$ for any $y \in \tilde{S}_g$ is, up to the action of G , equal to the pencil \mathcal{P}_{irr} defined in Example 6.9.

This pencil is not τ_1 regular, so u is not τ_1 -nearly geodesic. However because of Theorem 7.6 the domain $\Omega_u^{\tau_1}$ fibers over \tilde{S}_g with base $\mathcal{B}_{\tau_1}^*(\mathcal{P}_{\text{irr}})$, which is a point in $\mathcal{F}_{\tau_1} \simeq \mathbb{R}\mathbb{P}^2$. This domain is the disk of positive vectors for the chosen bilinear form of signature $(1, 2)$ on \mathbb{R}^3 . In this example the regular points of $\mathcal{B}_{\tau_1}(\mathcal{P}_{\text{irr}})$ form a connected component so the fibration is proper. If we consider a point $\ell \in \mathbb{R}\mathbb{P}^2$ outside of the closure of this disk, the associated Busemann function is minimal in $u(\tilde{S}_g)$ on a full geodesic line. If ℓ is in the boundary of the disk, the associated Busemann function is not bounded from below on $u(\tilde{S}_g)$.

7.3 Comparison with metric thickenings

In this section we consider the case when $M = \tilde{N}$ for some compact manifold N with fundamental group Γ and u is equivariant with respect to a representation $\rho: \Gamma \rightarrow G$. In other words ρ is a τ -nearly Fuchsian representation, as we defined in Section 5.4.

We show that if we have a τ -nearly Fuchsian ρ -equivariant map $u: \tilde{N} \rightarrow \mathbb{X}$ for a representation $\rho: \Gamma \rightarrow G$ the domain Ω_u^τ coincides with a domain of discontinuity associated to Anosov representations constructed by Kapovich, Leeb and Porti [30].

The domain $\Omega_\rho^\tau := \Omega_u^\tau$ depends on τ and ρ but not on u . Indeed a 1-Lipschitz function on \mathbb{X} is proper on the image of u if and only if it is proper on any $\rho(\Gamma)$ orbit in \mathbb{X} .

Even though this will be a consequence of Theorem 7.11, one can easily check that the existence of the fibration of Ω_ρ^τ implies that it is a cocompact domain of discontinuity.

Theorem 7.8 *Let $\rho: \Gamma \rightarrow G$ be a representation that admits an equivariant τ -nearly geodesic immersion $u: \tilde{N} \rightarrow \mathbb{X}$. The action of Γ on Ω_ρ^τ via ρ is properly discontinuous and cocompact.*

Proof Let π_u be the fibration from Theorem 7.3. Let A be a compact subset of Ω_ρ^τ . Its image $\pi_u(A) \subset \tilde{N}$ is compact on \tilde{N} . Since Γ acts properly on \tilde{N} , all but finitely many $\gamma \in \Gamma$ satisfy $\pi_u(A) \cap \gamma \pi_u(A) = \emptyset$. Hence for all but finitely many $\gamma \in \Gamma$, $A \cap \rho(\gamma)A = \emptyset$. The action of Γ via ρ is therefore properly discontinuous.

Let D be a compact fundamental domain for the action of Γ on \tilde{N} , namely a compact set that satisfies

$$\bigcup_{\gamma \in \Gamma} D = \tilde{N}.$$

The set $\pi_u^{-1}(D)$ is a fundamental domain for the action of Γ on Ω_ρ^τ by ρ by the equivariance of π_u . It is closed in Ω_ρ^τ . Moreover $\pi_u^{-1}(D)$ is closed in \mathcal{F}_τ . Indeed if we consider a sequence (a_n) of elements of Ω_ρ^τ that converge to $a \in \mathcal{F}_\tau$ such that $\pi_u(a_n)$ always belong to D , one can assume that $\pi_u(a_n)$ converges to

$y_0 \in D$ up to taking a subsequence. In the limit, $b_{a,o} \circ u(y_0) \leq b_{a,o} \circ u(y)$ for all $y \in \tilde{N}$. Hence y_0 is a critical point for $b_{a,o} \circ u$, so by Lemma 7.2, $a \in \Omega_\rho^\tau$.

Hence Ω_ρ^τ admits a compact fundamental domain for the action of Γ via ρ ; therefore this action is cocompact. □

We consider the domains of discontinuity constructed by *metric thickenings*, which are particular instances of the domains of discontinuity associated with a *Tits–Bruhat ideal*, defined in [30].

Let (τ, τ_0) be a pair of elements in $\mathbb{S}\mathfrak{a}^+$. This pair will be called *balanced* if τ_0 is τ -regular, namely

$$\tau_0 \notin \bigcup_{w \in W} (w \cdot \tau)^\perp.$$

Note that this is equivalent to τ being τ_0 -regular. Using the Tits angle $\angle_{\text{Tits}}: \partial_{\text{vis}}\mathbb{X}^2 \rightarrow [0, \pi]$ we associate to any $b \in \partial_{\text{vis}}\mathbb{X}$ a thickening $K_b \subset \mathcal{F}_\tau$ defined as

$$K_b = \{a \in \mathcal{F}_\tau \mid \angle_{\text{Tits}}(a, b) \leq \frac{1}{2}\pi\}.$$

Recall that the Tits angle was defined in Section 2, and is defined for points in $\partial_{\text{vis}}\mathbb{X}$.

Lemma 7.9 *Let b_1 and b_2 belong to a common maximal facet in $\partial_{\text{vis}}\mathbb{X}$, and suppose that their Cartan projections lie in the same connected component of*

$$\mathbb{S}\mathfrak{a}^+ \setminus \bigcup_{w \in W} (w \cdot \tau)^\perp.$$

Then $K_{b_1} = K_{b_2}$.

Proof Let $f \in \mathcal{F}_\Delta$ be a maximal facet that contains b_1 and b_2 . Let $a \in \mathcal{F}_\tau$. Then there exists a maximal flat of \mathbb{X} such that f and a belong to its visual boundary. This flat can be identified with \mathfrak{a} so that f corresponds to $\partial_{\text{vis}}\mathfrak{a}^+$.

Hence as long as b lies in the visual boundary of this flat, $\angle_{\text{Tits}}(a, b)$ is equal to the Euclidean angle in the flat, so the sign of its cosine does not vary as long as the Cartan projection of b does not lie in $(w \cdot \tau)^\perp$ for any $w \in W$. Therefore if the Cartan projections of b_1 and b_2 are in the same connected component of the complement, $K_{b_1} = K_{b_2}$. □

Recall that the flag manifold \mathcal{F}_{τ_0} is G -equivariantly diffeomorphic to $\mathcal{F}_{\Theta(\tau_0)} = G/P_{\Theta(\tau_0)}$ where $\Theta(\tau_0)$ is the set of simple root that do not vanish on τ_0 . Hence given a flag $f \in \mathcal{F}_{\Theta(\tau_0)}$ we define $K_f^{\tau_0} = K_b \subset \mathcal{F}_\tau$ for the unique $b \in \mathcal{F}_{\tau_0}$ corresponding to f .

Given a pair (τ, τ_0) and a $\Theta(\tau_0)$ -Anosov representation (see Definition 3.4), Kapovich, Leeb and Porti define a domain of discontinuity in \mathcal{F}_τ .

Theorem 7.10 [30, Theorem 1.10] *Let (τ, τ_0) be a pair of elements of $\mathbb{S}a^+$. Let $\rho: \Gamma \rightarrow G$ be a $\Theta(\tau_0)$ -Anosov representation. The following is a domain of discontinuity for ρ :*

$$\Omega_\rho^{(\tau, \tau_0)} = \mathcal{F}_\tau \setminus \bigcup_{\zeta \in \partial\Gamma} K_{\xi_\rho^{\Theta(\tau_0)}(\zeta)}^{\tau_0}.$$

Moreover if (τ, τ_0) is balanced, then the action of Γ via ρ on $\Omega_\rho^{(\tau, \tau_0)}$ is cocompact.

In this statement $\xi_\rho^{\Theta(\tau_0)}(\zeta) = f \in \mathcal{F}_{\Theta(\tau_0)}$ is the image of $\zeta \in \partial\Gamma$ by the boundary map $\xi_\rho^{\Theta(\tau_0)}$ associated to ρ , introduced in Theorem 3.7.

This theorem is a particular case of their result concerning Tits–Bruhat ideals. In [30] it is explained how a pair (τ, τ_0) yields a Tits–Bruhat ideal, defined via a *metric thickening*. This ideal is balanced if and only if the pair is balanced in our sense.

For a τ -nearly Fuchsian representation, the domain Ω_ρ^τ is always equal to some domain obtained by metric thickening. More precisely:

Theorem 7.11 *Let ρ be a τ -nearly Fuchsian representation of a nonelementary hyperbolic group. Recall that σ_ρ^τ and $\Theta(\sigma_\rho^\tau)$ were defined in Section 5.4. Let $\tau_0 \in \sigma_\rho^\tau$ be any element such that $\Theta(\tau_0) = \Theta(\sigma_\rho^\tau)$, whose existence is provided by Lemma 5.21. Then*

$$\Omega_\rho^\tau = \Omega_\rho^{(\tau, \tau_0)}.$$

The domain from Theorem 7.10 is a domain of discontinuity since ρ is $\Theta(\sigma_\rho^\tau)$ -Anosov by Theorem 5.22, and this domain is cocompact since the pair (τ, τ_0) is balanced.

Proof Let us write $\Theta = \Theta(\sigma_\rho^\tau) = \Theta(\tau_0)$. Let $a \in \mathcal{F}_\tau \setminus \Omega_\rho^\tau$ and let $(y_n)_{n \in \mathbb{N}}$ be a diverging sequence of points in M such that $(b_{a,o}(u(y_n)))_{n \in \mathbb{N}}$ is bounded from above. Up to taking a subsequence let us assume that it converges to a point $\zeta \in \partial\Gamma \simeq \partial\tilde{N}$. We consider the geodesic segments $[o, u(y_n)] \subset \mathbb{X}$ for $n \in \mathbb{N}$. Since ρ is τ -nearly Fuchsian it is a quasi-isometric embedding by Proposition 5.16, so in particular the length of these segments goes to $+\infty$. Up to taking a subsequence, we can assume that these geodesic segments converge to a geodesic ray $\eta: \mathbb{R}_{\geq 0} \rightarrow \mathbb{X}$ with $\eta(0) = o$. Let $[\eta] \in \partial_{\text{vis}}\mathbb{X}$ be the point corresponding to the class of η .

Busemann functions are convex, so the function $b_{a,o}$ is bounded from above on all the geodesic segments $[o, u(y_n)]$ for $n \in \mathbb{N}$, and hence $b_{a,o} \circ \eta$ is bounded from above. Therefore $\angle_{\text{Tits}}(a, b) \leq \frac{1}{2}\pi$ by Lemma 4.8, so $a \in K_{[\eta]}$. Let $b \in \mathcal{F}_{\tau_0}$ be an element such that b and $[\eta]$ belong to a common maximal facet and whose Cartan projection lies in σ_ρ^τ . Lemma 7.9 implies that $K_{[\eta]} = K_b$. Theorem 3.7 implies that $K_b = K_{\xi_\rho^{\Theta(\tau_0)}(\zeta)}^{\tau_0}$. Therefore if $a \in \mathcal{F}_\tau \setminus \Omega_\rho^\tau$ then $a \in \mathcal{F}_\tau \setminus \Omega_\rho^{(\tau, \tau_0)}$.

Conversely let $a \in \Omega_\rho^\tau$ and let $\zeta \in \partial\Gamma$. Consider a geodesic ray $\eta: \mathbb{R}_{> 0} \rightarrow \tilde{N}$ for the metric $u^*(g_{\mathbb{X}})$ converging to ζ . Theorem 3.8 implies that there exists $D > 0$ such that for all $t > 0$, there exists a geodesic ray $\eta_t: \mathbb{R}_{> 0} \rightarrow \mathbb{X}$ such that $\eta_t(0) = u \circ \eta(0)$, $\eta_t(t)$ is at distance at most D from $u \circ \eta(t)$ and $[\eta_t] \in \partial_{\text{vis}}\mathbb{X}$ belongs to a common maximal facet with $\xi_\rho^{\Theta(\tau_0)}(\zeta)$. Since $\mathcal{C}_\rho \subset \sigma_\rho^\tau$, for all t large enough, $K_{[\eta_t]} = K_{\xi_\rho^{\Theta(\tau_0)}(\zeta)}^{\tau_0}$.

The Busemann function $b_{a,o}$ is proper on η ; hence for t large enough $b_{a,o} \circ \eta_t(t) > b_{a,o} \circ \eta_t(0)$. Since $b_{a,o}$ is convex, this implies that $b_{a,o}$ is growing at least linearly on η_t , so $a \notin K_{[\eta_t]}$ by Lemma 4.8. Therefore $a \in \Omega_\rho^{(\tau, \tau_0)}$. \square

7.4 Invariance of the topology

In this section we prove that the topology of the quotient of the domains of discontinuity considered by Kapovich, Leeb and Porti does not vary when the representation is deformed continuously. Guichard and Wienhard proved this already for the domains of discontinuity that they consider in [24].

Let Γ be a torsion-free finitely generated group and \mathcal{F} a G -homogeneous space. Let $(\rho_t)_{t \in [0,1]}$ be a smooth family of representations from Γ to G . Consider for every $t \in [0, 1]$ an open $\rho_t(\Gamma)$ -invariant domain $\Omega_t \subset \mathcal{F}$.

Lemma 7.12 *Suppose that these domains are **uniformly** cocompact domains of discontinuity for (ρ_t) , namely the domain $\Omega = \{(t, a) \mid a \in \Omega_t\} \subset [0, 1] \times \mathcal{F}$ is open and the action of Γ via ρ is properly discontinuous and cocompact where $\rho(\gamma) \cdot (t, a) = (t, \rho_t(\gamma) \cdot a)$. The quotient $\Omega_0/\rho_0(\Gamma)$ is diffeomorphic to $\Omega_1/\rho_1(\Gamma)$.*

Proof The projection onto the first factor in $[0, 1] \times \mathcal{F}$ descends to a submersion $p: \Omega/\rho(\Gamma) \rightarrow \mathbb{R}$. Since $\Omega/\rho(\Gamma)$ is compact and the base is connected, Ehresmann’s fibration theorem implies that the proper submersion p is a fibration. Hence $p^{-1}(0)$ and $p^{-1}(1)$ are diffeomorphic. \square

Remarks 7.13 A concrete way to construct this diffeomorphism is to pick a Riemannian metric on $\Omega/\rho(\Gamma)$ and consider the flow of the gradient of p . If we consider two different Riemannian metrics, the diffeomorphisms obtained are isotopic, since the space of Riemannian metrics is path connected. Hence this operation constructs a unique diffeomorphism up to isotopy.

Note that a family of cocompact domains of discontinuity could be nonuniformly cocompact, for instance a family of representations such that $\rho_t: \mathbb{Z} \rightarrow \text{PSL}(2, \mathbb{R})$ is hyperbolic for $0 \leq t < 1$ and parabolic for $t = 1$. These representations admit a unique maximal domain of discontinuity in $\mathbb{R}\mathbb{P}^1$ with two connected components for $t < 1$ and one for $t = 1$. The quotient of the corresponding domain Ω is homeomorphic to the noncompact space $S^1 \times [0, 1] \sqcup S^1 \times [0, 1)$.

In order to apply Lemma 7.12, we need a slight adaptation of Theorem 7.10 from [30]. Let Γ be any Gromov-hyperbolic group. We check that the domains constructed by Kapovich, Leeb and Porti are uniformly cocompact domains of discontinuity for any smooth path of Anosov representations.

Proposition 7.14 (adaptation of [30, Theorem 1.10]) *Let (τ, τ_0) be a balanced pair as in Section 7.3. Let $\rho: [0, 1] \rightarrow \text{Hom}(\Gamma, G)$, $t \mapsto \rho_t$ be a continuous path such that the family $(\rho_t)_{t \in [0,1]}$ consists only of $\Theta(\tau_0)$ -Anosov representations.*

The family of domains $\Omega_t = \Omega_{\rho_t}^{(\tau, \tau_0)}$ for $t \in [0, 1]$ are uniformly cocompact domains of discontinuity for the family of representations ρ .

We check that the arguments from Kapovich, Leeb and Porti are uniform on neighborhoods of Anosov representations. The same proof holds if one considers more generally domains of discontinuity constructed with balanced Tits–Bruhat ideals as in [30].

Proof The domain Ω is the complement in $[0, 1] \times \mathcal{F}_\tau$ of

$$K_\rho = \bigcup_{t \in [0,1]} \{t\} \times K_{\rho,t}, \quad K_{\rho,t} = \bigcup_{x \in \partial\Gamma} K_{\xi_\rho^\ominus(x)}^{\tau_0}.$$

Since the boundary maps ξ_ρ^\ominus are continuous and vary continuously when ρ varies continuously in the space of Θ -Anosov representations (see [6, Section 6]), and since $K_{\xi_\rho^\ominus(x)}^{(\tau, \tau_0)}$ is compact, K_ρ is compact, so $\Omega = \{(t, a) \mid a \in \Omega_t\} \subset [0, 1] \times \mathcal{F}$ is open.

Let us fix a Riemannian distance d on \mathcal{F}_τ . Let $A = \{(t, a) \mid t \in [0, 1], a \in A_t \subset \Omega\}$ be a compact set and let $(\gamma_n) \in \Gamma$ be a diverging sequence. Corollary 6.8 of [30] implies that given $t \in [0, 1]$, for any $\epsilon > 0$, for all n large enough, if $d(a, K_{\rho,t}) \geq \epsilon$ then $d(\rho_t(\gamma_n) \cdot a, K_{\rho,t}) \leq \epsilon$, where the minimal value of n needed depends on the constants b and c that come into play in the definition of Anosov representations (Definition 3.4).

Since we consider a compact set of Anosov representations, and since these constants can be chosen locally uniformly around a given Anosov representation (see [29, Theorem 7.18]), these constants can be chosen uniformly for all representations $(\rho_t)_{t \in [0,1]}$. Moreover the compact sets A_t are at uniform distance from $K_{\rho,t}$. Therefore, for all $n \in \mathbb{N}$ large enough, for all $t \in \mathbb{R}$, $\rho_t(\gamma_n) \cdot A_t \cap A_t = \emptyset$. So for all n large enough $\rho(\gamma_n) \cdot A \cap A = \emptyset$. We have proven that the action of ρ on Ω is properly discontinuous.

In order to prove the cocompactness of the action on Ω , we will check that the transverse expansion holds uniformly. It follows from [30, Proposition 7.7] that for every $t \in [0, 1]$, the action of ρ_t is *transversely expanding* at the limit set of ρ_t as in [30, Definition 5.21], namely for all $x \in \partial\Gamma$, there exists $\gamma \in \Gamma$, an open neighborhood U of $K_{\xi_\rho^\ominus(x)}^{\tau_0}$ in \mathcal{F}_τ and a constant $\lambda > 1$ such that for all $a \in U$ and $y \in \partial_{\text{vis}}\Gamma$ that satisfy $K_{\xi_\rho^\ominus(y)}^{\tau_0} \subset U$, one has

$$d(\rho_t(\gamma) \cdot a, \rho_t(\gamma) \cdot K_{\xi_\rho^\ominus(y)}^{\tau_0}) \geq \lambda d(a, K_{\xi_\rho^\ominus(y)}^{\tau_0}).$$

Let $g \in G$, $f \in \mathcal{F}_\Theta$ $\lambda > 1$ and $U \subset \mathcal{F}_\tau$ be an open set. We say that g is *expanding* at $K_f^{\tau_0}$ over U with factor λ if for all $a \in U$,

$$d(g \cdot a, g \cdot K_f^{\tau_0}) \geq \mu d(a, K_f^{\tau_0}).$$

This property is open in the following sense: if g is expanding at $K_f^{\tau_0}$ over U with factor λ , then for any $1 < \lambda' < \lambda$ and any open subset E such that $\bar{E} \subset U$ there exists a neighborhood U_g of g in G and U_f of f in \mathcal{F}_Θ such that for all $g' \in U_g$ and $f' \in U_f$, g' is expanding at $K_{f'}^{\tau_0}$ over E with factor λ' .

This implies that the action of ρ on $\mathcal{F}_\tau \times [0, 1]$ satisfies the transverse expansion property, where K_ρ is considered as a bundle over $[0, 1] \times K$. Namely for all $t \in [0, 1]$ and $x \in \partial\Gamma$, there exists $\gamma \in \Gamma$, an open

neighborhood U_0 of $K_{\xi_{\rho_t}^\Theta(x)}^{\tau_0}$ in $\mathcal{F}_\tau \times [0, 1]$ and a constant $\lambda' > 1$ such that for all $a \in U$ and $y \in \partial\Gamma$ satisfying $K_{\xi_{\rho_t}^\Theta(y)}^{\tau_0} \times \{t\} \subset U_0$, one has

$$d(\rho(\gamma) \cdot a, \rho(\gamma) \cdot K_{\xi_{\rho_t}^\Theta(y)}^{\tau_0} \times \{t\}) \geq \lambda' d(a, K_{\xi_{\rho_t}^\Theta(y)}^{\tau_0} \times \{t\}).$$

Therefore by [30, Proposition 5.26] the action of ρ is cocompact on Ω . □

From Proposition 7.14 and Lemma 7.12 we get the following corollary:

Corollary 7.15 *Assume that Γ is torsion-free. Let $\mathcal{C} \subset \text{Hom}(\Gamma, G)$ be an open and connected set consisting only of Θ -Anosov representations for some $\Theta \subset \Delta$. Let (τ, τ_0) be a balanced pair such that for all $\alpha \in \Delta \setminus \Theta$, $\alpha(\tau_0) = 0$. The diffeomorphism type of $\Omega_\rho^{(\tau, \tau_0)} / \rho(\Gamma)$ is independent of $\rho \in \mathcal{C}$.*

If moreover \mathcal{C} is simply connected, the diffeomorphism provided by Lemma 7.12 between $\Omega_{\rho_1}^{(\tau, \tau_0)} / \rho_1(\Gamma)$ and $\Omega_{\rho_2}^{(\tau, \tau_0)} / \rho_2(\Gamma)$ for $\rho_1, \rho_2 \in \mathcal{C}$ is uniquely determined up to isotopy.

Connectedness via paths that are smooth by parts for an open set in $\text{Hom}(\Gamma, G)$ is equivalent to connectedness since $\text{Hom}(\Gamma, G)$ is locally a real algebraic variety.

8 Applications

In this section we apply our results to prove that all representations in some connected components of Anosov representations are the restricted holonomy of a geometric structure on a fiber bundle over a manifold. For these applications we only consider nearly geodesic surfaces that are totally geodesic. We will mostly focus on surface groups, but in Section 8.5 we also describe two applications for representations of fundamental groups of higher-dimensional compact hyperbolic manifolds.

8.1 Totally geodesic immersions

Totally geodesic surfaces provide examples of τ -nearly geodesic surfaces if these surfaces are τ -regular (Proposition 5.8). The study of totally geodesic surfaces in \mathbb{X} is related to the study of representation of semisimple Lie algebras in \mathfrak{g} . We recall here a classical fact:

Proposition 8.1 *Let $\mathfrak{h} \subset \mathfrak{g}$ be a semisimple Lie subalgebra of noncompact type. Let H be the closed Lie subgroup of G with Lie algebra \mathfrak{h} and \mathbb{Y} its associated symmetric space of noncompact type. There exists an H -equivariant and totally geodesic embedding $u_{\mathfrak{h}}: \mathbb{Y} \rightarrow \mathbb{X}$. The image of this embedding is unique up to the action of the centralizer:*

$$\mathcal{C}_G(\mathfrak{h}) = \{g \in G \mid \forall h \in \mathfrak{h}, \text{Ad}_g(h) = h\}.$$

If $y \in \mathbb{Y}$, let $K \subset G$ be the stabilizer of $u_{\mathfrak{h}}(y)$ in G . Every element in $\mathcal{C}_K(\mathfrak{h})$ of \mathfrak{h} in G fixes $u_{\mathfrak{h}}(\mathbb{Y})$ pointwise. If the centralizer $\mathcal{C}_G(\mathfrak{h})$ in G is compact, then $\mathcal{C}_K(\mathfrak{h}) = \mathcal{C}_G(\mathfrak{h})$, so the totally geodesic submanifold $u_{\mathfrak{h}}(\mathbb{Y}) \subset \mathbb{X}$ is uniquely determined.

Proof Let $\mathfrak{h} = \mathfrak{t} + \mathfrak{p}$ be a Cartan decomposition of \mathfrak{h} associated with the Cartan involution θ_y for $y \in \mathbb{Y}$. Let $\hat{\mathfrak{h}} = \mathfrak{t} + i\mathfrak{p}$ be the associated compact real form of $\mathfrak{h} \otimes \mathbb{C}$. Since $\hat{\mathfrak{h}}$ is the Lie algebra of a semisimple compact Lie group, it is the Lie algebra of a compact Lie subgroup of $G^{\mathbb{C}}$. In particular there exists a Cartan involution $\theta^{\mathbb{C}}$ of $\mathfrak{g} \otimes \mathbb{C}$ such that $\theta^{\mathbb{C}}(v) = v$ for all $v \in \hat{\mathfrak{h}}$.

Let $\overline{\theta^{\mathbb{C}}}$ be the Cartan involution conjugate to $\theta^{\mathbb{C}}$ in $\mathfrak{g} \otimes \mathbb{C}$. Let θ be the Cartan involution of $\mathfrak{g} \otimes \mathbb{C}$ corresponding to the midpoint in the symmetric space associated with $\mathfrak{g} \otimes \mathbb{C}$ of $\overline{\theta^{\mathbb{C}}}$ and $\theta^{\mathbb{C}}$. It is invariant by conjugation, so it descends to a Cartan involution on \mathfrak{g} , associated with a point $x \in \mathbb{X}$. There exists a transvection along the geodesic between $\overline{\theta^{\mathbb{C}}}$ and $\theta^{\mathbb{C}}$ in the symmetric space associated to $\mathfrak{g} \otimes \mathbb{C}$ that induces a unique inner automorphism ϕ of $\mathfrak{g} \otimes \mathbb{C}$ such that $\phi\theta^{\mathbb{C}}\phi^{-1} = \theta$ and $\phi^2\theta^{\mathbb{C}}\phi^{-2} = \overline{\theta^{\mathbb{C}}}$. Moreover ϕ is symmetric positive with respect to $B^{\mathbb{C}}(\cdot, \theta^{\mathbb{C}}(\cdot))$, where $B^{\mathbb{C}}$ is the Killing form on $\mathfrak{g} \otimes \mathbb{C}$ and the transvection ϕ^4 is equal to a composition of symmetries $\theta^{\mathbb{C}}\overline{\theta^{\mathbb{C}}}$. The transvection ϕ^4 stabilizes $\mathfrak{h} \otimes \mathbb{C}$, so ϕ also stabilizes $\mathfrak{h} \otimes \mathbb{C}$. Therefore $\theta = \phi\theta^{\mathbb{C}}\phi^{-1}$ stabilizes $\mathfrak{h} \otimes \mathbb{C}$. But θ is a real Cartan involution as it is preserved by conjugation, so it stabilizes \mathfrak{h} . Let $x \in \mathbb{X}$ be the point corresponding to θ in \mathbb{X} .

The map $u_{\mathfrak{h}}: \mathbb{Y} \rightarrow \mathbb{X}$ such that for all $h \in H$, $u_{\mathfrak{h}}(h \cdot y) = h \cdot x$ is well defined since $\mathfrak{t} \subset \mathfrak{t}_x$, so the image of the stabilizer of $y \in \mathbb{Y}$ by H lies in the stabilizer of $x \in \mathbb{X}$. Moreover it is totally geodesic, since $\eta(\mathfrak{p}) \subset \mathfrak{p}_x$; see [26, Chapter 4, Section 7]. This map is by definition H -equivariant.

Suppose that there is another H -equivariant and totally geodesic embedding $u'_{\mathfrak{h}}$ such that $u'_{\mathfrak{h}}(y) = x'$. Let θ' be the corresponding Cartan involution of \mathfrak{g} . Then $\theta' \circ \theta$ is the identity on \mathfrak{h} and is equal to the adjoint action of $\exp(z)$ for some $z \in \mathfrak{p}_x$. Therefore z is in the centralizer of \mathfrak{h} in \mathfrak{g} . Hence $g = \exp(\frac{1}{2}z) \in \mathcal{C}_G(\mathfrak{h})$ satisfies $\text{Ad}_g \circ u'_{\mathfrak{h}} = u_{\mathfrak{h}}$. Conversely let $g' \in \mathcal{C}_K(\mathfrak{h})$. Then g' fixes $u_{\mathfrak{h}}(y) \in \mathbb{X}$, and it fixes \mathfrak{h} , so it fixes $du_{\mathfrak{h}}(T_y \mathbb{Y})$. Therefore it preserves and acts trivially on $u_{\mathfrak{h}}(\mathbb{Y})$.

Now assume that $\mathcal{C}_G(\mathfrak{h})$ is compact; there cannot be any element $z \in \mathfrak{p}_x$ in the centralizer of \mathfrak{h} in \mathfrak{g} , so the totally geodesic and H -equivariant embedding $u_{\mathfrak{h}}$ is unique and $\mathcal{C}_K(\mathfrak{h}) = \mathcal{C}_G(\mathfrak{h})$. \square

Remark 8.2 A totally geodesic embedding $u: \mathbb{Y} \rightarrow \mathbb{X}$ can only be a τ -nearly geodesic immersion if $\text{rank}(\mathbb{Y}) = 1$, because otherwise it cannot be τ -regular for any $\tau \in \mathbb{S}\mathfrak{a}^+$ (see Proposition 6.11). When $\text{rank}(\mathbb{Y}) = 1$, all the unit tangent vectors to this embedded surface have the same Cartan projection, so the embedding is τ -regular for all τ in the complement in $\mathbb{S}\mathfrak{a}^+$ of a finite collection of hyperplanes.

We illustrate Proposition 8.1 in the following example for some special \mathfrak{sl}_2 Lie subalgebras in $\mathfrak{sl}_n(\mathbb{R})$.

Example 8.3 Here we construct representations ι from $\text{SL}(2, \mathbb{R})$ into $\text{SL}(n, \mathbb{K})$ that stabilize some totally geodesic hyperbolic planes inside the symmetric space $\mathbb{X} = \mathcal{S}_n$ associated with $G = \text{SL}(n, \mathbb{K})$ for $\mathbb{K} = \mathbb{R}$ or \mathbb{C} .

Let $\mathbb{K} = \mathbb{R}$ or \mathbb{C} . Let $V_n(\mathbb{K})$ be the space of homogeneous polynomial with coefficients in \mathbb{K} of degree $n-1$ in two variables X and Y . To an element $g \in \text{SL}_2(\mathbb{R})$ one can associate an element $\iota_{\text{irr}}(g) \in \text{SL}(V_n(\mathbb{K}))$ that acts by a change of variable on $V_n(\mathbb{K})$, ie that associates to $P \in V_n(\mathbb{K})$ the polynomial $P \circ g^{-1}$. Let \mathfrak{h} be the corresponding \mathfrak{sl}_2 -Lie subalgebra of $\mathfrak{sl}_n(\mathbb{R})$; note that $\iota_{\text{irr}} = \iota_{\mathfrak{h}}$.

Let q the Euclidean or Hermitian metric on $V_n(\mathbb{K})$ such that for all $0 \leq a, b \leq n - 1$:

$$q(X^a Y^{n-1-a}, X^b Y^{n-1-b}) = \begin{cases} \binom{n-1}{a}^{-1} & \text{if } a = b, \\ 0 & \text{otherwise.} \end{cases}$$

Consider the following basis of the lie algebra $\mathfrak{sl}_2(\mathbb{R})$:

$$\mathbf{h} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

which satisfies $[\mathbf{g}, \mathbf{h}] = -2\mathbf{f}$, $[\mathbf{h}, \mathbf{f}] = 2\mathbf{g}$ and $[\mathbf{f}, \mathbf{g}] = 2\mathbf{h}$. Fix the following orthonormal basis for q :

$$(e_a)_{0 \leq a \leq n-1} = \left(X^a Y^{n-1-a} \binom{n-1}{a}^{\frac{1}{2}} \right)_{0 \leq a \leq n-1}.$$

For $0 \leq a \leq n - 1$, one has

$$\begin{aligned} d\iota_{\text{irr}}(\mathbf{f})(e_a) &= (2a - n + 1)e_a, \\ d\iota_{\text{irr}}(\mathbf{e})(e_a) &= aX^{a-1}Y^{n-a} \binom{n-1}{a}^{\frac{1}{2}} - (n-1-a)X^{a+1}Y^{n-2-a} \binom{n-1}{a}^{\frac{1}{2}}, \\ d\iota_{\text{irr}}(\mathbf{e})(e_a) &= \sqrt{a(n-a)}e_{a-1} - \sqrt{(a+1)(n-1-a)}e_{a+1}, \end{aligned}$$

and moreover $\mathbf{g} = \frac{1}{2}[\mathbf{f}, \mathbf{h}]$. In particular $d\iota_{\text{irr}}(\mathbf{f})$ and $d\iota_{\text{irr}}(\mathbf{g})$ are symmetric or Hermitian and $d\iota_{\text{irr}}(\mathbf{h})$ is antisymmetric or anti-Hermitian with respect to q .

Hence as in the proof of Proposition 8.1, there is a totally geodesic map $u_{\mathfrak{g}}: \mathbb{H}^2 \rightarrow \mathcal{S}_n$ such that the image of the Cartan involution $M \mapsto -M^t$ is equal to the point in \mathcal{S}_n corresponding to q .

This representation ι_{irr} is the unique irreducible representation of $\text{SL}(2, \mathbb{R})$ into $\text{SL}(n, \mathbb{R})$ up to conjugation by elements of $\text{GL}(n, \mathbb{R})$. The image of ι_{irr} lies in the subgroup $\text{Sp}(2k, \mathbb{R})$ for some symplectic form on \mathbb{R}^{2k} when $n = 2k$ is even, and in $\text{SO}(k, k + 1)$ for some quadratic form on \mathbb{R}^{2k+1} when $n = 2k + 1$ is odd.

One can construct other totally geodesic hyperbolic planes in \mathcal{S}_n by considering representations of $\text{SL}(2, \mathbb{R})$ that can be decomposed into a direct sum of irreducible representations. Equivalently one can consider other reducible \mathfrak{sl}_2 -subalgebras of $\mathfrak{sl}_n(\mathbb{R})$.

For instance one can define $\iota_{\text{red}}: \text{SL}(2, \mathbb{R}) \rightarrow \text{SL}(2n, \mathbb{R})$ that associates to a matrix $M \in \text{SL}(2, \mathbb{R})$ the block diagonal matrix

$$\iota_{\text{red}}(M) = \begin{pmatrix} M & & \\ & \ddots & \\ & & M \end{pmatrix}.$$

The image of ι_{red} lies in $\text{Sp}(2n, \mathbb{R})$ for some symplectic form ω on \mathbb{R}^{2n} .

8.2 Geometric structures on fiber bundles

Using the projection defined by Busemann functions from Theorem 7.3, one can show that Anosov deformations of a representation that admits an equivariant nearly geodesic immersion are holonomies of (G, X) structures on a fiber bundle over S_g .

A (G, X) -structure on a manifold N , for a Lie group G and a G -homogeneous space X on which G acts faithfully, is a maximal atlas of charts of M valued in X whose transition functions are the restriction of the action of some elements in G . A more developed introduction to this notion can be found in [1].

To a (G, X) -structure one can associate a *developing map* $\text{dev}: \tilde{N} \rightarrow X$ that is a local diffeomorphism compatible with the atlas defining the (G, X) -structure on N , and a *holonomy* $\text{hol}: \pi_1(N) \rightarrow G$ such that dev is hol -equivariant. This pair is unique up to the action of G by conjugation of the holonomy and postcomposition of the developing map.

Let N be a manifold and Γ its fundamental group. In what follows, we say that a (G, X) -structure *on a fiber bundle* F over N is a (G, X) -structure on F for which the fundamental group of the fibers is included in the kernel of the holonomy. Hence one can define the *restricted holonomy* of the structure as the quotient map $\rho: \pi_1(N) \rightarrow G$ induced by the holonomy.

Constructing domains of discontinuity allows us to construct geometric structures.

Proposition 8.4 *Let $\rho: \Gamma \rightarrow G$ be a representation and $\Omega \subset X$ a cocompact nonempty domain of discontinuity which fibers ρ -equivariantly over \tilde{N} . Any connected component of the quotient $\Omega/\rho(\Gamma)$ inherits a (G, X) -structure **on a fiber bundle**, with restricted holonomy ρ .*

Note that even if Ω is disconnected, the quotient $\Omega/\rho(\Gamma)$ can be connected.

From now on, we assume that G is center-free, so it acts faithfully on its flag manifolds. Let N be a compact manifold whose fundamental group Γ is Gromov hyperbolic and torsion-free. Let $\tau \in \mathbb{S}\mathfrak{a}^+$.

Theorem 8.5 *Let $\rho_0: \Gamma \rightarrow G$ be a representation that admits an equivariant τ -nearly geodesic surface $u: \tilde{N} \rightarrow \mathbb{X}$ such that $\Omega_u^\tau \neq \emptyset$. Let \mathcal{C} be the connected component of the space of $\Theta(\sigma_{\rho_0}^\tau)$ -Anosov representations in $\text{Hom}(\Gamma, G)$ containing ρ_0 . Every representation in \mathcal{C} is the restricted holonomy of a (G, \mathcal{F}_τ) -structure **on a fiber bundle** F over N .*

Proof Theorem 7.8 implies that the domain Ω_u^τ admits a ρ_0 -equivariant fibration over \tilde{N} . The domain $\Omega_{\rho_0}^\tau$ coincides with a domain obtained as a metric thickening $\Omega_{\rho'}^{(\tau, \tau_0)}$ for some $\tau_0 \in \mathbb{S}\mathfrak{a}^+$ such that $\Theta(\tau_0) = \Theta(\sigma_{\rho_0}^\tau)$. Let F be a connected component of $\Omega_{\rho_0}^\tau/\rho_0(\Gamma)$. Corollary 7.15 implies that for every representation $\rho \in \mathcal{C}$, a connected component F_ρ of $\Omega_\rho^{(\tau, \tau_0)}/\rho(\Gamma)$ is diffeomorphic to F , which is a fiber bundle over M . The covering map $\Omega_\rho^{(\tau, \tau_0)} \rightarrow \Omega_{\rho_0}^{(\tau, \tau_0)}/\rho_0(\Gamma)$ induces the covering $\hat{F}_\rho \rightarrow F_\rho \simeq F$ associated to the subgroup of the fundamental group of F corresponding to the fundamental group of the fiber of F over M .

Note that F is assumed to be nonempty. The (G, \mathcal{F}_τ) -structure on $F_\rho \simeq F$ is such that the holonomy of the fundamental group of each fiber is trivial. Indeed the developing map $\text{dev}: \tilde{F}_\rho \rightarrow \mathcal{F}_\tau$ descends to the inclusion $\hat{F}_\rho \rightarrow \mathcal{F}_\tau$, so the fundamental group of the fiber belongs to the kernel of the holonomy. \square

Let $M = S_g$ be a closed orientable surface of genus g and $\Gamma = \Gamma_g$ be its fundamental group. We apply Theorem 8.5 in cases when the nearly geodesic surface is totally geodesic and we describe the fibration that is obtained.

Let $\mathfrak{h} \subset \mathfrak{g}$ be an \mathfrak{sl}_2 Lie subalgebra, namely a Lie subalgebra isomorphic to $\mathfrak{sl}_2(\mathbb{R})$. Note that if G is a quotient of its adjoint form, which is the case since we assumed that G is center-free, the corresponding Lie group H is isomorphic to $SL(2, \mathbb{R})$ or $PSL(2, \mathbb{R})$. We write $\iota_{\mathfrak{h}}: SL(2, \mathbb{R}) \rightarrow G$ for the corresponding Lie group representation, and $u_{\mathfrak{h}}: \mathbb{H}^2 \rightarrow \mathbb{X}$ for a corresponding equivariant totally geodesic embedding.

Definition 8.6 We say that a representation $\rho: \Gamma_g \rightarrow G$ is \mathfrak{h} -generalized Fuchsian if it preserves and acts cocompactly on $u_{\mathfrak{h}}(\mathbb{H}^2)$.

If G is center-free, an \mathfrak{h} -generalized Fuchsian representation can be written $\gamma \mapsto \iota_{\mathfrak{h}}(\rho_0(\gamma)) \times \chi(\gamma)$ for some Fuchsian representation $\rho_0: \Gamma_g \rightarrow SL(2, \mathbb{R})$ and some associated character $\chi: \Gamma_g \rightarrow \mathcal{C}_K(\mathfrak{h})$.

Let $y \in \mathbb{H}^2$ be a basepoint, and let K be the stabilizer in G of $u_{\mathfrak{h}}(y)$. We write $\mathcal{B}^{\tau}(\mathfrak{h})$ for the τ -base of the pencil of tangent vectors $du_{\mathfrak{h}}(T_y \mathbb{H}^2)$. Note that this pencil is stabilized by $\iota_{\mathfrak{h}}(SO(2, \mathbb{R})) \times \mathcal{C}_K(\mathfrak{h})$.

The quotient map $SL(2, \mathbb{R}) \rightarrow SL(2, \mathbb{R})/SO(2, \mathbb{R}) \simeq \mathbb{H}^2$ defines a principal $SO(2, \mathbb{R})$ -bundle \tilde{P} over \mathbb{H}^2 . Let P_{S_g} be its quotient via some Fuchsian representation. Given a character $\chi: \Gamma_g \rightarrow \mathcal{C}_K(\mathfrak{h})$, let $P_{S_g, \chi} \rightarrow S_g$ be the $(SO(2, \mathbb{R}) \times \mathcal{C}_K(\mathfrak{h}))$ -principal bundle obtained as the product of P_{S_g} and the flat $\mathcal{C}_K(\mathfrak{h})$ -bundle associated to χ .

Theorem 8.7 Suppose that \mathfrak{h} is τ -regular, namely $u_{\mathfrak{h}}$ is τ -regular, and that $\Omega_{u_{\mathfrak{h}}}^{\tau} \neq \emptyset$. Let $\rho: \Gamma_g \rightarrow G$ be an \mathfrak{h} -generalized Fuchsian representation with associated character χ and let $\tau \in \mathbb{S}\mathfrak{a}^+$.

Let \mathcal{C} be the connected component of the space of $\Theta(\sigma_{\rho}^{\tau})$ -Anosov representations that contains ρ . Every representations in \mathcal{C} is the restricted holonomy of a (G, \mathcal{F}_{τ}) -structure on a fiber bundle F over S_g .

The fiber bundle F is a connected component of the reduction of the $(SO(2, \mathbb{R}) \times \mathcal{C}_K(\mathfrak{h}))$ -principal bundle $P_{S_g, \chi}$ over S_g via the action of $\iota_{\mathfrak{h}}(SO(2, \mathbb{R})) \times \mathcal{C}_K(\mathfrak{h})$ on $\mathcal{B}^{\tau}(\mathfrak{h})$.

Note that when $\Theta \subset \Delta$ is a Weyl orbit of simple roots and $\tau = \tau_{\Theta}$, a τ_{Θ} -regular immersion is just a Θ -regular immersion and $\Theta(\sigma_{\rho}^{\tau}) = \Theta$.

Remark 8.8 If we replace G and \mathcal{F}_{τ} by finite covers \hat{G} and $\hat{\mathcal{F}}_{\tau}$ so that \hat{G} acts faithfully on $\hat{\mathcal{F}}_{\tau}$, Theorem 8.7 still applies. In this case one should replace $\mathcal{B}^{\tau}(\mathfrak{h})$ by its preimage $\widehat{\mathcal{B}^{\tau}(\mathfrak{h})}$ by the covering map $\hat{\mathcal{F}}_{\tau} \rightarrow \mathcal{F}_{\tau}$.

The only part of Theorem 8.7 that is not already contained in Theorem 8.5 is the description of the fibration.

Proof The embedding $u_{\mathfrak{h}}$ is totally geodesic. Moreover it is assumed to be τ -regular. Hence $u_{\mathfrak{h}}$ is τ -nearly geodesic, so one can apply Theorem 8.5. It only remains to describe the fibration.

The fibration $\pi: \Omega_{u_{\mathfrak{h}}}^{\tau} \rightarrow \mathbb{H}^2$ is $\iota_{\mathfrak{h}}(\mathrm{SL}(2, \mathbb{R}))$ -equivariant. The corresponding fiber bundle can be identified in an $\mathrm{SL}(2, \mathbb{R})$ -equivariant way as the reduction of the $\mathrm{SO}(2, \mathbb{R})$ -principal bundle \tilde{P} by the action of $\iota_{\mathfrak{h}}(\mathrm{SO}(2, \mathbb{R}))$ on the fiber $\mathcal{B}^{\tau}(\mathfrak{h})$. The quotient of this fiber bundle by any Fuchsian representation $\rho_0: \Gamma_g \rightarrow \mathrm{SL}(2, \mathbb{R})$ is hence the bundle induced by the principal bundle P_{S_g} . Once we quotient $\Omega_{u_{\mathfrak{h}}}^{\tau}$ by $\rho = \iota_{\mathfrak{h}} \circ \rho_0 \times \chi$ the quotient becomes the twisted fiber bundle induced by $P_{S_g, \chi}$. \square

8.3 Higher-rank Teichmüller spaces

In this section we apply Theorem 8.7 to Hitchin representations in $\mathrm{PSL}(n, \mathbb{R})$ and to maximal representations in $\mathrm{Sp}(2n, \mathbb{R})$. Then we explain how in general one can apply it to the connected components of Θ -positive representations containing at least one generalized Fuchsian representation associated to a Θ -principal \mathfrak{sl}_2 -Lie subalgebra.

8.3.1 Positive representations Let G be a connected simple Lie group of noncompact type with trivial center and Θ a set of its simple roots such that the pair (G, Θ) admits a notion of Θ -positivity in the sense of [25]. We moreover assume that G is not locally isomorphic to $\mathrm{PSL}(2, \mathbb{R})$. This means that one of the following holds:

- (i) G is split real and $\Theta = \Delta$.
- (ii) G is Hermitian of tube type.
- (iii) G is locally isomorphic to $\mathrm{SO}(p, q)$ and $\Theta = \{\alpha_1, \dots, \alpha_p\}$.
- (iv) G is a real form of the complex Lie group with Dynkin diagram F_4, E_6, E_7 or E_8 with restricted Dynkin diagram F_4 and Θ consists of the two larger roots.

For any of these pairs, Guichard and Wienhard [25] constructed a connected component U in the space and transverse triples of elements in G/P_{Θ} . They call such triples *positive triples*, and a representation $\rho: \Gamma_g \rightarrow G$ is called Θ -positive if it admits a continuous and ρ -equivariant map $\xi: \partial\Gamma \rightarrow G/P_{\Theta}$ such that for all distinct triples of points $(x, y, z) \in \partial\Gamma^{(3)}$, $(\xi(x), \xi(y), \xi(z)) \in U$. They prove in particular that such representations are Θ -Anosov. Moreover the space of Θ -positive representation is closed in the space of representation that do not virtually factor through a parabolic subgroup, by work of Guichard, Labourie and Wienhard [22].

A Θ -principal Lie subalgebra \mathfrak{h}_{Θ} for a pair (G, Θ) that admits a notion of Θ -positivity is a principal subalgebra of the split Lie subalgebra $\mathfrak{g}_{\Theta} \subset \mathfrak{g}$ generated by all the root spaces associated to Θ ; see [25]. These Lie subalgebra were introduced by Bradlow, Collier, Gothen and García -Prada as *magical triples* in [7]. Let \mathfrak{h}_{Θ} be a Θ -principal \mathfrak{sl}_2 Lie subalgebra of \mathfrak{g} .

Theorem 8.9 [7, Theorem 8.8] *There exists a union of connected components of $\rho: \Gamma \rightarrow G$, the **Cayley components**, consisting only of representations that do not factor through any parabolic subgroup. All \mathfrak{h}_Θ -generalized Fuchsian representations with respect to the principal \mathfrak{sl}_2 Lie subalgebra \mathfrak{h}_Θ lie in some Cayley component.*

Cayley components are conjectured to be all the connected components of Θ -positive representations [7], but there exist components consisting of positive representations that do not contain \mathfrak{h}_Θ -generalized representations, for instance the Gothen components for $G = \mathrm{Sp}(4, \mathbb{R})$. The results of Guichard, Labourie and Wienhard [22] imply the following:

Corollary 8.10 *Every connected component of representations $\rho: \Gamma_g \rightarrow G$ containing an \mathfrak{h}_Θ -generalized Fuchsian representation consists only of Θ -positive representations.*

The sets of simple roots $\Theta \subset \Delta$ that admit a notion of Θ -positivity always admit one or two subsets which are Weyl orbits of simple roots; see Figure 4. Let $\Theta' \subset \Theta$ be a Weyl orbit of simple roots. Let $G \neq \mathrm{PSL}(2, \mathbb{R})$. Theorem 8.7 implies:

Corollary 8.11 *Let $\rho: \Gamma_g \rightarrow G$ be a representation in a connected component of Θ -positive representations containing an \mathfrak{h}_Θ -generalized Fuchsian representation. It is the restricted holonomy of a $(G, \mathcal{F}_{\tau_{\Theta'}})$ -structure on a fiber bundle F over S_g .*

Let χ be the character associated to one of the \mathfrak{h}_Θ -generalized Fuchsian representations in the Cayley component. This fiber bundle is diffeomorphic to the reduction of the $(\mathrm{SO}(2, \mathbb{R}) \times \mathcal{C}_K(\mathfrak{h}))$ -bundle $P_{S_g, \chi}$ via its action on the base $\mathcal{B}^\tau(\mathfrak{h})$ of the pencil of tangent vectors associated to \mathfrak{h} .

The proof of the fact that the associated domains are nonempty as soon as G is not isomorphic to $\mathrm{PSL}(2, \mathbb{R})$ is delayed to Section 8.4.

8.3.2 Hitchin representations in $\mathrm{PSL}(n, \mathbb{R})$ Let \mathfrak{h} be a principal \mathfrak{sl}_2 Lie subalgebra in $\mathfrak{sl}_n(\mathbb{R})$. The associated representation $\iota_{\mathfrak{h}}$ is the representation ι_{irr} from Example 8.3.

Definition 8.12 A representation $\rho: \Gamma_g \rightarrow \mathrm{PSL}(n, \mathbb{R})$ is *Hitchin* if it is a deformation in $\mathrm{Hom}(\Gamma_g, G)$ of an \mathfrak{h} -generalized Fuchsian representation.

The centralizer of \mathfrak{h} in $\mathrm{PSL}(n, \mathbb{R})$ is trivial, so \mathfrak{h} -generalized Fuchsian representations can be written $\iota_{\mathrm{irr}} \circ \rho_0$.

Hitchin proved that the quotient of the space of Hitchin representations by conjugation in $\mathrm{PGL}(n, \mathbb{R})$ is a ball of dimension $(2g - 2)(n^2 - 1)$; see [27]. Labourie proved that Hitchin representations are Borel Anosov, namely Δ -Anosov [32].

The unique Weyl orbit of simple roots for $G = \mathrm{PSL}(n, \mathbb{R})$ is Δ . The flag manifold $\mathcal{F}_{\tau_\Delta}$ can be identified with the flag manifold $\mathcal{F}_{1, n-1}$ consisting of pairs of subspaces (ℓ, H) where $\ell \subset H \subset \mathbb{R}^2$, $\dim(\ell) = 1$ and $\dim(H) = n - 1$. The \mathfrak{sl}_2 Lie subalgebra \mathfrak{h} is Δ -regular. Theorem 8.7 implies the following:

Corollary 8.13 *Let $n \geq 3$. Every Hitchin representation $\rho: \Gamma_g \rightarrow \mathrm{PSL}(n, \mathbb{R})$ is the restricted holonomy of a $(\mathrm{PSL}(n, \mathbb{R}), \mathcal{F}_{1,n-1})$ -structure on a fiber bundle over S_g .*

When $n = 3$, the boundary map of any \mathfrak{h} -generalized Fuchsian representation is an ellipsoid $\mathcal{E} \subset \mathbb{RP}^2$. The domain $\Omega_\rho^{\tau\Delta} \subset \mathcal{F}_{1,2}$ admits three connected components: the set of $(\ell, H) \in \mathcal{F}_{1,2}$ with ℓ in the inside of the ellipsoid \mathcal{E} , with H completely outside of the ellipsoid, and finally with ℓ outside the ellipsoid and H crossing the ellipsoid in two points. One can see that the quotient of this domain is the union of three copies of the projectivization of the tangent bundle of S_g . If we apply Theorem 7.3 to $u_{\mathfrak{h}}$, we obtain a fibration where the model fiber $\mathcal{B}^\tau(\mathfrak{h})$ is the union of three circles described in Example 6.9. Also when $n = 3$ one can get a domain in projective space, as described in Example 7.7, that is the interior of an ellipse.

When n is even, Hitchin representations are also the holonomy of projective structures. We consider these structures in Appendix B.

In general, for any split simple Lie group G , Fock and Goncharov proved that all Δ -positive representations can be deformed into an \mathfrak{h}_Δ -generalized Fuchsian representation [18], namely they lie in a Hitchin component, so our method always applies.

8.3.3 Maximal representations in $\mathrm{PSp}(2n, \mathbb{R})$ Given an orientation of the surface S_g , one can define the *Toledo invariant* $\mathrm{Tol}: \mathrm{Hom}(\Gamma_g, \mathrm{PSp}(2n, \mathbb{R})) \rightarrow \mathbb{Z}$. This continuous map can be defined as the pullback by ρ of an element of the continuous group cohomology $H^2(G, \mathbb{Z})$ of G by ρ [11]. Reversing the orientation of S_g reverses the sign of the Toledo invariant.

A representation $\rho: \Gamma_g \rightarrow \mathrm{PSp}(2n, \mathbb{R})$ is called *maximal* if its Toledo invariant is maximal among all representations, namely if $\mathrm{Tol}(\rho) = n(2g - 2)$. A way to construct maximal representation is to use the representation $\iota_{\mathrm{red}}: \mathrm{PSL}(2, \mathbb{R}) \rightarrow \mathrm{PSp}(2n, \mathbb{R})$. Burger, Iozzi, Labourie and Wienhard proved that maximal representations are $\{\alpha_n\}$ -Anosov [10].

Let \mathfrak{h} be the \mathfrak{sl}_2 Lie subalgebra of $\mathfrak{sp}_{2n}(\mathbb{R})$ which is the image of $d\iota_{\mathfrak{h}}$. Every h -generalized representation is maximal for one of the two orientations of the surface S_g .

Theorem 8.14 [23] *If $n \geq 3$, every maximal representation $\rho: \Gamma_g \rightarrow \mathrm{PSp}(2n, \mathbb{R})$ can be deformed into an $\mathfrak{h}_{\{\alpha_n\}}$ -generalized Fuchsian representation in the sense of Definition 8.6.*

Theorem 8.7 implies:

Corollary 8.15 *Let $n \geq 2$. Every maximal representation $\rho: \Gamma_g \rightarrow \mathrm{PSp}(2n, \mathbb{R})$ is the holonomy of a contact projective structure, namely a $(\mathrm{PSp}(2n, \mathbb{R}), \mathbb{RP}^{2n-1})$ -structure on a fiber bundle.*

Indeed one can consider the Weyl orbit of simple roots $\Theta = \{\alpha_n\}$. The corresponding flag manifold $\mathcal{F}_{\tau_{\{\alpha_n\}}}$ can be identified with \mathbb{RP}^{2n-1} . The Lie subalgebra \mathfrak{h} is $\{\alpha_n\}$ -regular, so one can apply Theorem 8.7.

It is not clear if our method applies to the *Gothen components* of representations $\rho: \Gamma_g \rightarrow \mathrm{PSp}(4, \mathbb{R})$ that contain only Zariski-dense representations [8]. However since $\mathrm{PSp}(4, \mathbb{R}) \simeq \mathrm{SO}_o(2, 3)$, the case $n = 2$ of Corollary 8.15 is a consequence of the work of Collier, Tholozan and Toulisse [13]. We discuss in more depth the relation between their construction and our techniques in Appendix A.

The fiber obtained for Hitchin representations that are also $\{\alpha_n\}$ -positive is a union of connected components of $\mathcal{B}^{\tau\{\alpha_n\}}(\mathfrak{h}_\Delta)$, and for maximal representations one gets a union of connected components of $\mathcal{B}^{\tau\{\alpha_n\}}(\mathfrak{h}_\Theta)$. These two submanifolds of $\mathbb{R}\mathbb{P}^{2n-1}$ are diffeomorphic to the same Stiefel manifold, which is connected if $n \geq 3$; see Appendix B.

8.3.4 Positive representations in $\mathrm{PSO}(p, q)$ Let $G = \mathrm{PSO}(p, q)$ with $q > p$ and $\Theta = \Delta \setminus \{\alpha_p\}$. This pair admits a notion of Θ -positivity. The corresponding flag manifold $\mathcal{F}_{\tau\Theta}$ can be identified with the Grassmannian of isotropic planes in $\mathbb{R}^{p,q}$.

Representations satisfying the Θ -positive property were studied by Beyer and Pozzetti [5]. In particular they showed that all Θ -positive representations $\rho: \Gamma_g \rightarrow \mathrm{PSO}(p, q)$ can be deformed to an \mathfrak{h}_Θ -generalized Fuchsian representation when $q > p + 1$, so in this case Corollary 8.11 applies to all Θ -positive representations. However when $q = p + 1$, there are connected components of Θ -positive representations that are conjectured to contain only Zariski-dense representations; it is not clear if our techniques can be applied to these components.

8.4 Nonempty domains

In order to get a geometric structure associated to a domain of discontinuity, one needs to ensure that the domain is nonempty. Kapovich, Leeb and Porti have a condition that ensures that there exists a thickening such that the domain is not empty [30], and Guichard and Wienhard proved that the domains they considered were not empty by computing the dimension of their complement.

We will use the following criterion to prove that some domains of discontinuity for surface groups are nonempty. Remember that the groups Γ_g that we consider here are surface groups.

Lemma 8.16 *Let $\rho: \Gamma_g \rightarrow G$ be a Θ -Anosov representation and (τ, τ_0) be a balanced type pair of elements in $\mathbb{S}\mathfrak{a}^+$, namely such that τ_0 is τ -regular, satisfying $\Theta(\tau_0) = \Theta$. Suppose that \mathcal{F}_τ has finite fundamental group. The domain of discontinuity $\Omega_\rho^{(\tau, \tau_0)}$ is nonempty.*

Note that it is however not a necessary condition to have a nonempty domain, as we see later for $\mathrm{SO}(2, 3)$. This lemma is very similar to Proposition 6.10.

Proof If the domain is empty, it means that the flag manifold \mathcal{F}_τ can be written as the union for $x \in \partial\Gamma_g$ of the thickenings $K_{\xi_\rho^\Theta(x)}^{\tau_0}$. This union is disjoint since the limit map is transverse. Moreover the limit map ξ_ρ^Θ is continuous (see [6] for instance), which implies that \mathcal{F}_τ fibers over the circle with a compact base. As in Proposition 6.10, this implies that \mathcal{F}_τ has infinite fundamental group. \square

We prove that some domains of discontinuity are nonempty for representations of a surface group Γ_g . Let (G, Θ) be a pair that admits a notion of positivity and let $\Theta' \subset \Theta$ be a Weyl orbit of simple roots.

Proposition 8.17 *The flag manifold $\mathcal{F}_{\tau_{\Theta'}}$ has finite fundamental group, except if G is locally isomorphic to $\text{PSL}(2, \mathbb{R})$ or if G is locally isomorphic to $\text{SO}_o(2, 3)$, $\Theta = \Delta$ and $\Theta' = \{\alpha_2\}$.*

If G is locally isomorphic to $\text{SO}_o(2, 3)$, $\Theta = \Delta$ and $\Theta' = \{\alpha_2\}$ we work by hand using the notation from Section 2.5. The domain of discontinuity associated to a $\{\alpha_2\}$ -Anosov representation $\rho: \Gamma_g \rightarrow \text{SO}_o(2, 3)$ obtained by metric thickening for any $\{\alpha_2\}$ -regular $\tau_0 \in \mathbb{S}\mathfrak{a}^+$ is

$$\Omega_{\rho}^{(\tau_{\alpha_2}, \tau_0)} = \text{Ein}(\mathbb{R}^{2,3}) \setminus \bigcup_{x \in \partial\Gamma_g} \mathbb{P}(\xi_{\rho}^2(x)^{\perp}).$$

For any $x \in \partial\Gamma_g$, The submanifold $\mathbb{P}(\xi_{\rho}^2(x)^{\perp})$ has dimension 1 in $\text{Ein}(\mathbb{R}^{2,3})$, which has dimension 3. Therefore the domain $\Omega_{\rho}^{(\tau_{\alpha_2}, \tau_0)}$ is nonempty.

Lemma 8.16 together with Proposition 8.17 imply the following:

Corollary 8.18 *If G is not locally isomorphic to $\text{PSL}(2, \mathbb{R})$, the domain of discontinuity obtained by metric thickening $\Omega_{\rho}^{(\tau_{\Theta'}, \tau_0)} \subset \mathcal{F}_{\tau_{\Theta'}}$ for any Θ' -regular vector $\tau_0 \in \mathbb{S}\mathfrak{a}^+$ and any Θ -Anosov representation $\rho: \Gamma_g \rightarrow G$ is nonempty.*

The proof of Proposition 8.17 relies on a description of the fundamental groups of flag manifolds associated to real Lie groups.

Let us write $\Delta_0 \subset \Delta$ for the set of roots whose associated root-space is a line. Let $\alpha, \beta \in \Delta$. We define $\epsilon(\alpha, \beta) = (-1)^{(\alpha, \beta^{\vee})}$, namely $\epsilon(\alpha, \beta) = 1$ if α and β are linked by no edge or if they are linked by two edges and α is the longest root, and otherwise $\epsilon(\alpha, \beta) = -1$.

Theorem 8.19 [38, Theorem 1.1] *Let $A \subset \Delta$ be a set of simple roots. The fundamental group of $\mathcal{F}_A = G/P_A$ is the group generated by $(t_{\alpha})_{\alpha \in \Delta_0}$, defined by the relations $t_{\beta}t_{\alpha} = t_{\alpha}(t_{\beta})^{\epsilon(\alpha, \beta)}$ for $\alpha, \beta \in \Delta_0$ and $\alpha \neq \beta$, and $t_{\alpha} = e$ if $\alpha \in \Delta_0 \setminus A$.*

The following lemma will deal with most cases:

Lemma 8.20 *If the Dynkin diagram restricted to Δ_0 of the restricted root system of G has no connected component of type C_n or A_1 , every flag manifold of G has finite fundamental group.*

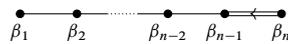
Proof Let $A \subset \Delta$. Using the relations, one may write any element of $\pi_1(\mathcal{F}_A)$ as a product of powers of generators so that each generator appears at most once. Therefore $\pi_1(\mathcal{F}_A)$ is a finite group if and only if for every $\alpha \in A$, t_{α} has finite order.

If no connected component of the Dynkin diagram restricted to Δ_0 is of type C_n or A_1 , then every $\alpha \in \Delta_0$ belongs to a subdiagram inside the Dynkin diagram of Δ_0 of type A_2 , G_2 or B_3 . In each of these cases the relations between the generators imply that they have finite order. Indeed the flag manifold associated to the Borel subgroup of $SL(2, \mathbb{R})$, $SO(3, 4)$ and the real split Lie group associated to G_2 have finite fundamental group; see [38]. \square

Proof of Proposition 8.17 If G is a split Lie group $\Delta_0 = \Delta$. If G is locally isomorphic to $SO(p, q)$ with $p \geq 3$ and $q > p + 1$, the Dynkin diagram restricted to Δ_0 is of type A_{p-1} . Finally if G is the real form of the complex Lie group associated to E_6, E_7 or E_8 whose restricted root system is of type F_4 , the Dynkin diagram restricted to Δ_0 is of type A_2 . This follows from [33, Table 9]. or E_8 whose restricted root system is of type F_4 , the Dynkin

Therefore if (G, Θ) is a pair that admits a notion of Θ -positivity, Lemma 8.20 applies and every flag manifold associated to G has finite fundamental group except in the following two cases: if G is of Hermitian type and of tube type with Θ consisting of only the longest simple root, or if G is split of type A_1 or C_n and $\Theta = \Delta$.

If G is of Hermitian type, the Dynkin diagram of the associated root system is C_n for $n \geq 2$. Suppose that $\Theta' = \{\beta_n\}$. Then $\mathcal{F}_{\tau_{\Theta'}} = \mathcal{F}_{\{\beta_1\}}$, as in Figure 4. Either $\beta_1 \notin \Delta_0$, in which case $\mathcal{F}_{\{\beta_1\}}$ is trivial, or G is split, by [33, Table 9]. If G is split, $\beta_1, \beta_2 \in \Delta_0$, so the generator t_{β_1} of $\pi_1(\mathcal{F})$ satisfies the relation $t_{\beta_1} t_{\beta_2} = t_{\beta_2} (t_{\beta_1})^{-1}$, and $t_{\beta_2} = e$ so $t_{\beta_1}^2 = e$. Therefore $\mathcal{F}_{\tau_{\Theta'}}$ has finite fundamental group.



It remains to consider the case where G is split with root system C_n for $n \geq 3$, $\Theta = \Delta$ and $\Theta' = \{\beta_1, \dots, \beta_{n-1}\}$. But in this case $\mathcal{F}_{\tau_{\Theta'}}$ is the flag manifold associated to the root β_2 , as shown in Figure 4. The root β_2 belongs to a subdiagram of type A_2 , so the fundamental group of $\mathcal{F}_{\Theta'}$ is finite. \square

8.5 Other applications

Theorem 8.5 can also be applied to Gromov hyperbolic groups that are not surface groups. In this subsection we consider representations of fundamental groups of hyperbolic manifolds.

For instance one can consider a compact hyperbolic 3-manifold M with fundamental group Γ and holonomy $\rho_0 : \Gamma \rightarrow \text{PSL}(2, \mathbb{C})$. Let $n \geq 3$, let $\iota_{\text{irr}} : \text{PSL}(2, \mathbb{C}) \rightarrow \text{PSL}(n, \mathbb{C})$ be the irreducible representation as in Example 8.3 and let $\mathfrak{h} = d\iota_{\text{irr}}(\mathfrak{sl}_2(\mathbb{C}))$. As for the real case, \mathfrak{h} is Δ -regular, so in particular $\iota_{\text{irr}} \circ \rho_0$ is Δ -Anosov. Here Δ is the only Weyl orbit of simple roots, and the corresponding flag manifold $\mathcal{F}_{\tau_{\Delta}}$ can be identified with the space $\mathcal{F}_{1,n-1}^{\mathbb{C}}$ of pairs (ℓ, H) where $\ell \subset H \subset \mathbb{C}^n$, ℓ is a line and H is a hyperplane.

The domain of discontinuity associated to a Δ -Anosov representation ρ for one, and hence any, balanced pair of the form (τ_{Δ}, τ_0) is the following, where the limit map of ρ decomposes as $\xi_{\rho}^{\Theta} = (\xi_{\rho}^1, \dots, \xi_{\rho}^{n-1})$:

$$\mathcal{F}_{1,n-1}^{\mathbb{C}} \setminus \bigcup_{x \in \partial\Gamma} \{(\ell, H) \mid \exists 1 \leq k \leq n-1 \text{ such that } \ell \subset \xi_{\rho}^k \subset H\}.$$

The topological dimension of the thickening $K_f^\tau \subset \mathcal{F}_{1,n-1}^\mathbb{C}$ for any flag $f \in G/P_\Delta$ is the maximum for $1 \leq k \leq n - 1$ of the real dimension of $\{(\ell, H) \in \mathcal{F}_{1,n-1}^\mathbb{C} \mid \ell \subset E \subset H\}$ for some $E \subset \mathbb{C}^n$ of dimension k . The dimension of the thickening equals $2n - 4$ and the dimension of the flag manifold equals $4n - 6$, so for $n \geq 3$ the domain is nonempty. Theorem 8.5 therefore implies the following:

Corollary 8.21 *The representation*

$$\iota_{\text{irr}} \circ \rho_0: \Gamma \rightarrow \text{PSL}(n, \mathbb{C})$$

is the restricted holonomy of a $(\text{PSL}(n, \mathbb{C}), \mathcal{F}_{1,n-1}^\mathbb{C})$ -structure on a fiber bundle over M .

One can also consider a hyperbolic n -manifold M for $n \geq 2$ with fundamental group Γ and holonomy $\rho_0: \Gamma \rightarrow \text{SO}_o(1, n)$. Let $\iota: \text{SO}(1, n) \rightarrow \text{SO}(p, pn)$ be the diagonal representation for $n \geq 1$ and let \mathfrak{h} be the image of $d\iota$. Here \mathfrak{h} is $\{\alpha_p\}$ -regular, so in particular $\iota_{\text{irr}} \circ \rho_0$ is $\{\alpha_p\}$ -Anosov. Note that $\{\alpha_p\}$ is a Weyl orbit of simple roots, and the corresponding flag manifold $\mathcal{F}_{\tau_{\{\alpha_p\}}}$ can be identified with the set of isotropic lines $\mathcal{I} \subset \mathbb{P}(\mathbb{R}^{p, pn})$.

The domain of discontinuity associated to a $\{\alpha_p\}$ -Anosov representation ρ for one, and hence any, balanced pair of the form $(\tau_{\{\alpha_p\}}, \tau_0)$ is the following:

$$\mathcal{I} \setminus \bigcup_{x \in \partial\Gamma} \{\ell \mid \ell \subset \xi_\rho^{\{\alpha_p\}}(x)\}.$$

The dimension of the complement equals $(p - 1) + n - 1$ and the dimension of the flag manifold equals $n(p + 1) - 2$, so for $n \geq 2$ and $p \geq 2$ the domain is nonempty. Theorem 8.5 therefore implies the following:

Corollary 8.22 *Let \mathcal{C} be the connected component of $\iota \circ \rho_0$ in the space of $\{\alpha_p\}$ -Anosov representations $\rho: \Gamma \rightarrow \text{PSL}(n, \mathbb{C})$. Every representation in \mathcal{C} is the restricted holonomy of an $(\text{SO}(p, pn), \mathcal{I})$ -structure on a fiber bundle over M .*

The fibers of this fiber bundle can be described as the set of isotropic lines in the intersection of p quadrics in $\mathbb{R}^{p, pn}$.

The centralizer of $\iota(\text{SO}(1, n - 1))$ in $\text{SO}(p, pn)$ has a larger dimension than the centralizer of $\iota(\text{SO}(1, n))$. Indeed let $E \subset \mathbb{R}^{p, pn}$ be the p -dimensional subspace preserved by $\iota(\text{SO}(1, n - 1))$. Any element $g \in \text{SO}(p, pn)$ acting trivially on E^\perp centralizes $\iota(\text{SO}(1, n - 1))$ but only finitely many such elements centralize $\iota(\text{SO}(1, n))$. Therefore if the hyperbolic manifold contains a totally geodesic embedded hypersurface, there exist nontrivial deformations of $\iota \circ \rho$ that one can construct using bending.

Appendix A Maximal representations of surface groups into $\text{SO}_o(2, n)$

In [13] Collier, Tholozan and Toulisse studied maximal representations into $\text{SO}_o(2, n)$, the connected component of the identity of $\text{SO}(2, n)$, for $n \geq 3$. In particular they showed that maximal representations are the restricted holonomies of photon structures which are geometric structures on a fiber bundle. They

also showed a similar result for Einstein structures and Hitchin representations in $SO_o(2, 3) \simeq \mathrm{PSp}(4, \mathbb{R})$. In this appendix we compare the fibration of the domain of discontinuity that they obtained with the one coming from the projection defined in Section 7, when the unique minimal equivariant immersion satisfies the corresponding nearly geodesic property.

Maximal representations are $\{\alpha_1\}$ -Anosov, where $\{\alpha_1\}$ is the root such that $P_{\{\alpha_1\}}$ is the stabilizer of an isotropic line in $\mathbb{R}^{2,n+1}$. The set $\{\alpha_1\} \subset \Theta$ is a Weyl orbit of simple roots. We will denote by $\Omega_\rho^{\tau_\Theta}$ this domain associated to ρ , which can be seen as an open subspace of the Grassmannian of isotropic planes on $\mathbb{R}^{2,n+1}$.

Let $\mathbb{R}^{2,n+1}$ be an $(n+3)$ -dimensional vector space equipped with a symmetric bilinear form q of signature $(2, n+1)$. Let $\mathrm{PSO}_o(2, n) \subset \mathrm{PSL}(n+3, \mathbb{R})$ be the connected component of the identity of the isometry group of q . Let $\mathbb{H}^{2,n} \subset \mathbb{P}(\mathbb{R}^{2,n+1})$ be the space of *timelike* lines, namely lines on which q is negative definite. The bilinear form q induces a pseudo-Riemannian metric of signature $(2, n)$ on $\mathbb{H}^{2,n}$.

For a subspace $V \subset \mathbb{R}^{2,n+1}$, we denote by $\mathrm{Pho}(V)$ the set of *photons* of V , namely the set of planes $P \subset V$ such that $q|_P = 0$. The flag manifold $\mathcal{F}_{\tau_{\{\alpha_2\}}}$ can be identified with the space $\mathrm{Pho}(\mathbb{R}^{2,n+1})$. We denote by $\mathrm{Ein}(V)$ the space of isotropic lines in V . When $n = 2$, the flag manifold $\mathcal{F}_{\tau_{\{\alpha_1\}}}$ can be identified with the space $\mathrm{Ein}(\mathbb{R}^{2,3})$.

Fix an orientation of S_g . A maximal representation $\rho: \Gamma \rightarrow \mathrm{PSO}_o(2, n+1)$ is a representation whose Toledo invariant is maximal. Collier, Tholozan and Toulisse proved the following:

Theorem A.1 [13] *Let $\rho: \Gamma \rightarrow \mathrm{SO}_o(2, n)$ be a maximal representation. There exists a unique ρ -equivariant maximal space-like embedding $\ell: \tilde{S}_g \rightarrow \mathbb{H}^{2,n}$ and a unique minimal ρ -equivariant embedding $u: \tilde{S}_g \rightarrow \mathbb{X}$. The following cocompact domain of discontinuity fibers over the surface:*

$$\Omega_\rho = \mathrm{Pho}(\mathbb{R}^{2,n}) \setminus \bigcup_{x \in \partial\Gamma_g} \{P \in \mathrm{Pho}(\mathbb{R}^{2,n}) \mid P \perp \xi_\rho^1(x)\}.$$

The fibration $\pi: \Omega_\rho \rightarrow \tilde{S}_g$ is such that for $x \in \tilde{S}$,

$$\pi^{-1}(x) = \mathrm{Pho}(\ell(x)^\perp).$$

When $n = 3$ and $\rho: \Gamma \rightarrow \mathrm{SO}_o(2, 3)$ is additionally a Hitchin representation, the following cocompact domain of discontinuity fibers over the surface:

$$\Omega_\rho = \mathrm{Ein}(\mathbb{R}^{2,3}) \setminus \bigcup_{x \in \partial\Gamma_g} \{\ell \in \mathrm{Ein}(\mathbb{R}^{2,3}) \mid \ell \perp \xi_\rho^2(x)\}.$$

The fibration $\pi: \Omega_\rho \rightarrow \tilde{S}_g$ is such that for $x \in \tilde{S}$,

$$\pi^{-1}(x) = \mathrm{Ein}(u(x) \oplus \ell(x)).$$

Theorem A.3 provides another description of the fibration from Theorem A.1; its purpose is to describe the difference with the one constructed in Theorem 7.3.

The symmetric space \mathbb{X} associated to $\text{PSO}_o(2, n + 1)$ can be identified with one connected component of the Grassmannian of oriented planes $U \subset \mathbb{R}^{2,n}$ that are *spacelike*, namely on which q is positive. Moreover it is of *Hermitian type*, so it admits a complex structure J .

Let \mathcal{P} be an oriented 2-pencil of tangent vectors based at $x \in \mathbb{X}$, ie a subspace of dimension 2 of $T_x\mathbb{X}$. It is *holomorphic* (respectively *antiholomorphic*) if J_x fixes \mathcal{P} and for all nonzero $v \in \mathcal{P}$, $(v, J_x(v))$ is a positively (respectively negatively) oriented basis of \mathcal{P} .

To an oriented 2-pencil \mathcal{P} at $x \in \mathbb{X}$ that is not antiholomorphic, with some orthonormal and oriented basis (e, f) , one can associate a holomorphic pencil:

$$\phi(\mathcal{P}) = \langle e - J_x(f), f + J_x(e) \rangle.$$

Note that the pencil $\phi(\mathcal{P})$ does not depend on the positively oriented orthonormal basis (e, f) . Similarly one can associate to a 2-pencil that is not holomorphic the antiholomorphic pencil:

$$\bar{\phi}(\mathcal{P}) = \langle e + J_x(f), f - J_x(e) \rangle.$$

Remark A.2 The pencils $\phi(\mathcal{P})$ and $\bar{\phi}(\mathcal{P})$ can be interpreted as the holomorphic and antiholomorphic parts of the complex 2-dimensional subspace generated by \mathcal{P} in $T_x\mathbb{X}$. In particular they are orthogonal to one another.

The fibers of the fibration from Theorem A.1 can be described as bases of pencils of tangent vectors in \mathbb{X} .

Theorem A.3 Let $\rho: \Gamma_g \rightarrow \text{PSO}_o(2, n)$ be a maximal representation. Let $u: \tilde{S}_g \rightarrow \mathbb{X}$ be the unique minimal ρ -equivariant embedding and $\ell: \tilde{S}_g \rightarrow \mathbb{H}^{2,n}$ the unique maximal ρ -equivariant spacelike embedding. For all $x \in \tilde{S}_g$, the submanifold

$$\text{Pho}(\ell(x)^\perp) \subset \text{Pho}(\mathbb{R}^{2,n}) \simeq \mathcal{F}_{\tau_{\{\alpha_1\}}}$$

is equal to the $\tau_{\{\alpha_1\}}$ -base of the holomorphic pencil of tangent vectors

$$\phi(du(T_x\tilde{S}_g)).$$

When $n = 3$ and $\rho: \Gamma \rightarrow \text{SO}_o(2, 3)$ is additionally a Hitchin representation, for all $x \in \tilde{S}_g$ the submanifold

$$\text{Ein}(\ell(x) \oplus u(x)) \subset \text{Ein}(\mathbb{R}^{2,3}) \simeq \mathcal{F}_{\tau_{\{\alpha_2\}}}$$

is equal to the $\tau_{\{\alpha_2\}}$ -base of the antiholomorphic pencil of tangent vectors

$$\bar{\phi}(du(T_x\tilde{S}_g)).$$

One can compare this fibration with the one provided by Theorem 7.3, if u is τ -nearly geodesic, for which the fiber at $y \in \tilde{S}_g$ was the τ -base of the pencil $du(T_y \tilde{S}_g)$. In particular, in the case of photon structures the two fibration coincide if u is holomorphic, which happens when $\rho: \Gamma_g \rightarrow \text{PSO}_o(2, n)$ factors through the reducible representation $\iota: \text{PSO}_o(2, 1) \rightarrow \text{PSO}_o(2, n)$. However they differ in general, and in particular if ρ factors through the irreducible representation $\iota: \text{PSO}_o(2, 1) \rightarrow \text{PSO}_o(2, 3)$.

From now on, this section is devoted to proving Theorem A.3. We use the notation from [13, Section 3.3]. We first need a preliminary result about the unique minimal immersion for Hitchin representations in $\text{SO}_o(2, 3)$:

Lemma A.4 *Let $\rho: \Gamma_g \rightarrow \text{SO}_o(2, 3)$ be a maximal and Hitchin representation. Its associated minimal immersion $u: \tilde{S}_g \rightarrow \mathbb{X}$ is never tangent to a totally geodesic submanifold $\mathbb{X}_0 \subset \mathbb{X}$ of the form*

$$\mathbb{X}_0 = \{U \in \mathbb{X} \mid U \subset E\}$$

for any hyperplane $E \subset \mathbb{R}^{2,3}$.

The proof of this fact relies on the parametrization of the minimal surface using Higgs bundles and the work of [13]. An introduction to Higgs bundles in $\text{SO}(2, n)$ can be found in [13, Section 2.3], and details on the Higgs bundles associated to Hitchin representations in $\text{SO}(2, 3)$ can be found in [13, Section 4.3].

Proof Given a Hitchin representation $\rho: \Gamma_g \rightarrow \text{SO}(2, 3)$, there exists a unique a conformal structure on S_g such that the corresponding harmonic map in the symmetric space is minimal, and the Higgs bundle (E_ρ, Φ) corresponding to ρ can be written as

$$E_\rho \simeq K^2 \oplus K^1 \oplus \mathcal{O} \oplus K^{-1} \oplus K^{-2}.$$

The Higgs field Φ is the following section of $K \otimes \text{End}(E_\rho)$, or in other words the following one-form valued in $\text{End}(E_\rho)$:

$$\Phi_x = \begin{pmatrix} 0 & 0 & 0 & \gamma & 0 \\ 1 & 0 & 0 & 0 & \gamma \\ 0 & \beta & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Note that 1 is the tautological one-form valued in K^{-1} , a section of $K \otimes K^{-1}$. Note also that for Hitchin representations, $\beta \neq 0$ is also the tautological section, but we denote it differently from 1 because their norm with respect to the harmonic metric are different in general. The harmonic metric on E_ρ is diagonal since the bundle is cyclic [13, Proposition 2.31]. As a consequence, the real form on the complex bundle E_ρ induced by the harmonic metric is the following for some sections h_1, h_2 of $\bar{K}^2 K^{-2}$ and $\bar{K}^1 K^{-1}$, respectively (see [13, Proposition 2.32]):

$$(z_1, z_2, z_3, z_4, z_5) \mapsto (h_1^{-1} \bar{z}_5, h_2^{-1} \bar{z}_4, \bar{z}_3, h_2 \bar{z}_2, h_1 \bar{z}_1).$$

We identify $\Phi + \Phi^{*h}$ with the differential of the minimal immersion u . Let us fix a locally defined holomorphic vector field ∂_z . Suppose that the associated minimal immersion is tangent to a totally geodesic submanifold $\mathbb{X}_0 \subset \mathbb{X}$ of the form $\mathbb{X}_0 = \{U \in \mathbb{X} \mid U \subset E\}$ for some hyperplane $E \subset \mathbb{R}^{2,3}$. If this was the case at a point $x \in \tilde{S}_g$, it would imply that the differential of u in every direction in $T_x \tilde{S}_g$ is identified with a symmetric endomorphism of E_ρ that preserves the complexification H of the real hyperplane in the fiber at x of E_ρ corresponding to $E \subset \mathbb{R}^{2,3}$ via the flat connection. In particular the endomorphisms $\Phi_x(\partial_z) + \Phi_x^{*h}(\partial_z)$ and $i\Phi_x(\partial_z) - i\Phi_x^{*h}(\partial_z)$ preserve H , so the endomorphism $\Phi_x(\partial_z)$ preserves H .

Since $\Phi_x(\partial_z)$ preserves a nondegenerate bilinear form (the one that determines the real form on E_ρ with the harmonic metric) and a real hyperplane, it must admit a real eigenvector $(z_1, z_2, z_3, z_4, z_5)$. This eigenvector being real, one has $z_3 = \bar{z}_3, z_4 = h_2 \bar{z}_2$ and $z_5 = h_1 \bar{z}_1$. The eigenvector equation implies that for some $\lambda \in \mathbb{C}$,

$$\begin{aligned} \lambda z_1 &= \gamma(\partial_z)h_2 \bar{z}_2, & \lambda z_2 &= 1(\partial_z)z_1 + \gamma(\partial_z)h_1 \bar{z}_1, & \lambda z_3 &= \beta(\partial_z)z_2, \\ \lambda h_2 \bar{z}_2 &= \beta(\partial_z)z_3, & \lambda h_1 \bar{z}_1 &= 1(\partial_z)h_2 \bar{z}_2. \end{aligned}$$

Moreover $\|\gamma\| < \|1\|$ for the metric h_1 on K^2 , or in other words $|1(\partial_z)| > h_1|\gamma(\partial_z)|$; see [13, Section 4.3]. Therefore Φ_x cannot admit a nonzero real eigenvector. Indeed if $\lambda = 0$ we have that $z_2, z_3 = 0$ and $0 = 1(\partial_z)z_1 + \gamma(\partial_z)h_1 \bar{z}_1$, so $z_1 = 0$. If $\lambda \neq 0$, $1(\partial_z)h_2 \bar{z}_2 = h_1\gamma(\partial_z)h_2 \bar{z}_2$ implies that $z_2 = 0$ and therefore $z_1, z_3 = 0$. □

Proof of Theorem A.3 Let $\mathcal{G}(\mathbb{H}^{2,n})$ be the space of triples (U, L, N) where $\mathbb{R}^{2,n} = U \oplus L \oplus N$ is an orthogonal splitting, L is a timelike line and U a spacelike plane. The tangent space of $\mathcal{G}(\mathbb{H}^{2,n})$ can be decomposed as

$$T_{(U,L,N)}\mathcal{G}(\mathbb{H}^{2,n}) \simeq \text{Hom}(L, U) \oplus \text{Hom}(L, N) \oplus \text{Hom}(U, N).$$

The generalized Gauss map $\mathcal{G}: \tilde{S} \rightarrow \mathcal{G}(\mathbb{H}^{2,n})$ associated to a spacelike immersion $\ell: \tilde{S}_g \rightarrow \mathbb{H}^{2,n}$ is the unique map $x \mapsto (u(x), \ell(x), N(x))$ such that

$$d\mathcal{G} = (d\ell, 0, A_\ell).$$

Here du is the differential of ℓ :

$$d\ell: T_x \tilde{S}_g \rightarrow \text{Hom}(\ell(x), u(x)) \subset \text{Hom}(\ell(x), \ell(x)^\perp) \simeq T_{\ell(x)}\mathbb{H}^{2,n}.$$

And A_ℓ is related to the second fundamental form Π_ℓ of the immersion ℓ :

$$\begin{aligned} A_\ell: T_x \tilde{S}_g &\rightarrow \text{Hom}(u(x), N(x)) \subset \text{Hom}(u(x), \ell(x)^\perp) \simeq T_{\ell(x)}\mathbb{H}^{2,n} \otimes (d\ell(T_x \tilde{S}_g))^*, \\ \Pi_\ell(v, w) &= A_\ell(v) \circ d\ell(w) \in T_{\ell(x)}\mathbb{H}^{2,n} \simeq \text{Hom}(\ell(x), \ell(x)^\perp). \end{aligned}$$

If ℓ is the unique equivariant maximal spacelike surface for a maximal representation $\rho: \Gamma_g \rightarrow \text{PSO}_o(2, n)$, the unique ρ -equivariant minimal surface $u: \tilde{S}_g \rightarrow \mathbb{X}$ is the map coming from the Gauss map of ℓ [13].

The differential of u in a direction $v \in T_x \tilde{S}_g$ is determined by $d\mathcal{G}$, and is equal to the element $du(v) \in \text{Hom}(u(x), u(x)^\perp) \simeq T_{u(x)}\mathbb{X}$ such that

$$du(v) = A_\ell(v) - (d\ell(v))^T.$$

In this expression $(d\ell(v))^T \in \text{Hom}(u(x), \ell(x))$ is obtained as the adjoint of the endomorphism $d\ell(v)$, for the restriction of the bilinear form q on $\ell(x)$ and $u(x)$.

The endomorphism $J = J_{u(x)}$ coming from the complex structure on \mathbb{X} acts by precomposition by the rotation of $u(x)$ of angle $-\frac{1}{2}\pi$ on $T_{u(x)}\mathbb{X} \simeq \text{Hom}(u(x), u(x)^\perp)$, with the orientation chosen on $u(x)$. The choice of an orientation for every spacelike plane is equivalent to the choice of a complex structure on \mathbb{X} . Since ρ is maximal for the orientation we fixed on S_g , this orientation is preserved by du . We write R for the corresponding rotation of $T_x \tilde{S}_g$, which is a rotation of angle $-\frac{1}{2}\pi$.

Let (v, w) be a positively oriented basis of $T_x \tilde{S}_g$. The endomorphism

$$f = du(v) - J(du(w)) \in \text{Hom}(u(x), u(x)^\perp) \simeq T_{u(x)}\mathbb{X}$$

satisfies

$$\begin{aligned} f(d\ell(v)) &= \Pi_\ell(v, v) - (d\ell(v))^T(d\ell(v)) - \Pi_\ell(w, R \cdot v) + (d\ell(w))^T(d\ell(R \cdot v)), \\ f(d\ell(w)) &= \Pi_\ell(v, w) - (d\ell(v))^T(d\ell(w)) - \Pi_\ell(w, R \cdot w) + (d\ell(w))^T(d\ell(R \cdot w)). \end{aligned}$$

One has $R \cdot v = -w$, $R \cdot w = v$. Also $\Pi_\ell(v, w) = \Pi_\ell(w, v)$, and since ℓ is maximal, $\Pi_\ell(v, v) + \Pi_\ell(w, w) = 0$. So the image of f is included in ℓ , and $f(d\ell(v)) \neq 0$.

This holds for any orthonormal positively oriented basis (v, w) , so $\phi(du(T_x \tilde{S}_g))$ can be identified with the dimension-2 subspace:

$$\text{Hom}(u(x), \ell(x)) \subset \text{Hom}(u(x), u(x)^\perp) \simeq T_{u(x)}\mathbb{X}.$$

It suffices now to show that for any $(U, L, N) \in \mathcal{G}(\mathbb{H}^{2,n})$, the $\tau_{\{\alpha_1\}}$ -base of the holomorphic pencil $\mathcal{P} \simeq \text{Hom}(U, L) \subset \text{Hom}(U, U^\perp) \simeq T_U\mathbb{X}$ is the photon $\text{Pho}(L^\perp)$, independently of U .

Let $a \in \text{Pho}(\mathbb{R}^{2,n})$ be a photon. The vector $v_{a,U} \in T_U\mathbb{X}$ pointing towards a can be identified with the homomorphism $f_a: U \rightarrow U^\perp$ whose graph $\{x + f_a(x) \mid x \in U\}$ is the photon a .

The Riemannian metric on \mathbb{X} induces the following scalar product on $T_{u(x)}\mathbb{X} \simeq \text{Hom}(U, U^\perp)$:

$$\langle f, g \rangle = \text{Tr}(g^\perp \circ f).$$

For this scalar product, f_a is orthogonal to $\text{Hom}(U, L) \subset \text{Hom}(U, U^\perp)$ if and only if its image is orthogonal to L , which is orthogonal to L if and only if $a \in \text{Pho}(L^\perp)$.

Now if $n = 2$ and ρ is maximal and Hitchin, the pencil $\bar{\phi}(du(T_x \tilde{S}_g))$ is orthogonal to $\text{Hom}(u(x), \ell(x)) \simeq \phi(du(T_x \tilde{S}_g))$. But every unit vector $f \in \text{Hom}(u(x), \ell(x))$ is identified to $v_{a,u(x)} \in T_{u(x)}\mathbb{X}$, where $a \in \text{Ein}(\mathbb{R}^{2,n})$ is the graph of $f^T: \ell(x) \rightarrow u(x)$. Therefore $\text{Ein}(u(x) \oplus \ell(x))$ belongs to the $\tau_{\{\alpha_2\}}$ -base

of $\bar{\phi}(du(T_x\tilde{S}_g))$. It only remains to check that this base is not larger. This is due in some sense to the genericity of $\bar{\phi}(du(T_x\tilde{S}_g))$.

Suppose that there exists $\ell_0 \in \text{Ein}(\mathbb{R}^{2,n}) \setminus \text{Ein}(u(x) \oplus \ell(x))$ that belongs to the $\tau_{\{\alpha_2\}}$ -base of $\bar{\phi}(du(T_x\tilde{S}_g))$. The associated element

$$v_{\ell_0, u(x)} \in T_x\mathbb{X} \simeq \text{Hom}(u(x), u(x)^\perp)$$

corresponds to a rank-1 homomorphism $v \in u(x) \mapsto \langle v, u_0 \rangle n_0$, where $u_0 \in u(x)$, $n_0 \in u(x)^\perp$ and $u_0 + n_0$ generates ℓ_0 . Since ℓ_0 belongs to the $\tau_{\{\alpha_2\}}$ -base of $\bar{\phi}(du(T_x\tilde{S}_g))$, one has $\text{Im}(f) \perp n_0$ for any

$$f \in \bar{\phi}(du(T_x\tilde{S}_g)) \subset T_x\mathbb{X} \simeq \text{Hom}(u(x), u(x)^\perp).$$

Let n'_0 be a nonzero vector in $(u(x) \oplus \ell(x) \oplus \ell'(x)) \cap N(x) \subset \mathbb{R}^{2,3}$. It is orthogonal to the image of every element in $\bar{\phi}(du(T_x\tilde{S}_g))$ since $\bar{\phi}(du(T_x\tilde{S}_g))$ is orthogonal to $\phi(du(T_x\tilde{S}_g))$. Moreover the image of every element in $\phi(du(T_x\tilde{S}_g))$ is orthogonal to $N(x)$, so n'_0 is orthogonal to the image of every element in $du(T_x\tilde{S}_g) \subset T_x\mathbb{X} \simeq \text{Hom}(u(x), u(x)^\perp)$. This cannot happen if ρ is Hitchin because of Lemma A.4. Therefore $\text{Ein}(u(x) \oplus \ell(x))$ is the $\tau_{\{\alpha_2\}}$ -base of $\bar{\phi}(du(T_x\tilde{S}_g))$. □

Appendix B Projective structures with (quasi-)Hitchin holonomy

In this appendix we compare the fibration obtained using Theorem 8.5 with the one constructed in [2].

Let $N \geq 3$ be an integer, and let \mathfrak{h} be a principal \mathfrak{sl}_2 -lie subalgebra of $\mathfrak{sl}(N, \mathbb{R})$. The associated totally geodesic immersion $u_{\mathfrak{h}}: \mathbb{H}^2 \rightarrow \mathbb{X}$ is τ_1 -regular if and only if $N = 2n$ is even. Indeed an element $\text{Diag}(\sigma_1, \dots, \sigma_n) \in \mathfrak{a}$ is τ_1 -regular if and only if none of the σ_i are equal to 0, and the Cartan projection of any tangent vector to $u_{\mathfrak{h}}$ is a multiple of $\text{Diag}(N - 1, N - 3, \dots, 3 - N, 1 - N)$. Recall that $\tau_1 \in \mathbb{S}\mathfrak{a}^+$ is such that \mathcal{F}_{τ_1} is identifiable with $\mathbb{R}\mathbb{P}^{2n-1}$.

Any $\{\alpha_n\}$ -Anosov representation ρ admits a cocompact domain of discontinuity

$$\Omega_\rho = \mathbb{R}\mathbb{P}^{2n-1} \setminus \bigcup_{x \in \partial\Gamma_\rho} \mathbb{P}(\xi_x^{\{\alpha_n\}}).$$

This is one of the domains constructed by Guichard and Wienhard [24], and is also the domain $\Omega_\rho^{(\tau_1, \tau_0)}$ as in Theorem 7.10 for any symmetric $\tau_0 \in \mathbb{S}\mathfrak{a}^+$ that is τ_1 -regular. For any Fuchsian representation $\rho_0: \Gamma_g \rightarrow \text{PSL}(2, \mathbb{R})$ the domain associated with $\iota_{\mathfrak{h}} \circ \rho_0$ coincides with the domain $\Omega_{u_{\mathfrak{h}}}^{\tau_1}$. This domain is nonempty for $n \geq 2$.

To understand the topology of the quotient, it is sufficient to understand it for generalized Fuchsian representations, so let $u_{\mathfrak{h}}: \mathbb{H}^2 \rightarrow \mathbb{X}$ be the totally geodesic embedding that is ι_{irr} -equivariant, as defined in Example 8.3.

We will compare two 2-pencils of quadrics. For this we will use an invariant associated to a regular 2-pencil of quadrics, which is called the Segre symbol [17].

A pencil of quadrics is *regular* if it admits a nondegenerate element. Let \mathcal{P} be a 2-pencil with a basis (q_1, q_2) such that q_2 is nondegenerate. There exists a unique q_2 -symmetric or q_2 -Hermitian endomorphism A_{q_1, q_2} of V such that $q_1(\cdot, \cdot) = q_2(A \cdot, \cdot)$.

Lemma B.1 *Let \mathcal{P} be a 2-pencil of quadrics, and (q_1, q_2) and (q'_1, q'_2) be two bases of the pencil with q_1, q_2, q'_1 and q'_2 nondegenerate quadratic forms. Let $A = A_{q_1, q_2}, B = A_{q'_1, q'_2}$ and $a, b, c, d \in \mathbb{R}$ be such that $q'_1 = aq_1 + bq_2$ and $q'_2 = cq_1 + dq_2$. Then $cA + d \text{Id}$ is invertible and $B = (aA + b \text{Id})(cA + d \text{Id})^{-1}$.*

Proof Let $M = aA + b \text{Id}$ and $N = cA + d \text{Id}$. Then $q'_2(\cdot, \cdot) = q_2(N \cdot, \cdot)$. But since q_2 and q'_2 are nondegenerate, then $N = A_{q'_2, q_2}$ is invertible.

To complete the proof of the lemma, we need to check that $q'_1(N \cdot, \cdot) = q'_2(M \cdot, \cdot)$. Let $v, w \in V$, then

$$\begin{aligned} q'_1(Nv, w) &= aq_1(Nv, w) + bq_2(Nv, w) \\ &= adq_1(v, w) + bcq_2(Av, w) + acq_1(Av, w) + bdq_2(v, w), \\ q'_2(Mv, w) &= cq_1(Mv, w) + dq_2(Mv, w) \\ &= caq_1(Av, w) + cbq_1(v, w) + daq_2(Av, w) + dbq_2(v, w), \end{aligned}$$

which concludes the proof since $q_1(\cdot, \cdot) = q_2(A \cdot, \cdot)$. □

We can associate to a 2-pencil of quadrics an equivalence class of Segre symbols in the following way.

Definition B.2 The *Segre symbol* associated to a pair (q_1, q_2) of nondegenerate symmetric or Hermitian forms is the set S of characteristic values of the endomorphism A_{q_1, q_2} , with the weights $w(s)$ given by the multiplicity of the characteristic values and the partition $p(s)$ given by the sizes of the blocks of the Jordan block decomposition of A_{q_1, q_2} .

To a nondegenerate 2-pencil of quadrics \mathcal{P} we can associate the class of all Segre symbols associated to any basis (q_1, q_2) of \mathcal{P} where $q_1, q_2 \in \mathcal{P}$ are nondegenerate, up to the action of $\text{PSL}(2, \mathbb{R})$ on the space of possible Segre symbols. This equivalence class is well defined because of Lemma B.1. The Segre symbol of a 2-pencil is invariant by the action of $GL(V)$, and hence two 2-pencils of quadrics with different Segre symbols are not isomorphic.

Theorem B.3 *The fiber of the fibration $\pi_{u_i}: \Omega_\rho \rightarrow \tilde{S}$ is the base of a nondegenerate 2-pencil of quadrics \mathcal{P} whose Segre symbol is equal to $\{i, -i\}$ with multiplicity n at each point, and a minimal partition $[n]$ of n at each point.*

For $\mathbb{K} = \mathbb{R}$ it is diffeomorphic to the Stiefel manifold

$$O(k)/(O(k-2) \times O(1)).$$

For $\mathbb{K} = \mathbb{C}$ it is diffeomorphic to the Stiefel manifold

$$O(2k)/(O(2k-2) \times U(1)).$$

- [2] **D Alessandrini, C Davalo, Q Li**, *Projective structures with (quasi-)Hitchin holonomy*, J. Lond. Math. Soc. 110 (2024) art. id. e13003 MR Zbl
- [3] **D Alessandrini, S Maloni, N Tholozan, A Wienhard**, *Fiber bundles associated with Anosov representations*, Forum Math. Sigma 13 (2025) art. id. e57 MR Zbl
- [4] **Y Benoist**, *Propriétés asymptotiques des groupes linéaires*, Geom. Funct. Anal. 7 (1997) 1–47 MR Zbl
- [5] **J Beyrer, M B Pozzetti**, *Positive surface group representations in $PO(p, q)$* , J. Eur. Math. Soc. (online publication November 2024) Zbl
- [6] **J Bochi, R Potrie, A Sambarino**, *Anosov representations and dominated splittings*, J. Eur. Math. Soc. 21 (2019) 3343–3414 MR Zbl
- [7] **S Bradlow, B Collier, O García-Prada, P Gothen, A Oliveira**, *A general Cayley correspondence and higher rank Teichmüller spaces*, Ann. of Math. 200 (2024) 803–892 MR Zbl
- [8] **S B Bradlow, O García-Prada, P B Gothen**, *Deformations of maximal representations in $Sp(4, \mathbb{R})$* , Q. J. Math. 63 (2012) 795–843 MR Zbl
- [9] **S Bronstein**, *Almost-Fuchsian structures on disk bundles over a surface*, preprint (2023) arXiv 2305.06665
- [10] **M Burger, A Iozzi, F Labourie, A Wienhard**, *Maximal representations of surface groups: symplectic Anosov structures*, Pure Appl. Math. Q. 1 (2005) 543–590 MR Zbl
- [11] **M Burger, A Iozzi, A Wienhard**, *Surface group representations with maximal Toledo invariant*, Ann. of Math. 172 (2010) 517–566 MR Zbl
- [12] **M Burger, M B Pozzetti**, *Maximal representations, non-Archimedean Siegel spaces, and buildings*, Geom. Topol. 21 (2017) 3539–3599 MR Zbl
- [13] **B Collier, N Tholozan, J Toulisse**, *The geometry of maximal representations of surface groups into $SO_0(2, n)$* , Duke Math. J. 168 (2019) 2873–2949 MR Zbl
- [14] **D Dumas, A Sanders**, *Geometry of compact complex manifolds associated to generalized quasi-Fuchsian representations*, Geom. Topol. 24 (2020) 1615–1693 MR Zbl
- [15] **P B Eberlein**, *Geometry of nonpositively curved manifolds*, Univ. Chicago Press (1996) MR Zbl
- [16] **C L Epstein**, *The hyperbolic Gauss map and quasiconformal reflections*, J. Reine Angew. Math. 372 (1986) 96–135 MR Zbl
- [17] **C Fevola, Y Mandelshtam, B Sturmfels**, *Pencils of quadrics: old and new*, Matematiche (Catania) 76 (2021) 319–335 MR Zbl
- [18] **V Fock, A Goncharov**, *Moduli spaces of local systems and higher Teichmüller theory*, Publ. Math. Inst. Hautes Études Sci. 103 (2006) 1–211 MR Zbl
- [19] **Ś R Gal**, *On normal subgroups of Coxeter groups generated by standard parabolic subgroups*, Geom. Dedicata 115 (2005) 65–78 MR Zbl
- [20] **M Gromov**, *Hyperbolic manifolds, groups and actions*, from “Riemann surfaces and related topics”, Ann. of Math. Stud. 97, Princeton Univ. Press (1981) 183–213 MR Zbl
- [21] **M Gromov, H B Lawson, Jr, W Thurston**, *Hyperbolic 4-manifolds and conformally flat 3-manifolds*, Inst. Hautes Études Sci. Publ. Math. 68 (1988) 27–45 MR
- [22] **O Guichard, F Labourie, A Wienhard**, *Positivity and representations of surface groups*, preprint (2021) arXiv 2106.14584

- [23] **O Guichard, A Wienhard**, *Topological invariants of Anosov representations*, J. Topol. 3 (2010) 578–642 MR Zbl
- [24] **O Guichard, A Wienhard**, *Anosov representations: domains of discontinuity and applications*, Invent. Math. 190 (2012) 357–438 MR Zbl
- [25] **O Guichard, A Wienhard**, *Generalizing Lusztig’s total positivity*, Invent. Math. 239 (2025) 707–799 MR Zbl
- [26] **S Helgason**, *Differential geometry, Lie groups, and symmetric spaces*, Pure Appl. Math. 80, Academic, New York (1978) MR Zbl
- [27] **N J Hitchin**, *Lie groups and Teichmüller space*, Topology 31 (1992) 449–473 MR Zbl
- [28] **M Kapovich, B Leeb**, *Finsler bordifications of symmetric and certain locally symmetric spaces*, Geom. Topol. 22 (2018) 2533–2646 MR Zbl
- [29] **M Kapovich, B Leeb, J Porti**, *Anosov subgroups: dynamical and geometric characterizations*, Eur. J. Math. 3 (2017) 808–898 MR Zbl
- [30] **M Kapovich, B Leeb, J Porti**, *Dynamics on flag manifolds: domains of proper discontinuity and cocompactness*, Geom. Topol. 22 (2018) 157–234 MR Zbl
- [31] **M Kapovich, B Leeb, J Porti**, *A Morse lemma for quasigeodesics in symmetric spaces and Euclidean buildings*, Geom. Topol. 22 (2018) 3827–3923 MR Zbl
- [32] **F Labourie**, *Anosov flows, surface groups and curves in projective space*, Invent. Math. 165 (2006) 51–114 MR Zbl
- [33] **A L Onishchik, È B Vinberg**, *Lie groups and algebraic groups*, Springer (1990) MR Zbl
- [34] **P Planche**, *Géométrie de Finsler sur les espaces symétriques*, PhD thesis, Université de Genève (1995) Zbl
- [35] **J M Riestenberg**, *A quantified local-to-global principle for Morse quasigeodesics*, Groups Geom. Dyn. 19 (2025) 37–107 MR Zbl
- [36] **K K Uhlenbeck**, *Closed minimal surfaces in hyperbolic 3-manifolds*, from “Seminar on minimal submanifolds”, Ann. of Math. Stud. 103, Princeton Univ. Press (1983) 147–168 MR Zbl
- [37] **A Wienhard**, *An invitation to higher Teichmüller theory*, from “Proceedings of the International Congress of Mathematicians, II”, World Sci., Hackensack, NJ (2018) 1013–1039 MR Zbl
- [38] **M Wiggerman**, *The fundamental group of a real flag manifold*, Indag. Math. 9 (1998) 141–153 MR Zbl

Mathematisches Institut, Ruprecht-Karls Universität Heidelberg
Heidelberg, Germany

cdavalo@mathi.uni-heidelberg.de

Proposed: Anna Wienhard
Seconded: David Fisher, Ian Agol

Received: 31 March 2023
Revised: 7 May 2024

Graded character sheaves, HOMFLY-PT homology, and Hilbert schemes of points on \mathbb{C}^2

QUOC P HO

PENGHUI LI

Using a geometric argument building on our new theory of graded sheaves, we compute the categorical trace and Drinfel'd center of the (graded) finite Hecke category $H_W^{\text{gr}} = \text{Ch}^b(\text{SBim}_W)$ in terms of the category of (graded) unipotent character sheaves, upgrading results of Ben-Zvi and Nadler, and Bezrukavnikov, Finkelberg, and Ostrik. In type A , we relate the categorical trace to the category of 2-periodic coherent sheaves on the Hilbert schemes $\text{Hilb}_n(\mathbb{C}^2)$ of points on \mathbb{C}^2 (equivariant with respect to the natural $\mathbb{C}^* \times \mathbb{C}^*$ action), yielding a proof of (a 2-periodized version of) a conjecture of Gorsky, Neţuţ, and Rasmussen which relates HOMFLY-PT link homology and the spaces of global sections of certain coherent sheaves on $\text{Hilb}_n(\mathbb{C}^2)$. As an important computational input, we also establish a conjecture of Gorsky, Hogancamp, and Wedrich on the formality of the Hochschild homology of H_W^{gr} .

14A30, 14F08, 20C15, 57K18; 18N25

1. Introduction	2463
2. Categorical traces and Drinfel'd centers of (graded) finite Hecke categories	2475
3. Formality of the Grothendieck–Springer sheaf	2496
4. A coherent realization of character sheaves via Hilbert schemes of points on \mathbb{C}^2	2503
5. HOMFLY-PT homology via Hilbert schemes of points on \mathbb{C}^2	2519
References	2543

1 Introduction

1.1 Motivation

Let G be a (split) reductive group, $T \subseteq B$ a fixed pair of a maximal torus and a Borel subgroup, and W the associated Weyl group. The geometric finite Hecke category associated to G , defined to be the category of B -bivequariant sheaves/ D -modules on G , plays a central role in representation theory. Of particular interest to us are results of Bezrukavnikov, Finkelberg, and Ostrik [12] (underived) and Ben-Zvi and Nadler [9] (derived) which identify its *categorical trace* and *Drinfel'd center* with the category of

character sheaves constructed by Lusztig in a series of seminal papers [41; 42; 43; 44; 45] as a tool for studying the characters of the finite group $G(\mathbb{F}_q)$.

The finite Hecke category has a *graded* cousin $\text{Ch}^b(\text{SBim}_{\mathcal{W}})$, the monoidal (DG-)category of bounded chain complexes of Soergel bimodules.¹ Like its geometric counterpart, (variations of) graded finite Hecke categories play an important role in geometric representation theory, especially in Koszul duality phenomena. See, for example, Achar, Makisumi, Riche, and Williamson [1], Beilinson, Ginzburg, and Soergel [6], Bezrukavnikov and Yun [14], and Lusztig and Yun [46]. Of particular interest to us is the role these categories play in one construction of the HOMFLY-PT link homology. See Bezrukavnikov and Tolmachov [13], Khovanov [32], Trinh [57], and Webster and Williamson [59].

It is thus natural to study the categorical trace and Drinfel'd center of $\text{Ch}^b(\text{SBim}_{\mathcal{W}})$ with an eye toward *both* categorified link invariants and character sheaves. Indeed, the categorical trace has been studied by Beliakova, Putyra, and Wehrli [7] and Queffelec, Rose, and Sartori [51; 52] (underived), and Gorsky, Hogancamp, and Wedrich [27] (derived) with applications to HOMFLY-PT link homology. Moreover, it was proposed in [27] that an in-depth study of the categorical trace should yield a proof of a conjecture by Gorsky, Neğu, and Rasmussen [29] which relates HOMFLY-PT link homology and global sections of coherent sheaves on the Hilbert schemes of points $\text{Hilb}_n(\mathbb{C}^2)$ on \mathbb{C}^2 .

Despite their importance in both categorified link homology and geometric representation theory, many questions are still left unanswered by these previous works:

- (i) What is the Drinfel'd center of $\text{Ch}^b(\text{SBim}_{\mathcal{W}})$?
- (ii) How are the categorical trace and Drinfel'd center of $\text{Ch}^b(\text{SBim}_{\mathcal{W}})$ related to the category of character sheaves?
- (iii) In type A , how are the trace and centers related to the category of coherent sheaves on $\text{Hilb}_n(\mathbb{C}^2)$?

Various aspects of these questions have been raised by the experts previously, but without a precise answer. See Ben-Zvi and Nadler [9, Section 1.1.5], Webster and Williamson [58, discussion after Theorem A], Gorsky, Hogancamp, and Wedrich [27, Section 1.5], and Gorsky, Kivinen, and Simental [28, Section 7.5].

Here we provide complete answers to the questions above (except the center case of (iii)) and as an application, give a proof of a version (see Remark 1.5.12 for a precise comparison) of the conjecture of Gorsky, Neğu, and Rasmussen.² Moreover, as the main computational input for answering (iii) above, we also establish a conjecture of Gorsky, Hogancamp, and Wedrich [27, Conjecture 1.7] on the formality of the Hochschild homology of $\text{Ch}^b(\text{SBim}_{\mathcal{W}})$, which is a consequence of the formality of the Grothendieck–Springer sheaf in our geometric setup. This formality conjecture is the main obstacle for [27] to make the trace of $\text{Ch}^b(\text{SBim}_{\mathcal{W}})$ more explicit.

¹Note that the notation is somewhat abusive here since $\text{SBim}_{\mathcal{W}}$ depends on the Coxeter system rather than just the Weyl group.

²The link between the Drinfel'd centers and Hilbert schemes of points on \mathbb{C}^2 requires an entirely different technique to establish. The details will appear in a forthcoming paper. We will not need this fact here; see also Remark 1.5.9.

1.2 The approach

1.2.1 Categorical traces and Drinfel'd centers Unlike Beliakova, Putyra, and Wehrli [7] and Gorsky, Hogancamp, and Wedrich [27], our computations of the categorical traces and Drinfel'd centers are completely geometric, inspired by the arguments of Ben-Zvi and Nadler found in [9] and our previous work [30]. This is possible thanks to our new theory of graded sheaves developed in [31], which provides a geometric realization of $\text{Ch}^b(\text{SBim}_W)$ (as the category $\text{Shv}_{\text{gr},c}(B \setminus G/B)$ of B -biequivariant constructible graded sheaves on G) within a sheaf theory with weight structures and perverse t -structures and sufficient functoriality to *do geometry and categorical algebra*. Such a theory is indispensable for our strategy since other geometric constructions of $\text{Ch}^b(\text{SBim}_W)$ via $B \setminus G/B$, such as those of Beilinson, Ginzburg, and Soergel [6], Eberhardt and Scholbach [19], Rider [53], and Soergel, Virk, and Wendt [55], have not been extended to include those spaces such as G/B and G/G (with conjugation actions), geometric objects that appear naturally in the study of categorical traces and Drinfel'd centers.

While the general idea for computing the categorical traces and Drinfel'd centers (see Theorem 1.5.2) is similar to the one in [9; 30], the implementation is more involved. This is because compared to quasicoherent sheaves or D -modules, the categorical Künneth formula fails much more frequently for constructible sheaves. But now, with the modifications done in this paper, the proof of Theorem 1.5.2 works equally well with other sheaf-theoretic settings, such as D -modules (which recovers the result of Ben-Zvi and Nadler in [9]) or sheaves with positive characteristic coefficients (either in the classical analytic topology when the geometric objects involved are defined over \mathbb{C} or in the étale topology when working over $\mathbb{F}_q = \mathbb{F}_{p^n}$ for some prime number p and natural number $n \geq 1$, as long as $\ell \neq p$). We choose to work with ℓ -adic sheaves throughout due to the availability of the theory of graded sheaves in this setting, crucial for applications to Soergel bimodules.

1.2.2 Formality of Hochschild homologies of Hecke categories An interesting feature of our approach to the formality problem conjectured by Gorsky, Hogancamp, and Wedrich in [27] is that its solution does not follow the usual “purity implies formality” route. This is not possible since the algebra of derived endomorphisms of the Grothendieck–Springer sheaf is not pure, already for the simplest case when G is a torus. Instead, it relies on a transcendental argument carried out by the second author in [36] and a spreading argument. In type A , once the categories of graded unipotent character sheaves are computed explicitly using the formality result, their relation to the Hilbert schemes of points on \mathbb{C}^2 can then be deduced as a combination of Koszul duality and results of Bridgeland, King, and Reid [17] and Krug [35].

1.2.3 Hilbert schemes and link invariants In terms of link invariants, the categorical trace of $H_{\text{GL}_n}^{\text{gr}}$ is a universal receptacle for (derived) annular link invariants coming out of $H_{\text{GL}_n}^{\text{gr}}$. As such, HOMFLY-PT homology factors through it. Using an argument involving weight structures in the sense of Bondarko and Pauksztello, we prove a corepresentability result for the degree- a parts of the HOMFLY-PT homology. Matching the corepresenting objects on the Hilbert scheme side, we obtain a proof of a version (see Remark 1.5.12 for a precise comparison) of a conjecture of Gorsky, Neğu, and Rasmussen.

1.2.4 The use of ∞ -categories and higher algebra The use of homotopical/categorical algebra and ∞ -categories as developed by Gaitsgory and Rozenblyum [25; 26] and Lurie [38; 39] is indispensable to our approach at many levels, from the definitions/constructions of the objects and the formulation of the statements to the actual proofs. While being sophisticated, the theory is packaged in such a clean and convenient way that our arguments can still be followed by readers who are not familiar with it. The readers might also consult Ho and Li [31, Appendix A], where most of the relevant background is reviewed.

Remark 1.2.5 Employing completely different techniques involving matrix factorizations, Oblomkov and Rozansky proved a 2-periodized version of work of Gorsky, Neguț, and Rasmussen [29] in a series of papers [48; 49; 50]. It is not clear to us how their work fits with known results in geometric representation theory revolving around *finite* (as opposed to *affine*) Hecke categories and character sheaves. The appearance of affine Hecke categories and their interactions with the finite ones in their work are themselves extremely interesting and deserve a closer look.

Our approach, on the other hand, draws a direct connection between finite Hecke categories, character sheaves, and categorified knot invariants, where the passage from the first to the last one is as in the original construction of HOMFLY-PT homology theory via Soergel bimodules by Khovanov in [32]. In fact, our main result is geometric representation theoretic: we compute both the trace and the Drinfel'd center of the graded finite Hecke categories and relate them to character sheaves, upgrading previous results of Ben-Zvi and Nadler in [9] and Bezrukavnikov, Finkelberg, and Ostrik in [12]. The relation to HOMFLY-PT homology is then deduced as a consequence. We note that the geometric description of the finite Hecke categories is still conjectural in their framework; see Oblomkov and Rozansky [49, Conjecture 7.3.1]).

Their work lives in the *coherent* world whereas our arguments happen in the *constructible* world. Under the Langlands philosophy, there should be a duality between the objects considered in their work and ours. We expect that the two approaches are related via a graded version of Bezrukavnikov's theorem on two geometric realizations of affine Hecke algebras [11]. We plan to revisit this question in a subsequent paper.

In the remainder of the introduction, we will recall the basic objects in Sections 1.3 and 1.4 and then provide the precise statements of the main results in Section 1.5.

1.3 Notation and conventions

Let us now quickly review the basic notation and conventions used throughout the paper.

1.3.1 Category theory We work within the framework of $(\infty, 1)$ -categories (or more briefly, ∞ -categories) as developed by Lurie in [38; 39]. By default, our categories are all ∞ -categories. In particular, we use the language of DG-categories as developed by Gaitsgory and Rozenblyum in this framework [25; 26]. See also Ho and Li [31, Appendix A] for a quick recap. All of our functors and categories are “derived” by default when it makes sense.

We let $\mathrm{DGCat}_{\mathrm{pres}, \mathrm{cont}}$ denote the category whose objects are presentable DG-categories and whose morphisms are continuous (ie *colimit*-preserving) functors. Similarly, we let $\mathrm{DGCat}_{\mathrm{idem}, \mathrm{ex}}$ denote the category

whose objects are idempotent complete DG-categories and whose morphisms are exact (ie finite colimit- and limit-preserving) functors. Finally, let $\text{DGCat}_{\text{pres,cont},c}$ denote the 1-full subcategory (see Gaitsgory and Rozenblyum [25, Chapter 1, Section 1.2.5]) of $\text{DGCat}_{\text{pres,cont}}$ whose objects are compactly generated and whose morphisms are quasiproper functors, ie compact-preserving functors. Each is equipped with a symmetric monoidal structure — the Lurie tensor product. The operation Ind of taking ind-completion is symmetric monoidal and factors as

$$\text{DGCat}_{\text{idem,ex}} \xrightarrow{\text{Ind}'} \text{DGCat}_{\text{pres,cont},c} \xrightarrow{\text{Ind}} \text{DGCat}_{\text{pres,cont}}.$$

Moreover, the first functor is an equivalence of categories.

We will informally refer to objects in $\text{DGCat}_{\text{pres,cont}}$ (resp. $\text{DGCat}_{\text{idem,ex}}$) as “large”/“big” (resp. “small”) categories.

1.3.2 Algebraic geometry Unless otherwise specified our stacks are Artin, with smooth affine stabilizers, and are of finite type over \mathbb{F}_q (or $\overline{\mathbb{F}}_q$, depending on the field of definition), where $q = p^k$ for some prime number p . We let Stk and $\text{Stk}_{\mathbb{F}_q}$ denote the categories of such Artin stacks over \mathbb{F}_q and $\overline{\mathbb{F}}_q$, respectively.

We write $\text{pt}_0 = \text{Spec } \mathbb{F}_q$ and $\text{pt} = \text{Spec } \overline{\mathbb{F}}_q$. Moreover, for any Artin stack \mathcal{Y} over $\overline{\mathbb{F}}_q$, we usually use \mathcal{Y}_0 to denote an \mathbb{F}_q -form of \mathcal{Y} . We will make heavy use of the theory of graded sheaves developed by Ho and Li [31], obtained by categorically semisimplifying Frobenius actions in the category of mixed sheaves; see Beilinson, Bernstein, Deligne, and Gabber [5]. All notation involving the theory of graded sheaves will be the same as in [31]. We will now recall only the main pieces of notation from there.

As the theory of graded sheaves is built on top of the theory of mixed ℓ -adic sheaves, we fix a prime number $\ell \neq p$ throughout this paper. For any Artin stack \mathcal{Y}_0 over \mathbb{F}_q and \mathcal{Y} its base change to $\overline{\mathbb{F}}_q$, we will use $\text{Shv}_{m,c}(\mathcal{Y}_0)$, $\text{Shv}_m(\mathcal{Y}_0)$, and $\text{Shv}_m(\mathcal{Y}_0)^{\text{ren}} := \text{Ind}(\text{Shv}_{m,c}(\mathcal{Y}_0))$ (resp. $\text{Shv}_{gr,c}(\mathcal{Y})$, $\text{Shv}_{gr}(\mathcal{Y})$, and $\text{Shv}_{gr}(\mathcal{Y})^{\text{ren}} := \text{Ind}(\text{Shv}_{gr,c}(\mathcal{Y})^{\text{ren}})$) to denote the DG-category of constructible mixed (resp. graded) sheaves, the DG-category of mixed (resp. graded) sheaves, and the renormalized DG-category of mixed (resp. graded) sheaves on \mathcal{Y}_0 (resp. \mathcal{Y}). The first one is an object in $\text{Shv}_{m,c}(\text{pt}_0)\text{-Mod}$ (resp. $\text{Vect}^{\text{gr},c}\text{-Mod}$) whereas the last two are objects in $\text{Shv}_m(\text{pt}_0)\text{-Mod}$ (resp. $\text{Vect}^{\text{gr}}\text{-Mod}$). All the usual operations on sheaves, when defined, are linear over these categories. Note that here, $\text{Shv}_{gr}(\text{pt}) \simeq \text{Vect}^{\text{gr}}$, the (∞ -derived) symmetric monoidal category of graded chain complexes over $\overline{\mathbb{Q}}_\ell$, and $\text{Shv}_{gr,c}(\text{pt}) \simeq \text{Vect}^{\text{gr},c}$, the full symmetric monoidal subcategory consisting of perfect complexes.

$\text{Shv}_{gr,c}(\mathcal{Y})$ is equipped with a six-functor formalism, a perverse t -structure, and a weight/co- t -structure in the sense of Bondarko and Pauksztello. Moreover, it fits into the diagram

$$\text{Shv}_{m,c}(\mathcal{Y}_0) \xrightarrow{\text{gr}_{\mathcal{Y}_0}} \text{Shv}_{gr,c}(\mathcal{Y}) \xrightarrow{\text{oblv}_{gr}} \text{Shv}_c(\mathcal{Y}),$$

which is compatible with the six-functor formalism, the perverse t -structures, and the Frobenius weights. Furthermore, $\text{oblv}_{gr} \circ \text{gr}_{\mathcal{Y}_0}$ is simply the pullback functor along $\mathcal{Y} \rightarrow \mathcal{Y}_0$. Roughly speaking, $\text{gr}_{\mathcal{Y}_0}$ turns Frobenius weights into an actual grading and oblv_{gr} forgets this grading.

For any $\mathcal{F}, \mathcal{G} \in \text{Shv}_{\text{gr},c}(\mathcal{Y})$, one can talk about the graded $\mathcal{H}\text{om}$ -space $\mathcal{H}\text{om}_{\text{Shv}_{\text{gr},c}(\mathcal{Y})}^{\text{gr}}(\mathcal{F}, \mathcal{G}) \in \text{Vect}^{\text{gr}}$, the category of graded (chain complexes) of vector spaces over $\overline{\mathbb{Q}}_{\ell}$, such that

$$\mathcal{H}\text{om}_{\text{Shv}_c(\mathcal{Y})}(\text{oblv}_{\text{gr}} \mathcal{F}, \text{oblv}_{\text{gr}} \mathcal{G}) \simeq \text{oblv}_{\text{gr}} \mathcal{H}\text{om}_{\text{Shv}_{\text{gr},c}(\mathcal{Y})}^{\text{gr}}(\mathcal{F}, \mathcal{G}) \simeq \bigoplus_k \mathcal{H}\text{om}_{\text{Shv}_{\text{gr},c}(\mathcal{Y})}^{\text{gr}}(\mathcal{F}, \mathcal{G})_k,$$

where for any $V \in \text{Vect}^{\text{gr}}$, $V_k \in \text{Vect}$ denotes the k^{th} graded component of V . We also write

$$\text{End}_{\text{Shv}_{\text{gr},c}(\mathcal{Y})}^{\text{gr}}(\mathcal{F}) := \mathcal{H}\text{om}_{\text{Shv}_{\text{gr},c}(\mathcal{Y})}^{\text{gr}}(\mathcal{F}, \mathcal{F}).$$

See Ho and Li [31, Appendix A.2.6] for a quick review on enriched Hom-spaces.

Since we always make use of the renormalized categories of sheaves (for example, $\text{Shv}_m(\mathcal{Y}_0)^{\text{ren}}$ and its graded counterpart $\text{Shv}_{\text{gr}}(\mathcal{Y})^{\text{ren}}$) introduced by Arinkin, Gaitsgory, Kazhdan, Raskin, Rozenblyum, and Varshavsky [4] and used extensively in [31] rather than the usual categories, we will omit *ren* from the notation of the various pull and push functors. For example, we will simply write f^* , f_* , $f^!$, and $f_!$ rather than f_{ren}^* , $f_{*,\text{ren}}$, $f_{\text{ren}}^!$, and $f_{!,\text{ren}}$ as in [31].

Remark 1.3.3 We deviate slightly from the convention used in [31] in two places:

- (i) The category $\text{Vect}^{\text{gr}}\text{-Mod}$ of Vect^{gr} -module categories is denoted by $\text{Mod}_{\text{Vect}^{\text{gr}}}$ there.
- (ii) An \mathbb{F}_q -form of \mathcal{Y} is denoted by \mathcal{Y}_1 there rather than \mathcal{Y}_0 as we do here. What we use here conforms to the standard convention employed by, for example, Kiehl and Weissauer [33]. More generally, in [31], \mathcal{Y}_n is used to denote an \mathbb{F}_{q^n} -form of \mathcal{Y} . This is necessary because we have to deal with different forms of the same stack over different fields of definition. For example, see [31, Section 2.7].

1.3.4 Grading conventions The different grading conventions appearing in the theory of HOMFLY-PT link homology can be a source of confusion (at least to the authors). We will now collect the various grading conventions we use and where they appear. Serving as a point of orientation, this subsection should thus be skipped, to be returned to only when the need arises.

In Vect^{gr} , we use X to denote the formal grading and C the cohomological grading. Sometimes, to emphasize the grading convention in the presence of other gradings, we also use $\text{Vect}^{\text{gr}X}$ in place of Vect^{gr} . As most DG-categories in this paper are module categories over Vect^{gr} , given two objects c_1 and c_2 in such a category \mathcal{C} , the Vect^{gr} -enriched Hom, $\mathcal{H}\text{om}_{\mathcal{C}}^{\text{gr}}(c_1, c_2) \in \text{Vect}^{\text{gr}}$ is a graded chain complex. The gradings X and C therefore also apply to $\mathcal{H}\text{om}_{\mathcal{C}}^{\text{gr}}(c_1, c_2)$. We refer to this as the singly graded situation (as the cohomological grading is not a formal grading). This is our *default* grading.

The 2-periodic construction in Section 4.3 allows one to exchange an extra grading with 2-periodization. It is used in Section 4.4 to introduce an extra grading on (the $\mathcal{H}\text{om}^{\text{gr}}$ in) any Vect^{gr} -module category by turning it into a $\text{Vect}^{\text{gr}X \cdot \text{gr}Y}$ -module category. As the notation suggests, the extra grading is denoted by Y . Moreover, all the cohomological shearing and 2-periodization in this paper will be with respect to this Y -grading. In other words, the Y -grading is artificially introduced and then gets “canceled out.”

	small	big/renormalized
mixed	$H_{G_0}^m := \text{Shv}_{m,c}(B_0 \setminus G_0 / B_0)$	$H_{G_0}^{m,\text{ren}} := \text{Shv}_m(B_0 \setminus G_0 / B_0)^{\text{ren}} := \text{Ind}(\text{Shv}_{m,c}(B_0 \setminus G_0 / B_0)) \simeq \text{Ind}(H_{G_0}^m)$
graded	$H_G^{\text{gr}} := \text{Shv}_{\text{gr},c}(B \setminus G / B)$	$H_G^{\text{gr,ren}} := \text{Shv}_{\text{gr}}(B \setminus G / B)^{\text{ren}} := \text{Ind}(\text{Shv}_{\text{gr},c}(B \setminus G / B)) \simeq \text{Ind}(H_G^{\text{gr}})$
ungraded	$H_G := \text{Shv}_c(B \setminus G / B)$	$H_G^{\text{ren}} := \text{Shv}(B \setminus G / B)^{\text{ren}} := \text{Ind}(\text{Shv}_c(B \setminus G / B)) \simeq \text{Ind}(H_G)$

Table 1: Different versions of finite Hecke categories.

In Section 4.4.6, the \tilde{X}, \tilde{Y} -grading is introduced, which is related to the X, Y -grading via a linear change of coordinates. This is the grading that is used in the statements of the main theorems regarding the conjecture of Gorsky, Neguţ and Rasmussen [29]. The switch is necessary to match with the usual grading convention on the HOMFLY-PT homology side.

Finally, on the HOMFLY-PT homology, the most natural gradings, from our geometric point of view, are given by $Q', A',$ and T' , defined in Section 5.2.4. The relations between $Q', A', T', Q, A, T,$ and q, a, t are also given there, where the last two grading conventions appear in work of Gorsky, Kivinen, and Simental [28]. See also Remark 5.2.7 for the source of the difference between Q', A', T' and Q, A, T .

1.4 The main players

We will now recall the definitions of the various objects/constructions that appear in our main results.

1.4.1 The various versions of finite Hecke categories Let G_0 be a split reductive group over \mathbb{F}_q , equipped with a pair $T_0 \subseteq B_0$ of a maximal torus and a Borel subgroup. Let W be the associated Weyl group. By convention, $G, T,$ and B are the pullbacks of $G_0, T_0,$ and B_0 to $\overline{\mathbb{F}}_q$.

While our primary interest lies in the graded finite Hecke category $H_G^{\text{gr}} := \text{Shv}_{\text{gr},c}(B \setminus G / B)$, which is equivalent to $\text{Ch}^b(\text{SBim}_W)$ by Ho and Li [31, Theorem 4.4.1],³ it is necessary to consider its “big” or renormalized version $H_G^{\text{gr,ren}} := \text{Ind}(H_G^{\text{gr}})$ because the “big” world possesses more functorial symmetries, such as the adjoint functor theorem (see Lurie [38, Corollary 5.5.2.9]) and the fact that compactly generated (presentable) stable ∞ -categories are dualizable; see Gaitsgory and Rozenblyum [25, Proposition 7.3.2].⁴ Consequently, we will most of the time start with the renormalized case, from which we deduce the corresponding result for the small variants.

Due to their geometric nature, our arguments apply equally well to other variants of the finite Hecke categories, such as the mixed and ungraded versions, which are of independent interest. For the reader’s convenience, we summarize all the different variants in Table 1 (see also Section 1.3.2 for our conventions regarding algebraic geometry).

³See also note 1.

⁴Here, “big” refers to the fact that we are working with presentable categories. On the other hand, “renormalized” refers to the fact that we use the renormalized sheaf theory on stacks.

We also will adopt the notation $H_G^{?,ren}$, $H_G^?$, $\text{Shv}_?(pt)$, and $\text{Shv}_?(Y)$ etc in statements where $?$ can be gr , m , or $\text{nothing}/\emptyset$, that is, the ungraded case, eg $H_G^? = H_G = \text{Shv}_c(B \setminus G/B)$. In this notation, $H_G^{?,ren} \in \text{Alg}(\text{Shv}_?(pt)\text{-Mod})$ and $H_G^? \in \text{Alg}(\text{Shv}_{?,c}(pt)\text{-Mod})$. Unless otherwise specified, $?$ can be any of the three.

Remark 1.4.2 (abuse of notation) The subscript 0 in, for example, G_0 , B_0 , T_0 , and pt_0 etc can be unwieldy and even conflicts with the use of $?$ introduced above. For example, $\text{Shv}_?(Y_0)$ is $\text{Shv}(Y_0)$ in the ungraded case, which does not really make sense, as it should be $\text{Shv}(Y)$ instead in this case. Therefore, we will, in most cases, drop it altogether without fear of confusion. For example, $\text{Shv}_m(BG)^{ren}$ and $\text{Shv}_m(pt)$ only make sense when we take BG and pt to be BG_0 and pt_0 , respectively. Similarly, we will use H_G^m and $H_G^{m,ren}$ rather than the more precise $H_{G_0}^m$ and $H_{G_0}^{m,ren}$.

Remark 1.4.3 The ungraded renormalized Hecke category H_G^{ren} is larger than that of Ben-Zvi and Nadler [9]. Ignoring the distinction between D -modules and constructible sheaves, their category corresponds to $\text{Shv}(B \setminus G/B)$, the usual (as opposed to *renormalized*) category of sheaves on $B \setminus G/B$.

1.4.4 Categorical trace and Drinfel'd center As mentioned earlier, our main goal is to study the categorical trace and Drinfel'd center of H_G^{gr} . Recall the following definition:

Definition 1.4.5 (Ben-Zvi, Francis and Nadler [8, Definition 5.1]) Let A be an associative algebra object in a closed symmetric monoidal ∞ -category \mathcal{C} .

- (i) The (derived) *trace* or *categorical Hochschild homology* of A , denoted by $\text{Tr}(A) \in \mathcal{C}$, is the relative tensor $A \otimes_{A \otimes A^{rev}} A$. It comes with a natural universal trace morphism $\text{tr}: A \rightarrow \text{Tr}(A)$ given by $\text{tr}(a) = a \otimes 1_A$.
- (ii) The (derived) *Drinfel'd center* or *categorical Hochschild cohomology* of A , denoted by $Z(A) \in \mathcal{C}$, is the endomorphism object $\underline{\text{End}}_{A \otimes A^{rev}}(A)$. It comes with a natural central morphism $z: Z(A) \rightarrow A$ given by $z(F) = F(1_A)$.

Here, we view A as a bimodule over itself. Moreover, A^{rev} denotes an associative algebra with the same underlying object as A but with reversed multiplications.⁵

Remark 1.4.6 The trace also has a natural S^1 -action. Moreover, the center is naturally equipped with an E_2 -structure (Deligne's conjecture), ie in the case of categories, it has a braided monoidal structure. See Lurie [39, Sections 5.3 and 5.5].

Note that $B \setminus G/B \simeq BB \times_{BG} BB$ and the monoidal structures on the various versions of finite Hecke categories are given by $g_! f^*$ (with the appropriate sheaf theory) in the following diagram:

$$(1.4.7) \quad BB \times_{BG} BB \times_{BG} BB \xleftarrow[\text{id} \times \Delta \times \text{id}]{f} BB \times_{BG} BB \times_{BG} BB \xrightarrow[\text{p}_{13}]{g} BB \times_{BG} BB.$$

⁵In the literature, A^{op} is also used to denote A^{rev} . We opt to use A^{rev} since for a category A , A^{op} is usually already understood as the *opposite* category.

Depending on whether we work with the small or big/renormalized versions, these Hecke categories are naturally algebra objects in $\text{Shv}_{\gamma,c}(\text{pt})\text{-Mod}$ or $\text{Shv}_{\gamma}(\text{pt})\text{-Mod}$, where the symmetric monoidal structures of the latter are given by *relative* tensors. See also Ho and Li [31, Appendices A.2 and A.7]. Their traces and Drinfel'd centers are computed as objects in these symmetric monoidal categories. In other words, the ambient categories \mathcal{C} in Definition 1.4.5 that are relevant to us are the various categories of module categories $\text{Shv}_{\gamma,c}(\text{pt})\text{-Mod}$ or $\text{Shv}_{\gamma}(\text{pt})\text{-Mod}$.

1.5 The main results

1.5.1 Trace and center of Hecke categories We start with the computations of the trace and Drinfel'd center. The result below upgrades the main result of Ben-Zvi and Nadler [9] to the graded setting.

Theorem 1.5.2 (Theorem 2.8.3) *The trace $\text{Tr}(H_G^{?,\text{ren}})$ and center $Z(H_G^{?,\text{ren}})$ of $H_G^{?,\text{ren}}$ coincide with the full subcategory $\text{Ch}_G^{u,?,\text{ren}}$ of $\text{Shv}_{\gamma}(G/G)^{\text{ren}}$ generated under colimits by the essential image of $H_G^{?,\text{ren}}$ under $q_!p^*$ in the correspondence*

$$\begin{array}{ccc} & G/B & \\ p \swarrow & & \searrow q \\ B \backslash G/B & & G/G \end{array}$$

Moreover, under this identification, the canonical trace and center maps are adjoint $\text{tr} \dashv z$ and are identified with the adjoint pair $q_!p^* \dashv p_*q^!$.

The trace $\text{Tr}(H_G^{?,\text{ren}})$ coincides with the full subcategory $\text{Ch}_G^{u,?,\text{ren}}$ of $\text{Ch}_G^{u,?,\text{ren}}$ spanned by compact objects. Namely,

$$\text{Ch}_G^{u,?,\text{ren}} := (\text{Ch}_G^{u,?,\text{ren}})^c = \text{Ch}_G^{u,?,\text{ren}} \cap \text{Shv}_{\gamma,c}\left(\frac{G}{G}\right).$$

Equivalently, they are generated under finite colimits and idempotent splittings by the essential image of $H_G^{?,\text{ren}}$ under $q_!p^*$.

The center $Z(H_G^{?,\text{ren}})$ is the full subcategory $\widetilde{\text{Ch}}_G^{u,?,\text{ren}}$ of $\text{Ch}_G^{u,?,\text{ren}}$ spanned by the **preimage** of $H_G^{?,\text{ren}}$ under the central functor z .

Remark 1.5.3 The Drinfel'd center of $H_G^{?,\text{ren}}$, that is, working within the context of small categories, is larger than its trace, unlike in the large category setting. This is true already in the case $G = T$ is a torus; see Remark 2.8.4. This seems to be new.

1.5.4 Formality of Hochschild homology By the description of the trace map above, the image of the unit $\text{tr}(1_{H_G^{\text{gr,ren}}}) \in \text{Ch}_G^{u,\text{gr}}$ is, by definition, the (graded) Grothendieck–Springer sheaf Spr_G^{gr} , a graded refinement of the usual Grothendieck–Springer sheaf, ie $\text{oblv}_{\text{gr}}(\text{Spr}_G^{\text{gr}}) \simeq \text{Spr}_G \in \text{Ch}_G^u \subseteq \text{Shv}_c(G/G)$. By Gaitsgory, Kazhdan, Rozenblyum, and Varshavsky [24, Theorem 3.8.5], the graded ring of endomorphisms of this object is the Hochschild homology $\text{HH}(H_G^{\text{gr,ren}})$ of $H_G^{\text{gr,ren}}$, where $H_G^{\text{gr,ren}}$ is viewed as a dualizable object in $\text{Vect}^{\text{gr}}\text{-Mod}$.

Theorem 1.5.5 (Theorems 2.10.11 and 3.2.1, Gorsky, Hogancamp, and Wedrich [27, Conjecture 1.7]) *The graded algebra $\mathrm{HH}(\mathrm{H}_G^{\mathrm{gr}, \mathrm{ren}}) \simeq \mathrm{End}_{\mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}, \mathrm{ren}}}(\mathrm{Spr}_G^{\mathrm{gr}}) \in \mathrm{Alg}(\mathrm{Vect}^{\mathrm{gr}})$ is formal. Moreover, we have an equivalence of algebras*

$$\mathrm{End}_{\mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}, \mathrm{ren}}}(\mathrm{Spr}_G^{\mathrm{gr}}) \simeq \mathrm{H}^*(\mathrm{End}_{\mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}, \mathrm{ren}}}(\mathrm{Spr}_G^{\mathrm{gr}})) \simeq (\mathrm{H}_{\mathrm{gr}}^*(BT) \otimes \mathrm{H}_{\mathrm{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_\ell[W],$$

where $\mathrm{H}_{\mathrm{gr}}^*(-)$ is the functor of taking graded cohomology, ie we remember the Frobenius weights on the cohomology (see Ho and Li [31, (4.2.1)]), and where the action of $\overline{\mathbb{Q}}_\ell[W]$ on the first factor is induced by the natural actions of W on T and BT .

In type A , a result of Lusztig states that $\mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}}$ is generated by $\mathrm{Spr}_G^{\mathrm{gr}}$. As a consequence, we obtain the following result:

Theorem 1.5.6 (Theorem 3.3.4) *When G is of type A , we have*

$$\mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}} \simeq \mathrm{End}_{\mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}}}(\mathrm{Spr}_G^{\mathrm{gr}})\text{-Mod}(\mathrm{Vect}^{\mathrm{gr}})^c \simeq (\mathrm{H}_{\mathrm{gr}}^*(BT) \otimes \mathrm{H}_{\mathrm{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_\ell[W]\text{-Mod}(\mathrm{Vect}^{\mathrm{gr}})^{\mathrm{perf}},$$

where for any graded ring $R \in \mathrm{Alg}(\mathrm{Vect}^{\mathrm{gr}})$, $R\text{-Mod}(\mathrm{Vect}^{\mathrm{gr}})$ denotes the category of R -module objects in $\mathrm{Vect}^{\mathrm{gr}}$, ie the category of graded R -modules, and the superscript perf denotes the full subcategory spanned by perfect complexes (or equivalently, compact objects).

1.5.7 The Hilbert scheme of points on \mathbb{C}^2 While the results above are of independent interest in representation theory, their relation to the HOMFLY-PT link homology and the conjecture of Gorsky, Neguț, and Rasmussen is one of our main motivations.

The first link is given by the following result, which relates the categories of unipotent character sheaves and Hilbert schemes⁶ of points on \mathbb{C}^2 . It is a consequence of Theorem 1.5.6, Koszul duality, and a result of Krug in [35].

Theorem 1.5.8 (Theorem 4.4.15) *When $G = \mathrm{GL}_n$, we have an equivalence of $\mathrm{Vect}^{\mathrm{gr}, \tilde{x}, \mathrm{gr}, \tilde{y}, 2\text{-per}}$ -module categories (thus, also as DG-categories):*

$$\Rightarrow \mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}, \mathrm{ren}} \simeq \mathbb{C}[\tilde{x}, \tilde{y}] \otimes \mathbb{C}[\mathrm{S}_n]\text{-Mod}_{\mathrm{nilp}_{\tilde{y}}}^{\mathrm{gr}, \tilde{x}, \mathrm{gr}, \tilde{y}, 2\text{-per}} \xrightarrow{\Psi^{2\text{-per}}} \mathrm{QCoh}(\mathrm{Hilb}_n / \mathbb{G}_m^2)_{\mathrm{Hilb}_{n, \tilde{x}}}^{2\text{-per}}.$$

Similarly, we have the small variant, which is an equivalence of $\mathrm{Vect}^{\mathrm{gr}, \tilde{x}, \mathrm{gr}, \tilde{y}, 2\text{-per}, c}$ -module categories:

$$\Rightarrow \mathrm{Ch}_G^{\mathrm{u}, \mathrm{gr}} \simeq \mathbb{C}[\tilde{x}, \tilde{y}] \otimes \mathbb{C}[\mathrm{S}_n]\text{-Mod}_{\mathrm{nilp}_{\tilde{y}}}^{\mathrm{gr}, \tilde{x}, \mathrm{gr}, \tilde{y}, 2\text{-per}} \xrightarrow{\Psi^{2\text{-per}}} \mathrm{Perf}(\mathrm{Hilb}_n / \mathbb{G}_m^2)_{\mathrm{Hilb}_{n, \tilde{x}}}^{2\text{-per}}.$$

Here,

- (i) $\mathrm{Vect}^{\mathrm{gr}, \tilde{x}, \mathrm{gr}, \tilde{y}, 2\text{-per}}$ denotes the category of 2-periodic bigraded chain complexes (see Definition 4.4.11 for the precise definition);
- (ii) \tilde{x} and \tilde{y} are multivariables (\tilde{x} denotes $\tilde{x}_1, \dots, \tilde{x}_n$ and similarly for \tilde{y}) living in cohomological degree 0 and bigraded degrees $(1, 0)$ and $(0, 2)$, respectively;

⁶Strictly speaking, the Hilbert schemes that appear in our paper are over $\overline{\mathbb{Q}}_\ell$. However, we have an abstract isomorphism $\mathbb{C} \simeq \overline{\mathbb{Q}}_\ell$ as fields. We will elide this inconsequential difference in the introduction section.

- (iii) similarly, the action of \mathbb{G}_m^2 on Hilb_n is induced by its action on \mathbb{A}^2 , scaling the two axes with weights $(1, 0)$ and $(0, 2)$, respectively;
- (iv) the subscript $\text{nil}_{\tilde{y}}$ indicates the fact that we only consider the full subcategories where the variables \tilde{y} act (ind-)nilpotently;
- (v) the subscript $\text{Hilb}_{n, \tilde{x}}$ indicates the fact that we only consider quasicoherent sheaves with set-theoretic supports on the closed subscheme of Hilb_n consisting of points supported on the first axis of \mathbb{A}^2 (see Section 4.2.9 for the precise definition); and
- (vi) the superscripts \Rightarrow and 2-per denote cohomological shear and 2-periodization constructions (see Sections 4.3 and 4.4 for an in-depth discussion on these constructions and Definition 4.4.10 for the definitions of the 2-periodized categories of sheaves on the Hilbert schemes).

Remark 1.5.9 Theorem 1.5.8 above establishes the relation between the categorical trace of $H_{GL_n}^{\text{gr}}$ and the category of coherent sheaves on Hilb_n set-theoretically supported “on the x -axis.” The center, on the other hand, corresponds to the whole category of coherent sheaves on Hilb_n without any support condition. The details will appear in a forthcoming paper.

1.5.10 HOMFLY-PT link homology Using an argument involving weight structures in the sense of Bondarko and Pauksztello, we prove a corepresentability result for the degree- a parts of the HOMFLY-PT homology. Matching the corepresenting objects on the Hilbert scheme side, we obtain a proof of a version of a conjecture of Gorsky, Neguț, and Rasmussen.

Theorem 1.5.11 (Theorem 5.4.6, Gorsky, Kivinen, and Simental [28, Conjecture 7.2(a)]) *For any $R_\beta \in H_n^{\text{gr}}$ associated to a braid β , there exists a natural $\mathcal{F}_\beta \in \text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{x}}}^{2\text{-per}}$ such that the a -degree α component of the HOMFLY-PT homology of β is given by*

$$\mathcal{H}\text{om}_{\text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)^{2\text{-per}}}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}}(\wedge^\alpha \mathcal{T}^{2\text{-per}}, \mathcal{F}_\beta),$$

where \mathcal{T} is the tautological bundle. The two formal gradings (denoted by \tilde{X} and \tilde{Y}) coming from the \mathbb{G}_m^2 action coincide with q and \sqrt{t} . Moreover, the cohomological degree corresponds to \sqrt{qt} .

See Theorem 5.4.6 for a more precise formulation.

Remark 1.5.12 Theorem 1.5.11 differs from the conjecture of Gorsky, Neguț, and Rasmussen as formulated in [28] in two important aspects. First, we work with the 2-periodized version of the Hilbert scheme of points on \mathbb{C}^2 rather than the usual version; see Remark 1.5.13 below for brief discussion of why this is necessary for us. And second, the degrees \tilde{X} and \tilde{Y} correspond to q and \sqrt{t} for us rather than q and t ; we believe that this is the correct torus action to consider. Moreover, compared to the conjecture of Gorsky, Neguț, and Rasmussen as formulated in [29], our version and the version formulated in [28] use the usual rather than the flagged version of the Hilbert scheme; see Remark 1.5.14 for a speculation on how to relate the two.

Remark 1.5.13 Theorem 1.5.8 relates the category of graded unipotent character sheaves for GL_n with the 2-periodic category of (quasi)coherent sheaves on Hilb_n , equivariant with respect to the scaling \mathbb{G}_m^2 -action. As there is only *one* formal grading on the character sheaf side compared to the two \mathbb{G}_m actions on the Hilbert scheme side, 2-periodization is necessary to relate the two. Because of this, our theorem regarding link invariants above relates HOMFLY-PT homology and 2-periodic quasicohherent sheaves on Hilbert schemes rather than the precise version in [28]. The same phenomenon also appeared in the work of Oblomkov and Rozansky.

A recent work of Elias [21] introduces an extra formal grading on the Hecke category. Even though it is highly suggestive, we do not know how this helps remove the necessity of 2-periodization.

Remark 1.5.14 The main conjecture of [29] relates H_n^{gr} and the category of coherent sheaves on the (derived) flag Hilbert scheme FHilb_n of \mathbb{C}^2 via a pair of adjoint functors. More precisely, they relate unbounded versions of these categories, which we expect to correspond to $H_n^{\mathrm{gr}, \mathrm{ren}}$ and $\mathrm{IndCoh}(\mathrm{FHilb}_n)$, respectively. We also expect that their pair of adjoint functors fits with ours in the diagram

$$H_n^{\mathrm{gr}, \mathrm{ren}} \rightleftarrows \mathrm{IndCoh}(\mathrm{FHilb}_n)_{\mathrm{FHilb}_n, \mathbb{X}} \rightleftarrows \mathrm{IndCoh}(\mathrm{Hilb}_n)_{\mathrm{Hilb}_n, \mathbb{X}} \simeq \mathrm{QCoh}(\mathrm{Hilb}_n)_{\mathrm{Hilb}_n, \mathbb{X}},$$

where we suppress 2-periodization and equivariant parameter(s) altogether. Here, the compositions are our trace and central functors. Moreover, the second adjoint pair should be given by pulling and pushing along the natural map $\mathrm{FHilb}_n \rightarrow \mathrm{Hilb}_n$.

In work of Elias [20], the category of coherent sheaves over FHilb_n is speculated to be related to a subcategory of the Hecke category generated by the Jucys–Murphy elements, which are themselves certain relative centralizers. The methods developed in (the first part of) this paper could be extended to study these relative centers geometrically. But it is unclear to us at the moment how to put *all* the Jucys–Murphy elements *together* in a geometric way.

1.5.15 Support of \mathcal{F}_β Thanks to the Hilbert–Chow morphism, $\mathrm{Hilb}_n \rightarrow \mathbb{A}^{2n} // \mathrm{S}_n$, the geometric realization of HOMFLY-PT link homology given in Theorem 1.5.11 implies that it admits an action of the algebra of symmetric functions in n variables. More algebraically, the definition of HOMFLY-PT homology via Soergel bimodules also affords such an action (see, for example, Gorsky, Kivinen, and Simental [28, Section 5.1]). We show that these two are the same.

Theorem 1.5.16 (Theorem 5.5.1, part of [28, Conjecture 7.2(b)]) *The two actions above coincide.*

Theorem 1.5.16 is a statement about the “global” property of \mathcal{F}_β as it contains information about \mathcal{F}_β after pushing forward to $\mathbb{A}^{2n} // \mathrm{S}_n$. More precisely, by [28, Section 5.1], the supports of the various a -degree parts of the HOMFLY-PT homology of a braid β , as a sheaf over $\mathbb{A}^n // \mathrm{S}_n$, can be bounded above using the number of connected components of the link associated to β .

The support of \mathcal{F}_β over Hilb_n itself is, however, a more local (and hence more refined) property. To prove a similar statement, we have to start by transporting the corresponding support condition on the Rouquier

complex R_β . However, this does not quite make sense since there is no quasicoherent sheaf involved at this stage yet. To circumvent this difficulty, we formulate this support condition using module category structures and the notion of support developed by Benson, Iyengar, and Krause [10] and Arinkin and Gaitsgory [3]. Transporting this structure around to the Hilbert scheme side, where it coincides with the usual notion of support, we obtain the desired statement.

Theorem 1.5.17 (Theorem 5.6.4, [28, Conjecture 7.2(b) and (c)]) *Let β be a braid on n strands. Then an upper bound for the support of \mathcal{F}_β (as a sheaf over Hilb_n) given in Theorem 1.5.11 is determined by the number of connected components of the link obtained by closing up β . See Theorem 5.6.4 for a more precise formulation.*

Outline

We will now give an outline of the paper. In Section 2, we study the categorical traces and Drinfel'd centers of the various variants of the finite Hecke categories and relate them to the theory of character sheaves. In Section 3, we prove the formality result for the graded Grothendieck–Springer sheaf, of which formality of the Hochschild homologies of the graded finite Hecke categories is a consequence. In type A , this result gives an explicit description of the categorical trace of $H_n^{\text{gr,ren}}$. This is followed by Section 4, where the relation between unipotent character sheaves and Hilbert schemes of points on \mathbb{C}^2 is established. We conclude with a proof of aversion (see Remark 1.5.12 for a precise comparison) of the conjectures by Gorsky, Neguț, and Rasmussen in Section 5.

We note that the different sections are largely independent of each other, both in terms of techniques and results. More precisely, the sections can be grouped as follows: (Section 2), (Section 3), (Sections 4 and 5). The reader can read each group mostly independently, referring to the other parts of the paper mostly to look up the notation.

2 Categorical traces and Drinfel'd centers of (graded) finite Hecke categories

This section is dedicated to the proof of the first main theorem, Theorem 1.5.2 (which appears as Theorem 2.8.3 below), which relates the categorical trace and Drinfel'd center of a finite Hecke category and the category of unipotent character sheaves. We work with mixed, graded, and ungraded versions and in both settings of “big” and “small” categories. One interesting feature is that the trace and center of a “big” (also known as *renormalized*) finite Hecke category coincide, whereas the center of the “small” version is larger than its trace.

In what follows, Section 2.1 reviews basic definitions and outlines the basic strategy. In Sections 2.2 and 2.3, we study the categorical analog of Künneth formula for finite orbit stacks and the Beck–Chevalley condition, respectively. As the trace/center is computed by the colimit/limit of a certain

simplicial/cosimplicial category, the categorical Künneth formula allows one to interpret each term in the (co)simplicial diagram algebro-geometrically, whereas the Beck–Chevalley condition acts as a descent-type result which allows one to realize the colimit/limit more concretely in terms of sheaves on some space. In Section 2.4, we formulate the monoidal structure of finite Hecke categories in terms of 1-manifolds, which is then used in Section 2.5 to formulate the cyclic bar simplicial category computing the trace of a finite Hecke category in terms of the geometry of a circle. The geometric formulation developed thus far is applied in Section 2.6 to verify the Beck–Chevalley conditions in a uniform way. The computations of the traces and centers conclude in Section 2.7. In Section 2.8, we introduce the various versions of the category of (graded) unipotent character sheaves and state our main results thus far in terms of these categories. In Section 2.9, we show that $\mathrm{Tr}(\mathrm{H}_G^{\mathrm{gr}})$ inherits the weight and perverse t -structure from the ambient category $\mathrm{Shv}_{\mathrm{gr},c}(G/G)$. Finally, Section 2.10, which is mostly independent of the rest of the paper, deduces several consequences for the trace and center of a finite Hecke category from the rigidity of the latter.

2.1 Preliminaries

We will now review the basic definitions of the objects involved and outline the strategy employed to prove Theorem 1.5.2.

2.1.1 Generalities regarding big vs small categories In the situation of Definition 1.4.5, the trace $\mathrm{Tr}(A) = A \otimes_A A^{\mathrm{rev}} A$ can be computed as the geometric realization (ie colimit) of the simplicial object obtained from the cyclic bar construction. Namely,

$$\mathrm{Tr}(A) = |A^{\otimes(\bullet+1)}| := \mathrm{colim} A^{\otimes(\bullet+1)},$$

where the last face map is given by multiplying the last and first factors of A and where the other face maps are given by multiplying adjacent A factors; see also [8, Section 5.1.1; 39, Remark 5.5.3.13].

The small and big category settings are related by taking Ind via the following standard result:

Proposition 2.1.2 [9, Corollary 3.6 and Proposition 3.3] *Let $\mathcal{A} \in \mathrm{Alg}(\mathrm{DGCat}_{\mathrm{pres},\mathrm{cont}})$ be a compactly generated rigid monoidal category (see [9, Definition 3.1]). Taking Ind -completion (resp. passing to right adjoints, resp. passing to the full subcategory $(-)^c$ spanned by compact objects) induces a canonical equivalence (i)→(ii) (resp. (ii)→(iii), resp. (ii)→(i)), between*

- (i) *the category $\mathcal{A}^c\text{-Mod}$ of \mathcal{A}^c -module categories in $\mathrm{DGCat}_{\mathrm{idem},\mathrm{ex}}$, where \mathcal{A}^c is full subcategory of \mathcal{A} spanned by compact objects,*
- (ii) *the category $\mathcal{A}\text{-Mod}_c$ of \mathcal{A} -module categories in $\mathrm{DGCat}_{\mathrm{pres},\mathrm{cont},c}$,*
- (iii) *the opposite of the category whose objects are \mathcal{A} -module categories in the category of compactly generated DG-categories and whose morphisms are functors that are both cocontinuous (ie limit preserving) and continuous.*

As mentioned earlier, we will work mostly in the setting of Proposition 2.1.2(ii) even though we are most interested in the “small” setting described in Proposition 2.1.2(i) because the former has more functoriality.

Remark 2.1.3 This result was also proved (but not explicitly stated) in [25, Chapter 1, Section 9]. Note also that [9] proved the result above more generally for semirigid monoidal categories. We do not need this level of generality here.

Remark 2.1.4 In the literature, the property of being rigid is usually applied to a “small” monoidal DG-category. By definition, a category $\mathcal{C}_0 \in \text{Alg}(\text{DGCat}_{\text{idem,ex}})$ is *rigid* if all objects $c \in \mathcal{C}_0$ have left and right duals. By [25, Chapter 1, Lemma 9.1.5], such a category \mathcal{C}_0 is rigid if and only if $\mathcal{C} := \text{Ind}(\mathcal{C}_0)$ is rigid in the sense of [25, Chapter 1, Definition 9.1.2]. In this case, \mathcal{C} is compactly generated by definition.

We use the term rigid for both “big” and “small” categories. The context should make it clear in which sense we use the term.

2.1.5 Computing colimits Recall that $\text{Ind}: \text{DGCat}_{\text{idem,ex}} \rightarrow \text{DGCat}_{\text{pres,cont}}$ preserves all colimits by [25, Chapter 1, Corollary 7.2.7] and that the forgetful functor $\mathcal{A}\text{-Mod} \rightarrow \text{DGCat}_{\text{pres,cont}}$ preserves all colimits by [39, Corollaries 4.2.3.5]. Proposition 2.1.2 then implies that the colimit $\text{colim}_{i \in I} \mathcal{C}_i^0$ of a diagram $I \rightarrow \mathcal{A}^c\text{-Mod}$ can be computed as

$$\text{colim}_{i \in I} \mathcal{C}_i^0 \simeq \left(\text{colim}_{i \in I} \text{Ind}(\mathcal{C}_i^0) \right)^c,$$

where $(-)^c$ denotes the procedure of taking the full subcategory spanned by compact objects. Moreover, the category underlying the colimit on the right-hand side can be computed in $\text{DGCat}_{\text{pres,cont}}$.⁷

2.1.6 Computing limits The forgetful functors $\mathcal{A}\text{-Mod} \rightarrow \text{DGCat}_{\text{pres,cont}}$ and $\text{DGCat}_{\text{pres,cont}} \rightarrow \text{Cat}$ preserve all limits by [39, Corollary 4.2.3.3] and [38, Proposition 5.5.3.13], where Cat denotes the category of all $(\infty, 1)$ -categories. The underlying category of a limit in $\mathcal{A}\text{-Mod}$ can thus be computed in the category of all categories.

However, Ind does not preserve limits. In fact, this is the reason why the renormalized category of sheaves on a stack differs from the usual category. Moreover, as we shall see, this will also be responsible for the fact that the Drinfel’d center of the “small” version of a finite Hecke category differs from its trace.

2.1.7 The strategy The trace of $H_G^{?,ren}$ as an algebra object in $\text{Shv}_?(pt)\text{-Mod}$ is given by the geometric realization of the cyclic bar construction

$$(2.1.8) \quad \text{Tr}(H_G^{?,ren})_{\bullet} := (H_G^{?,ren})^{\otimes_{\text{Shv}_?(pt)} (\bullet+1)} \\ \simeq H_G^{?,ren} \otimes_{\text{Shv}_?(pt)} H_G^{?,ren} \otimes_{\text{Shv}_?(pt)} \cdots \otimes_{\text{Shv}_?(pt)} H_G^{?,ren} \in (\text{Shv}_?(pt)\text{-Mod})^{\Delta^{\text{op}}}.$$

In the next two subsections, we will discuss the two main technical ingredients used to understand the geometric realization of the simplicial object above. First, the categorical Künneth formula for finite orbit

⁷Note, however, that colimits in the presentable setting are different from colimits in the category of all ∞ -categories.

stacks allows one to realize the terms in the simplicial object above as the categories of sheaves on certain stacks. Second, the Beck–Chevalley condition on a simplicial category is essentially a descent-type result that allows one to realize its geometric realization as a full subcategory of a more concrete category.

By a duality argument, we show that the Drinfel’d center of $H_G^{?,ren}$ coincides with its trace. Finally, the traces and Drinfel’d centers of the small versions $H_G^?$ are obtained from those of the renormalized versions.

2.2 The categorical Künneth formula for finite orbit stacks

This subsection is dedicated to the following result:

Proposition 2.2.1 *Let \mathcal{Y}_0 and \mathcal{Z}_0 be Artin stacks over pt_0 . Suppose that \mathcal{Y}_0 is a finite orbit stack. Then the external tensor product*

$$(2.2.2) \quad \text{Shv}_?(Y)^{ren} \otimes_{\text{Shv}_?(pt)} \text{Shv}_?(Z)^{ren} \xrightarrow[\cong]{\boxtimes} \text{Shv}_?(Y \times Z)^{ren}$$

induces an equivalence of categories.

Proof We will treat the mixed case, ie $? = m$. The graded and ungraded cases can be treated similarly.

By [4, (F.18)], we know that (2.2.2) is fully faithful. By [31, Corollary A.4.13], we know that the left-hand side of (2.2.2) is generated by objects of the form $\mathcal{F}_0 \boxtimes \mathcal{G}_0$ where \mathcal{F}_0 and \mathcal{G}_0 are constructible. Since (2.2.2) is continuous, it suffices to show that $\mathcal{F}_0 \boxtimes \mathcal{G}_0 \in \text{Shv}_m(\mathcal{Y}_0 \times_{pt_0} \mathcal{Z}_0)^{ren}$ generates the category. Using excision and the fact that \mathcal{Y}_0 is a finite orbit stack, we reduce to the case where $\mathcal{Y}_0 = BH_0$ for a (smooth) algebraic group H_0 over \mathbb{F}_q . This case is treated in Lemma 2.2.3 below. □

Lemma 2.2.3 *Proposition 2.2.1 holds when $\mathcal{Y}_0 = BH_0$, where H_0 is a smooth algebraic group over pt_0 .*

Proof It suffices to show that

$$\text{Shv}_{m,c}(BH_0) \otimes_{\text{Shv}_{m,c}(pt_0)} \text{Shv}_{m,c}(Z_0) \xrightarrow{\boxtimes} \text{Shv}_{m,c}(BH_0 \times_{pt_0} Z_0)$$

is an equivalence of categories.

By descent along the surjective smooth map $Z_0 \rightarrow BH_0 \times_{pt_0} Z_0$ (using the $(-)^!$ -functor), we have

$$\text{Shv}_{m,c}(BH_0 \times_{pt_0} Z_0) \simeq \text{Tot}(\text{Shv}_{m,c}(H_0^{\times_{pt_0} \bullet} \times_{pt_0} Z_0)).$$

Then [39, Proposition 4.7.5.1] provides an equivalence of categories

$$(2.2.4) \quad \text{Shv}_{m,c}(BH_0 \times_{pt_0} Z_0) \simeq ((\pi_0)_!(\pi_0)^! \overline{\mathbb{Q}}_\ell)\text{-Mod}(\text{Shv}_{m,c}(Z_0)),$$

where $\pi_0: H_0 \rightarrow pt_0$ and where $(\pi_0)_!(\pi_0)^! \overline{\mathbb{Q}}_\ell \in \text{Alg}(\text{Shv}_{m,c}(pt_0))$. Note that here, the algebra structure on $(\pi_0)_!(\pi_0)^! \overline{\mathbb{Q}}_\ell$ is obtained from the multiplication structure of H : indeed, $(\pi_0)_!(\pi_0)^! \overline{\mathbb{Q}}_\ell$ is the *homology*

$C_*(H) := \pi_! \pi^! \overline{\mathbb{Q}}_\ell$ of H along with a Frobenius action, where $\pi: H \rightarrow \text{pt}$ is the pullback of π_0 . Equation (2.2.4) can thus be rewritten in the following more familiar form:

$$\text{Shv}_{m,c}(BH_0 \times_{\text{pt}_0} \mathcal{Z}_0) \simeq C_*(H)\text{-Mod}(\text{Shv}_{m,c}(\mathcal{Z}_0)).$$

Note also that the action of $\text{Shv}_{m,c}(\text{pt}_0)$ on $\text{Shv}_{m,c}(\mathcal{Z}_0)$ is used to make sense of the right side of (2.2.4).

Arguing similarly (see also [25, Chapter 1, Proposition 8.5.4]),

$$\begin{aligned} \text{Shv}_{m,c}(BH_0) \otimes_{\text{Shv}_{m,c}(\text{pt}_0)} \text{Shv}_{m,c}(\mathcal{Z}_0) &\simeq C_*(H)\text{-Mod}(\text{Shv}_{m,c}(\text{pt}_0)) \otimes_{\text{Shv}_{m,c}(\text{pt}_0)} \text{Shv}_{m,c}(\mathcal{Z}_0) \\ &\simeq C_*(H)\text{-Mod}(\text{Shv}_{m,c}(\mathcal{Z}_0)). \end{aligned} \quad \square$$

2.3 The Beck–Chevalley condition

We will now turn to the Beck–Chevalley condition, a technical condition that allows one to realize geometric realizations (ie colimits of simplicial objects) and totalizations (ie limits of cosimplicial objects) of categories in more concrete terms.

Definition 2.3.1 [39, Definition 4.7.4.13] Suppose we have a diagram of ∞ -categories

$$\begin{array}{ccc} \mathcal{C} & \xrightarrow{H} & \mathcal{D} \\ \downarrow V & & \downarrow V' \\ \mathcal{C}' & \xrightarrow{H'} & \mathcal{D}' \end{array}$$

which commutes up to a specified equivalence $\alpha: V' \circ H \xrightarrow{\cong} H' \circ V$.

We say that this diagram is *horizontally left* (resp. *right adjointable*) if H and H' admit left (resp. right) adjoints H^L and H'^R (resp. H^R and H'^L), respectively, and if the composite transformation

$$\begin{aligned} H'^L \circ V' \rightarrow H'^L \circ V' \circ H \circ H^L &\xrightarrow{\alpha} H'^L \circ H' \circ V \circ H^L \rightarrow V \circ H^L \\ \text{(resp. } V \circ H^R \rightarrow H'^R \circ H' \circ V \circ H^R &\xrightarrow{\alpha^{-1}} H'^R \circ V' \circ H \circ H^R \rightarrow H'^R \circ V') \end{aligned}$$

is an equivalence.

We say that this diagram is *vertically left* (resp. *right adjointable*) if the transposed diagram

$$\begin{array}{ccc} \mathcal{C} & \xrightarrow{V} & \mathcal{C}' \\ \downarrow H & & \downarrow H' \\ \mathcal{D} & \xrightarrow{V'} & \mathcal{D}' \end{array}$$

is *horizontally left* (resp. *right adjointable*).

Remark 2.3.2 What we call horizontally left (resp. right) adjointable is simply called left (resp. right) adjointable in [39]. This condition on a commutative square of categories is also commonly called the Beck–Chevalley condition.

We are now ready to give the main statement of this subsection.

Proposition 2.3.3 (Lurie) *Let $\mathcal{C}_\bullet : \Delta_+^{\text{op}} \rightarrow \text{DGCat}_{\text{pres,cont}}$ be an augmented simplicial diagram of DG-categories. Let $\mathcal{C} = \mathcal{C}_{-1}$, $F : \mathcal{C}_0 \rightarrow \mathcal{C}$ be the obvious functor, and G its right adjoint. Suppose that for every morphism $\alpha : [n] \rightarrow [m]$ in Δ_+ which induces a morphism $\alpha_+ : [n+1] \simeq [0] * [n] \rightarrow [0] * [m] \simeq [m+1]$, the diagram*

$$(2.3.4) \quad \begin{array}{ccc} \mathcal{C}_{m+1} & \xrightarrow{d_0} & \mathcal{C}_m \\ \downarrow & & \downarrow \\ \mathcal{C}_{n+1} & \xrightarrow{d_0} & \mathcal{C}_n \end{array}$$

is vertically right adjointable. Then the functor $\theta : |\mathcal{C}_\bullet|_{\Delta^{\text{op}}} \rightarrow \mathcal{C}$ is fully faithful. When G is conservative, θ is an equivalence of categories.

Proof We will deduce this proposition from its dual version [39, Corollary 4.7.5.3].

Passing to right adjoints, we get an augmented cosimplicial object \mathcal{C}^\bullet . Note that for each n , $\mathcal{C}^n \simeq \mathcal{C}_n$ as only the functors change. By [38, Corollary 5.5.3.4] (see also [25, Chapter 1, Proposition 2.5.7]), we have a canonical equivalence of categories

$$|\mathcal{C}_\bullet|_{\Delta^{\text{op}}} \simeq \text{Tot}(\mathcal{C}^\bullet|_\Delta).$$

Moreover, the canonical functor $\mathcal{C} \rightarrow \text{Tot}(\mathcal{C}^\bullet|_\Delta)$ is the right adjoint θ^R of θ .

Observe that

$$\begin{array}{ccc} \mathcal{C}^{m+1} & \xleftarrow{d^0} & \mathcal{C}^m \\ \uparrow & & \uparrow \\ \mathcal{C}^{n+1} & \xleftarrow{d^0} & \mathcal{C}^n \end{array}$$

is horizontally left adjointable as this is equivalent to (2.3.4) being vertically right adjointable. Thus [39, Corollary 4.7.5.3] implies that θ is a fully faithful embedding, and moreover, it is an equivalence of categories when G is conservative. □

Remark 2.3.5 The same proof goes through when we replace $\text{DGCat}_{\text{pres,cont}}$ by $\mathcal{A}\text{-Mod}$, where \mathcal{A} is a rigid monoidal category, for example, when \mathcal{A} is $\text{Shv}_m(\text{pt}_0)$ or $\text{Vect}^{\text{gr}} \simeq \text{Shv}_{\text{gr}}(\text{pt})$. This is because the forgetful functor $\mathcal{A}\text{-Mod} \rightarrow \text{DGCat}_{\text{pres,cont}}$ preserves all limits and colimits, by [39, Corollaries 4.2.3.3 and 4.2.3.5], and adjoints of an \mathcal{A} -linear functor are automatically \mathcal{A} -linear by [31, Corollary A.4.7].

2.4 $H_G^{?,\text{ren}}$ as a functor out of 1-manifolds

While it is straightforward how to apply Proposition 2.2.1, the verification of the adjointability conditions (aka Beck–Chevalley conditions) necessary for the application of Proposition 2.3.3 is more subtle. The

argument can be made transparent by reformulating the simplicial object (2.1.8) geometrically in terms of 1-manifolds. In preparation for that, we will, in this subsection, upgrade $H_G^{?,ren}$ to a right-lax symmetric monoidal functor coming out of the category of 1-manifolds. The construction found in this subsection is a simplification of the one found in [30].

2.4.1 1-manifolds and their boundaries Let $Mnfd_1$ denote the symmetric monoidal ∞ -category of 1-dimensional manifolds with finitely many connected components whose morphisms are given by embeddings and whose symmetric monoidal structure is given by disjoint unions. Note that the objects are simply finite disjoint unions of lines and circles. Although these manifolds are, technically speaking, without boundary, we will make use of their “boundaries” in our construction. We will now explain what this means.

Let $Mnfd'_1$ be the ∞ -category of compact 1-manifolds, usually denoted by M , with possibly nonempty boundary, denoted by ∂M , such that $\overset{\circ}{M} := M \setminus \partial M \in Mnfd_1$. Moreover, morphisms are given by (necessarily closed) embeddings. Taking the interior gives a natural functor of ∞ -categories

$$F: Mnfd'_1 \rightarrow Mnfd_1, \quad M \mapsto \overset{\circ}{M}.$$

Lemma 2.4.2 [30, Lemma 2.4.10] *The functor F is an equivalence of categories. We write $M \mapsto \overline{M}$ to denote an inverse of F .*

Because of this equivalence, we will generally not make a distinction between 1-manifolds without boundaries and compact 1-manifolds with possibly nonempty boundaries, unless confusion is likely to occur. For instance, when $M \in Mnfd_1$, by abuse of notation,

$$\partial M := \partial \overline{M} := \overline{M} \setminus M$$

is used to denote the boundary of \overline{M} .

We also use $Disk_1$ to denote the full subcategory of $Mnfd_1$ consisting of just lines. Moreover, the category $Disk'_1 := Disk_1 \times_{Mnfd_1} Mnfd'_1$ consisting of compact line segments is equivalent to $Disk_1$ itself.

2.4.3 $H_G^{?,ren}$ as a functor out of $Mnfd_1$ Following [30, Section 3.1.2], we will now construct a right-lax symmetric monoidal functor

$$H_G^{?,ren}: Mnfd_1 \rightarrow Shv_?(pt)\text{-Mod},$$

whose restriction to $Disk_1$ is symmetric monoidal. The functor $H_G^{?,ren}$ is obtained as a composition⁸

$$(2.4.4) \quad Mnfd_1 \xrightarrow{M} \text{Corr}(\text{Stk})_{\text{prop};\text{sm}} \xrightarrow{Shv_{?,!}^{*,ren}} Shv_?(pt)\text{-Mod}.$$

We will now explain the various functors and categories that appear in the diagram above. We start with the functor M , which is defined as the composition

$$(2.4.5) \quad Mnfd_1 \xrightarrow{\mathfrak{B}} \text{Corr}((\text{Spc}_{\text{fin}}^{\Delta^1})^{\text{op}}) \xrightarrow{\mathcal{M}\text{ap}} \text{Corr}(\text{Stk}).$$

⁸Even though we write Stk , it should be understood as $\text{Stk}_{\mathbb{F}_q}$ when we consider the mixed variant.

2.4.6 Correspondences For any category \mathcal{C} , $\text{Corr}(\mathcal{C})$ is the category of correspondences in \mathcal{C} . A morphism from c_1 to c_2 in $\text{Corr}(\mathcal{C})$ is illustrated, equivalently, by diagrams of the form

$$c_1 \xleftarrow{h} c \begin{array}{c} \downarrow v \\ c_2 \end{array} \quad \text{or} \quad \begin{array}{ccc} & c & \\ h \swarrow & & \searrow v \\ c_1 & & c_2 \end{array} \quad \text{or} \quad c_1 \xleftarrow{h} c \xrightarrow{v} c_2.$$

Here, h and v stand for horizontal and vertical, respectively. As usual, compositions are given by pullbacks. More generally, let vert and horiz be two collections of morphisms in \mathcal{C} such that vert (resp. horiz) is closed under pulling back along a morphism in horiz (resp. vert). Then we let $\text{Corr}(\mathcal{C})_{\text{vert};\text{horiz}}$ be the 1-full subcategory of $\text{Corr}(\mathcal{C})$ consisting of the same objects, but for morphisms we require that $v \in \text{vert}$ and $h \in \text{horiz}$. See [30, Section 2.8.1] for more details.

2.4.7 The functor \mathfrak{B} When $\mathcal{C} = \text{Spc}_{\text{fin}}$, the category of finite CW-complexes, the functor

$$\mathfrak{B}: \text{Spc}_{\text{fin}} \rightarrow \text{Corr}((\text{Spc}_{\text{fin}}^{\Delta^1})^{\text{op}}),$$

(\mathfrak{B} stands for *boundary*) at the level of objects, is given by

$$\mathfrak{B}(M) = (\partial M \rightarrow \overline{M}) \in \text{Spc}_{\text{fin}}^{\Delta^1}.$$

Moreover, \mathfrak{B} sends an open embedding $N \hookrightarrow M$ to the following morphism in $\text{Corr}((\text{Spc}_{\text{fin}}^{\Delta^1})^{\text{op}})$:

$$(2.4.8) \quad \begin{array}{ccccc} \partial N & \longrightarrow & \overline{M} \setminus N & \longleftarrow & \partial M \\ \downarrow & & \downarrow & & \downarrow \\ \overline{N} & \longrightarrow & \overline{M} & \xleftarrow{\cong} & \overline{M} \end{array}$$

We will often suppress the morphisms and simply use, for example, $(\partial M, \overline{M})$ to denote an object in $\text{Spc}_{\text{fin}}^{\Delta^1}$ to make diagrams and formulas more compact.

2.4.9 The functor $\mathfrak{M}\text{ap}$ We have a natural functor

$$\mathfrak{M}\text{ap}: (\text{Spc}^{\Delta^1})^{\text{op}} \rightarrow \text{Stk}$$

which assigns to each object $(N \rightarrow M) \in (\text{Spc}^{\Delta^1})^{\text{op}}$ an object $(BB, BG)^{N, M} := BB^N \times_{BG^N} BG^M$ (see also Remark 1.4.2 and note 8), which is precisely the stack of commutative squares

$$\begin{array}{ccc} N & \longrightarrow & BB \\ \downarrow & & \downarrow \\ M & \longrightarrow & BG \end{array}$$

This functor upgrades naturally to an eponymous functor

$$\mathfrak{M}\text{ap}: \text{Corr}((\text{Spc}^{\Delta^1})^{\text{op}}) \rightarrow \text{Corr}(\text{Stk}),$$

used in (2.4.5) above.

By construction, $\mathfrak{M}\text{ap}$ turns colimits in Spc^{Δ^1} to limits in Stk .

2.4.10 The functor M The functor M in (2.4.4) is given by the following result:

Lemma 2.4.11 *The composition $\mathcal{M}\text{ap} \circ \mathfrak{B}$ factors through $\text{Corr}(\text{Stk})_{\text{prop}; \text{sm}}$, ie the horizontal (resp. vertical) maps are smooth (resp. schematic and proper).*

Proof Let $N \hookrightarrow M$ be a morphism in Mnfd_1 , ie it is an open embedding of 1-manifolds. We will now show that the map

$$(2.4.12) \quad (BB, BG)^{\overline{M} \setminus N, \overline{M}} \simeq BB^{\overline{M} \setminus N} \times_{BG^{\overline{M} \setminus N}} BG^{\overline{M}} \rightarrow BB^{\partial N} \times_{BG^{\partial N}} BG^{\overline{N}} \simeq (BB, BG)^{\partial N, \overline{N}},$$

which is induced by the left square of (2.4.8), is smooth. Since the left square of (2.4.8) is a pushout, we have

$$(BB, BG)^{\overline{M} \setminus N, \overline{M}} \simeq (BB, BG)^{\partial N, \overline{N}} \times_{BB^{\partial N}} BB^{\overline{M} \setminus N}$$

and the map (2.4.12) identifies with the vertical map on the left of the following pullback diagram:

$$\begin{array}{ccc} (BB, BG)^{\overline{M} \setminus N, \overline{M}} & \longrightarrow & BB^{\overline{M} \setminus N} \\ \downarrow & & \downarrow \\ (BB, BG)^{\partial N, \overline{N}} & \longrightarrow & BB^{\partial N} \end{array}$$

It thus suffices to show that the map

$$(2.4.13) \quad BB^{\overline{M} \setminus N} \rightarrow BB^{\partial N}$$

is smooth. Without loss of generality, we can assume that M is connected, in which case, M is either a circle or a line.

When N is empty, it is clear that (2.4.13) is smooth since it is simply

$$BB^{\overline{M}} \rightarrow \text{pt},$$

where $BB^{\overline{M}}$ is either BB or B/B depending on whether M is a line or a circle. Thus, it remains to treat the case where N is nonempty.

When M is a line, the desired statement follows from the fact that (2.4.13) is a product of (copies of) the smooth maps id_{BB} and $\Delta_{BB}: BB \rightarrow BB \times BB$. When M is a circle, since the only embedding of a circle into itself is a homeomorphism, N can only be a circle or a disjoint union of lines. In the first case, the map under consideration is an equivalence, and hence it is smooth. In the second case, we also see that (2.4.13) is a product of (copies of) the diagonal map $\Delta_{BB}: BB \rightarrow BB \times BB$, which is smooth.

Next, we will show that the map

$$(2.4.14) \quad (BB, BG)^{\overline{M} \setminus N, \overline{M}} \simeq BB^{\overline{M} \setminus N} \times_{BG^{\overline{M} \setminus N}} BG^{\overline{M}} \rightarrow BB^{\partial M} \times_{BG^{\partial M}} BG^{\overline{M}} \simeq (BB, BG)^{\partial M, \overline{M}},$$

induced by the right square of (2.4.8), is proper. As above, we can (and we will) assume that M is connected. Note that when N is empty, (2.4.14) becomes

$$BB^{\overline{M}} \rightarrow BB^{\partial M} \times_{BG^{\partial M}} BG^{\overline{M}},$$

which is equivalent to

$$\frac{B}{B} \simeq BB^{S^1} \rightarrow \frac{G}{G}$$

when $M \simeq S^1$, and to

$$BB \rightarrow BB \times_{BG} BB$$

when $M \simeq \mathbb{R}$. Both of these are easily seen to be proper.

When N is not empty, $\bar{M} \setminus N$ and ∂M are homotopically equivalent to a (possibly empty) disjoint union of points. Moreover, $\partial M \rightarrow \bar{M} \setminus N$ is homotopically equivalent to an inclusion of connected components. We will thus treat them as disjoint unions of points in what follows. In particular, we can write $\bar{M} \setminus N \simeq \partial M \sqcup ((\bar{M} \setminus N) \setminus \partial M)$. The map (2.4.14) factors as

$$\begin{aligned} BB^{\bar{M} \setminus N} \times_{BG^{\bar{M} \setminus N}} BG^{\bar{M}} &\simeq (BB^{\partial M} \times_{BG^{\partial M}} BG^{\bar{M}}) \times_{BG^{\bar{M}}} (BB^{(\bar{M} \setminus N) \setminus \partial M} \times_{BG^{(\bar{M} \setminus N) \setminus \partial M}} BG^{\bar{M}}) \\ &\rightarrow BB^{\partial M} \times_{BG^{\partial M}} BG^{\bar{M}}, \end{aligned}$$

where the last map is induced by $BB^{(\bar{M} \setminus N) \setminus \partial M} \rightarrow BG^{(\bar{M} \setminus N) \setminus \partial M}$, which is proper since it is just a product of maps of the form $BB \rightarrow BG$. □

2.4.15 The functor $H_G^{?,ren}$ Applying the functor $\text{Shv}_{?,!}^{ren,*}$ of [31, (2.8.13) and Theorem 2.8.20], which is right-lax symmetric monoidal, we complete (2.4.4). Note that in our case, we only need $\text{Shv}_{?,!}^{ren,*}$ to encode (renormalized) $*$ -pullbacks along smooth morphisms and (renormalized) $!$ -pushforwards along proper morphisms rather than the more general case described in [31].

2.4.16 Renormalized finite Hecke categories Being a right-lax symmetric monoidal functor, $H_G^{?,ren}$ sends algebra objects to algebra objects. In particular,

$$H_G^{?,ren}(\mathbb{R}) \simeq \text{Shv}_?((BB \times BB) \times_{BG \times BG} BG)^{ren} \simeq \text{Shv}_?(BB \times_{BG} BB)^{ren}$$

has an algebra structure, induced by the algebra structure on $\mathbb{R} \in \text{Mnfd}_1$, whose multiplication is given by $\mathbb{R}^{\sqcup 2} \hookrightarrow \mathbb{R}$. Chasing through (2.4.8), we see that this is precisely the monoidal structure on $\text{Shv}_?(BB \times_{BG} BB)^{ren}$ defined earlier via the correspondence (1.4.7). This justifies the abuse of notation where we use $H_G^{?,ren}$ to denote both the functor and its value on \mathbb{R} .

2.4.17 Horocycle correspondence We will now explain how the horocycle correspondence appears naturally from this point of view. First, observe that $\partial S^1 \simeq \emptyset$, and hence $H_G^{?,ren}(S^1) \simeq \text{Shv}_?(G/G)^{ren}$ since $BG^{S^1} \simeq G/G$, where the quotient is taken using the conjugation action.

Let $\psi: \mathbb{R} \rightarrow S^1$ be an embedding. The functor \mathfrak{B} carries ψ to the diagram

$$\begin{array}{ccccc} \text{pt} \sqcup \text{pt} & \longrightarrow & \text{pt} & \longleftarrow & \emptyset \\ \downarrow & & \downarrow & & \downarrow \\ \mathbb{R} & \longrightarrow & S^1 & \longleftarrow & S^1 \end{array}$$

which is sent to

$$B \backslash G / B \simeq BB \times_{BG} BB \leftarrow \frac{G}{B} \rightarrow \frac{G}{G}$$

under Map , where we have used $BB \times_{BG} (G/G) \simeq G/B$. This is the horocycle correspondence appearing in Theorem 1.5.2.

2.5 Augmented cyclic bar construction via Mnfd_1

Using the geometric picture of Section 2.4, we will now produce an augmented simplicial object that will be used as the input for Proposition 2.3.3. The underlying simplicial object is the same as the one that computes the categorical trace of $H_G^{?,\text{ren}}$ introduced in (2.1.8).

2.5.1 Circle geometry and augmented simplicial sets Consider the over-category $(\text{Mnfd}_1)_{/S^1}$. Fix a morphism $\psi: \mathbb{R} \rightarrow S^1$, and consider $(\text{Mnfd}_1)_{\psi//S^1} := ((\text{Mnfd}_1)_{/S^1})_{\psi/}$. We note that $(\text{Mnfd}_1)_{/S^1}$ has a final object, by construction, and moreover, it is precisely the category obtained from $(\text{Disk}_1)_{/S^1} := \text{Disk}_1 \times_{\text{Mnfd}_1} (\text{Mnfd}_1)_{/S^1}$ by adjoining a final object. Similarly, the category $(\text{Mnfd}_1)_{\psi//S^1}$ also has a final object and is equivalent to the category obtained from $(\text{Disk}_1)_{\psi//S^1}$ by adjoining a final object.

We will now relate $(\text{Mnfd}_1)_{\psi//S^1}$ and Δ_+^{op} using a variation of the construction in [39, Section 5.5.3]. Note that an object of $(\text{Mnfd}_1)_{\psi//S^1}$ is given by a diagram

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{j} & U \\ \psi \searrow & & \swarrow \psi' \\ & S^1 & \end{array}$$

which commutes up to isotopy, where U is either S^1 or a finite disjoint union of copies of \mathbb{R} . The set of components $\pi_0(S^1 \setminus \psi'(U))$ is finite: empty when U is S^1 and equal to the number of components of U when it is a disjoint union of copies of \mathbb{R} . Fix an orientation of the circle. We define a linear ordering \leq on $\pi_0(S^1 \setminus \psi'(U))$: if $x, y \in S^1$ belong to different components of $S^1 \setminus \psi'(U)$, then we write $x < y$ if the three points $(x, y, \psi'(j(0)))$ are arranged in clockwise order around the circle, and $y < x$ otherwise. This construction determines a functor from $(\text{Mnfd}_1)_{\psi//S^1}$ to the opposite of the category of finite linearly ordered sets, which is Δ_+^{op} .

Lemma 2.5.2 *The above construction determines an equivalence of ∞ -categories*

$$\theta_+ : (\text{Mnfd}_1)_{\psi//S^1} \rightarrow \Delta_+^{\text{op}},$$

which restricts to an equivalence of ∞ -categories

$$\theta := \theta_+|_{(\text{Disk}_1)_{\psi//S^1}} : (\text{Disk}_1)_{\psi//S^1} \rightarrow \Delta_+^{\text{op}}.$$

Proof The second equivalence is [39, Lemma 5.5.3.10]. The first part is a direct extension of the second by adjoining a final object. □

2.5.3 The augmented simplicial category We are now ready to construct the augmented simplicial category. Let $\overline{\text{Tr}}(\mathbb{H}_G^{?,\text{ren}})_\bullet$ be the augmented simplicial category given by the following composition of functors:

$$(2.5.4) \quad \Delta_+^{\text{op}} \xrightarrow{\theta_+^{-1}} (\text{Mnfd}_1)_{\psi//S^1} \xrightarrow{\varphi} \text{Mnfd}_1 \xrightarrow{\mathbb{H}_G^{\text{ren}}} \text{Shv}_?(pt)\text{-Mod}.$$

Here φ is the obvious forgetful functor.

Since $BB \times_{BG} BB \simeq B \backslash G / B$ is a finite orbit stack by the Bruhat decomposition, Proposition 2.2.1 implies that $\overline{\text{Tr}}(\mathbb{H}_G^{?,\text{ren}})_\bullet|_{\Delta^{\text{op}}} \simeq \text{Tr}(\mathbb{H}_G^{?,\text{ren}})_\bullet$ of (2.1.8); see also [39, Remarks 5.5.3.13 and 5.5.3.14]. In particular,

$$\overline{\text{Tr}}(\mathbb{H}_G^{?,\text{ren}})_n \simeq \text{Tr}(\mathbb{H}_G^{?,\text{ren}})_n \simeq \text{Shv}_?((BB \times_{BG} BB)^{n+1})^{\text{ren}} \quad \text{for } n \geq 0$$

and

$$\overline{\text{Tr}}(\mathbb{H}_G^{?,\text{ren}})_{-1} \simeq \text{Shv}_?\left(\frac{G}{G}\right)^{\text{ren}}$$

where, by convention, G/G denotes the quotient of G by itself under the adjoint action.

2.5.5 “Small” variants We have “small” variants $\mathbb{H}_G^?$ and $\overline{\text{Tr}}(\mathbb{H}_G^?)_\bullet$ of $\mathbb{H}_G^{?,\text{ren}}$ and $\overline{\text{Tr}}(\mathbb{H}_G^{?,\text{ren}})_\bullet$, given by the following diagram:

$$\begin{array}{ccccccc} & & & & \mathbb{H}_G^? & & \\ & & & & \curvearrowright & & \\ \Delta_+^{\text{op}} & \xrightarrow{\theta_+^{-1}} & (\text{Mnfd}_1)_{\psi//S^1} & \xrightarrow{\varphi} & \text{Mnfd}_1 & \xrightarrow{M} & \text{Corr}(\text{Stk})_{\text{prop};\text{sm}} \xrightarrow{\text{Shv}_{?,c,!}^*} & \text{Shv}_{?,c}(pt)\text{-Mod} \\ & & & & \curvearrowleft & & \\ & & & & \overline{\text{Tr}}(\mathbb{H}_G^?)_\bullet & & \end{array}$$

The only difference between the two versions is that in the last step, we use $\text{Shv}_{?,c,!}^*$ rather than $\text{Shv}_{?,!}^{\text{ren},*}$.

2.6 Adjointability

We are now ready to show that $\overline{\text{Tr}}(\mathbb{H}_G^{?,\text{ren}})_\bullet$ satisfies the adjointability condition of Proposition 2.3.3. The result follows from a simple geometric statement about topological 1-manifolds.

2.6.1 Adjointable squares in a category of correspondences For any category \mathcal{C} , a commutative square in $\text{Corr}(\mathcal{C})$, which illustrates two morphisms from x to y' to be equivalent, has the shape

$$(2.6.2) \quad \begin{array}{ccccc} x & \longleftarrow & c_{xy} & \longrightarrow & y \\ \uparrow & & 2 & \uparrow & 1 & \uparrow \\ c_{xx'} & \longleftarrow & c_{xy'} & \longrightarrow & c_{yy'} \\ \downarrow & & 3 & \downarrow & 4 & \downarrow \\ x' & \longleftarrow & c_{x'y'} & \longrightarrow & y' \end{array}$$

where 1 and 3 are pullback squares.

Definition 2.6.3 The commutative diagram in $\text{Corr}(\mathcal{C})$ (2.6.2) is called *adjointable* if 2 and 4 are also pullback squares, ie all squares are pullback squares.

2.6.4 Let $\alpha: [n] \rightarrow [m]$ be a morphism in Δ_+ and $M \rightarrow N$ the corresponding morphism in Mnfd_1 via $\varphi \circ \theta_+^{-1}$; see (2.5.4). Let $M_+ \rightarrow N_+$ be the morphism associated to $\alpha_+: [n+1] \rightarrow [m+1]$. M_+ can be obtained from M by deleting the image of $[-\varepsilon, \varepsilon] \subset \mathbb{R}$ in M for some fixed ε , and similarly for N . This construction is functorial, and hence we obtain the map $M_+ \rightarrow N_+$.

We have the commutative diagram

$$\begin{array}{ccc} M_+ & \longrightarrow & M \\ \downarrow & & \downarrow \\ N_+ & \longrightarrow & N \end{array}$$

in Mnfd_1 , which induces, via the \mathfrak{B} construction of Section 2.4.7, the following commutative diagram in $\text{Corr}((\text{Spc}_{\text{fin}}^{\Delta^1})^{\text{op}})$:

$$(2.6.5) \quad \begin{array}{ccccc} (\partial M_+, \bar{M}_+) & \longrightarrow & (\bar{M} \setminus M_+, \bar{M}) & \longleftarrow & (\partial M, \bar{M}) \\ \downarrow & & \downarrow & & \downarrow \\ (\bar{N}_+ \setminus M_+, \bar{N}_+) & \longrightarrow & (\bar{N} \setminus M_+, \bar{N}) & \longleftarrow & (\bar{N} \setminus M, \bar{N}) \\ \uparrow & & \uparrow & & \uparrow \\ (\partial N_+, \bar{N}_+) & \longrightarrow & (\bar{N} \setminus N_+, \bar{N}) & \longleftarrow & (\partial N, \bar{N}) \end{array}$$

Applying the $\mathcal{M}\text{ap}$ construction, we obtain the following commutative diagram in $\text{Corr}(\text{Stk}_{\mathbb{F}_q})_{\text{prop};\text{sm}}$:

$$(2.6.6) \quad \begin{array}{ccccc} (BB, BG)^{\partial M_+, \bar{M}_+} & \xleftarrow{f} & (BB, BG)^{\bar{M} \setminus M_+, \bar{M}} & \xrightarrow{g} & (BB, BG)^{\partial M, \bar{M}} \\ p \uparrow & 2 & p \uparrow & 1 & p \uparrow \\ (BB, BG)^{\bar{N}_+ \setminus M_+, \bar{N}_+} & \xleftarrow{f} & (BB, BG)^{\bar{N} \setminus M_+, \bar{N}} & \xrightarrow{g} & (BB, BG)^{\bar{N} \setminus M, \bar{N}} \\ q \downarrow & 3 & q \downarrow & 4 & q \downarrow \\ (BB, BG)^{\partial N_+, \bar{N}_+} & \xleftarrow{f} & (BB, BG)^{\bar{N} \setminus N_+, \bar{N}} & \xrightarrow{g} & (BB, BG)^{\partial N, \bar{N}} \end{array}$$

Lemma 2.6.7 The diagram (2.6.5) is adjointable in $\text{Corr}((\text{Spc}_{\text{fin}}^{\Delta^1})^{\text{op}})$. As a result, (2.6.6) is adjointable in $\text{Corr}(\text{Stk})_{\text{prop};\text{sm}}$.

Proof The second part follows from the first part because $\mathcal{M}\text{ap}$ turns colimits to limits, and hence, in particular, it sends pushout squares to pullback squares. The first part is an explicit and elementary statement about gluing 1-manifolds that is easier to check directly than to describe. We leave the details to the reader. □

Proposition 2.6.8 *The augmented simplicial object $\overline{\mathrm{Tr}}(\mathrm{H}_G^{?,\mathrm{ren}})_\bullet$ in $\mathrm{Shv}_?(pt)\text{-Mod}$ satisfies the condition of Proposition 2.3.3. Namely, for every $\alpha: [n] \rightarrow [m]$ in Δ_+ , the diagram*

$$(2.6.9) \quad \begin{array}{ccc} \overline{\mathrm{Tr}}(\mathrm{H}_G^{?,\mathrm{ren}})_{m+1} & \xrightarrow{d_0} & \overline{\mathrm{Tr}}(\mathrm{H}_G^{?,\mathrm{ren}})_m \\ \downarrow & & \downarrow \\ \overline{\mathrm{Tr}}(\mathrm{H}_G^{?,\mathrm{ren}})_{n+1} & \xrightarrow{d_0} & \overline{\mathrm{Tr}}(\mathrm{H}_G^{?,\mathrm{ren}})_n \end{array}$$

is vertically right adjointable.

Proof Note that (2.6.9) is obtained from (2.6.6) by applying $\mathrm{Shv}_{?,!}^{*,\mathrm{ren}}$. By Lemma 2.4.11, in (2.6.6) the morphisms f and p (resp. g and q) are smooth (resp. proper). To prove that (2.6.9) is vertically right adjointable, it suffices to show that we have the natural equivalence

$$g_! f^* p_* q^! \simeq p_* q^! g_! f^*,$$

where we start from the bottom left of (2.6.6). Indeed, we have

$$g_! f^* p_* q^! \simeq g_! p_* f^* q^! \simeq p_* g_! q^! f^* \simeq p_* q^! g_! f^*,$$

where the first, second, and third equivalences are due to the following reasons, respectively:

- smooth base change for square 2, using the fact that f is smooth,
- commuting upper $!$ and upper $*$ using the fact that f is smooth, and commuting lower $!$ and lower $*$ using the fact that g is proper, and
- using the fact that g is proper, the desired equivalence follows from $g_* q^! \simeq q^! g_*$, which is the Verdier dual of the usual proper base change result. □

Remark 2.6.10 Unlike the “big” version, the “small” variant $\overline{\mathrm{Tr}}(\mathrm{H}_G^?)_\bullet$ discussed in Section 2.5.5 does not satisfy the adjointability condition of Proposition 2.3.3. This is because the right adjoint to the unit map does not preserve constructibility. Indeed, the right adjoint is given by $p_* q^!$ in the following diagram:

$$\begin{array}{ccc} & BB & \\ q \swarrow & & \searrow p \\ BB \times_{BG} BB & & pt. \end{array}$$

But now, note that p_* does not preserve constructibility.

2.7 Traces and Drinfel’d centers of finite Hecke categories

We are now ready to complete the proof of the first main result. The trace case follows directly from the discussion above and thus will be handled at the beginning of this subsection. We will then deduce the Drinfel’d center case from the trace case.

2.7.1 Traces of finite Hecke categories We will now prove the trace part of the first main result:

Theorem 2.7.2 *The trace of $H_G^{?,ren}$ (resp. $H_G^?$) where $H_G^{?,ren}$ (resp. $H_G^?$) is viewed as an algebra object of $\text{Shv}_?(pt)\text{-Mod}$ (resp. $\text{Shv}_{?,c}(pt)\text{-Mod}$) coincides with the full subcategory of $\text{Shv}_?(G/G)^{ren}$ (resp. $\text{Shv}_{?,c}(G/G)$) generated under colimits (resp. finite colimits and idempotent completion) by the image of $H_G^{?,ren}$ (resp. $H_G^?$) via $q_!p^*$ in the horocycle correspondence (see Remark 1.4.2)*

(2.7.3)

$$\begin{array}{ccc}
 & G/B & \\
 p \swarrow & & \searrow q \\
 B \backslash G/B & & G/G
 \end{array}$$

Moreover, under this identification, the natural functor $\text{tr}: H_G^{?,ren} \rightarrow \text{Tr}(H_G^{?,ren})$ (resp. $\text{tr}: H_G^? \rightarrow \text{Tr}(H_G^?)$) is identified with $q_!p^*$.

Proof We start with the “big” variant $H_G^{?,ren}$. Propositions 2.6.8 and 2.3.3 imply that the natural functor $\text{Tr}(H_G^{?,ren}) \rightarrow \text{Shv}_?(G/G)^{ren}$ is fully faithful. Moreover, the functor $H_G^{?,ren} \rightarrow \text{Tr}(H_G^{?,ren}) \hookrightarrow \text{Shv}_?(G/G)^{ren}$ identifies with $q_!p^*$ in the horocycle correspondence.

By [25, Chapter 1, Proposition 8.7.4], we know that $\text{Tr}(H_G^{?,ren})$ is compactly generated by the essential image of $H_G^? \rightarrow H_G^{?,ren} \rightarrow \text{Tr}(H_G^{?,ren})$. Note that the conditions required to apply this result amount to the existence of the “small” variant introduced in Section 2.5.5.

Combining the two statements, we conclude that $\text{Tr}(H_G^{?,ren})$ is identified with the full subcategory of $\text{Shv}_?(G/G)^{ren}$ compactly generated by the image of $H_G^? = \text{Shv}_{?,c}(B \backslash G/B)$ under the horocycle correspondence. In particular, it is generated by the image of $H_G^{?,ren}$ under colimits.

The “small” variant is obtained from the first by taking the full subcategories of compact objects, using Proposition 2.1.2. The statement regarding generation under finite colimits and idempotent splittings follows from [25, Chapter 1, Lemma 7.2.4(1'')]. □

2.7.4 Drinfel’d centers of the finite Hecke categories The case of Drinfel’d center is slightly more subtle. Consider the following versions $H_G^{?,ren,!}$ (resp. $H_G^{?,!}$) of the Hecke categories where instead of using $\text{Shv}_{?,!}^{ren,*}$ (resp. $\text{Shv}_{?,c,!}^*$), we use $\text{Shv}_{?,*}^{ren,!}$ (resp. $\text{Shv}_{?,c,*}^!$). More concretely, we use $g_*f^!$ in the correspondence (1.4.7) to define the convolution monoidal structure.

Since f is smooth of relative dimension $\dim B$, $f^! \simeq f^*[2 \dim B](\dim B)$ in the mixed case and $f^! \simeq f^*[2 \dim B]\langle 2 \dim B \rangle$ in the graded case, where $(-)$ denotes the Tate twist and $\langle - \rangle$ denotes the grading shift defined in [31, Section 2.4.7].⁹ Note that in the ungraded setting, we simply have $f^! \simeq f^*[2 \dim B]$. Thus, by cohomologically shifting and Tate twisting $[-2 \dim B](-\dim B)$ for the mixed case (resp. cohomologically shifting and grade shifting $[-2 \dim B]\langle -2 \dim B \rangle$ for the graded case, resp. cohomologically shifting $[-2 \dim B]$ for the ungraded case), we obtain an equivalence of monoidal categories

$$H_G^{?,ren} \cong H_G^{?,ren,!} \quad (\text{resp. } H_G^? \cong H_G^{?,!}).$$

⁹The factor 2 is there because weight is twice the Tate twist.

Even though the $!$ -version is equivalent to the usual version, it is, as we shall see, technically more advantageous to use when studying the center.

2.7.5 Observe that the category $H_G^{?,ren,!}$ is self-dual as an object in $Shv_?(pt)\text{-Mod}$. Indeed, first note that any object $\mathcal{Y} \in \text{Corr}(\text{Stk})$ is self-dual, with duality datum given by

$$pt \leftarrow \mathcal{Y} \rightarrow \mathcal{Y} \times \mathcal{Y} \quad \text{and} \quad \mathcal{Y} \times \mathcal{Y} \leftarrow \mathcal{Y} \rightarrow pt.$$

Thus, if \mathcal{Y} is a finite orbit stack, such as $\mathcal{Y} = BB \times_{BG} BB$, $Shv_{?,*}^{ren,!}$ turns the self-duality datum above into a self-duality datum of $Shv_?(X)^{ren}$, using Proposition 2.2.1. In particular, $H_G^{?,ren,!}$ is self dualizable.

Suppose

$$\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y}$$

is a morphism from \mathcal{X} to \mathcal{Y} in $\text{Corr}(\text{Stk})$. The dual of this morphism is precisely

$$\mathcal{Y} \leftarrow \mathcal{Z} \rightarrow \mathcal{X}.$$

Thus, if \mathcal{X} and \mathcal{Y} are finite orbit stacks, then the functors $Shv_?(X)^{ren} \rightarrow Shv_?(Y)^{ren}$ and $Shv_?(Y)^{ren} \rightarrow Shv_?(X)^{ren}$ associated to the two correspondences above are dual to each other. Note that when following the correspondences here, we use $(-)_*(-)^!$.

2.7.6 We are now ready to compute the Drinfel'd centers of $H_G^{?,ren}$.

Theorem 2.7.7 *The Drinfel'd center $Z(H_G^{?,ren})$ of $H_G^{?,ren}$, where the latter is viewed as an object in the category $\text{Alg}(Shv_?(pt)\text{-Mod})$, coincides with its trace $\text{Tr}(H_G^{?,ren})$. Moreover, under the identification of $\text{Tr}(H_G^{?,ren})$ as a full subcategory of $Shv_?(G/G)^{ren}$, the natural functor $z: Z(H_G^{?,ren}) \rightarrow H_G^{?,ren}$ can be identified with $p_*q^!$ in the horocycle correspondence (2.7.3). In particular, we have a pair of adjoint functors $\text{tr} \dashv z$.*

Proof To simplify the notation, we will write $\mathcal{A} = H_G^{?,ren,!}$ and $\mathcal{B} = H_G^{?,ren}$. Moreover, unless otherwise specified, all tensors and $\underline{\mathcal{H}om}$ are over $Shv_?(pt)$. We will therefore omit $Shv_?(pt)$ from the notation.

Recalling from Definition 1.4.5, we have

$$Z(\mathcal{A}) \simeq \underline{\mathcal{H}om}_{\mathcal{A} \otimes \mathcal{A}^{rev}}(\mathcal{A}, \mathcal{A}),$$

ie the category of continuous $\mathcal{A} \otimes \mathcal{A}^{rev}$ -linear functors from \mathcal{A} to itself. As in [8, Section 5.1.1], we have an equivalence of categories $\mathcal{A} \simeq |\mathcal{A}^{\otimes(\bullet+2)}|$. Thus,

$$\begin{aligned} \underline{\mathcal{H}om}_{\mathcal{A} \otimes \mathcal{A}^{rev}}(\mathcal{A}, \mathcal{A}) &\simeq \underline{\mathcal{H}om}_{\mathcal{A} \otimes \mathcal{A}^{rev}}(|\mathcal{A}^{\otimes(\bullet+2)}|, \mathcal{A}) \\ &\simeq \text{Tot}(\underline{\mathcal{H}om}_{\mathcal{A} \otimes \mathcal{A}^{rev}}(\mathcal{A}^{\otimes(\bullet+2)}, \mathcal{A})) \simeq \text{Tot}(\underline{\mathcal{H}om}(\mathcal{A}^{\otimes\bullet}, \mathcal{A})) \\ \text{(2.7.8)} \quad &\simeq \text{Tot}(\mathcal{A}^{\otimes(\bullet+1)}) \end{aligned}$$

$$\begin{aligned} \text{(2.7.9)} \quad &\simeq |\mathcal{B}^{\otimes(\bullet+1)}| \\ &\simeq \text{Tr}(\mathcal{B}). \end{aligned}$$

Here (2.7.8) is obtained by passing to the duals, using the discussion in Section 2.7.5. Moreover, (2.7.9) is obtained by passing to left adjoints [38, Corollary 5.5.3.4] (see also [25, Chapter 1, Proposition 2.5.7]), and noticing that the resulting diagram is precisely the diagram used to compute the trace of $\mathcal{B} := H_G^{?,ren}$. We deduce the corresponding statement for $Z(\mathcal{B})$ using the equivalence of monoidal categories $\mathcal{A} \simeq \mathcal{B}$.

The second part regarding the identification of the functor z follows from the argument above. □

2.7.10 We will now study the centers of the “small” version $H_G^?$, which is more subtle than the case of traces above.

Theorem 2.7.11 *The Drinfel’d center $Z(H_G^?)$ of $H_G^? \in \text{Alg}(\text{Shv}_{\gamma,c}(\text{pt})\text{-Mod})$ is equivalent to the full subcategory of $Z(H_G^{?,ren})$ consisting of objects whose images under the natural central functor z are constructible as sheaves on $B \setminus G/B$.*

Proof For brevity’s sake, we write $\mathcal{A} = H_G^{?,ren,!}$ and $\mathcal{A}_0 = H_G^{?,!}$. Moreover, all tensors and $\underline{\mathcal{H}om}$, unless otherwise specified, will be over $\text{Shv}_{\gamma}(\text{pt})$ or $\text{Shv}_{\gamma,c}(\text{pt})$ depending on whether we are working with “big” or “small” categories, which should be clear from the context. We will thus not include $\text{Shv}_{\gamma}(\text{pt})$ or $\text{Shv}_{\gamma,c}(\text{pt})$ in the notation.

By definition,

$$Z(\mathcal{A}_0) \simeq \underline{\mathcal{H}om}_{\mathcal{A}_0 \otimes \mathcal{A}_0^{rev}}(\mathcal{A}_0, \mathcal{A}_0),$$

the category of exact $\mathcal{A}_0 \otimes \mathcal{A}_0^{rev}$ -linear functors from \mathcal{A}_0 to itself. As in the proof of Theorem 2.7.7, we have

$$Z(\mathcal{A}_0) \simeq \text{Tot}(\underline{\mathcal{H}om}(\mathcal{A}_0^{\otimes \bullet}, \mathcal{A}_0)),$$

which naturally embeds into

$$Z(\mathcal{A}) \simeq \text{Tot}(\underline{\mathcal{H}om}(\mathcal{A}^{\otimes \bullet}, \mathcal{A})),$$

whose essential image is the full subcategory of $Z(\mathcal{A})$ consisting of objects whose images in $\underline{\mathcal{H}om}(\mathcal{A}^{\otimes \bullet}, \mathcal{A})$ are compact-preserving functors. Observe that all the functors between $\mathcal{A}^{\otimes \bullet}$ (which induces functors between $\text{Tot}(\underline{\mathcal{H}om}(\mathcal{A}^{\otimes \bullet}, \mathcal{A}))$) are compact preserving, since we only push along schematic (in fact proper) morphisms, by Lemma 2.4.11. Thus $Z(\mathcal{A}_0)$ is, equivalently, the full subcategory of $Z(\mathcal{A})$ spanned by objects whose images in

$$\underline{\mathcal{H}om}(\mathcal{A}^{\otimes 0}, \mathcal{A}) \simeq \underline{\mathcal{H}om}(\text{Shv}_{\gamma}(\text{pt}), \mathcal{A}) \simeq \mathcal{A} \simeq \text{Shv}_{gr}(B \setminus G/B)^{ren}$$

are compact, ie constructible. But this functor is precisely the central functor z and is identified with $p_*q^!$ in the horocycle correspondence (2.7.3), as in the proof of Theorem 2.7.7 above.

The statement for $H_G^?$ follows since $H_G^? \simeq H_G^{?,!}$ and the proof concludes. □

2.8 Character sheaves

We will now state our results in terms of character sheaves. We start with the following definition:

Definition 2.8.1 The *renormalized category of mixed (resp. graded, resp. ungraded) unipotent character sheaves* of G , denoted by $\text{Ch}_G^{u,?,\text{ren}}$ where $? = \text{m}$ (resp. $? = \text{gr}$, resp. $? = \emptyset$), is the full subcategory of $\text{Shv}_?(G/G)^{\text{ren}}$ generated by colimits by the image of $H_G^{?,\text{ren}}$ under $q_!p^*$ in the horocycle correspondence (2.7.3). Equivalently, it is compactly generated by the image of $H_G^?$ under this functor.

The *category of mixed (resp. graded, resp. ungraded) unipotent character sheaves* of G is the full subcategory spanned by compact objects, ie

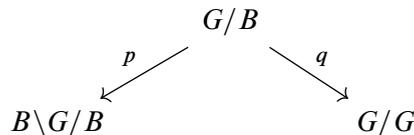
$$\text{Ch}_G^{u,?,c} := (\text{Ch}_G^{u,?,\text{ren}})^c = \text{Ch}_G^{u,?,\text{ren}} \cap \text{Shv}_{?,c}\left(\frac{G}{G}\right).$$

Finally, the *category of mixed (resp. graded, resp. ungraded) monodromic character sheaves*, denoted by $\widetilde{\text{Ch}}_G^{u,?,c}$, is the full subcategory of $\text{Ch}_G^{u,?,\text{ren}}$ consisting of objects whose images under $p_*q^!$ are constructible, where p and q are defined in the horocycle correspondence (2.7.3).

Remark 2.8.2 The last equality in the definition of $\text{Ch}_G^{u,?,c}$ uses the following fact: if we have a fully faithful and compact-preserving functor $A \hookrightarrow B$, then $A^c = A \cap B^c$. Indeed, $A^c \subseteq A \cap B^c$ due to the compact preservation assumption. On the other hand, $A \cap B^c \subseteq A^c$ is always true by the very definition of compactness. In the case of interest, the fact that the functor involved is compact preserving is proved in Theorem 2.7.2 above.

With the new notation, Theorems 2.7.2, 2.7.7, and 2.7.11 above can be summarized:

Theorem 2.8.3 The trace $\text{Tr}(H_G^{?,\text{ren}})$ and center $Z(H_G^{?,\text{ren}})$ of $H_G^{?,\text{ren}}$ coincide with the full subcategory $\text{Ch}_G^{u,?,\text{ren}}$ of $\text{Shv}_?(G/G)^{\text{ren}}$ generated under colimits by the essential image of $H_G^{?,\text{ren}}$ under $q_!p^*$ in the correspondence



Moreover, under this identification, the canonical trace and center maps are adjoint $\text{tr} \dashv z$ and are identified with the adjoint pair $q_!p^* \dashv p_*q^!$.

The trace $\text{Tr}(H_G^?)$ coincides with the full subcategory $\text{Ch}_G^{u,?,c}$ of $\text{Ch}_G^{u,?,\text{ren}}$ spanned by compact objects. Moreover, the center $Z(H_G^?)$ is the full subcategory $\widetilde{\text{Ch}}_G^{u,?,c}$ of $\text{Ch}_G^{u,?,\text{ren}}$ spanned by the **preimage** of $H_G^?$ under the central functor z .

Remark 2.8.4 It is easy to see that $\text{Ch}_G^{u,\text{gr}}$ is contained in $\widetilde{\text{Ch}}_G^{u,\text{gr}}$. In fact, the latter is strictly larger, already when $G = \mathbb{G}_m$, the multiplicative group. Indeed, in this case, $H_{\mathbb{G}_m}^{\text{gr},\text{ren}} \simeq \text{Shv}_{\text{gr}}(B\mathbb{G}_m)^{\text{ren}}$ and $\text{Ch}_{\mathbb{G}_m}^{u,\text{gr},\text{ren}}$ is the full subcategory of $\text{Shv}_{\text{gr}}(\mathbb{G}_m/\mathbb{G}_m)^{\text{ren}} \simeq \text{Shv}_{\text{gr}}(\mathbb{G}_m \times B\mathbb{G}_m)^{\text{ren}}$ generated by the constant sheaf. Moreover, the natural central functor is simply given by p_* where $p: \mathbb{G}_m \times B\mathbb{G}_m \rightarrow B\mathbb{G}_m$ is the projection onto the second factor. Thus, to see that $\widetilde{\text{Ch}}_G^{u,\text{gr}}$ is larger than $\text{Ch}_G^{u,\text{gr}}$, it suffices to realize

that $\pi_* : \text{LS}_{\text{gr}}^u(\mathbb{G}_m) \rightarrow \text{Vect}^{\text{gr}}$ does not reflect constructibility, where $\text{LS}_{\text{gr}}^u(\mathbb{G}_m)$ is the full subcategory of $\text{Shv}_{\text{gr}}(\mathbb{G}_m)$ consisting of unipotent local systems, ie it is generated by (grading shifts of) the constant sheaves, and where $\pi : \mathbb{G}_m \rightarrow \text{pt}$ is the structure morphism.

By construction,¹⁰

$$\text{LS}_{\text{gr}}^u(\mathbb{G}_m) \xrightarrow[\simeq]{\mathcal{H}\text{om}^{\text{gr}}(\overline{\mathbb{Q}}_\ell, -) \simeq \pi_*} \mathcal{E}\text{nd}^{\text{gr}}(\overline{\mathbb{Q}}_\ell)^{\text{op}}\text{-Mod}(\text{Vect}^{\text{gr}}) \simeq \overline{\mathbb{Q}}_\ell[\alpha]\text{-Mod}(\text{Vect}^{\text{gr}}),$$

where $\alpha^2 = 0$ and α lives in cohomological degree 1 and graded degree 2. Under this equivalence, π_* corresponds to the forgetful functor $\overline{\mathbb{Q}}_\ell[\alpha]\text{-Mod}(\text{Vect}^{\text{gr}}) \rightarrow \text{Vect}^{\text{gr}} \simeq \text{Shv}_{\text{gr}}(\text{pt})$. But now, on the one hand, the object $\overline{\mathbb{Q}}_\ell \in \overline{\mathbb{Q}}_\ell[\alpha]\text{-Mod}(\text{Vect}^{\text{gr}})$ is not compact, and hence it corresponds to a nonconstructible sheaf on the left. On the other hand, its image in Vect^{gr} is compact.

Remark 2.8.5 The restriction of the adjunction $\text{tr} : \text{H}_G^{?,\text{ren}} \rightleftarrows \text{Ch}_G^{u,?,\text{ren}} : z$ to small categories yields the commutative diagram

$$\begin{array}{ccc} & & \text{Ch}_G^{u,?} \\ & \nearrow \text{tr} & \downarrow \iota \\ \text{H}_G^? & \xleftarrow{\tilde{\text{tr}}} & \tilde{\text{Ch}}_G^{u,?} \\ & \xleftarrow{z} & \end{array}$$

where $\tilde{\text{tr}} \simeq \iota \circ \text{tr}$ and $\bar{z} \simeq z \circ \iota$.

2.9 Weight structure and perverse t -structure on $\text{Ch}_G^{u,\text{gr}}$

In this subsection, we will show that $\text{Ch}_G^{u,\text{gr}}$ inherits the weight structure and perverse t -structure from the ambient category $\text{Shv}_{\text{gr},c}(G/G)$.

2.9.1 The perverse t -structure on $\text{Ch}_G^{u,?}$ We will now show that for $? \in \{\text{gr}, \emptyset\}$, the category $\text{Ch}_G^{u,?}$ inherits the perverse t -structure from $\text{Shv}_{?,c}(G/G)$. Moreover, $\text{Ch}_G^{u,\text{gr}}$ inherits the weight structure from $\text{Shv}_{\text{gr},c}(G/G)$.

We start with a general lemma concerning t -structures.

Lemma 2.9.2 *Let \mathcal{D} be a triangulated category equipped with a bounded t -structure whose heart is Artinian. Let $\mathcal{C} := \langle \mathcal{L}_i \rangle_{i \in I}$ be the smallest full DG-subcategory of \mathcal{D} containing a collection of simple objects $\{\mathcal{L}_i\}_{i \in I}$ for some indexing set I . Then \mathcal{C} is stable under the truncation of \mathcal{D} , and hence it inherits the t -structure on \mathcal{D} . Consequently, \mathcal{C} is idempotent complete.*

Proof The argument is similar to that of [31, Proposition 3.6.2]. For $\mathcal{F} \in \text{Shv}_{?,c}(\mathcal{Y})$, we will show that the following conditions are equivalent:

- (i) $\mathcal{F} \in \mathcal{C}$.
- (ii) The simple constituents of $\text{H}^i(\mathcal{F})$ belong to $\{\mathcal{L}_i\}_{i \in I}$.

¹⁰The superscript gr in $\mathcal{H}\text{om}^{\text{gr}}$ and $\mathcal{E}\text{nd}^{\text{gr}}$ denotes the Vect^{gr} -enriched Hom-spaces; see [31, Appendix A.2.6].

Indeed, suppose \mathcal{F} satisfies (ii). Then \mathcal{F} can be built from successive extensions of \mathcal{L}_i 's, which then implies that $\mathcal{F} \in \mathcal{C}$, ie \mathcal{F} satisfies (i). The other direction is obtained by observing that (ii) is closed under finite direct sums, shifts, and cones.

(ii) is clearly stable under the perverse truncations of $\text{Shv}_{\gamma,c}(\mathcal{Y})$, and hence \mathcal{C} inherits the perverse t -structure of $\text{Shv}_{\gamma,c}(\mathcal{Y})$. Because this t -structure is bounded, idempotent completeness of \mathcal{C} follows from [2, Corollary 2.14]. □

Corollary 2.9.3 *For $\gamma \in \{\text{gr}, \emptyset\}$, that is, we are working in the graded or ungraded setting, the category $\text{Ch}_G^{\text{u},\gamma}$ is stable under the perverse truncations of $\text{Shv}_{\gamma,c}(G/G)$, and hence it inherits the perverse t -structure on $\text{Shv}_{\gamma,c}(G/G)$.*

Proof We will argue for the graded case as the ungraded case is identical and is in fact well known.

By definition, $\text{Ch}_G^{\text{u},\text{gr}}$ is the smallest idempotent complete full DG-subcategory of $\text{Shv}_{\text{gr},c}(G/G)$ containing the images $\text{tr}(\mathcal{K})$ of $\mathcal{K} \in \text{H}_G^{\text{gr}}$ that are irreducible perverse sheaves, ie they are (grading shifts of) Kazhdan–Lusztig elements. Since tr is obtained by pulling back along a smooth map and pushing forward along a proper map, $\text{tr}(\mathcal{K})$ decomposes into a finite direct sum of simple perverse sheaves.¹¹ Let \mathcal{C} be the smallest full DG-subcategory containing these simple perverse sheaves. By Lemma 2.9.2, \mathcal{C} inherits the perverse t -structure on $\text{Shv}_{\text{gr},c}(G/G)$. It remains to show that $\mathcal{C} = \text{Ch}_G^{\text{u},\text{gr}}$.

Clearly, $\mathcal{C} \subseteq \text{Ch}_G^{\text{u},\text{gr}}$. Moreover, $\text{Ch}_G^{\text{u},\text{gr}}$ is the idempotent completion of \mathcal{C} . But by Lemma 2.9.2, \mathcal{C} is already idempotent complete. □

2.9.4 Weight structure on $\text{Ch}_G^{\text{u},\text{gr}}$

Lemma 2.9.5 *The category $\text{Ch}_G^{\text{u},\text{gr}}$ inherits the weight structure on $\text{Shv}_{\text{gr},c}(G/G)$. Consequently, the trace functor $\text{tr}: \text{H}_G^{\text{gr}} \rightarrow \text{Ch}_G^{\text{u},\text{gr}}$ is weight exact.*

Proof As above, $\text{Ch}_G^{\text{u},\text{gr}}$ is generated as a DG-category by $\text{tr}(\mathcal{K})$ where $\mathcal{K} \in \text{H}_G^{\text{gr}}$ is pure of weight 0. Note that $\text{tr}(\mathcal{K})$ is pure of weight 0 in $\text{Shv}_{\text{gr},c}(G/G)$ since in the horocycle correspondence, we only pull back along a smooth map and push forward along a proper map. Thus, the objects $\text{tr}(\mathcal{K})$ form a negatively self-orthogonal collection in $\text{Shv}_{\text{gr},c}(G/G)$, and hence also in $\text{Ch}_G^{\text{u},\text{gr}}$. By [16, Corollary 2.1.2] (see also [22, Remark 2.2.6]), they form the weight heart of a weight structure on $\text{Ch}_G^{\text{u},\text{gr}}$. It is clear from the definition of this weight structure that it is compatible with the one on $\text{Shv}_{\text{gr},c}(G/G)$. □

2.10 Rigidity and consequences

In this subsection, we show that the finite Hecke categories are (compactly generated) rigid monoidal categories from which we deduce various interesting consequences. This subsection is mostly independent of the rest of the paper.

¹¹Note that this is not necessarily the case if we work with the mixed sheaves, ie when $\gamma = m$.

2.10.1 Rigidity of finite Hecke categories The rigidity of Hecke categories has been established in [9].

Proposition 2.10.2 [9, Theorem 6.2] *The category $H_G^{?,ren}$ is compactly generated rigid monoidal.*

Proof $H_G^{?,ren}$ is compactly generated by construction. Rigidity follows from the same proof as that of [9, Theorem 6.2]. Indeed, the same argument as over there implies that $H_G^{?,ren}$ is semirigid. But since we are working with the renormalized category of sheaves, all constructible sheaves are compact by definition. In particular, the monoidal unit is compact. Thus, they are rigid, by [9, Proposition 3.3]. \square

Remark 2.10.3 As a consequence of Proposition 2.10.2, $H_G^?$ is rigid as “small” monoidal categories in the sense of Remark 2.1.4.

2.10.4 Rigidity of Drinfel’d centers The rigidity of finite Hecke categories implies that for their Drinfel’d centers:

Proposition 2.10.5 *The (braided) monoidal category $Z(H_G^{?,ren})$ is semirigid. Moreover, the (braided) monoidal category $Z(H_G^?)$ is rigid (in the sense of Remark 2.1.4).*

Proof The “small” case follows from [34, Remark 2.4.2]. For the “big” case, first note that by construction, we have a (braided) monoidal functor $Z(H_G^?) \hookrightarrow Z(H_G^{?,ren})$ whose image contains the compact generators of $Z(H_G^{?,ren})$; see also Remark 2.8.4. But since all objects of $Z(H_G^?)$ are dualizable (as the category is rigid), the compact generators of $Z(H_G^{?,ren})$ are also dualizable. In other words, $Z(H_G^{?,ren})$ is compactly generated by dualizable objects, and hence is semirigid by definition; see [9, Definition 3.1]. \square

Directly from Definition 1.4.5, we see that the Drinfel’d center always acts on the original category and its trace. We will use \otimes to denote this action. In the case of Hecke categories, we have the following compatibility:

Corollary 2.10.6 *Using the same notation as in Remark 2.8.5, let $a \in H_G^{?,ren}$ (resp. $a \in H_G^?$) and $b \in Ch_G^{u,?,ren}$ (resp. $b \in \tilde{Ch}_G^{u,?}$). Then we have a natural equivalence*

$$\text{tr}(a \otimes z(b)) \simeq \text{tr}(a) \otimes b \quad (\text{resp. } \tilde{\text{tr}}(a \otimes z(b)) \simeq \tilde{\text{tr}}(a) \otimes b).$$

Proof This is equivalent to stating that the functor tr (resp. $\tilde{\text{tr}}$) is a functor of $Ch_G^{u,?,ren}$ - (resp. $\tilde{Ch}_G^{u,?}$ -) modules. By [9, Lemma 3.5] (resp. [34, Remark 2.4.2]), this follows from the fact that the central functor z is monoidal and that $Ch_G^{u,ren}$ (resp. Ch_G^u) is semirigid (resp. rigid). \square

2.10.7 Hochschild homology Let c be a dualizable object, with dual c^\vee , in a symmetric monoidal category \mathcal{C} (see [25, Chapter 1, Section 4] for an extended discussion on this topic).¹² Then the trace

¹²Note that since we are working in a symmetric monoidal category, the left and right duals coincide.

of c , also known as the *Hochschild homology* of c and denoted¹³ by $\mathrm{HH}(c)$, is an element of $\mathrm{End}_{\mathcal{C}}(1_{\mathcal{C}})$, where $1_{\mathcal{C}}$ is the monoidal unit of \mathcal{C} , given by the composition

$$1_{\mathcal{C}} \xrightarrow{\mathrm{coev}_c} c \otimes c^{\vee} \simeq c^{\vee} \otimes c \xrightarrow{\mathrm{ev}_c} 1_{\mathcal{C}},$$

where coev_c and ev_c are, respectively, the coevaluation and evaluation maps coming from the duality datum between c and c^{\vee} .

The category \mathcal{C} of interest to us is $\mathrm{Shv}_{\gamma}(\mathrm{pt})\text{-Mod}$. Since $\mathrm{Shv}_{\gamma}(\mathrm{pt})$ is a rigid symmetric monoidal category, [25, Chapter 1, Propositions 7.3.2 and 9.4.4] imply that any compactly generated category in $\mathrm{Shv}_{\gamma}(\mathrm{pt})\text{-Mod}$ is dualizable. Given such an object \mathcal{A} in $\mathrm{Shv}_{\gamma}(\mathrm{pt})\text{-Mod}$, its Hochschild homology $\mathrm{HH}(\mathcal{A})$ is an object in $\mathrm{Shv}_{\gamma}(\mathrm{pt})$.

When $\mathcal{A} \in \mathrm{Alg}(\mathrm{Shv}_{\gamma}(\mathrm{pt})\text{-Mod})$ is such that the underlying object $\mathcal{A} \in \mathrm{Shv}_{\gamma}(\mathrm{pt})\text{-Mod}$ is dualizable, $\mathrm{HH}(\mathcal{A})$ acquires a natural algebra structure, ie $\mathrm{HH}(\mathcal{A}) \in \mathrm{Alg}(\mathrm{Shv}_{\gamma}(\mathrm{pt}))$, by [24, Section 3.3.2].

2.10.8 We have the following result from [24, Theorem 3.8.5]. Note that the meanings of HH and Tr in that paper are switched compared to ours. Note also that they prove it for the case where the ambient category is $\mathrm{DGCat}_{\mathrm{pres}, \mathrm{cont}}$. However, the same proof carries through for $\mathrm{Shv}_{\gamma}(\mathrm{pt})\text{-Mod}$.

Theorem 2.10.9 [24, Theorem 3.8.5] *Let \mathcal{A} be a rigid monoidal category in $\mathrm{Shv}_{\gamma}(\mathrm{pt})\text{-Mod}$ such that \mathcal{A} is dualizable. Then there is a canonical equivalence of associative algebras*

$$\mathrm{HH}(\mathcal{A}) \simeq \mathcal{E}nd_{\mathrm{Tr}(\mathcal{A})}^?(\mathrm{tr}(1_{\mathcal{A}})) \in \mathrm{Alg}(\mathrm{Shv}_{\gamma}(\mathrm{pt})),$$

where the superscript $?$ in $\mathcal{E}nd$, which is a placeholder for m (resp. gr , resp. nothing/ \emptyset), denotes the $\mathrm{Shv}_m(\mathrm{pt}_0)$ - (resp. $\mathrm{Vect}^{\mathrm{gr}}$ -, resp. Vect -) enriched Hom-spaces. See [31, Appendix A.2.6].

2.10.10 Let $\mathrm{Spr}_G^? = \mathrm{tr}(1_{\mathrm{H}_G^?}) \in \mathrm{Ch}_G^{u,?}$ denote the image of the monoidal unit of the finite Hecke category via the natural functor. In the ungraded setting, this is known as the *Grothendieck–Springer sheaf*. In the mixed (resp. graded) case, we will thus refer to this object as the *mixed (resp. graded) Grothendieck–Springer sheaf*. Proposition 2.10.2 and Theorem 2.10.9 imply the following result:

Theorem 2.10.11 *We have a natural equivalence of algebras*

$$\mathrm{HH}(\mathrm{H}_G^{?, \mathrm{ren}}) \simeq \mathcal{E}nd^?(\mathrm{Spr}_G^?) \in \mathrm{Alg}(\mathrm{Shv}_{\gamma}(\mathrm{pt})).$$

3 Formality of the Grothendieck–Springer sheaf

The first main result of this section, Theorem 3.2.1, states that $\mathcal{E}nd^{\mathrm{gr}}(\mathrm{Spr}_G^{\mathrm{gr}}) \in \mathrm{Alg}(\mathrm{Shv}_{\mathrm{gr}}(\mathrm{pt})) \simeq \mathrm{Alg}(\mathrm{Vect}^{\mathrm{gr}})$ is formal. It is obtained by a spreading argument, taking as input the ungraded case, proved by the second author using transcendental methods. Combining with a generation result of Lusztig in type A ,

¹³This is not to be confused with the notion of trace used above. This is why we will exclusively use the term Hochschild homology for the type of trace discussed here.

the formality result provides a concrete realization of the category of graded unipotent character sheaves, Theorem 3.3.4.

Below, the ungraded case is recalled in Section 3.1, followed by the proof of the graded case in Section 3.2 using a spreading argument. The section then concludes with Section 3.3, where everything can be made more explicit, especially in type A .

3.1 The ungraded case via transcendental method

In this subsection, we will work over \mathbb{C} ; namely, the geometric objects are stacks over \mathbb{C} and the sheaves are in \mathbb{C} -vector spaces. Let $p: B/B \rightarrow G/G$, $q: B/B \rightarrow T/T$, $\text{Ind}_{T \subset B}^G := p_!q^*$, and its right adjoint $\text{Res}_{T \subset B}^G := q_*p^!$. We have $\text{Spr}_G = p_!\mathbb{C}_{B/B} = \text{Ind}_{T \subset B}^G \mathbb{C}_{T/T} \in \text{Ch}_G^u$. Here, all the quotients are obtained by using the adjoint actions.

The category of *all* character sheaves Ch_G was explicitly computed in [36] (for simply connected groups) and [37] (for reductive groups) by using a complex analytic cover of the adjoint quotient G/G to reduce the calculation to generalized Springer theory. We need the following particular statement:

Theorem 3.1.1 [37] *Let G be a reductive group.*

- (i) *There is an equivalence of DG-algebras*

$$\text{End}(\text{Spr}_G) \simeq (\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \mathbb{C}[W].$$

In particular, $\text{End}(\text{Spr}_G)$ is a formal DG-algebra.

- (ii) *There is a natural equivalence $\text{Res}_{T \subset B}^G(\text{Spr}_G) \simeq \mathbb{C}_{T/T}^{\oplus W}$. Moreover, the natural DG-algebra homomorphism $\text{End}(\text{Spr}_G) \rightarrow \text{End}(\text{Res}_{T \subset B}^G(\text{Spr}_G)) \simeq \text{End}(\mathbb{C}_{T/T}^{\oplus W})$ can be expressed explicitly via the commutative diagram*

$$\begin{array}{ccc} \text{End}(\text{Spr}_G) & \xrightarrow{\hspace{10em}} & \text{End}(\text{Res}_{T \subset B}^G(\text{Spr}_G)) \\ \downarrow \simeq & & \downarrow \simeq \\ (\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \mathbb{C}[W] & \xrightarrow{\hspace{10em}} & (\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \text{End}_{\mathbb{C}}(\mathbb{C}[W]) \\ \downarrow \simeq & & \downarrow \simeq \\ \text{End}_{(\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \mathbb{C}[W]}((\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \mathbb{C}[W]) & \rightarrow & \text{End}_{\text{H}^*(BT) \otimes \text{H}^*(T)}((\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \mathbb{C}[W]) \end{array}$$

where the bottom arrow is induced by the restriction of module structure along

$$\text{H}^*(BT) \otimes \text{H}^*(T) \rightarrow (\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \mathbb{C}[W].$$

- (iii) *Similarly, the natural DG-algebra homomorphism $\text{End}(\mathbb{C}_{T/T}) \rightarrow \text{End}(\text{Spr}_G)$ can be identified with*

$$\text{H}^*(BT) \otimes \text{H}^*(T) \rightarrow (\text{H}^*(BT) \otimes \text{H}^*(T)) \otimes \mathbb{C}[W].$$

Picking an abstract isomorphism of fields $\mathbb{C} \cong \overline{\mathbb{Q}}_\ell$, we see that Theorem 3.1.1 holds equally well for cohomology with coefficients in $\overline{\mathbb{Q}}_\ell$. In the rest of this section, we will work with $\overline{\mathbb{Q}}_\ell$ coefficients.

3.2 Formality of the graded Grothendieck–Springer sheaf

The main goal of the current subsection is a graded version of Theorem 3.1.1 formulated in the following theorem, whose proof will conclude in Section 3.2.11, after some preliminary preparation.

Theorem 3.2.1 *Let G be a reductive group over $\overline{\mathbb{F}}_q$, $T \subset B$ and W as before.*

(i) *There is an equivalence of DG-algebras*

$$\mathcal{E}nd^{\text{gr}}(\text{Spr}_G^{\text{gr}}) \simeq (H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_{\ell}[W] \simeq \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \otimes \overline{\mathbb{Q}}_{\ell}[W],$$

where \underline{x} and $\underline{\theta}$ are generators of the cohomology rings $H_{\text{gr}}^*(BT)$ and $H_{\text{gr}}^*(T)$, respectively. In particular, $\mathcal{E}nd^{\text{gr}}(\text{Spr}_G^{\text{gr}})$ is formal. Moreover, \underline{x} and $\underline{\theta}$ have degrees $(2, 2)$ and $(2, 1)$, respectively, with the first (resp. second) index indicating graded (resp. cohomological) degrees.

(ii) *There is a natural equivalence $\text{Res}_{T \subset B}^G(\text{Spr}_G^{\text{gr}}) \simeq \overline{\mathbb{Q}}_{\ell, T/T}^{\oplus W}$. Moreover, after taking cohomology, the natural DG-algebra homomorphism $\mathcal{E}nd^{\text{gr}}(\text{Spr}_G^{\text{gr}}) \rightarrow \mathcal{E}nd^{\text{gr}}(\text{Res}_{T \subset B}^G(\text{Spr}_G^{\text{gr}})) \simeq \mathcal{E}nd^{\text{gr}}(\overline{\mathbb{Q}}_{\ell, T/T}^{\oplus W})$ can be expressed explicitly via the commutative diagram*

$$\begin{array}{ccc} H^*(\mathcal{E}nd^{\text{gr}}(\text{Spr}_G^{\text{gr}})) & \xrightarrow{\quad\quad\quad} & H^*(\mathcal{E}nd^{\text{gr}}(\text{Res}_{T \subset B}^G(\text{Spr}_G^{\text{gr}}))) \\ \downarrow \simeq & & \downarrow \simeq \\ (H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_{\ell}[W] & \xrightarrow{\quad\quad\quad} & (H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \mathcal{E}nd_{\overline{\mathbb{Q}}_{\ell}}(\overline{\mathbb{Q}}_{\ell}[W]) \\ \downarrow \simeq & & \downarrow \simeq \\ \mathcal{E}nd_{(H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_{\ell}[W]}^{\text{gr}}((H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_{\ell}[W]) & \rightarrow & \mathcal{E}nd_{H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)}^{\text{gr}}((H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_{\ell}[W]) \end{array}$$

where the bottom arrow is induced by the restriction of module structure along

$$H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T) \rightarrow (H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_{\ell}[W].$$

(iii) *Similarly, after taking cohomology, the natural DG-algebra homomorphism $\mathcal{E}nd^{\text{gr}}(\overline{\mathbb{Q}}_{\ell, T/T}) \rightarrow \mathcal{E}nd^{\text{gr}}(\text{Spr}_G)$ can be identified with*

$$H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T) \rightarrow (H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T)) \otimes \overline{\mathbb{Q}}_{\ell}[W].$$

Combining with Theorem 2.10.11, we obtain the following result:

Corollary 3.2.2 *The DG-algebra $\text{HH}(H_G^{\text{gr}, \text{ren}})$ is formal, and we have an equivalence of DG-algebras*

$$\text{HH}(H_G^{\text{gr}, \text{ren}}) \simeq \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \otimes \overline{\mathbb{Q}}_{\ell}[W],$$

where \underline{x} and $\underline{\theta}$ have degrees $(2, 2)$ and $(2, 1)$, respectively, where the first (resp. second) index indicates graded (resp. cohomological) degree.

3.2.3 The strategy for proving Theorem 3.2.1 The passage from Theorem 3.1.1 to Theorem 3.2.1 is via a spreading argument that we will now explain. Let R be a strictly Henselian discrete valuation ring between $\mathbb{Z}[1/(IN)]$ and \mathbb{C} , where $N \gg 0$. Let $\iota: \text{Spec } \overline{\mathbb{F}}_q \rightarrow \text{Spec } R$ and $j: \text{Spec } \mathbb{C} \rightarrow \text{Spec } R$ be

geometric points over the special and generic points of $\text{Spec } R$, respectively. Then we have the symmetric monoidal equivalences of categories

$$\text{Vect} \simeq \text{Shv}(\text{Spec } \overline{\mathbb{F}}_q) \xleftarrow{\simeq} \text{LS}(\text{Spec } R) \xrightarrow{\simeq} \text{Shv}(\text{Spec } \mathbb{C}) \simeq \text{Vect},$$

where $\text{LS}(\text{Spec } R)$ denotes the category of $\overline{\mathbb{Q}}_\ell$ -local systems on $\text{Spec } R$. This induces equivalences of categories

$$(3.2.4) \quad \text{Alg}(\text{Vect}) \simeq \text{Alg}(\text{Shv}(\text{Spec } \overline{\mathbb{F}}_q)) \xleftarrow{\simeq} \text{Alg}(\text{LS}(\text{Spec } R)) \xrightarrow{\simeq} \text{Alg}(\text{Shv}(\text{Spec } \mathbb{C})) \simeq \text{Alg}(\text{Vect}).$$

Thus, if we have an algebra in $\text{Alg}(\text{Shv}(\text{Spec } \overline{\mathbb{F}}_q))$ whose formality we would like to establish, it suffices to produce a natural candidate in $\text{Alg}(\text{LS}(\text{Spec } R))$ whose image under j^* is known to be formal in $\text{Alg}(\text{Shv}(\text{Spec } \mathbb{C}))$.

The algebra in question is $\text{End}(\text{Spr}_G)$, which we will now spread out to $\text{Spec } R$.

3.2.5 Spreading out Let G_R denote the split reductive group over $\text{Spec } R$ given by the same root datum as that of G . Fix $T_R \subset B_R$ a pair of a maximal torus and a Borel subgroup. All the objects considered above have natural relative versions over R . We will use the subscript R (resp. $\overline{\mathbb{F}}_q$, resp. \mathbb{C}) in the notation, for example, Spr_{G_R} (resp. $\text{Spr}_{G_{\overline{\mathbb{F}}_q}}$, resp. $\text{Spr}_{G_{\mathbb{C}}}$), when it is necessary to emphasize where these objects live over, ie over R (resp. $\overline{\mathbb{F}}_q$, resp. \mathbb{C}).

Let $\text{St}_{G_R} = B_R/B_R \times_{G_R/G_R} B_R/B_R$. Then we have the following composition of correspondences:

$$(3.2.6) \quad \begin{array}{ccccc} & & \text{St}_{G_R} & & \\ & \swarrow r & & \searrow s & \\ & B_R/B_R & & B_R/B_R & \\ \swarrow q & & & & \searrow q \\ T_R/T_R & & G_R/G_R & & T_R/T_R \end{array}$$

Lemma 3.2.7 (Mackey filtration) St_{G_R} has a locally closed stratification indexed by $w \in W$,

$$\text{St}_{G_R} = \bigsqcup_{w \in W} \text{St}_{G_R}^w,$$

such that on $\text{St}_{G_R}^w$, (3.2.6) is identified with

$$\begin{array}{ccccc} & & B_R^w/B_R^w & & \\ & \swarrow i & & \searrow \text{Ad}_w & \\ & B_R/B_R & & B_R/B_R & \\ \swarrow q & & & & \searrow q \\ T_R/T_R & & G_R/G_R & & T_R/T_R \end{array}$$

where $B_R^w = B_R \cap \text{Ad}_w^{-1}(B_R)$.

Proof Observe that for any stack \mathcal{Y} , $\text{Map}(S^1, \mathcal{Y}) \simeq \mathcal{Y} \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Y}$. Hence, if \mathcal{Z} is a locally closed substack of \mathcal{Y} , then

$$\text{Map}(S^1, \mathcal{Z}) \simeq \mathcal{Z} \times_{\mathcal{Z} \times \mathcal{Z}} \mathcal{Z} \simeq \mathcal{Z} \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Z}$$

is a locally closed substack of $\mathcal{Y} \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Y}$. Moreover, if $\mathcal{Y} = \bigsqcup_i \mathcal{Z}_i$ is a stratification of \mathcal{Y} where the \mathcal{Z}_i are locally closed substacks of \mathcal{Y} , then

$$\text{Map}(S^1, \mathcal{Y}) \simeq \mathcal{Y} \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Y} \simeq \bigsqcup_{i,j} \mathcal{Z}_i \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Z}_j \simeq \bigsqcup_i \mathcal{Z}_i \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Z}_i \simeq \bigsqcup_i \mathcal{Z}_i \times_{\mathcal{Z}_i \times \mathcal{Z}_i} \mathcal{Z}_i \simeq \bigsqcup_i \text{Map}(S^1, \mathcal{Z}_i)$$

is a stratification of $\mathcal{Y} \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Y}$ by locally closed substacks. Here, the third equivalence is due to the fact that when $i \neq j$, $\mathcal{Z}_i \times_{\mathcal{Y} \times \mathcal{Y}} \mathcal{Z}_j$ is empty.

Applying this to the case where $\mathcal{Y} = BB_R \times_{BG_R} BB_R$ and the Bruhat stratification, we obtain

$$\begin{aligned} \text{St}_{G_R} &= \text{Map}(S^1, BB_R \times_{BG_R} BB_R) = \text{Map}(S^1, B_R \backslash G_R / B_R) = \bigsqcup_{w \in W} \text{Map}(S^1, B_R \backslash B_R w B_R / B_R) \\ &= \bigsqcup_{w \in W} \text{Map}(S^1, BB_R^w) = \bigsqcup_{w \in W} B_R^w / B_R^w. \end{aligned} \quad \square$$

Corollary 3.2.8 $\text{Res}_{T_R \subset B_R}^{G_R} \text{Spr}_{G_R} \simeq \text{Res}_{T_R \subset B_R}^{G_R} \text{Ind}_{T_R \subset B_R}^{G_R} \overline{\mathbb{Q}}_{\ell, T_R/T_R} \simeq \overline{\mathbb{Q}}_{\ell, T_R/T_R}^{\oplus W}$.

Proof We first show that $\text{Res}_{T_R \subset B_R}^{G_R} \text{Spr}_{G_R}$ is a local system on T_R/T_R concentrated in cohomological degree 0. The filtration in Lemma 3.2.7 implies that $\text{Res}_{T_R \subset B_R}^{G_R} \text{Spr}_{G_R}$ has a filtration whose associated pieces are given by $(q \circ i)_* \overline{\mathbb{Q}}_{\ell, B_R^w/B_R^w}$, which is simply $\overline{\mathbb{Q}}_{\ell, T_R/T_R}$. Thus, $\text{Res}_{T_R \subset B_R}^{G_R} \text{Spr}_{G_R}$ is a complex of local systems. To see that it concentrates in degree 0, it suffices to check the stalk at a point in $T_{\mathbb{C}}/T_{\mathbb{C}}$. But this is now a well-known statement; see, for example, [18, Proposition 3.2].

Generically on T_R/T_R , the map $\text{St}_{G_R} \rightarrow T_R/T_R$ is a trivial W -cover. Thus, on this open dense subset, $\text{Res}_{T_R \subset B_R}^{G_R} \text{Spr}_{G_R}$ is a trivial local system of rank $|W|$. This implies the same statement for $\text{Res}_{T_R \subset B_R}^{G_R} \text{Spr}_{G_R}$ itself, as it is the IC-extension of the local system on an open dense subset. \square

Corollary 3.2.9 We have that $\pi_{G_R/G_R, *} \underline{\text{End}}(\text{Spr}_{G_R}) \simeq \pi_{\text{St}_{G_R}, *} \overline{\mathbb{Q}}_{\ell, \text{St}_{G_R}}$ is a complex of local systems on $\text{Spec } R$, ie an object in $\text{LS}(\text{Spec } R)$. Here $\underline{\text{End}}$ denotes the sheaf of endomorphisms, and for any R -stack \mathcal{Y}_R , $\pi_{\mathcal{Y}_R} : \mathcal{Y}_R \rightarrow \text{Spec } R$ denotes the structure map of \mathcal{Y}_R .

Proof The equivalence $\pi_{G_R/G_R, *} \underline{\text{End}}(\text{Spr}_{G_R}) \simeq \pi_{\text{St}_{G_R}, *} \overline{\mathbb{Q}}_{\ell, \text{St}_{G_R}}$ is a standard statement. The second claim follows from the first since

$$\pi_{\text{St}_{G_R}, *} \overline{\mathbb{Q}}_{\ell, \text{St}_{G_R}} \simeq \pi_{T_R/T_R, *} q_* i^* \overline{\mathbb{Q}}_{\ell, \text{St}_{G_R}} \simeq \pi_{T_R/T_R, *} \overline{\mathbb{Q}}_{\ell, T_R/T_R}^{\oplus W} \simeq (\pi_{T_R/T_R, *} \overline{\mathbb{Q}}_{\ell, T_R/T_R})^{\oplus W},$$

where the second equivalence is due to Corollary 3.2.8, and

$$\pi_{T_R/T_R, *} \overline{\mathbb{Q}}_{\ell, T_R/T_R, *} \simeq \pi_{T_R \times B_T, *} \overline{\mathbb{Q}}_{\ell, T_R \times B_T, *},$$

which is a complex of local systems. \square

Corollary 3.2.10 We have $i^* \pi_{G_R/G_R,*} \underline{\mathcal{E}nd}(\text{Spr}_{G_R}) \simeq \mathcal{E}nd(\text{Spr}_{G_{\overline{\mathbb{F}}_q}})$ and $J^* \pi_{G_R/G_R,*} \underline{\mathcal{E}nd}(\text{Spr}_{G_R}) \simeq \mathcal{E}nd(\text{Spr}_{G_{\mathbb{C}}})$.

Proof We will prove the statement for J^* only; the proof for i^* is identical. By adjunction, we have a natural map of algebras

$$J^* \pi_{G_R/G_R,*} \underline{\mathcal{E}nd}(\text{Spr}_{G_R}) \rightarrow \pi_{G_{\mathbb{C}}/G_{\mathbb{C}},*} \underline{\mathcal{E}nd}(\text{Spr}_{G_{\mathbb{C}}}) \simeq \mathcal{E}nd(\text{Spr}_{G_{\mathbb{C}}}).$$

It suffices to show that this is an equivalence of the underlying chain complexes. But now, we have

$$\begin{aligned} J^* \pi_{G_R/G_R,*} \underline{\mathcal{E}nd}(\text{Spr}_{G_R}) &\simeq J^* \pi_{\text{St}_{G_R},*} \overline{\mathcal{Q}}_{\ell, \text{St}_{G_R}} \simeq J^* \pi_{T_R/T_R,*} \overline{\mathcal{Q}}_{\ell, T_R/T_R}^{\oplus W} \simeq \pi_{T_{\mathbb{C}}/T_{\mathbb{C}},*} \overline{\mathcal{Q}}_{\ell, T_{\mathbb{C}}/T_{\mathbb{C}}}^{\oplus W} \\ &\simeq \mathcal{E}nd(\text{Spr}_{G_{\mathbb{C}}}). \end{aligned}$$

Here, the third equivalence is by direct computation (as it only involves the torus) and the other equivalences are due to Corollary 3.2.9, which holds equally over $\overline{\mathbb{F}}_q$ and \mathbb{C} . □

3.2.11 Proof of Theorem 3.2.1 Applying the discussion in Section 3.2.3 to $\pi_{G_R/G_R,*} \underline{\mathcal{E}nd}(\text{Spr}_{G_R})$ using Corollaries 3.2.9 and 3.2.10 and the result over \mathbb{C} , Theorem 3.1.1(i), we obtain the formality of the algebra $\mathcal{E}nd(\text{Spr}_{G_{\overline{\mathbb{F}}_q}})$. But since the formality of a graded DG-algebra can be detected at the ungraded level, we obtain the formality of $\mathcal{E}nd^{\text{gr}}(\text{Spr}_G^{\text{gr}})$ as well. In particular, we have an equivalence of DG-algebras

$$\mathcal{E}nd^{\text{gr}}(\text{Spr}_G^{\text{gr}}) \simeq H^*(\mathcal{E}nd^{\text{gr}}(\text{Spr}_G^{\text{gr}})) \simeq \overline{\mathcal{Q}}_{\ell}[\underline{x}, \underline{\theta}] \otimes \overline{\mathcal{Q}}_{\ell}[W]$$

for *some* variables \underline{x} and $\underline{\theta}$ whose cohomology degrees we know (ie 2 and 1, respectively) but whose graded degrees still need to be computed.

We note that in the case of a torus, $\text{Spr}_T^{\text{gr}} \simeq \overline{\mathcal{Q}}_{\ell, T/T}$ and we have

$$\mathcal{E}nd^{\text{gr}}(\text{Spr}_T^{\text{gr}}) \simeq H_{\text{gr}}^*(BT) \otimes H_{\text{gr}}^*(T) \simeq \overline{\mathcal{Q}}_{\ell}[\underline{x}, \underline{\theta}],$$

where \underline{x} and $\underline{\theta}$ have degrees (2, 2) and (2, 1), respectively. We will deduce the general case from this case.

Over $\overline{\mathbb{F}}_q$, we still have the equivalence $\text{Res}_{T \subset B}^G(\text{Spr}_G) \simeq \overline{\mathcal{Q}}_{\ell, T/T}^{\oplus W}$, which is the precise statement of [18, Proposition 3.2]. In fact, the ungraded version¹⁴ of Theorem 3.2.1(ii) follows from (3.2.4), Corollary 3.2.10 and the complex version, which is Theorem 3.1.1(ii). In particular, at the ungraded level, the natural algebra map

$$\mathcal{E}nd(\text{Spr}_G) \rightarrow \mathcal{E}nd(\text{Res}_{T \subset B}^G \text{Spr}_G)$$

is identified with

$$\overline{\mathcal{Q}}_{\ell}[\underline{x}, \underline{\theta}] \otimes \overline{\mathcal{Q}}_{\ell}[W] \hookrightarrow \mathcal{E}nd_{\overline{\mathcal{Q}}_{\ell}}(\overline{\mathcal{Q}}_{\ell}[W]) \otimes (H^*(BT) \otimes H^*(T)) \simeq \overline{\mathcal{Q}}_{\ell}[\underline{x}, \underline{\theta}] \otimes \mathcal{E}nd_{\overline{\mathcal{Q}}_{\ell}}(\overline{\mathcal{Q}}_{\ell}[W]).$$

The grading on the right is known, as this is the case of a torus. By degree considerations, \underline{x} and $\underline{\theta}$ on the left must have degrees (2, 2) and (2, 1), respectively.

The identifications of the algebra morphisms in Theorem 3.2.1(ii) and (iii) follow from the ungraded case. □

¹⁴That is, we still work over $\overline{\mathbb{F}}_q$, but forget the grading. In other words, we work with Shv rather than Shv_{gr} .

Remark 3.2.12 In the above, we used the fact that the formality of a graded DG-algebra (ie a DG-algebra with an *extra* formal grading) can be detected after forgetting the grading. This can be seen, for example, by applying the main result of [47] and the fact that the formation of a certain spectral sequence in loc. cit. commutes with forgetting the formal grading.

We expect that the same statement also holds for the *formality of a morphism* between formal graded DG-algebras. However, we know neither how to prove this statement nor a place where it is proved. Because of that, the statements appearing in Theorem 3.2.1(ii) and (iii) are only at the level of cohomology groups.

3.3 An explicit presentation of $\text{Ch}_G^{\text{u,gr}}$

The main goal of this subsection is to give an explicit presentation of $\text{Ch}_G^{\text{u,gr}}$. We will only fully work out type A case.

3.3.1 A generation statement in type A We first recall the following result, which is well known to experts. We include it here for the reader's convenience.

Proposition 3.3.2 *Let G be a reductive group of type A over $\overline{\mathbb{F}}_q$. Then Ch_G^{u} is generated by the Grothendieck–Springer sheaf Spr_G .*

Proof Let $\mathcal{F} \in \text{Ch}_G^{\text{u}}$. Then, by Corollary 2.9.3, we know that \mathcal{F} can be built from successive extensions of irreducible character sheaves. It thus suffices to show that when \mathcal{F} is irreducible, it is a direct summand of Spr_G .

By [41, Theorem 4.4(a)], \mathcal{F} is a summand of $\text{Ind}_{L \subset P}^G(\mathcal{K})$ for some cuspidal character sheaf \mathcal{K} on L , the Levi factor of some parabolic subgroup P of G . The group $A(G) := Z(G)/Z^0(G)$ acts on \mathcal{F} via a character $\chi_{\mathcal{F}}$ which is the composition $A(G) \twoheadrightarrow A(L) \xrightarrow{\chi_{\mathcal{K}}} \overline{\mathbb{Q}}_{\ell}$, where $\chi_{\mathcal{K}}$ is the action of $A(L)$ on \mathcal{K} . Since \mathcal{F} is a unipotent character sheaf, $\chi_{\mathcal{F}}$ is trivial, and hence so is $\chi_{\mathcal{K}}$.

We claim that if L is not a torus, then \mathcal{K} as above with trivial $\chi_{\mathcal{K}}$ must be zero, which would force \mathcal{F} to be a summand of $\text{Ind}_{T \subset B}^G(\overline{\mathbb{Q}}_{\ell}) = \text{Spr}_G$. Indeed, by [42, (7.1.3)], $\mathcal{K} \simeq \text{IC}(\overline{\Sigma}, \mathcal{E})[\dim \Sigma]$, where (Σ, \mathcal{E}) is a cuspidal pair of L in the sense of [40, Definition 2.4]. Moreover, by [40, Section 2.10], the classification of cuspidal pairs of L is further reduced to the classification of unipotent cuspidal pairs of H , where $H = Z_{L'}(s)$, with the group L' being the simply connected cover of $L/Z^0(L)$ and s a semisimple element in an isolated conjugacy class of L' . Now L is of type A , and hence H is also of type A (and is not isomorphic to a torus). Therefore, by the classification of unipotent cuspidal pairs [40, beginning of page 206], we see that $\chi_{\mathcal{K}}$ must be nontrivial if \mathcal{K} is nonzero (it is easy to see that $\chi_{\mathcal{K}}$'s (non)triviality is preserved under the above reductions). \square

We now turn to the graded version of the result above:

Corollary 3.3.3 *Let G be a reductive group of type A over $\overline{\mathbb{F}}_q$. Then $\text{Ch}_G^{\text{u,gr}}$ is generated by grading shifts $\text{Spr}_G^{\text{gr}}\langle - \rangle$ of the graded Grothendieck–Springer sheaf. In other words, $\text{Ch}_G^{\text{u,gr}}$ is generated by Spr_G^{gr} as a $\text{Vect}^{\text{gr},c}$ -module category. Consequently, $\text{Ch}_G^{\text{u,gr,ren}}$ is compactly generated by Spr_G^{gr} as a Vect^{gr} -module category.*

Proof As above, it suffices to show that any irreducible perverse graded character sheaf $\mathcal{K} \in \text{Ch}_G^{\text{u,gr}}$ appears as a direct summand of Spr_G^{gr} up to a grading shift. However, this can be checked after forgetting the grading since simplicity implies purity, and moreover we have a complete description of simple pure graded perverse sheaves by [31, Theorem 3.2.19].

The last statement follows immediately since $\text{Ch}_G^{\text{u,gr,ren}} \simeq \text{Ind}(\text{Ch}_G^{\text{u,gr}})$. □

Theorem 3.3.4 *Let G be a reductive group of type A over $\overline{\mathbb{F}}_q$. Then taking $\mathcal{H}\text{om}_{\text{Ch}_G^{\text{u,gr,ren}}}(\text{Spr}_G^{\text{gr}}, -)$ induces an equivalence of Vect^{gr} -module categories*

$$\begin{aligned} \text{Ch}_G^{\text{u,gr,ren}} &\simeq \overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[W]\text{-Mod}(\text{Vect}^{\text{gr}}) =: (\text{H}_{\text{gr}}^*(BT) \otimes \text{H}_{\text{gr}}^*(T)) \boxtimes \overline{\mathbb{Q}}_\ell[W]\text{-Mod}^{\text{gr}} \\ &\simeq \text{HH}(\text{H}_G^{\text{gr,ren}})\text{-Mod}^{\text{gr}}. \end{aligned}$$

Taking the full subcategory of compact objects, we get

$$\begin{aligned} \text{Ch}_G^{\text{u,gr}} &\simeq \overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[W]\text{-Mod}(\text{Vect}^{\text{gr},c})^{\text{perf}} =: (\text{H}_{\text{gr}}^*(BT) \otimes \text{H}_{\text{gr}}^*(T)) \boxtimes \overline{\mathbb{Q}}_\ell[W]\text{-Mod}^{\text{gr,perf}} \\ &\simeq \text{HH}(\text{H}_G^{\text{gr,ren}})\text{-Mod}^{\text{gr,perf}}. \end{aligned}$$

Here, *perf* denotes the full subcategory consisting of perfect complexes.

Proof The first statement can be obtained by a standard Barr–Beck–Lurie argument using the generation result in Corollary 3.3.3 and the identification of $\text{HH}(\text{H}_G^{\text{gr,ren}})$ in Corollary 3.2.2. The second statement follows from the first by taking compact objects. □

4 A coherent realization of character sheaves via Hilbert schemes of points on \mathbb{C}^2

Working in type A , ie $G = \text{GL}_n$ for some n , this section’s main result, Theorem 4.4.15, relates the category of graded unipotent character sheaves $\text{Ch}_G^{\text{u,gr}} \simeq \text{Tr}(\text{H}_G^{\text{gr}})$ and the Hilbert scheme of n points on \mathbb{C}^2 . The passage from the categorical trace to the Hilbert scheme of points on \mathbb{C}^2 is given by Koszul duality and a result of Krug [35], both of which will be reviewed in Sections 4.1 and 4.2. In Section 4.3, we recall the categorical constructions of cohomological shearing and 2-periodization. The relation between the two allows us to introduce an extra formal grading at the cost of having to work with 2-periodic objects. All of these results are then used in Section 4.4 to prove the main result of this section, Theorem 4.4.15. Finally, in Section 4.5, we match various objects on the two sides, to be used in Section 5 to realize the HOMFLY-PT homology of a link geometrically via Hilbert schemes of points.

In this section, we will work exclusively with the case $G = \text{GL}_n$ and adopt the notation $\text{H}_n^{\text{gr}} := \text{H}_{\text{GL}_n}^{\text{gr}} \simeq \text{Ch}^b(\text{SBim}_n)$.

4.1 Koszul duality: a recollection

We will now recall an equivalence of categories coming out of Koszul duality. The materials presented here are classical, but it can also be viewed as a particularly simple case of the theory developed by Arinkin and Gaitsgory in [3]; for example, see Section 1.3.5 therein.

4.1.1 A t -structure on $\text{Ch}_G^{\text{u,gr,ren}}$ By shearing, we obtain an equivalence of DG-categories

$$\text{Ch}_G^{\text{u,gr,ren}} \simeq \overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr}} \simeq \overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{\theta}] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr}},$$

where \tilde{x} and $\tilde{\theta}$ live in degrees $(2, 0)$ and $(2, -1)$, respectively. Here we follow the same convention as in Theorem 3.2.1.

The latter category has a natural t -structure which makes the forgetful functor to Vect^{gr} t -exact, where Vect^{gr} is equipped with the standard t -structure. The equivalence of categories above then endows $\text{Ch}_G^{\text{u,gr,ren}} \simeq \overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[W]\text{-Mod}^{\text{gr}}$ with a t -structure compatible with the *sheared* t -structure on Vect^{gr} , namely, the t -heart of Vect^{gr} in this t -structure consisting of complexes whose support lie in degrees (k, k) for $k \in \mathbb{Z}$.

Note that the algebra $\overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]$ is connective with respect to this sheared t -structure on Vect^{gr} , whereas the algebra $\overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{\theta}] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]$ is connective in the usual t -structure. In fact, by shearing, we have a t -exact equivalence of DG-categories

$$\text{Ch}_G^{\text{u,gr,ren}} \simeq \text{QCoh}((\mathbb{A}^n \times \mathbb{A}^n[-1]) / (\mathbb{G}_m \times S_n)),$$

where \mathbb{G}_m scales all coordinates with degree 2.

4.1.2 Coherent objects Following [23], we define

$$\overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr,coh}} \subseteq \overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr}} \simeq \text{Ch}_G^{\text{u,gr,ren}}$$

to be the full subcategory spanned by objects of bounded cohomological amplitude and coherent cohomologies (with respect to the t -structure defined in the previous subsection). This category is generated by $\overline{\mathbb{Q}}_\ell[x] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]$ under finite (co)limits, idempotent splittings, and grading shifts.

4.1.3 Relative Koszul duality Consider the functor of taking θ -invariants,

$$\text{inv}_\theta := \mathcal{H}\text{om}_{\overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(\overline{\mathbb{Q}}_\ell[x] \boxtimes \overline{\mathbb{Q}}_\ell[S_n], -) : \overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr,coh}} \rightarrow \text{Vect}^{\text{gr}}.$$

Since the category on the left is generated by $\overline{\mathbb{Q}}_\ell[x] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]$, an application of the Barr–Beck–Lurie theorem as in Theorem 3.3.4 above implies that we have an equivalence of categories

$$\begin{aligned} \overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr,coh}} &\xrightarrow[\simeq]{\text{inv}_\theta^{\text{enh}}} \mathcal{E}\text{nd}_{\overline{\mathbb{Q}}_\ell[x, \theta] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(\overline{\mathbb{Q}}_\ell[x] \boxtimes \overline{\mathbb{Q}}_\ell[S_n])\text{-Mod}^{\text{gr,perf}} \\ &\simeq \overline{\mathbb{Q}}_\ell[x, \underline{y}] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr,perf}}, \end{aligned}$$

where the \underline{y} live in degrees $(-2, 0)$.

We refer to this equivalence as the (relative) Koszul duality.

4.1.4 To simplify the notation, we let V denote a graded vector space of dimension n living in graded degree 2 and cohomological degree 0, equipped with a basis and hence also a permutation representation of $W = S_n$. Then

$$\overline{\mathbb{Q}}_\ell[\underline{\theta}] \simeq \text{Sym } V_\theta[-1], \quad \overline{\mathbb{Q}}_\ell[\underline{x}] \simeq \text{Sym } V_x[-2], \quad \overline{\mathbb{Q}}_\ell[\underline{y}] \simeq \text{Sym } V_y^\vee,$$

where the subscripts x , θ , and y are there just to make it easy to keep track of the names of the variables. To further simplify the notation, we define

$$\mathcal{A}_n := \text{Sym}(V_x[-2] \oplus V_\theta[-1]) \otimes \overline{\mathbb{Q}}_\ell[S_n]$$

and

$$\mathcal{B}_n := \text{Sym}(V_x[-2] \oplus V_y^\vee) \otimes \overline{\mathbb{Q}}_\ell[S_n].$$

The equivalence of categories above then becomes

$$\text{inv}_\theta^{\text{enh}} : \mathcal{A}_n\text{-Mod}^{\text{gr,coh}} \xrightarrow{\simeq} \mathcal{B}_n\text{-Mod}^{\text{gr,perf}}.$$

4.1.5 Let

$$\text{triv}_\theta := \text{Sym } V_x[-2] \otimes \overline{\mathbb{Q}}_\ell[S_n] \in \mathcal{A}_n\text{-Mod}^{\text{gr,coh}},$$

where the θ act trivially, ie by 0. Directly from the construction, we have

$$\text{inv}_\theta^{\text{enh}}(\text{triv}_\theta) \simeq \mathcal{B}_n.$$

Moreover, the inverse functor to $\text{inv}_\theta^{\text{enh}}$ is given by taking \underline{y} -coinvariants:

$$\text{coinv}_y^{\text{enh}} := \overline{\mathbb{Q}}_\ell \otimes_{\text{Sym } V_y^\vee} -.$$

4.1.6 Twisting For our purposes, it is convenient to twist $\text{inv}_\theta^{\text{enh}}$ and $\text{coinv}_y^{\text{enh}}$ by the sign representation of S_n . We let

$$\widetilde{\text{inv}}_\theta^{\text{enh}} := \text{Sym}^n V^\vee[1] \otimes \text{inv}_\theta^{\text{enh}} \quad \text{and} \quad \widetilde{\text{coinv}}_y^{\text{enh}} := \text{Sym}^n V[-1] \otimes \text{coinv}_y^{\text{enh}}$$

be mutually inverse functors

$$(4.1.7) \quad \widetilde{\text{inv}}_\theta^{\text{enh}} : \mathcal{A}_n\text{-Mod}^{\text{gr,coh}} \rightleftarrows \mathcal{B}_n\text{-Mod}^{\text{gr,perf}} : \widetilde{\text{coinv}}_y^{\text{enh}}.$$

4.1.8 Restricting to $\mathcal{A}_n\text{-Mod}^{\text{gr,perf}}$ Let

$$\text{triv}_y := \text{Sym}(V_x[-2]) \otimes \overline{\mathbb{Q}}_\ell[S_n] \in \mathcal{B}_n\text{-Mod}^{\text{gr,perf}},$$

where \underline{y} act trivially, ie by 0. An easy computation using the self-duality of the Koszul complex shows that

$$\widetilde{\text{coinv}}_y^{\text{enh}}(\text{triv}_y) \simeq \mathcal{A}_n,$$

and hence,

$$\widetilde{\text{inv}}_\theta^{\text{enh}}(\mathcal{A}_n) \simeq \text{triv}_y \simeq \text{Sym}(V_x[-2]) \otimes \overline{\mathbb{Q}}_\ell[S_n].$$

In other words, $\widetilde{\text{inv}}_\theta^{\text{enh}}$ simply “kills” the variables $\underline{\theta}$ in \mathcal{A}_n .

This implies that $\widetilde{\text{inv}}_{\theta}^{\text{enh}}$ and $\widetilde{\text{coinv}}_y^{\text{enh}}$ restrict to a pair of (eponymous) mutually inverse functors

$$(4.1.9) \quad \widetilde{\text{inv}}_{\theta}^{\text{enh}} : \mathcal{A}_n\text{-Mod}^{\text{gr,perf}} \rightleftarrows \mathcal{B}_n\text{-Mod}_{\text{nilp}_y}^{\text{gr,perf}} : \widetilde{\text{coinv}}_y^{\text{enh}},$$

where the subscript nilp_y denotes the fact that the variables y act nilpotently. This is because on the one hand, the left side is generated by \mathcal{A}_n under finite (co)limits, idempotent splittings, and grading shifts and on the other hand, \mathcal{A}_n is sent to triv_y on which the y act by 0.

The equivalence (4.1.9) is in fact a particularly simple case of the theory of singular support developed in [3], where perfect complexes have 0-singular support.

4.1.10 Unipotent character sheaves The discussion above gives the following Koszul dual descriptions of $\text{Ch}_G^{\text{u,gr}}$ and $\text{Ch}_G^{\text{u,gr,ren}}$ when $G = \text{GL}_n$:

Proposition 4.1.11 *Let $G = \text{GL}_n$ be the general linear group of rank n over $\overline{\mathbb{F}}_q$. Then we have an equivalence of $\text{Vect}^{\text{gr},c}$ -module categories*

$$\text{Ch}_G^{\text{u,gr}} \simeq \mathcal{B}_n\text{-Mod}_{\text{nilp}_y}^{\text{gr,perf}}.$$

Taking Ind-completion, we get an equivalence of Vect^{gr} -module categories

$$\text{Ch}_G^{\text{u,gr,ren}} \simeq \mathcal{B}_n\text{-Mod}_{\text{nilp}_y}^{\text{gr}}.$$

Proof The second statement follows from the first by taking the Ind-completion on both sides. The first statement follows from Theorem 3.3.4 and (4.1.9). □

4.2 The Hilbert scheme of points on \mathbb{C}^2

We will now recall the main results of [35], which give an explicit presentation of the categories of quasicohherent sheaves on the Hilbert schemes of points on \mathbb{C}^2 . This will allow us to relate Hilbert schemes of points on \mathbb{C}^2 , $\text{Ch}_G^{\text{u,gr}}$, and $\text{Ch}_G^{\text{u,gr,ren}}$ (when $G = \text{GL}_n$), which will be discussed in Section 4.3.

As we have been working over $\overline{\mathbb{Q}}_{\ell}$ rather than \mathbb{C} , our Hilbert schemes are in fact $\text{Hilb}_n(\overline{\mathbb{Q}}_{\ell}^2)$, which live over $\overline{\mathbb{Q}}_{\ell}$. Although the two are isomorphic as abstract varieties due to the isomorphism $\overline{\mathbb{Q}}_{\ell} \simeq \mathbb{C}$, we will keep using $\overline{\mathbb{Q}}_{\ell}$ for the sake of consistency. In fact, all varieties appearing in this subsection are over $\overline{\mathbb{Q}}_{\ell}$, and hence we will employ base-field-agnostic notation as much as possible; for example, \mathbb{A}^2 will be used to denote the 2-dimensional affine space over $\overline{\mathbb{Q}}_{\ell}$.

4.2.1 The isospectral Hilbert scheme Let Hilb_n denote the Hilbert scheme of points on \mathbb{A}^2 , $\mathbb{A}^{2n}/S_n = (\mathbb{A}^2)^n/S_n$ the stack quotient of \mathbb{A}^{2n} by the permutation action of S_n , and $\mathbb{A}^{2n} // S_n$ the GIT quotient.

There is a natural map $g : \mathbb{A}^{2n}/S_n \rightarrow \mathbb{A}^{2n} // S_n$ and the Hilbert–Chow morphism $H : \text{Hilb}_n \rightarrow \mathbb{A}^{2n} // S_n$. The isospectral Hilbert scheme is the *reduced* pullback

$$\begin{array}{ccc} \text{Iso}_n & \longrightarrow & \mathbb{A}^{2n} \\ \downarrow & & \downarrow \\ \text{Hilb}_n & \longrightarrow & \mathbb{A}^{2n} // S_n \end{array}$$

Iso_n has a natural S_n -action compatible with the permutation S_n -action on \mathbb{A}^{2n} and the trivial S_n -action on Hilb_n . Thus, we obtain the following commutative diagram:

(4.2.2)
$$\begin{array}{ccc} \text{Iso}_n/S_n & \xrightarrow{p} & \mathbb{A}^{2n}/S_n \\ q \downarrow & & \downarrow g \\ \text{Hilb}_n & \xrightarrow{H} & \mathbb{A}^{2n} // S_n. \end{array}$$

4.2.3 A derived equivalence In the notation above, one of the main results of [35] (which is itself a variant of [17] but has a more convenient form for us) takes the following form:

Theorem 4.2.4 [35, Proposition 2.8] *The functor $\Psi := q_* p^* : \text{QCoh}(\mathbb{A}^{2n}/S_n) \rightarrow \text{QCoh}(\text{Hilb}_n)$ is an equivalence of categories, where all functors are derived and where $\text{QCoh}(-)$ denotes the (∞) -derived category of quasicoherent sheaves.*

Passing to the full subcategories spanned by compact objects, we obtain an equivalence

$$\Psi : \text{Perf}(\mathbb{A}^{2n}/S_n) \xrightarrow{\cong} \text{Perf}(\text{Hilb}_n).$$

4.2.5 \mathbb{G}_m -equivariant structures Each of the terms in (4.2.2) has a natural \mathbb{G}_m^2 -action induced by the action of \mathbb{G}_m^2 on \mathbb{A}^2 where the first (resp. second) factor of \mathbb{G}_m^2 scales the first (resp. second) factor of \mathbb{A}^2 by weight 1 (resp. 2). Moreover, all maps are compatible with this action. We thus obtain an equivariant form of the theorem above. By abuse of notation, we will employ the same notation for the equivariant case as the nonequivariant case above. Note also that the category $\text{QCoh}(\text{Hilb}_n / (\mathbb{G}_m^2))$ below also appears as $\text{QCoh}_{\mathbb{G}_m^2}(\text{Hilb}_n)$ (or some variant thereof) in the literature.

Corollary 4.2.6 *The functor $\Psi := q_* p^* : \text{QCoh}(\mathbb{A}^{2n}/(S_n \times \mathbb{G}_m^2)) \rightarrow \text{QCoh}(\text{Hilb}_n / (\mathbb{G}_m^2))$ is an equivalence of categories. The same statement applies when restricted to the full subcategories of compact objects (ie perfect complexes).*

This equivalence allows us to have an explicit presentation of $\text{QCoh}(\text{Hilb}_n / (\mathbb{G}_m^2))$ as a plain (as opposed to a symmetric monoidal) DG-category.

Proposition 4.2.7 *We have an equivalence of categories*

$$\text{QCoh}(\text{Hilb}_n / (\mathbb{G}_m^2)) \xrightarrow{\Psi^{-1}} \text{QCoh}(\mathbb{A}^{2n}/(S_n \times \mathbb{G}_m^2)) \simeq \overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}] \boxtimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{\text{gr}, \text{gr}},$$

where the superscript gr, gr indicates the fact that we are working with bigraded complexes. Moreover, \tilde{x} and \tilde{y} have cohomological degree 0 and bidegrees (with respect to the superscript gr, gr) $(1, 0)$ and $(0, 2)$, respectively.

Proof Only the last equivalence needs to be proved. Applying Barr–Beck–Lurie to the adjunction

$$\mathrm{QCoh}(\mathbb{A}^{2n}/(\mathbb{S}_n \times \mathbb{G}_m^2)) \xrightleftharpoons[p_*]{p^*} \mathrm{QCoh}(B(\mathbb{S}_n \times \mathbb{G}_m^2)) \simeq \overline{\mathbb{Q}}_\ell[\mathbb{S}_n]\text{-Mod}^{gr,gr},$$

where $p: \mathbb{A}^{2n}/(\mathbb{S}_n \times \mathbb{G}_m^2) \rightarrow B(\mathbb{S}_n \times \mathbb{G}_m^2)$ is the natural map, we obtain an equivalence of categories

$$\mathrm{QCoh}(\mathbb{A}^{2n}/(\mathbb{S}_n \times \mathbb{G}_m^2)) \simeq \overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}]\text{-Mod}(\overline{\mathbb{Q}}_\ell[\mathbb{S}_n]\text{-Mod}^{gr,gr}).$$

But the latter is equivalent to

$$\overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}] \otimes \overline{\mathbb{Q}}_\ell[\mathbb{S}_n]\text{-Mod}^{gr,gr}. \quad \square$$

Remark 4.2.8 Instead of $(1, 0)$ and $(0, 2)$, we could have chosen any other bigradings for the \tilde{x} and \tilde{y} variables. As we will see, these specific choices are made because of the connection to character sheaves and HOMFLY-PT homology. For example, see Section 4.4.6 for the relation to the grading on the category of graded unipotent character sheaves.

4.2.9 Supports Let $i_{\tilde{x}}: \mathbb{A}_{\tilde{x}}^n/\mathbb{S}_n \rightarrow \mathbb{A}^{2n}/\mathbb{S}_n$ denote the closed subscheme defined by the vanishing of all the \tilde{y} -coordinates. Let $\mathrm{Hilb}_{n,\tilde{x}}$ be the pullback

$$\begin{array}{ccc} \mathrm{Hilb}_{n,\tilde{x}} & \longrightarrow & \mathrm{Hilb}_n \\ \downarrow & & \downarrow \\ \mathbb{A}_{\tilde{x}}^n/\mathbb{S}_n & \xrightarrow{i_{\tilde{x}}} & \mathbb{A}^{2n}/\mathbb{S}_n \end{array}$$

For any closed embedding of schemes (or in fact stacks) $Z \subset X$, we let $\mathrm{QCoh}(X)_Z$ denote the full subcategory of $\mathrm{QCoh}(X)$ consisting of objects whose supports lie in Z . Similarly, we let $\mathrm{Perf}(X)_Z$ denote the full subcategory of $\mathrm{Perf}(X)$ spanned by objects whose supports lie in Z .

Lemma 4.2.10 *The equivalence Ψ above restricts to the following equivalences of categories:*

$$\overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}] \otimes \overline{\mathbb{Q}}_\ell[\mathbb{S}_n]\text{-Mod}_{\mathrm{nilp}_{\tilde{y}}}^{gr,gr} \simeq \mathrm{QCoh}(\mathbb{A}^{2n}/(\mathbb{S}_n \times \mathbb{G}_m^2))_{\mathbb{A}_{\tilde{x}}^n/(\mathbb{S}_n \times \mathbb{G}_m^2)} \xrightarrow{\Psi} \mathrm{QCoh}(\mathrm{Hilb}_n/(\mathbb{G}_m^2))_{\mathrm{Hilb}_{n,\tilde{x}}/(\mathbb{G}_m^2)}.$$

The same statement applies when restricted to the full subcategories of compact objects (ie perfect complexes).

Proof For brevity’s sake, we suppress the \mathbb{G}_m -equivariant structures (ie the gradings) from the notation.

From (4.2.2), we see that Ψ is compatible with the action of $\mathrm{QCoh}(\mathbb{A}^{2n}/\mathbb{S}_n)$. Now, we note that the full subcategories of interest are cut out precisely by the condition that the variables \tilde{y} (from $\mathbb{A}^{2n}/\mathbb{S}_n$) act nilpotently.

Alternatively (and more categorically), one can use the theory of support as discussed in, for example, [3, Section 3.5], to conclude:

$$\begin{aligned} \mathrm{QCoh}(\mathbb{A}^{2n}/S_n)_{\mathbb{A}_{\tilde{x}}^n/S_n} &\simeq \mathrm{QCoh}(\mathbb{A}^{2n}/S_n) \otimes_{\mathrm{QCoh}(\mathbb{A}^{2n}/S_n)} \mathrm{QCoh}(\mathbb{A}^{2n}/S_n)_{\mathbb{A}_{\tilde{x}}^n/S_n} \\ &\simeq \mathrm{QCoh}(\mathrm{Hilb}_n) \otimes_{\mathrm{QCoh}(\mathbb{A}^{2n}/S_n)} \mathrm{QCoh}(\mathbb{A}^{2n}/S_n)_{\mathbb{A}_{\tilde{x}}^n/S_n} \simeq \mathrm{QCoh}(\mathrm{Hilb}_n)_{\mathrm{Hilb}_n, \tilde{x}}. \quad \square \end{aligned}$$

4.3 Cohomological shearing and 2-periodization

The algebra $\overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}] \otimes \overline{\mathbb{Q}}_\ell[S_n]$ appearing in Proposition 4.2.7 and the algebra $\mathcal{B}_n = \overline{\mathbb{Q}}_\ell[x, y] \otimes \overline{\mathbb{Q}}_\ell[S_n]$ appearing in Section 4.1.4 are almost the same, except for the mismatch in the cohomological gradings and the number of formal gradings.

In this subsection, we will discuss two general categorical constructions which will allow us, in Section 4.4, to relate the categories of modules over these two rings, and hence also to relate the category of (quasi)coherent sheaves on Hilb_n and $\mathrm{Ch}_G^{u, \mathrm{gr}, \mathrm{ren}}$ for $G = \mathrm{GL}_n$. One of them, known as cohomological shearing, reduces the number of formal gradings whereas the other, 2-periodization, increases the number of gradings. As it turns out, these two are equivalent, which is the content of Proposition 4.3.16 below.

The materials presented here are more or less standard, at least among the experts. We include it here since we cannot find a place where everything is written down in a way that is convenient for us.

The discussion in this subsection applies to both small and large categories. For brevity’s sake, we will only discuss the large category case. As usual, the small case can be obtained by passing to compact objects.

4.3.1 Shearing functors on $\mathrm{Vect}^{\mathrm{gr}}$

Consider the shear functors

$$\mathrm{sh}^{\leftarrow} : \mathrm{Vect}^{\mathrm{gr}} \rightarrow \mathrm{Vect}^{\mathrm{gr}}, \quad (V_i)_i \mapsto (V_i[2i])_i \quad \text{and} \quad \mathrm{sh}^{\rightarrow} : \mathrm{Vect}^{\mathrm{gr}} \rightarrow \mathrm{Vect}^{\mathrm{gr}}, \quad (V_i)_i \mapsto (V_i[-2i])_i.$$

Note that these are equivalences of symmetric monoidal categories.

For $A^{\mathrm{gr}} \in \mathrm{Vect}^{\mathrm{gr}}$, we let $A^{\mathrm{gr}, \leftarrow} := \mathrm{sh}^{\leftarrow}(A^{\mathrm{gr}}) \in \mathrm{Vect}^{\mathrm{gr}}$ and $A^{\mathrm{gr}, \rightarrow} := \mathrm{sh}^{\rightarrow}(A^{\mathrm{gr}}) \in \mathrm{Vect}^{\mathrm{gr}}$. Since sh^{\leftarrow} and $\mathrm{sh}^{\rightarrow}$ are symmetric monoidal equivalences, $A^{\mathrm{gr}, \leftarrow}$ (resp. $A^{\mathrm{gr}, \rightarrow}$) is equipped with a natural algebra structure if and only if A^{gr} is.

4.3.2 Shearing a $\mathrm{Vect}^{\mathrm{gr}}$ -module structure

Since $\mathrm{Vect}^{\mathrm{gr}}$ is symmetric monoidal, any $\mathcal{C}^{\mathrm{gr}} \in \mathrm{Vect}^{\mathrm{gr}}\text{-Mod}$ can be upgraded naturally to a $\mathrm{Vect}^{\mathrm{gr}}$ -bimodule category $\mathcal{C} \in \mathrm{Vect}^{\mathrm{gr}}\text{-BiMod}$. We visualize them as $\mathrm{Vect}^{\mathrm{gr}}$ acting on the left and on the right even though they all come from the left module structure.

Remark 4.3.3 For us, the right $\mathrm{Vect}^{\mathrm{gr}}$ -module structure is important only for the purpose of taking relative tensor products. We will generally ignore this structure unless the formation of relative tensor products is involved.

For each of the left/right module structures, we can precompose the action of Vect^{gr} with sh^{\leftarrow} to obtain a new module structure. For example, $\Rightarrow \mathcal{C}^{\text{gr}, \leftarrow}$ has the following Vect^{gr} -bimodule structure:¹⁵

$$\overline{\mathbb{Q}}_{\ell}(i) \boxtimes c \boxtimes \overline{\mathbb{Q}}_{\ell}(j) \mapsto c(i + j)[2i - 2j].$$

Here, the positions of the arrows with respect to the category indicate which of the left or right module structure we are shearing, and the directions of the arrows indicate which shear we use. A missing arrow indicates that shearing does not occur; for example, $\mathcal{C}^{\text{gr}, \leftarrow}$ has a sheared Vect^{gr} -action on the right while keeping the action on the left unchanged. Note that these modifications only change the action of Vect^{gr} while keeping the underlying DG-category intact.

4.3.4 The following diagram demonstrates how the functors sh^{\leftarrow} and sh^{\Rightarrow} interact with Vect^{gr} -bimodule structures:

$$\Rightarrow \text{Vect}^{\text{gr}, \Rightarrow} \begin{matrix} \xleftarrow{\text{sh}^{\leftarrow}} \\ \xrightarrow{\text{sh}^{\Rightarrow}} \end{matrix} \text{Vect}^{\text{gr}} \begin{matrix} \xleftarrow{\text{sh}^{\leftarrow}} \\ \xrightarrow{\text{sh}^{\Rightarrow}} \end{matrix} \leftarrow \text{Vect}^{\text{gr}, \leftarrow}.$$

Here all functors are equivalences as morphisms in Vect^{gr} -BiMod. Similarly, we can mix and match the directions of the arrows. For example,

$$\text{sh}^{\leftarrow} : \text{Vect}^{\text{gr}, \Rightarrow} \xrightarrow{\cong} \leftarrow \text{Vect}^{\text{gr}}.$$

Shearing Vect^{gr} -bimodule structures can be written naturally in terms of relative tensors. For example,

$$\mathcal{C}^{\text{gr}, \leftarrow} \simeq \mathcal{C}^{\text{gr}} \otimes_{\text{Vect}^{\text{gr}}} \text{Vect}^{\text{gr}, \leftarrow} \quad \text{and} \quad \Rightarrow \mathcal{C}^{\text{gr}} \simeq \Rightarrow \text{Vect}^{\text{gr}} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}}.$$

4.3.5 Shearing A^{gr} -Mod^{gr} Let A^{gr} be a graded DG-algebra, ie $A^{\text{gr}} \in \text{Alg}(\text{Vect}^{\text{gr}})$. Then the category of graded A^{gr} -modules, $A^{\text{gr}}\text{-Mod}^{\text{gr}} := A^{\text{gr}}\text{-Mod}(\text{Vect}^{\text{gr}}) \in \text{Vect}^{\text{gr}}\text{-Mod}$, is naturally a Vect^{gr} -module category, and hence also an object in Vect^{gr} -BiMod, as discussed above. A similar discussion as in the case of Vect^{gr} gives the following equivalences of objects in $\text{Vect}^{\text{gr}}\text{-Mod}$:

$$\Rightarrow (A^{\text{gr}, \Rightarrow}\text{-Mod}^{\text{gr}, \Rightarrow}) \begin{matrix} \xleftarrow{\text{sh}^{\leftarrow}} \\ \xrightarrow{\text{sh}^{\Rightarrow}} \end{matrix} A^{\text{gr}}\text{-Mod}^{\text{gr}} \begin{matrix} \xleftarrow{\text{sh}^{\leftarrow}} \\ \xrightarrow{\text{sh}^{\Rightarrow}} \end{matrix} \leftarrow (A^{\text{gr}, \leftarrow}\text{-Mod}^{\text{gr}, \leftarrow}).$$

Moreover, as above, we can mix and match the directions of the arrows.

Replacing A^{gr} by $A^{\text{gr}, \leftarrow}$, we obtain the following equivalences of objects in Vect^{gr} -BiMod:

$$A^{\text{gr}}\text{-Mod}^{\text{gr}, \Rightarrow} \begin{matrix} \xleftarrow{\text{sh}^{\leftarrow}} \\ \xrightarrow{\text{sh}^{\Rightarrow}} \end{matrix} \leftarrow (A^{\text{gr}, \leftarrow}\text{-Mod}^{\text{gr}}) \quad \text{and} \quad A^{\text{gr}}\text{-Mod}^{\text{gr}, \leftarrow} \begin{matrix} \xleftarrow{\text{sh}^{\leftarrow}} \\ \xrightarrow{\text{sh}^{\Rightarrow}} \end{matrix} \Rightarrow (A^{\text{gr}, \Rightarrow}\text{-Mod}^{\text{gr}}).$$

4.3.6 Sheared degrading Let $\mathcal{C}^{\text{gr}} \in \text{Vect}^{\text{gr}}\text{-Mod}$. Then we can form the degraded version \mathcal{C} of \mathcal{C}^{gr} ,

$$\mathcal{C} := \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}},$$

¹⁵Note that the formula below is compatible with the definitions of sh^{\Rightarrow} and sh^{\leftarrow} since, for example, $\overline{\mathbb{Q}}_{\ell}(i)$ lives in graded degree $-i$.

where the Vect^{gr} -module structure on Vect is given by the symmetric monoidal functor

$$\text{oblv}_{\text{gr}}: \text{Vect}^{\text{gr}} \rightarrow \text{Vect}, \quad (V_i)_i \mapsto \bigoplus_i V_i,$$

which forgets the grading on Vect^{gr} .

We note that the functor of forgetting the grading is also defined for \mathcal{C}^{gr} :

$$\text{oblv}_{\text{gr}}: \mathcal{C}^{\text{gr}} \simeq \text{Vect}^{\text{gr}} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}} \xrightarrow{\text{oblv}_{\text{gr}} \otimes \text{id}_{\mathcal{C}^{\text{gr}}}} \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}} \simeq \mathcal{C}.$$

4.3.7 The sheared degradings of \mathcal{C}^{gr} , denoted by \mathcal{C}^{\Leftarrow} and $\mathcal{C}^{\Rightarrow}$, are defined to be the usual degradings of $\Rightarrow \mathcal{C}^{\text{gr}}$ and $\Leftarrow \mathcal{C}^{\text{gr}}$, respectively (note the reversal of the arrows!):

$$\mathcal{C}^{\Rightarrow} := \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \Leftarrow \mathcal{C}^{\text{gr}} \quad \text{and} \quad \mathcal{C}^{\Leftarrow} := \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \Rightarrow \mathcal{C}^{\text{gr}}.$$

Remark 4.3.8 In general, \mathcal{C} , \mathcal{C}^{\Leftarrow} , and $\mathcal{C}^{\Rightarrow}$ are different as DG-categories. However, the case where $\mathcal{C}^{\text{gr}} = \text{Vect}^{\text{gr}}$ is special as we always have an equivalence of DG-categories

$$(4.3.9) \quad \begin{aligned} \text{Vect} &\xrightarrow[\simeq]{V \mapsto V \boxtimes \overline{\mathbb{Q}}_\ell} \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \Leftarrow \text{Vect}^{\text{gr}} =: \text{Vect}^{\Rightarrow}, \\ \text{Vect} &\xrightarrow[\simeq]{V \mapsto V \boxtimes \overline{\mathbb{Q}}_\ell} \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \Rightarrow \text{Vect}^{\text{gr}} =: \text{Vect}^{\Leftarrow}. \end{aligned}$$

Since Vect has a natural Vect^{gr} -module category structure, we can shear this action and obtain, for example, Vect^{\Leftarrow} and $\text{Vect}^{\Rightarrow}$. These two are in fact equivalent to Vect^{\Leftarrow} and $\text{Vect}^{\Rightarrow}$, respectively, which explains the reversal of the arrows in the definition of the sheared degrading procedure.

Indeed, the right action of Vect^{gr} on $\text{Vect}^{\Rightarrow}$ is as follows: $\overline{\mathbb{Q}}_\ell \langle i \rangle [j]$ sends $V \boxtimes \overline{\mathbb{Q}}_\ell$ to

$$V \boxtimes (\overline{\mathbb{Q}}_\ell \langle i \rangle [j]) \simeq V \boxtimes (\overline{\mathbb{Q}}_\ell \langle i \rangle [-2i] [j + 2i]) \simeq (V[j + 2i]) \boxtimes \overline{\mathbb{Q}}_\ell,$$

which corresponds to $V[j + 2i] \in \text{Vect}$ under the equivalence of categories (4.3.9). But this is precisely the \Rightarrow -sheared action of Vect^{gr} on Vect on the right.

4.3.10 We can write sheared degradings in a natural way as relative tensors of categories

$$\mathcal{C}^{\Leftarrow} \simeq \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \Rightarrow \mathcal{C}^{\text{gr}} \simeq \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \Rightarrow \text{Vect}^{\text{gr}} \otimes_{\text{Vect}^{\text{gr}}} \Leftarrow \mathcal{C}^{\text{gr}} \simeq \text{Vect}^{\Leftarrow} \otimes_{\text{Vect}^{\text{gr}}} \Leftarrow \mathcal{C}^{\text{gr}} \simeq \text{Vect}^{\Leftarrow} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}}.$$

Similarly,

$$\mathcal{C}^{\Rightarrow} \simeq \text{Vect}^{\Rightarrow} \otimes_{\text{Vect}^{\text{gr}}} \Leftarrow \mathcal{C}^{\text{gr}} \simeq \text{Vect}^{\Rightarrow} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}}.$$

4.3.11 Sheared degrading $A^{\text{gr}}\text{-Mod}$ Let $A^{\text{gr}} \in \text{Alg}(\text{Vect}^{\text{gr}})$ be as above and consider the category of (ungraded) A^{gr} -modules, denoted by $\text{oblv}_{\text{gr}}(A^{\text{gr}})\text{-Mod}$. By abuse of notation, when confusion is unlikely, we also write $A^{\text{gr}}\text{-Mod} := \text{oblv}_{\text{gr}}(A^{\text{gr}})\text{-Mod}$.

Clearly, $A^{\text{gr}}\text{-Mod} \simeq \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} A^{\text{gr}}\text{-Mod}^{\text{gr}}$ is a degrading of $A^{\text{gr}}\text{-Mod}^{\text{gr}}$. But we also have sheared degradings as defined in Section 4.3.6. Namely,

$$A^{\text{gr}}\text{-Mod}^{\Leftarrow} := \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} \Rightarrow (A^{\text{gr}}\text{-Mod}^{\text{gr}}) \xrightarrow[\simeq]{\text{id}_{\text{Vect}} \otimes \text{sh}^{\Leftarrow}} \text{Vect} \otimes_{\text{Vect}^{\text{gr}}} A^{\text{gr}, \Leftarrow}\text{-Mod}^{\text{gr}, \Leftarrow} \simeq \text{oblv}_{\text{gr}}(A^{\text{gr}, \Leftarrow})\text{-Mod}^{\Leftarrow},$$

and similarly for $A^{\text{gr}}\text{-Mod}^{\Rightarrow}$.

Ignoring the right Vect^{gr} -actions (see also Remark 4.3.3), we have the following equivalences of DG-categories:

$$(4.3.12) \quad A^{\text{gr}}\text{-Mod}^{\Leftarrow} \simeq \text{oblv}_{\text{gr}}(A^{\text{gr},\Leftarrow})\text{-Mod} \quad \text{and} \quad A^{\text{gr}}\text{-Mod}^{\Rightarrow} \simeq \text{oblv}_{\text{gr}}(A^{\text{gr},\Rightarrow})\text{-Mod}.$$

4.3.13 2-periodization We start with the simplest example of 2-periodization and its interaction with sheared degrading.

Let $\overline{\mathbb{Q}}_{\ell}[u, u^{-1}] \in \text{ComAlg}(\text{Vect}^{\text{gr}})$ be a commutative algebra object where u is in graded degree 1 and cohomological degree 0. It is easy to see that the functor of extracting the graded degree 0 part

$$\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]\text{-Mod}^{\text{gr}} \xrightarrow[\simeq]{(M_i)_{i \rightarrow M_0}} \text{Vect}$$

is an equivalence of Vect^{gr} -module categories where the actions of Vect^{gr} are the obvious ones (ie no shearing). The inverse functor is given by $V \mapsto \overline{\mathbb{Q}}_{\ell}[u, u^{-1}] \otimes V$.

The discussion above thus gives us the equivalences

$$\text{Vect}^{\Leftarrow} \simeq \text{Vect}^{\Leftarrow} \xrightarrow[\simeq]{\overline{\mathbb{Q}}_{\ell}[u, u^{-1}] \otimes -} \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]\text{-Mod}^{\text{gr}, \Leftarrow} \xrightarrow[\simeq]{\text{sh}^{\Rightarrow}} \Rightarrow (\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow}\text{-Mod}^{\text{gr}})$$

$\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow} \otimes -$

and

$$\text{Vect}^{\Rightarrow} \simeq \text{Vect}^{\Rightarrow} \xrightarrow[\simeq]{\overline{\mathbb{Q}}_{\ell}[u, u^{-1}] \otimes -} \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]\text{-Mod}^{\text{gr}, \Rightarrow} \xrightarrow[\simeq]{\text{sh}^{\Leftarrow}} \Leftarrow (\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}\text{-Mod}^{\text{gr}}).$$

$\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow} \otimes -$

Here, by construction, $\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow}$ (resp. $\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}$) is the algebra of Laurent polynomials generated by one element living in graded degree 1 and cohomological degree 2 (resp. -2).

In particular, we have

$$(4.3.14) \quad \Leftarrow \text{Vect}^{\Leftarrow} \simeq \Leftarrow \text{Vect}^{\Leftarrow} \simeq \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow}\text{-Mod}^{\text{gr}}, \quad \Rightarrow \text{Vect}^{\Rightarrow} \simeq \Rightarrow \text{Vect}^{\Rightarrow} \simeq \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}\text{-Mod}^{\text{gr}}.$$

4.3.15 We will now turn to the general case.

Let $\mathcal{C}^{\text{gr}} \in \text{Vect}^{\text{gr}}\text{-Mod}$ be as above. Then the positive and negative 2-periodizations of \mathcal{C}^{gr} are defined to be

$$\mathcal{C}^{\text{gr}, 2\text{-per}^+} := \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow}\text{-Mod}^{\text{gr}} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}} \simeq \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow}\text{-Mod}(\mathcal{C}^{\text{gr}})$$

and

$$\mathcal{C}^{\text{gr}, 2\text{-per}^-} := \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}\text{-Mod}^{\text{gr}} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}} \simeq \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}\text{-Mod}(\mathcal{C}^{\text{gr}}),$$

respectively.

Note that $\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}$ (resp. $\overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow}$) is equivalent to $\overline{\mathbb{Q}}_{\ell}[\beta, \beta^{-1}]$ where β lives in graded degree 1 and cohomological degree -2 (resp. 2). The category of module objects over $\overline{\mathbb{Q}}_{\ell}[\beta, \beta^{-1}]$ is thus the category of (graded) 2-periodic objects appearing in the literature.

The grading allows one to absorb 2-periodicity. In fact, the 2-periodization construction agrees with the sheared degrading construction from above.

Proposition 4.3.16 *Let $\mathcal{C}^{\text{gr}} \in \text{Vect}^{\text{gr}}\text{-Mod}$. Then we have natural equivalences of Vect^{gr} -module categories*

$$\mathcal{C}^{\text{gr},2\text{-per}_-} \simeq \Rightarrow (\mathcal{C}^{\Rightarrow}) \quad \text{and} \quad \mathcal{C}^{\text{gr},2\text{-per}_+} \simeq \Leftarrow (\mathcal{C}^{\Leftarrow}).$$

Proof We treat the case of 2-per_- , as the other case is similar. We have

$$\mathcal{C}^{\text{gr},2\text{-per}_-} := \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}\text{-Mod}^{\text{gr}} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}} \simeq \Rightarrow \text{Vect}^{\Rightarrow} \otimes_{\text{Vect}^{\text{gr}}} \mathcal{C}^{\text{gr}} \simeq \Rightarrow \mathcal{C}^{\Rightarrow},$$

where the second and third equivalences are due to (4.3.14) and Section 4.3.10, respectively. □

Remark 4.3.17 We have the following commutative diagrams of symmetric monoidal categories:

$$\begin{array}{ccc} \text{Vect}^{\text{gr}} & \xrightarrow{\text{oblv}_{\text{gr}}^{\text{osh}} \Rightarrow} & \text{Vect} \\ \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow} \otimes - \downarrow & \nearrow M \mapsto M_0 & \uparrow \\ \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Leftarrow}\text{-Mod}^{\text{gr}} & \simeq & \text{Vect}^{\text{gr},2\text{-per}_-} \end{array} \qquad \begin{array}{ccc} \text{Vect}^{\text{gr}} & \xrightarrow{\text{oblv}_{\text{gr}}^{\text{osh}} \Leftarrow} & \text{Vect} \\ \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow} \otimes - \downarrow & \nearrow M \mapsto M_0 & \uparrow \\ \overline{\mathbb{Q}}_{\ell}[u, u^{-1}]^{\Rightarrow}\text{-Mod}^{\text{gr}} & \simeq & \text{Vect}^{\text{gr},2\text{-per}_+} \end{array}$$

Thus, by construction, the action of Vect^{gr} on $\Rightarrow (\mathcal{C}^{\Rightarrow})$ (resp. $\Leftarrow (\mathcal{C}^{\Leftarrow})$) factors through $\text{Vect}^{\text{gr},2\text{-per}_-}$ (resp. $\text{Vect}^{\text{gr},2\text{-per}_+}$). On the other hand, $\mathcal{C}^{\text{gr},2\text{-per}_-}$ (resp. $\mathcal{C}^{\text{gr},2\text{-per}_+}$) is equipped with a natural action of $\text{Vect}^{\text{gr},2\text{-per}_-}$ (resp. $\text{Vect}^{\text{gr},2\text{-per}_+}$).

Chasing through the definitions, it is easy to see that the two actions are compatible. In other words, the two equivalences in Proposition 4.3.16 are equivalences of $\text{Vect}^{\text{gr},2\text{-per}_-}$ -(resp. $\text{Vect}^{\text{gr},2\text{-per}_+}$)-module categories.

4.3.18 More formal gradings In the above, the “background category” is Vect in the sense that all categories are DG-categories, ie they are Vect -module categories. Then we work with categories with one extra grading, ie with Vect^{gr} -module categories, and consider various categorical operations using this structure. We could have started with Vect^{gr} -module categories but still added another grading and worked with $\text{Vect}^{\text{gr},\text{gr}}$ -module categories. Everything discussed above still goes through, except that now everything has one more grading. For example, degrading goes from $\text{Vect}^{\text{gr},\text{gr}}$ -module categories to Vect^{gr} -categories.¹⁶ This is the setting that we will work with below.

4.4 2-periodic Hilbert schemes and character sheaves

We will now relate Hilbert schemes of points and character sheaves, using the procedure of 2-periodization discussed above to make the precise statement.

4.4.1 Introduce an extra grading By default, all of our categories are singly graded in the sense that they are Vect^{gr} -(or, if we work with the small variant, $\text{Vect}^{\text{gr},c}$)-module categories. Note that any DG-category can be viewed as a Vect^{gr} -module category via the symmetric monoidal functor $\text{oblv}_{\text{gr}}: \text{Vect}^{\text{gr}} \rightarrow \text{Vect}$. In particular, any Vect^{gr} -module category is a $\text{Vect}^{\text{gr},\text{gr}}$ -module category, where the second Vect^{gr} acts by

¹⁶In fact, we could, more generally, work with $(\text{Vect}^{\text{gr}})^{\otimes n}$ -module categories and then add one more grading to get $n + 1$ gradings. We will not need this generality here.

forgetting the formal grading. By convention, we use X to denote the default grading and Y the extra grading we just introduced. We will also use the notation $\text{Vect}^{\text{gr}_X, \text{gr}_Y}$ if we want to make it clear which grading convention we are using.

In this subsection, unless otherwise specified, all the shearing and 2-periodization constructions will be with respect to the Y -grading, even when we start with $\mathcal{C}^{\text{gr}} \in \text{Vect}^{\text{gr}_X}\text{-Mod}$, which has only *one* grading. In this case, we simply view \mathcal{C} as an object in $\text{Vect}^{\text{gr}_X, \text{gr}_Y}\text{-Mod}$, where $\text{Vect}^{\text{gr}_Y}$ acts via the symmetric monoidal functor $\text{oblv}_{\text{gr}_Y} : \text{Vect}^{\text{gr}_Y} \rightarrow \text{Vect}$. In particular, if we forget the $\text{Vect}^{\text{gr}_Y}$ -action, $\mathcal{C}^{\text{gr}}, \leftarrow \mathcal{C}^{\text{gr}},$ and $\Rightarrow \mathcal{C}^{\text{gr}}$ are equivalent as objects in $\text{Vect}^{\text{gr}_X}\text{-Mod}$ (and hence also as DG-categories). The difference in the $\text{Vect}^{\text{gr}_Y}$ -module structures can be seen by looking at $\text{Vect}^{\text{gr}_X, \text{gr}_Y}$ -enriched Hom, denoted by $\mathcal{H}\text{om}^{\text{gr}_X, \text{gr}_Y}$. More precisely, for $c_1, c_2 \in \mathcal{C}^{\text{gr}}$, on the one hand, we have

$$\mathcal{H}\text{om}_{\mathcal{C}^{\text{gr}}}^{\text{gr}_X, \text{gr}_Y}(c_1, c_2) \simeq \bigoplus_k \mathcal{H}\text{om}_{\mathcal{C}^{\text{gr}}}^{\text{gr}_X}(c_1, c_2)\langle k \rangle_Y \simeq \overline{\mathbb{Q}}_\ell[u, u^{-1}] \otimes \mathcal{H}\text{om}_{\mathcal{C}^{\text{gr}}}^{\text{gr}_X}(c_1, c_2) \in \text{Vect}^{\text{gr}_X, \text{gr}_Y}.$$

Here, in the tensor formula, the gr_X -enriched Homs are put in Y -degree 0. In other words, it is simply copies of the $\text{Vect}^{\text{gr}_X}$ -enriched Hom, put in all Y -degrees. On the other hand, a simple argument using adjunctions gives the following identification, which is essentially the same as Proposition 4.3.16 but with one extra grading (see also Remark 4.3.17):

Lemma 4.4.2 *In the situation above, we have natural equivalences*

$$\mathcal{H}\text{om}_{\leftarrow \mathcal{C}^{\text{gr}}}^{\text{gr}_X, \text{gr}_Y}(c_1, c_2) \simeq \text{sh}^{\Rightarrow}(\mathcal{H}\text{om}_{\mathcal{C}^{\text{gr}}}^{\text{gr}_X, \text{gr}_Y}(c_1, c_2)) \simeq \overline{\mathbb{Q}}_\ell[u, u^{-1}]^{\Rightarrow} \otimes \mathcal{H}\text{om}_{\mathcal{C}^{\text{gr}}}^{\text{gr}_X}(c_1, c_2) \in \text{Vect}^{\text{gr}_X, \text{gr}_Y, 2\text{-per}+}$$

and

$$\mathcal{H}\text{om}_{\Rightarrow \mathcal{C}^{\text{gr}}}^{\text{gr}_X, \text{gr}_Y}(c_1, c_2) \simeq \text{sh}^{\Leftarrow}(\mathcal{H}\text{om}_{\mathcal{C}^{\text{gr}}}^{\text{gr}_X, \text{gr}_Y}(c_1, c_2)) \simeq \overline{\mathbb{Q}}_\ell[u, u^{-1}]^{\Leftarrow} \otimes \mathcal{H}\text{om}_{\mathcal{C}^{\text{gr}}}^{\text{gr}_X}(c_1, c_2) \in \text{Vect}^{\text{gr}_X, \text{gr}_Y, 2\text{-per}-},$$

where sh^{\Rightarrow} and sh^{\Leftarrow} are with respect to the Y -grading. Moreover, in the tensor formula, the gr_X -enriched Homs are put in Y -degree 0.

4.4.3 The algebra \mathcal{B}_n and variants Let $\mathcal{B}_n^{\text{gr}} = \overline{\mathbb{Q}}_\ell[\underline{x}, \underline{y}] \otimes \overline{\mathbb{Q}}_\ell[\mathbf{S}_n] \in \text{Alg}(\text{Vect}^{\text{gr}_X, \text{gr}_Y})$ be a bigraded DG-algebra, where the variables \underline{x} (resp. \underline{y}) live in cohomological degree 2 (resp. 0) and graded degrees (2, 1) (resp. (−2, 0)). Here, the first (resp. second) coordinate represents the X - (resp. Y -)grading. Note that the Y -grading is the extra grading whereas the X -grading came from the above.

By construction, we have an equivalence of (singly graded) DG-algebras $\text{oblv}_{\text{gr}_Y}(\mathcal{B}_n^{\text{gr}}) \simeq \mathcal{B}_n \in \text{Alg}(\text{Vect}^{\text{gr}_X})$, where \mathcal{B}_n is the graded DG-algebra defined in Section 4.1.4. Let $\tilde{\mathcal{B}}_n^{\text{gr}} := \mathcal{B}_n^{\text{gr}, \Leftarrow}$ and $\hat{\mathcal{B}}_n := \text{oblv}_{\text{gr}_Y}(\tilde{\mathcal{B}}_n^{\text{gr}})$, where the shear is with respect to the Y -grading. Then

$$\tilde{\mathcal{B}}_n^{\text{gr}} \simeq \overline{\mathbb{Q}}_\ell[\underline{\tilde{x}}, \underline{\tilde{y}}] \otimes \overline{\mathbb{Q}}_\ell[\mathbf{S}_n] \in \text{Alg}(\text{Vect}^{\text{gr}_X, \text{gr}_Y}),$$

where the variables $\underline{\tilde{x}}$ (resp. $\underline{\tilde{y}}$) live in cohomological degree 0 and graded degrees (2, 1) (resp. (−2, 0)). Moreover,

$$\hat{\mathcal{B}}_n := \text{oblv}_{\text{gr}_Y}(\mathcal{B}_n^{\text{gr}}) \simeq \overline{\mathbb{Q}}_\ell[\underline{\tilde{x}}, \underline{\tilde{y}}] \otimes \overline{\mathbb{Q}}_\ell[\mathbf{S}_n] \in \text{Alg}(\text{Vect}^{\text{gr}_X}),$$

where the $\underline{\tilde{x}}$ (resp. $\underline{\tilde{y}}$) live in cohomological degree 0 and graded degree 2 (resp. −2).

Lemma 4.4.4 We have the following equivalence of $\text{Vect}^{\text{gr}X}$ -categories:

$$\mathcal{B}_n\text{-Mod}^{\text{gr}X} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}^{\text{gr}X, \Rightarrow}.$$

Proof We have

$$\mathcal{B}_n \simeq \text{oblv}_{\text{gr}Y}(\mathcal{B}_n^{\text{gr}}) \simeq \text{oblv}_{\text{gr}Y}(\tilde{\mathcal{B}}_n^{\text{gr}, \Rightarrow})$$

and hence, by (4.3.12), we have the equivalence of $\text{Vect}^{\text{gr}X}$ -module categories

$$\mathcal{B}_n\text{-Mod}^{\text{gr}X} \simeq \text{oblv}_{\text{gr}Y}(\tilde{\mathcal{B}}_n^{\text{gr}, \Rightarrow})\text{-Mod}^{\text{gr}X} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}^{\text{gr}X, \Rightarrow},$$

where, by convention, all shears are with respect to the Y -grading. □

Corollary 4.4.5 We have a natural equivalence of $\text{Vect}^{\text{gr}X, \text{gr}Y}$ -module categories

$$\Rightarrow(\mathcal{B}_n\text{-Mod}^{\text{gr}X}) \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}^{\text{gr}X, \text{gr}Y, 2\text{-per}_-},$$

where the $\text{Vect}^{\text{gr}Y}$ -module structure on the right-hand side is the obvious one. Moreover, the shear on the left-hand side is with respect to the $\text{Vect}^{\text{gr}Y}$ -module structure. Namely, $\text{Vect}^{\text{gr}Y}$ acts on the left via

$$\text{Vect}^{\text{gr}Y} \xrightarrow{\text{sh}^{\leftarrow}} \text{Vect}^{\text{gr}Y} \xrightarrow{\text{oblv}_{\text{gr}Y}} \text{Vect},$$

as in Proposition 4.3.16. Consequently, by Remark 4.3.17, the equivalence above is an equivalence of $\text{Vect}^{\text{gr}X, \text{gr}Y, 2\text{-per}_-}$ -module categories, where 2-per_- is with respect to the Y -grading.

Proof We have

$$\Rightarrow(\mathcal{B}_n\text{-Mod}^{\text{gr}X}) \simeq \Rightarrow(\tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}^{\text{gr}X, \Rightarrow}) \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}^{\text{gr}X, \text{gr}Y, 2\text{-per}_-},$$

where the first and second equivalences are due to Lemma 4.4.4 and Proposition 4.3.16, respectively. By convention, shearing and 2-periodization are with respect to the Y -grading.

The last statement is due to Remark 4.3.17. □

4.4.6 Change of gradings The gradings X and Y used above need to be changed to match with the one used in [28]. We denote the new gradings by \tilde{X} and \tilde{Y} , where

$$\tilde{X} = X^2Y \quad \text{and} \quad \tilde{Y} = X^{-1},$$

or equivalently,

$$X = \tilde{Y}^{-1} \quad \text{and} \quad Y = \tilde{X}\tilde{Y}^2.$$

Under this change of gradings, we have

$$\tilde{\mathcal{B}}_n^{\text{gr}} \simeq \overline{\mathbb{Q}}_\ell[\underline{\tilde{x}}, \underline{\tilde{y}}] \otimes \overline{\mathbb{Q}}_\ell[S_n] \in \text{Alg}(\text{Vect}^{\text{gr}\tilde{X}, \text{gr}\tilde{Y}}),$$

where the variables \tilde{x} (resp. \tilde{y}) live in cohomological degree 0 and graded degrees (1, 0) (resp. (0, 2)). We note that this algebra matches precisely with the algebra appearing in Proposition 4.2.7.

Remark 4.4.7 In the rest of this paper, unless otherwise specified, all the shears and 2-periodizations are still with respect to the Y -grading (in the (X, Y) -grading system), even when we are working with the (\tilde{X}, \tilde{Y}) -grading. As different grading conventions in the HOMFLY-PT homology theory cause a lot of confusion, at least, to the authors of the current paper, we write down explicit examples below.

4.4.8 2-per₋ in terms of (\tilde{X}, \tilde{Y}) -grading In terms of the (\tilde{X}, \tilde{Y}) -grading, negative periodization 2-per₋ in the Y -direction with respect to the (X, Y) -grading (such as the one appearing in Corollary 4.4.5) takes the following form. For $\mathcal{C} \in \text{Vect}^{\text{gr}_X, \text{gr}_Y}\text{-Mod} \simeq \text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}}\text{-Mod}$,

$$\mathcal{C}^{2\text{-per}_-} \simeq \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}(\mathcal{C}),$$

where β lives in cohomological degree -2 and graded degrees $(0, 1)$ in terms of the (X, Y) -grading or equivalently, $(1, 2)$ in terms of the (\tilde{X}, \tilde{Y}) -grading. Since this is the only 2-periodization that we will use in connection to the (\tilde{X}, \tilde{Y}) -grading, we will adopt the notation

$$\mathcal{C}^{2\text{-per}} := \mathcal{C}^{2\text{-per}_-} \simeq \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}(\mathcal{C})$$

in the rest of the paper, where, as above, β lives in cohomological degree -2 and graded degrees $(0, 1)$ in terms of the (X, Y) -grading or, equivalently, $(1, 2)$ in terms of the (\tilde{X}, \tilde{Y}) -grading.

Below are a couple of examples that are important to us.

Definition 4.4.9 (2-periodizing an object) For $\mathcal{C} \in \text{Vect}^{\text{gr}_X, \text{gr}_Y}\text{-Mod} \simeq \text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}}\text{-Mod}$ and $c \in \mathcal{C}$, the corresponding 2-periodized object $c^{2\text{-per}} \in \mathcal{C}^{2\text{-per}}$ is defined to be $\overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}] \otimes c$.

Definition 4.4.10 (2-periodized Hilbert schemes) Let \mathbb{G}_m^2 act on \mathbb{A}^2 by scaling the coordinate axes \tilde{x} and \tilde{y} with weights $(1, 0)$ and $(0, 2)$, respectively (in the (\tilde{X}, \tilde{Y}) -grading). This induces an action of \mathbb{G}_m^2 on $\text{Hilb}_n := \text{Hilb}_n(\mathbb{A}^2)$ for any n . The 2-periodized category of quasicohherent sheaves on Hilb_n is defined to be

$$\text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)^{2\text{-per}} := \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}(\text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)),$$

where β lives in cohomological degree -2 and bigraded degrees $(1, 2)$. Taking the compact objects, we define¹⁷

$$\text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)^{2\text{-per}} := (\overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}(\text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)))^{\text{perf}}.$$

The full subcategories $\text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_n, \tilde{X}}^{2\text{-per}} / \mathbb{G}_m^2$ and $\text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_n, \tilde{X}}^{2\text{-per}} / \mathbb{G}_m^2$ are defined analogously (see also Section 4.2.9).

Definition 4.4.11 (2-periodized bigraded chain complexes) We define

$$\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}} := \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}},$$

¹⁷Note that since the Hilbert scheme is smooth, we could replace Perf by Coh.

where β lives in cohomological degree -2 and bigraded degrees $(1, 2)$. The full subcategory spanned by compact objects is defined analogously and is denoted by

$$\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}, c} := \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, \text{perf}}.$$

More generally, for any algebra $\mathcal{B} \in \text{Alg}(\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}})$, we let

$$\mathcal{B}\text{-Mod}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}} := \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}(\mathcal{B}\text{-Mod}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}}),$$

and similarly for the full subcategory spanned by the compact objects

$$\mathcal{B}\text{-Mod}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}, \text{perf}} := \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}(\mathcal{B}\text{-Mod}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}})^{\text{perf}}.$$

4.4.12 Shearing in terms of \tilde{X} , \tilde{Y} -grading Let $\mathcal{C} \in \text{Vect}^{\text{gr}_X, \text{gr}_Y}\text{-Mod} \simeq \text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}}\text{-Mod}$. Then, unless otherwise specified, all shears are considered to be with respect to the Y -grading. For example, the action of $\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}}$ on $\Rightarrow \mathcal{C}$ is given by

$$\overline{\mathbb{Q}}_\ell\langle k \rangle_{\tilde{X}} \boxtimes \overline{\mathbb{Q}}_\ell\langle l \rangle_{\tilde{Y}} \boxtimes c \mapsto \overline{\mathbb{Q}}_\ell\langle 2k - l \rangle_X \boxtimes \overline{\mathbb{Q}}_\ell\langle k \rangle_Y \boxtimes c \mapsto c\langle 2k - l \rangle_X \langle k \rangle_Y [2k].$$

Here $c\langle m \rangle_X$ denotes the result obtained by $\overline{\mathbb{Q}}_\ell\langle m \rangle_X$ acting on c (and similarly for $c\langle m \rangle_Y$). Moreover, the square bracket denotes cohomological shift, which appears here due to the shear.

By Remark 4.3.17, $\Rightarrow \mathcal{C}$ is equipped with a natural action of $\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}}$ (see Definition 4.4.11).

4.4.13 The Hilbert scheme of points and graded unipotent character sheaves The relation between Hilbert schemes of points and graded unipotent character sheaves can now be deduced in a straightforward way from the discussion above.

We start with a 2-periodized form of Proposition 4.2.7:

Lemma 4.4.14 *We have the equivalences of $\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}}$ -module categories*

$$\text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)^{2\text{-per}} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}} \quad \text{and} \quad \text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{X}}}^{2\text{-per}} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\tilde{Y}}}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}},$$

and similarly for the full subcategories of perfect complexes.

Proof The second equivalence follows from the first. The first follows from applying the 2-periodization construction 2-per described in Section 4.4.8 to Proposition 4.2.7. □

Theorem 4.4.15 *When $G = \text{GL}_n$, we have the equivalence of $\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}}$ -module categories*

$$\Rightarrow \text{Ch}_G^{\text{u, gr, ren}} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\tilde{Y}}}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}} \xrightarrow{\Psi^{2\text{-per}}} \text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{X}}}^{2\text{-per}},$$

and similarly, we have the small variant, which is an equivalence of $\text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}, c}$ -module categories,

$$\Rightarrow \text{Ch}_G^{\text{u, gr}} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\tilde{Y}}}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}} \xrightarrow{\Psi^{2\text{-per}}} \text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{X}}}^{2\text{-per}}.$$

Proof The second statement follows from the first one by taking the full subcategories spanned by compact objects. For the first one, we have

$$\begin{aligned} \Rightarrow \text{Ch}_G^{u, \text{gr}, \text{ren}} &\simeq \Rightarrow (\mathcal{B}_n\text{-Mod}_{\text{nilp}_{\underline{y}}}^{\text{gr}X}) \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\underline{y}}}^{\text{gr}X, \text{gr}Y, 2\text{-per}} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\underline{y}}}^{\text{gr}\tilde{X}, \text{gr}\tilde{Y}, 2\text{-per}} \\ &\simeq \text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{X}}}^{2\text{-per}}, \end{aligned}$$

where

- the first equivalence is due to Proposition 4.1.11 (after adding a sheared action of $\text{Vect}^{\text{gr}Y}$),
- the second equivalence is due to Corollary 4.4.5,
- the third equivalence is by definition (see also Section 4.4.8),
- and finally, the last equivalence is due to Lemma 4.4.14. □

4.5 Matching objects

Let \mathcal{T} denote the tautological bundle over Hilb_n . By abuse of notation, we will use the same symbol to denote its equivariant version, ie a vector bundle of rank n over $\text{Hilb}_n / \mathbb{G}_m^2$. We will now match its exterior powers $\wedge^\alpha \mathcal{T}$ under the equivalences of categories stated in Proposition 4.2.7 and Lemma 4.4.14. This subsection is merely a recollection of the results proved in [35], stated in our notation.

4.5.1 Exterior powers of the permutation representation

Let

$$T \in \tilde{\mathcal{B}}_n\text{-Mod}^{\text{gr}\tilde{X}, \text{gr}\tilde{Y}} \simeq \text{QCoh}(\mathbb{A}^{2n} / (\mathbb{S}_n \times \mathbb{G}_m^2))$$

denote the tensor of the structure sheaf with the permutation representation. More explicitly,

$$T := \overline{\mathbb{Q}}_\ell[\tilde{X}, \tilde{Y}] \otimes P \in \tilde{\mathcal{B}}_n\text{-Mod}^{\text{gr}\tilde{X}, \text{gr}\tilde{Y}},$$

where $P \in \text{Rep } \mathbb{S}_n$ is the permutation representation. More geometrically, if we let

$$p: \mathbb{A}^{2n} / (\mathbb{S}_n \times \mathbb{G}_m^2) \rightarrow B(\mathbb{S}_n \times \mathbb{G}_m^2),$$

then $T := p^*(P)$, where $P \in \text{QCoh}(B(\mathbb{S}_n \times \mathbb{G}_m^2))$ is the permutation of \mathbb{S}_n put in bigraded degrees $(0, 0)$ and cohomological degree 0. More generally, we have $\wedge^\alpha T \simeq p^*(\wedge^\alpha P)$.

We have the following elementary lemma:

Lemma 4.5.2 *For any $0 \leq \alpha \leq n$, we have a natural isomorphism of \mathbb{S}_n -representations*

$$\text{Ind}_{\mathbb{S}_\alpha \times \mathbb{S}_{n-\alpha}}^{\mathbb{S}_n} (\text{sign}_\alpha \boxtimes \text{triv}_{n-\alpha}) \simeq \wedge^\alpha P,$$

where sign_α is the sign representation of \mathbb{S}_α and $\text{triv}_{n-\alpha}$ is the 1-dimensional trivial representation of $\mathbb{S}_{n-\alpha}$.

Proof Let v_1, \dots, v_n be the natural basis of P as the permutation representation of \mathbb{S}_n and v a basis of $\text{sign}_\alpha \boxtimes \text{triv}_{n-\alpha}$. We have a natural morphism between $(\mathbb{S}_\alpha \times \mathbb{S}_{n-\alpha})$ -representations

$$\text{sign}_\alpha \boxtimes \text{triv}_{n-\alpha} \rightarrow \wedge^\alpha P, \quad v \mapsto v_1 \wedge \dots \wedge v_\alpha.$$

Since $v_1 \wedge \cdots \wedge v_\alpha$ generates $\wedge^\alpha P$ as a representation of S_n , we obtain a surjective map

$$\mathrm{Ind}_{S_\alpha \times S_{n-\alpha}}^{S_n} (\mathrm{sign}_\alpha \boxtimes \mathrm{triv}_{n-\alpha}) \rightarrow \wedge^\alpha P.$$

Comparing the dimensions, we see that this must be an isomorphism. □

4.5.3 Wedges of the tautological bundle and wedges of the permutation representation We will now match the wedges on both sides.

Theorem 4.5.4 [35, Theorem 3.9] *Under the equivalence of categories stated in Proposition 4.2.7, $\wedge^\alpha T \in \widetilde{\mathcal{B}}_n\text{-Mod}^{\mathrm{gr} \tilde{x}, \mathrm{gr} \tilde{y}} \simeq \mathrm{QCoh}(\mathbb{A}^{2n}/(S_n \times \mathbb{G}_m^2))$ corresponds to $\wedge^\alpha \mathcal{T} \in \mathrm{QCoh}(\mathrm{Hilb}_n/\mathbb{G}_m^2)$.*

Proof This is [35, Theorem 3.9] in the case where the line bundle involved is just the structure sheaf. Lemma 4.5.2 allows us to match the wedges $\wedge^\alpha T$ with the $W^\alpha(-)$ construction in [35, Definition 3.4]. □

4.5.5 2-periodization We will need the 2-periodized version of Theorem 4.5.4 below.

Corollary 4.5.6 *Under the equivalence of categories stated in Lemma 4.4.14,*

$$(\wedge^\alpha T)^{2\text{-per}} \in \widetilde{\mathcal{B}}_n\text{-Mod}^{\mathrm{gr} \tilde{x}, \mathrm{gr} \tilde{y}, 2\text{-per}} \simeq \mathrm{QCoh}(\mathbb{A}^{2n}/(S_n \times \mathbb{G}_m^2))^{2\text{-per}}$$

corresponds to $(\wedge^\alpha \mathcal{T})^{2\text{-per}} \in \mathrm{QCoh}(\mathrm{Hilb}_n/\mathbb{G}_m^2)^{2\text{-per}}$. Here the superscript 2-per denotes the procedure of 2-periodizing an object (see Definition 4.4.9).

Proof This follows from 2-periodizing Theorem 4.5.4. □

5 HOMFLY-PT homology via Hilbert schemes of points on \mathbb{C}^2

We will now use the results above to obtain a proof of a version of a conjecture of Gorsky, Neguț, and Rasmussen [29] which predicts a remarkable way to realize HOMFLY-PT homology as the cohomology of a certain coherent sheaf on the Hilbert scheme of points on \mathbb{C}^2 . The precise version of the conjecture that we prove appeared as [28, Conjecture 7.2] with some modifications emphasized in Remark 1.5.12. In particular, we will prove parts (a), (b), and (c) of [28, Conjecture 7.2].

Below, in Section 5.1, we use weight structures to describe a general mechanism to turn a cohomological grading into a formal grading and interpret the construction of the HOMFLY-PT homology theory in these terms (this is what happens with the a -degree). It is followed by Section 5.2, where we describe how HOMFLY-PT can be obtained from the categorical trace of H_n^{gr} . In Section 5.3, we explicitly compute all the functors involved in the factorization of HOMFLY-PT homology through the categorical trace of H_n^{gr} in terms of the explicit description of the category of the latter, which was obtained in Theorem 3.3.4. In Section 5.4, transporting these computations to the Hilbert scheme side using Koszul duality and Krug’s result (which are encapsulated in Section 4), we arrive at Theorem 5.4.6 (part (a) of [28, Conjecture 7.2]) which associates to each braid β on n strands a (2-periodic) coherent sheaf \mathcal{F}_β on Hilb_n whose global

sections (after being tensored by the dual of various exterior powers of the tautological bundle) recover the HOMFLY-PT homology of the link associated by β . Finally, in Sections 5.5 and 5.6 we study the actions of symmetric functions on HOMFLY-PT homology and the support of \mathcal{F}_β on Hilb_n in relation to the number of connected components the braid closure of β has. These appear as Theorems 5.5.1 and 5.6.4, which are parts (b) and (c) of [28, Conjecture 7.2].

5.1 Restricting to and extending from the weight heart

Let $R_\beta \in \text{Ch}^b(\text{SBim}_n)$ be the Rouquier complex associated to a braid β . Recall that the HOMFLY-PT homology of the associated braid closure of β is obtained by taking Hochschild homology *termwise* and then taking the cohomology of the resulting complexes. The final answer thus has *three* gradings: internal grading, complex grading (from $\text{Ch}^b(\text{SBim}_n)$), and the Hochschild grading. From the point of view of homological algebra, this is quite unusual: one does not normally apply a derived functor termwise to a complex.

There have been many interpretations of this construction, for example, [13; 57; 59]. In this subsection, we offer a general paradigm to understand this type of construction using weight structures and then apply it to the case of HOMFLY-PT homology. This allows one to see more clearly what is happening conceptually, especially on the Koszul dual side, which will be discussed in Section 5.4 below.

As we will use weight structures in a crucial way, the reader might find it beneficial to take a quick look at [31, Section 3.1] for a quick review of the theory.

5.1.1 Extending from the weight heart One salient feature of the theory of weights is that given an idempotent complete stable ∞ -category \mathcal{C} with a bounded weight structure, the category \mathcal{C} along with the weight structure can be reconstructed from its weight heart $\mathcal{C}^{\heartsuit_w}$.¹⁸ In a precise sense, \mathcal{C} is the free stable ∞ -category generated by its weight heart. Namely, $\mathcal{C} \simeq (\mathcal{C}^{\heartsuit_w})^{\text{fin}}$, where for any additive category \mathcal{A} , \mathcal{A}^{fin} is the stabilization of the sifted-completion of \mathcal{A} (see [22] for more details). We note that the $(-)^{\text{fin}}$ procedure generalizes the construction $\text{Ch}^b(-)$ in the following sense: if \mathcal{A} is a classical additive category, then $\mathcal{A}^{\text{fin}} \simeq \text{Ch}^b(\mathcal{A})$.

Lemma 5.1.2 *Let \mathcal{C} and \mathcal{D} be idempotent complete stable ∞ -categories such that \mathcal{C} is equipped with a bounded weight structure. Then restricting along $\iota: \mathcal{C}^{\heartsuit_w} \rightarrow \mathcal{C}$ gives an equivalence of categories*

$$(5.1.3) \quad \text{Fun}^{\text{ex}}(\mathcal{C}, \mathcal{D}) \xrightarrow{\iota^!} \text{Fun}^{\text{add}}(\mathcal{C}^{\heartsuit_w}, \mathcal{D}),$$

where Fun^{ex} denotes the category of exact functors (ie those that preserve all finite (co)limits) and Fun^{add} denotes the category of additive functors (ie those that preserve all finite direct sums).

¹⁸A DG-structure is not necessary for this discussion in this subsection. However, the reader should feel free to replace all occurrences of stable ∞ -categories with DG-categories if they prefer.

Proof This is essentially [56, Proposition 3.3, part 2] (but see also [22, Theorem 2.2.9] for a variant). We will thus only indicate the main steps.

We will first show the equivalence of categories

$$(5.1.4) \quad \text{Fun}^{\text{ex}}(\mathcal{C}, \text{Ind}(\mathcal{D})) \xrightarrow{\iota^!} \text{Fun}^{\text{add}}(\mathcal{C}^{\heartsuit w}, \text{Ind}(\mathcal{D})),$$

where $\text{Ind}(\mathcal{D})$ is the Ind-completion of \mathcal{D} . Here $\iota^!$ admits a left adjoint $\iota_!$ given by left Kan extending along ι . Since left Kan extending along a fully faithful embedding is fully faithful, $\iota_!$ is fully faithful. It remains to show that $\iota^!$ is also fully faithful. Namely, we want to show that the natural map $\iota_! \iota^! F \rightarrow F$ is an equivalence for any $F \in \text{Fun}^{\text{ex}}(\mathcal{C}, \mathcal{D})$.

Since $\iota^! \iota_! \iota^! F \simeq \iota^! F$ by full faithfulness of $\iota_!$, we see that $\iota_! \iota^! F(c) \simeq F(c)$ for all $c \in \mathcal{C}^{\heartsuit w}$. But since both sides of $\iota_! \iota^! F \rightarrow F$ preserve all finite (co)limits and since \mathcal{C} is generated by $\mathcal{C}^{\heartsuit w}$ under finite (co)limits, we are done.

To obtain (5.1.3) from (5.1.4) we only need to show that $\iota^!$ and $\iota_!$ restrict to the corresponding full subcategories on both sides. But this is immediate using the fact that \mathcal{C} is generated by $\mathcal{C}^{\heartsuit w}$ under finite (co)limits. □

We will also need the following result, which is a consequence of the lemma above:

Corollary 5.1.5 [22, Theorem 2.2.9] *Let \mathcal{C} and \mathcal{D} be idempotent complete stable ∞ -categories such that both \mathcal{C} and \mathcal{D} are each equipped with a bounded weight structure. Then restricting along $\iota: \mathcal{C}^{\heartsuit w} \rightarrow \mathcal{C}$ gives an equivalence of categories*

$$\text{Fun}^{\text{ex}, w\text{-ex}}(\mathcal{C}, \mathcal{D}) \xrightarrow{\iota^!} \text{Fun}^{\text{add}}(\mathcal{C}^{\heartsuit w}, \mathcal{D}^{\heartsuit w}),$$

where $w\text{-ex}$ denotes weight exactness, ie we consider the category of exact functors which send weight heart to weight heart.

Consequently, when $\mathcal{D}^{\heartsuit w}$ is classical, we have the following equivalences:

$$\text{Fun}^{\text{ex}, w\text{-ex}}(\mathcal{C}, \mathcal{D}) \simeq \text{Fun}^{\text{add}}(\mathcal{C}^{\heartsuit w}, \mathcal{D}^{\heartsuit w}) \simeq \text{Fun}^{\text{add}}(\text{h}\mathcal{C}^{\heartsuit w}, \mathcal{D}^{\heartsuit w}) \simeq \text{Fun}^{\text{ex}, w\text{-ex}}(\text{Ch}^b(\text{h}\mathcal{C}_w^{\heartsuit}), \mathcal{D}).$$

5.1.6 Turning cohomological grading into a formal grading In situations where \mathcal{D} also has a t -structure, the procedure of restricting and extending along $\mathcal{C}^{\heartsuit w} \hookrightarrow \mathcal{C}$ also allows one to create an extra grading. Indeed, let \mathcal{C} and \mathcal{D} be as in Lemma 5.1.2 and $F: \mathcal{C} \rightarrow \mathcal{D}$ be an exact functor. Suppose that \mathcal{D} is equipped with a t -structure. Then we obtain a functor

$$\text{H}^*(F|_{\mathcal{C}^{\heartsuit w}}): \mathcal{C}^{\heartsuit w} \xrightarrow{F|_{\mathcal{C}^{\heartsuit w}}} \mathcal{D} \xrightarrow{\text{H}^*} \mathcal{D}^{\heartsuit t, \text{gr}},$$

where $\mathcal{D}^{\heartsuit t, \text{gr}}$ is the category of \mathbb{Z} -graded objects in $\mathcal{D}^{\heartsuit t}$. Applying the $(-)^{\text{fin}}$ construction and using Corollary 5.1.5, we obtain a functor

$$\tilde{F}: \mathcal{C} \rightarrow (\mathcal{D}^{\heartsuit t, \text{gr}})^{\text{fin}} \simeq \text{Ch}^b(\mathcal{D}^{\heartsuit t, \text{gr}}) \rightarrow \text{Ch}^b(\mathcal{D}^{\heartsuit t})^{\text{gr}},$$

where the equivalence is due to the fact that $\mathcal{D}^{\heartsuit_t, \text{gr}}$ is *classical*, ie it does not have negative Exts. In fact, since $\mathcal{D}^{\heartsuit_t, \text{gr}}$ is classical, $\mathcal{C}^{\heartsuit_w} \rightarrow \mathcal{D}^{\heartsuit_t, \text{gr}}$ factors through $\mathcal{C}^{\heartsuit_w} \rightarrow \text{h}\mathcal{C}^{\heartsuit_w}$ and hence \tilde{F} factors as follows:

$$(5.1.7) \quad \begin{array}{ccc} \mathcal{C} & \xrightarrow{\tilde{F}} & \text{Ch}^b(\mathcal{D}^{\heartsuit_t})^{\text{gr}} \\ & \searrow \text{wt} & \nearrow \bar{F} \\ & & \text{Ch}^b(\text{h}\mathcal{C}^{\heartsuit_w}) \end{array}$$

Here $\text{h}\mathcal{C}^{\heartsuit_w}$ is the homotopy category of $\mathcal{C}^{\heartsuit_w}$, obtained from $\mathcal{C}^{\heartsuit_w}$ by killing negative Exts,¹⁹ and wt denotes the weight complex functor (see [31, Remark 3.1.12] for a quick review). Note that $\text{h}\mathcal{C}^{\heartsuit_w}$ is always a classical category whereas $\mathcal{C}^{\heartsuit_w}$ might have nontrivial ∞ -categorical structures. Moreover, \bar{F} is given by applying $H^*(F)$ termwise.

5.1.8 When $\mathcal{C}^{\heartsuit_w}$ is already classical, ie $\mathcal{C}^{\heartsuit_w} \simeq \text{h}\mathcal{C}^{\heartsuit_w}$, the weight complex functor is an equivalence of categories

$$\text{wt}: \mathcal{C} \xrightarrow{\simeq} \text{Ch}^b(\mathcal{C}^{\heartsuit_w}).$$

Under this identification, the functor \tilde{F} is identified with \bar{F} , which is given by applying $H^*(F)$ termwise to the chain complexes in $\mathcal{C}^{\heartsuit_w}$.

5.1.9 The category \mathcal{D} that is of interest to us is $\mathcal{D} \simeq \text{Vect}^{\text{gr}}$. Given an exact functor $F: \mathcal{C} \rightarrow \text{Vect}^{\text{gr}}$, the construction above thus produces a functor $\tilde{F}: \mathcal{C} \rightarrow \text{Vect}^{\text{gr}, \text{gr}}$ whose target is the DG-category of chain complexes in doubly graded vector spaces. We can take cohomology another time and obtain triply graded vector spaces. As we will soon see, further specializing to the case where $\mathcal{C} = H_n^{\text{gr}} \simeq \text{Ch}^b(H_n^{\text{gr}, \heartsuit_w}) \simeq \text{Ch}^b(\text{SBim}_n)$ and F is the functor of taking Hochschild homology, we recover the usual construction of the triply graded HOMFLY-PT homology theory.

5.1.10 The last observation is a tautology, but a powerful one when combined with Lemma 5.1.2 which allows one to compare functors by restricting to the weight heart of the source. This point of view is especially useful when the weight structure is not evident, as is the case on the Hilbert scheme side. Indeed, as we shall see below, Lemma 5.1.2 allows us to identify HH_α and the functor corepresented by the α^{th} exterior of the tautological bundle on the Hilbert scheme.

5.2 HOMFLY-PT homology via $\text{Tr}(H_n^{\text{gr}})$

We will now describe how HOMFLY-PT homology can be obtained from $\text{Tr}(H_n^{\text{gr}}) \simeq \text{Ch}_G^{u, \text{gr}}$. Everything in this subsection has essentially been proved in [59], but in different language: they use the chromatographic complex construction (which is not known to be functorial) of which the weight complex functor is a functorial upgrade. See also [54, Section 6.4] for a presentation that is closer to ours (but still does not use weight structures).

¹⁹For an additive ∞ -category \mathcal{A} , the homotopy category $\text{h}\mathcal{A}$ of \mathcal{A} is obtained from \mathcal{A} by taking π_0 of the Hom spaces. This is not to be confused with the homotopy category $\mathcal{K}^b(\mathcal{B}) := \text{h}\text{Ch}^b(\mathcal{B})$ of bounded chain complexes of a (classical) additive category \mathcal{B} , which is also sometimes referred to as, somewhat confusingly, the homotopy category of \mathcal{B} .

5.2.1 A geometric interpretation of HOMFLY-PT homology We start with a formulation of HOMFLY-PT homology in terms of graded sheaves on $BB \times_{BG} BB = B \backslash G/B$. Consider the commutative diagram

$$\begin{array}{ccccc} BB \times_{BG} BB & \xleftarrow{p} & G/B & \xrightarrow{q} & G/G & \longrightarrow & \text{pt} \\ \downarrow \pi & & \downarrow \pi & & & & \\ BB \times BB & \xleftarrow{p} & BB & & & & \end{array}$$

where the square is Cartesian. We have the following functor:

(5.2.2) $\text{HH}: H_n^{\text{gr}} \simeq \text{Shv}_{\text{gr},c}(BB \times_{BG} BB) \xrightarrow{\pi_*} \text{Shv}_{\text{gr},c}(BB \times BB) \xrightarrow{p^*} \text{Shv}_{\text{gr},c}(BB) \xrightarrow{C_{\text{gr}}^*(BB,-)} \text{Vect}^{\text{gr}}.$

Applying the construction in Section 5.1.6, we obtain a functor

$$\widetilde{\text{HH}}: H_n^{\text{gr}} \rightarrow \text{Vect}^{\text{gr},\text{gr}}.$$

Taking cohomology one more time, we get

$$\text{HHH} := H^*(\widetilde{\text{HH}}): H_n^{\text{gr}} \rightarrow \text{Vect}^{\heartsuit_{t,\text{gr},\text{gr},\text{gr}}},$$

the category of triply graded vector spaces.

Lemma 5.2.3 *Let $R_\beta \in H_n^{\text{gr}}$ be associated to a braid β . Then up to a change of grading (to be discussed in Section 5.2.4 below), $H^*(\widetilde{\text{HH}}(R_\beta))$ is the triply graded HOMFLY-PT link homology of β .*

Proof (sketch) This is the main theorem of [59], phrased in our language. See also [54, Section 6.4] for a presentation that is closer to ours. We will thus only indicate the main steps here.

The construction in Section 5.1.6 amounts to saying that $\widetilde{\text{HH}}(R_\beta)$ is obtained by applying $p^* \pi_*$ termwise to $R_\beta \in H_n^{\text{gr}} \simeq \text{Ch}^b(H_n^{\text{gr},\heartsuit_w})$. By [31, Proposition 4.3.5], $\text{Shv}_{\text{gr},c}(BB \times BB)$ (resp. $\text{Shv}_{\text{gr},c}(BB)$) corresponds to the category of (perfect complexes of) graded bimodules (resp. modules) over $C_{\text{gr}}^*(BB) \simeq C_*^{\text{gr}}(BT) \simeq \text{Sym}(V[-2])$, where V is the graded vector space of dimension n living in graded degree 2 and cohomological degree 0 (see also Section 4.1.4). Moreover, by [31, Section 4.3.4], pulling back along $p: BB \rightarrow BB \times BB$ corresponds to taking Hochschild homology. Finally, by [31, (4.4.4)], $\pi_*|_{H_n^{\text{gr},\heartsuit_w}}$ is identified with the forgetful functor from Soergel bimodules to bimodules over $C_*^{\text{gr}}(BB) \simeq C_*^{\text{gr}}(BT)$. \square

5.2.4 Gradings We let Q' and A' denote the two formal gradings on $\text{Vect}^{\text{gr},\text{gr}}$ (the target of the functor $\widetilde{\text{HH}}$) and T' the cohomological grading. To keep track of these gradings, we will from now on adopt the notation $\text{Vect}^{\text{gr},Q',\text{gr},A'}$ rather than simply $\text{Vect}^{\text{gr},\text{gr}}$ as above. We use $[n]_{Q'}$, $\langle n \rangle_{A'}$, and $\{n\}_{T'}$ to denote the grading shifts.

The relations between the Q' , A' , and T' gradings and the Q , A , and T of [28] are as follows (written multiplicatively):

$$Q' = A^{-1}Q, \quad A' = A, \quad T' = T.$$

Note, however, that [28] uses Hochschild *cohomology* rather than homology to define the HOMFLY-PT link homology. Thus, for the unknot we get $(1 + Q'^2 A')/(1 - Q'^2 A'^2) = (1 + Q^2 A^{-1})/(1 - Q^2)$ whereas they get $(1 + Q^{-2} A)/(1 - Q^2)$.

We also use the q , t , and a gradings, where

$$q = Q^2 = Q'^2 A'^2, \quad a = Q^2 A^{-1} = Q'^2 A', \quad t = T^2 Q^{-2} = T'^2 Q'^{-2} A'^{-2}.$$

In terms of q , a , and t , the unknot gives $(1 + a)/(1 - q)$.

5.2.5 On H_n^{gr} , $\text{Ch}_G^{\text{u,gr}} \simeq \mathcal{A}_n\text{-Mod}^{\text{gr,perf}}$, and $\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}$ there are two grading shift operators for each $n \in \mathbb{Z}$, the cohomological shift $[n]$ and the grading shift $\langle n \rangle$, and both are compatible with the various functors between these categories. In terms of the grading convention of Section 4.4, this formal grading is the X -grading over there.

5.2.6 Unwinding the construction, we see that the functor $\widetilde{\text{HH}}$ relates the grading shift operators on H_n^{gr} and those on $\text{Vect}^{\text{gr}_{Q', \text{gr}_{A'}}$ as follows:²⁰

$$[n] \rightsquigarrow \{n\}_{T'}, \quad \langle n \rangle \rightsquigarrow [n]_{A'} \langle n \rangle_{Q'} \{-n\}_{T'}, \quad [n] \langle n \rangle \rightsquigarrow [n]_{A'} \langle n \rangle_{Q'}.$$

Remark 5.2.7 The difference between Q' , A' , T' and Q , A , T comes from the fact that the identification $H_n^{\text{gr}} := \text{Shv}_{\text{gr},c}(B \setminus G/B) \simeq \text{Ch}^b(\text{SBim}_n)$ involves a shear. More concretely, objects in H_n^{gr} are most naturally viewed as graded $C_{\text{gr}}^*(BB)$ -bimodules, where the generators of $C_{\text{gr}}^*(BB) \simeq C_{\text{gr}}^*(BT) \simeq \text{Sym } V[-2]$ live in graded (resp. cohomological) degree 2 (resp. 2). On the other hand, the generators of the polynomial ring used to define Soergel bimodules are, by convention, put in graded (resp. cohomological) degrees 2 (resp. 0). See [31, Section 4] for a more in-depth discussion.

5.2.8 HOMFLY-PT homology via truncated categorical trace By smooth base change, we see that the functor HH of (5.2.2) admits an alternative construction

$$\text{HH}: H_n^{\text{gr}} \simeq \text{Shv}_{\text{gr},c}(BB \times_{BG} BB) \xrightarrow{p^*} \text{Shv}_{\text{gr},c}\left(\frac{G}{B}\right) \xrightarrow{q_* \simeq q'_!} \text{Shv}_{\text{gr},c}\left(\frac{G}{G}\right) \xrightarrow{C_{\text{gr}}^*(G/G, -)} \text{Vect}^{\text{gr}},$$

which is the same as

$$(5.2.9) \quad \text{HH}: H_n^{\text{gr}} \xrightarrow{\text{tr}} \text{Tr}(H_n^{\text{gr}}) \simeq \text{Ch}_G^{\text{u,gr}} \hookrightarrow \text{Shv}_{\text{gr},c}(G/G) \xrightarrow{C_{\text{gr}}^*(G/G, -)} \text{Vect}^{\text{gr}}.$$

Since the functor $\text{tr}: H_n^{\text{gr}} \rightarrow \text{Ch}_G^{\text{u,gr}}$ is weight exact, it commutes with the weight complex functors. Thus, to construct $\widetilde{\text{HH}}$ from HH , we can apply the construction from Section 5.1.6 to the last step of (5.2.9). As a result, we obtain the factorization of $\widetilde{\text{HH}}$

$$(5.2.10) \quad \begin{array}{ccccc} & & \widetilde{\text{HH}} & & \\ & \nearrow & & \searrow & \\ H_n^{\text{gr}} & \xrightarrow{\text{tr}} & \text{Ch}_G^{\text{u,gr}} & \xrightarrow{\Gamma} & \text{Vect}^{\text{gr}_{Q', \text{gr}_{A'}}} \\ & & \downarrow \text{wt} & \nearrow \bar{\Gamma} & \\ & & \text{Ch}^b(\text{h Ch}_G^{\text{u,gr}, \heartsuit_w}) & & \end{array}$$

²⁰The first and last are easiest to see from the definition, from which the middle one could be deduced. Alternatively, the middle one can also be seen by looking at the weight complex functor.

where Γ is obtained by applying the construction from Section 5.1.6 to the functor $C_{gr}^*(G/G, -)$ and $\bar{\Gamma}$ is the induced functor; see (5.1.7). The category $Ch^b(h Ch_G^{u,gr,\heartsuit_w})$ (resp. $h Ch_G^{u,gr,\heartsuit_w}$) can be thought of as a truncation of $Ch_G^{u,gr}$ (resp. Ch^{u,gr,\heartsuit_w}), which justifies the name *truncated trace*. It is also referred to as the *underived trace* and is denoted by $Tr_0(H_G^{gr})$ in [27].

5.3 Explicit computation of functors

In what follows, we will compute the functors $\Gamma, \bar{\Gamma}$, and wt in (5.2.9) explicitly in terms of the identification

$$Ch_G^{u,gr} \simeq \overline{\mathbb{Q}}_\ell[x, \theta] \otimes \overline{\mathbb{Q}}_\ell[S_n]\text{-Mod}^{gr,perf} = \mathcal{A}_n\text{-Mod}^{gr,perf}$$

of Theorem 3.3.4. In particular, we apply Lemma 5.1.2 to deduce a corepresentability statement similar to the main result of [13; 57].²¹ Moreover, our computation identifies the weight complex functor wt with the functor of restricting to the nilpotent cone, explaining conceptually why the latter appears in [57]; see Section 5.3.5.

5.3.1 Identifying the weight complex functor wt for $Ch_G^{u,gr}$ We will now give a concrete interpretation of the functor wt appearing in (5.2.10) in terms of modules over \mathcal{A}_n . Consider the functor

$$\widetilde{inv}_\theta := \text{Sym}^n V^\vee[1] \otimes inv_\theta : \mathcal{A}_n\text{-Mod}^{gr,perf} \rightarrow \bar{\mathcal{A}}_n\text{-Mod}^{gr,perf},$$

where inv_θ is defined in Section 4.1 and where we define

$$\bar{\mathcal{A}}_n := \overline{\mathbb{Q}}_\ell[x] \otimes \overline{\mathbb{Q}}_\ell[S_n] \simeq \text{Sym } V_x[-2] \otimes \overline{\mathbb{Q}}_\ell[S_n].$$

See also Section 4.1.4 for the definitions of V and \mathcal{A}_n .

Note that inv_θ is originally defined on $\mathcal{A}_n\text{-Mod}^{gr,coh}$. So strictly speaking, in the above, we restrict this functor to the full subcategory of perfect complexes. It is easy to see that \widetilde{inv}_θ is given precisely by applying $\widehat{inv}_\theta^{enh}$ followed by the functor of forgetting the action by the variables \underline{y} .

The goal now is to identify $wt : Ch_G^{u,gr} \rightarrow Ch^b(h Ch_G^{u,gr,\heartsuit_w})$ with \widetilde{inv}_θ .

5.3.2 By Lemma 2.9.5, we see that in type A , $Ch_G^{u,gr} \simeq \mathcal{A}_n\text{-Mod}^{gr,perf}$ is equipped with a weight structure whose weight heart is spanned under finite direct sums of direct summands of $\text{Spr}_G^{gr}[k]\langle k \rangle$ (for all $k \in \mathbb{Z}$), which corresponds to $\mathcal{A}_n[k]\langle k \rangle$ (for all $k \in \mathbb{Z}$) under this equivalence of categories. Here $[k]$ and $\langle k \rangle$ denote cohomological and grading shifts, respectively.

A similar statement is true for $\bar{\mathcal{A}}_n\text{-Mod}^{gr,perf}$.

Lemma 5.3.3 *The category $\bar{\mathcal{A}}_n\text{-Mod}^{gr,perf}$ has a natural weight structure whose weight heart is spanned by finite direct sums of direct summands of $\bar{\mathcal{A}}_n[k]\langle k \rangle$ for all $k \in \mathbb{Z}$.*

²¹It is in fact possible to reprove the corepresentability statement of [13] using the computation presented here. The details will appear in a forthcoming paper.

Moreover, the weight heart $\overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}$ is classical, and hence the weight complex functor induces an equivalence of categories

$$\text{wt}: \overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}} \xrightarrow{\cong} \text{Ch}^b(\overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}).$$

Proof It is easy to see that the direct summands of $\overline{\mathcal{A}}_n[k]\langle k \rangle$ generate the category under finite (co)limits. Moreover, as there are no positive Exts between them, the weight structure is obtained by invoking [15, Theorem 4.3.2.II]. The second part follows by observing that there are no negative Exts between these objects either; see also [31, Section 3.1.9]. \square

We are now ready to identify wt and $\widetilde{\text{inv}}_\theta$.

Proposition 5.3.4 *The functor $\widetilde{\text{inv}}_\theta$ induces an equivalence of categories $\overline{\text{inv}}_\theta$ as given in the following commutative diagram:*

$$\begin{array}{ccc} \mathcal{A}_n\text{-Mod}^{\text{gr,perf}} & \xrightarrow{\widetilde{\text{inv}}_\theta} & \overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}} \\ & \searrow \text{wt} & \nearrow \cong \\ & & \text{Ch}^b(\mathfrak{h}\mathcal{A}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}) \end{array}$$

Proof The computation in Section 4.1.8 shows that $\widetilde{\text{inv}}_\theta(\mathcal{A}_n) \simeq \overline{\mathcal{A}}_n$ and hence $\widetilde{\text{inv}}_\theta$ is weight exact. By Corollary 5.1.5, $\widetilde{\text{inv}}_\theta$ fits into the following commutative diagram:

$$\begin{array}{ccc} \mathcal{A}_n\text{-Mod}^{\text{gr,perf}} & \xrightarrow{\widetilde{\text{inv}}_\theta} & \overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}} \\ \downarrow \text{wt} & & \downarrow \cong \\ \text{Ch}^b(\mathfrak{h}\mathcal{A}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}) & \xrightarrow{\overline{\text{inv}}_\theta} & \text{Ch}^b(\overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}) \end{array}$$

It remains to show that $\overline{\text{inv}}_\theta$ is an equivalence of categories.

The diagram above is determined by the following commutative diagram involving the weight hearts:

$$\begin{array}{ccc} \mathcal{A}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w} & \xrightarrow{\widetilde{\text{inv}}_\theta^{\heartsuit_w}} & \overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w} \\ \downarrow \pi_0 & & \parallel \\ \mathfrak{h}\mathcal{A}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w} & \xrightarrow{\overline{\text{inv}}_\theta^{\heartsuit_w}} & \overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w} \end{array}$$

By Corollary 5.1.5, it suffices to show that $\overline{\text{inv}}_\theta^{\heartsuit_w}$ is an equivalence of categories. But this is clear since both

$$\mathcal{H}\text{om}_{\mathcal{A}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}}(\mathcal{A}_n, \mathcal{A}_n[k]\langle k \rangle) \rightarrow \mathcal{H}\text{om}_{\mathfrak{h}\mathcal{A}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}}(\pi_0\mathcal{A}_n, \pi_0\mathcal{A}_n[k]\langle k \rangle)$$

and

$$\mathcal{H}\text{om}_{\overline{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}}(\overline{\mathcal{A}}_n, \overline{\mathcal{A}}_n[k]\langle k \rangle)$$

realize the right side by killing off θ from the left side. \square

5.3.5 Restricting to the nilpotent cone The discussion above has a geometric interpretation in terms of character sheaves and sheaves on the nilpotent cone of G . It gives an explanation for why the nilpotent cone should appear at all in [57]. The materials presented here are of independent interest and are not needed anywhere else in the paper; we include them here mostly for the sake of completeness. The reader should feel free to skip to Section 5.3.11.

Let \mathcal{N} denote the nilpotent cone of the group $G = \mathrm{GL}_n$, equipped with the conjugation action of G . Let $\iota_{\mathcal{N}}: \mathcal{N} \hookrightarrow G$ denote the closed embedding. Then we have the following functor:

$$\iota_{\mathcal{N}}^*: \mathrm{Ch}_G^{\mathrm{u},\mathrm{gr}} \rightarrow \mathrm{Shv}_{\mathrm{gr},c}(\mathcal{N}/G).$$

We let $\overline{\mathrm{Spr}}_G^{\mathrm{gr}} := \iota_{\mathcal{N}}^* \mathrm{Spr}_G^{\mathrm{gr}}$ denote the (graded) Springer sheaf.

The starting point is the following formality result of Rider, which was formulated in different language:

Theorem 5.3.6 [53, Theorem 7.9] *The functor $\mathcal{H}\mathrm{om}_{\mathrm{Shv}_{\mathrm{gr},c}(\mathcal{N}/G)}^{\mathrm{gr}}(\overline{\mathrm{Spr}}_G^{\mathrm{gr}}, -)$ induces an equivalence of $\mathrm{Vect}^{\mathrm{gr},c}$ -categories*

$$\mathrm{Shv}_{\mathrm{gr},c}(\mathcal{N}/G) \simeq \mathcal{E}\mathrm{nd}^{\mathrm{gr}}(\overline{\mathrm{Spr}}_G^{\mathrm{gr}})\text{-Mod}^{\mathrm{gr},\mathrm{perf}} \simeq \overline{\mathcal{A}}_n\text{-Mod}^{\mathrm{gr},\mathrm{perf}}.$$

The next input is a theorem of Trinh which compares the graded derived endomorphism rings of $\mathrm{Spr}_G^{\mathrm{gr}}$ and $\overline{\mathrm{Spr}}_G^{\mathrm{gr}}$:

Theorem 5.3.7 [57, Theorem 1] *The morphism of algebras*

$$\mathcal{E}\mathrm{nd}^{\mathrm{gr}}(\mathrm{Spr}_G^{\mathrm{gr}}) \rightarrow \mathcal{E}\mathrm{nd}^{\mathrm{gr}}(\iota_{\mathcal{N}}^* \mathrm{Spr}_G^{\mathrm{gr}}) \simeq \mathcal{E}\mathrm{nd}^{\mathrm{gr}}(\overline{\mathrm{Spr}}_G^{\mathrm{gr}})$$

identifies with the natural quotient $\mathcal{A}_n \rightarrow \overline{\mathcal{A}}_n$.

The following is thus a direct consequence of the two theorems above and Theorem 3.3.4:

Corollary 5.3.8 *We have a commutative diagram*

$$\begin{array}{ccc} \mathrm{Ch}_G^{\mathrm{u},\mathrm{gr}} & \xrightarrow{\iota_{\mathcal{N}}^*} & \mathrm{Shv}_{\mathrm{gr},c}(\mathcal{N}/G) \\ \mathcal{H}\mathrm{om}_{\mathrm{Ch}_G^{\mathrm{u},\mathrm{gr}}}^{\mathrm{gr}}(\mathrm{Spr}_G^{\mathrm{gr}}, -) \Big\downarrow \simeq & & \simeq \Big\downarrow \mathcal{H}\mathrm{om}_{\mathrm{Shv}_{\mathrm{gr},c}(\mathcal{N}/G)}^{\mathrm{gr}}(\overline{\mathrm{Spr}}_G^{\mathrm{gr}}, -) \\ \mathcal{A}_n\text{-Mod}^{\mathrm{gr},\mathrm{perf}} & \xrightarrow{\widetilde{\mathrm{inv}}_{\theta}} & \overline{\mathcal{A}}_n\text{-Mod}^{\mathrm{gr},\mathrm{perf}} \end{array}$$

5.3.9 The category $\mathrm{Shv}_{\mathrm{gr},c}(\mathcal{N}/G)$ has a natural weight structure whose weight heart is spanned by finite direct sums of summands of $\overline{\mathrm{Spr}}_G^{\mathrm{gr}}[k]\langle k \rangle$, $k \in \mathbb{Z}$. This corresponds to the natural weight structure on $\overline{\mathcal{A}}_n\text{-Mod}^{\mathrm{gr},\mathrm{perf}}$ which is spanned by finite direct sums of summands of $\overline{\mathcal{A}}_n[k]\langle k \rangle$ for $k \in \mathbb{Z}$. Since $\iota_{\mathcal{N}}^*(\mathrm{Spr}_G^{\mathrm{gr}}) \simeq \overline{\mathrm{Spr}}_G^{\mathrm{gr}}$, $\iota_{\mathcal{N}}^*$ is weight exact (as can also be seen at the level of $\widetilde{\mathrm{inv}}_{\theta}$).

Proposition 5.3.10 We have a commutative diagram

$$\begin{array}{ccc}
 \text{Ch}_G^{u,gr} & \xrightarrow{l_N^*} & \text{Shv}_{gr,c}(\mathcal{N}/G) \\
 \searrow \text{wt} & & \nearrow \simeq \\
 & \text{Ch}^b(\text{h Ch}_G^{u,gr,\heartsuit_w}) &
 \end{array}$$

where all functors are weight-exact.

Proof This follows from Corollary 5.3.8 and the identification of wt with $\widetilde{\text{inv}}_\theta$ in Proposition 5.3.4. \square

In [59], the HOMFLY-PT homology groups are given via the chromatographic complex construction, which is a triangulated (as opposed to DG) version of the weight complex functor; see also [31, Remark 3.1.12]. In [57], the HOMFLY-PT homology construction is then proved to factor through the nilpotent cone. Proposition 5.3.10 above puts these two results into context by identifying the weight complex functor itself and the functor l_N^* of restricting to the nilpotent cone, explaining why the nilpotent cone appears. Moreover, the equivalence $\text{h Ch}_G^{u,gr,\heartsuit_w} \simeq \text{Shv}_{gr,c}(\mathcal{N}/G)^{\heartsuit_w}$ formulates precisely the sense in which $\text{Shv}_{gr,c}(\mathcal{N}/G)$ is the truncated trace of H_n^{gr} , as was speculated in [57].

5.3.11 Identifying the functor $\bar{\Gamma}$ We will now compute $\bar{\Gamma}$ appearing in (5.2.10) more explicitly: we decompose it into a direct sum (according to the a -degrees) of corepresentable functors. In other words, the α^{th} piece captures the part of Γ that is α away from being pure.

For each integer α , consider

$$\tilde{\varphi}_\alpha : \mathbb{Z} \rightarrow \mathbb{Z}^2$$

such that

$$\tilde{\varphi}_\alpha(X) := \tau_\alpha + \varphi(X) := (2\alpha, \alpha) + (X, X), \quad X \in \mathbb{Z}.$$

We define the associated functors $\tilde{\iota}_\alpha, \iota_\alpha : \text{Vect}^{gr} \rightarrow \text{Vect}^{gr_{Q'}, gr_{A'}}$ given by $\tilde{\varphi}_{\alpha,*}$ and φ_* , both followed by a cohomological shear to the left whose amount is given by the X -degree. More precisely, for $(V_X)_{X \in \mathbb{Z}} \in \text{Vect}^{gr}$ (note that T' is the cohomological degree of $\text{Vect}^{gr_{Q'}, gr_{A'}}$, see Section 5.2.4),

$$\begin{aligned}
 \tilde{\iota}_\alpha((V_X)_{X \in \mathbb{Z}})_{Q', A'} &= \begin{cases} V_X \{X\}_{T'} & \text{if } (Q', A') = \tilde{\varphi}_\alpha(X), \\ 0 & \text{otherwise,} \end{cases} \\
 \iota_\alpha((V_X)_{X \in \mathbb{Z}})_{Q', A'} &= \begin{cases} V_X \{X\}_{T'} & \text{if } (Q', A') = \varphi(X), \\ 0 & \text{otherwise.} \end{cases}
 \end{aligned}$$

Note that the shear to the left by X is there to account for the fact that something in graded degree X and cohomological degree X in Vect^{gr} is sent to something of cohomological degree 0 when we apply the construction in Section 5.1.6 to turn a cohomological grading to a formal grading.

Proposition 5.3.12 The functor $\bar{\Gamma} : \bar{\mathcal{A}}_n\text{-Mod}^{gr, perf} \rightarrow \text{Vect}^{gr_{Q'}, gr_{A'}}$ breaks up into a finite direct sum

$$\bar{\Gamma} \simeq \bigoplus_\alpha \tilde{\iota}_\alpha(\mathcal{H}om_{\bar{\mathcal{A}}_n\text{-Mod}^{gr, perf}}^{gr}(\wedge^\alpha P \otimes \bar{\mathbb{Q}}_\ell[X], -)),$$

where $P \in \text{Rep } S_n$ is the permutation representation.

Proof By Lemma 5.1.2, $\bar{\Gamma}$ is determined by its restriction to the source’s weight heart,

$$\bar{\Gamma}^{\heartsuit_w} : \bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w} \rightarrow \text{Vect}^{\heartsuit_t, \text{gr}_{Q'}, \text{gr}_{A'}} \rightarrow \text{Vect}^{\text{gr}_{Q'}, \text{gr}_{A'}}.$$

This functor, in turn, is determined by its action on $\bar{\mathcal{A}}_n$. Under the equivalence of categories in Theorem 3.3.4, the constant sheaf corresponds to $\bar{\mathbb{Q}}_\ell[x, \theta]$, from which we obtain the natural equivalence

$$\begin{aligned} \bar{\Gamma}^{\heartsuit_w}(\bar{\mathcal{A}}_n) &\simeq \Gamma(\mathcal{A}_n) \simeq \bigoplus_{\alpha} \tilde{t}_{\alpha}(\wedge^{\alpha} P \otimes \bar{\mathbb{Q}}_\ell[x]) \\ &\simeq \bigoplus_{\alpha} \tilde{t}_{\alpha}(\mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}}^{\text{gr}}(\wedge^{\alpha} P \otimes \bar{\mathbb{Q}}_\ell[x], \bar{\mathbb{Q}}_\ell[x] \boxtimes \bar{\mathbb{Q}}_\ell[S_n])) \\ &\simeq \bigoplus_{\alpha} \tilde{t}_{\alpha}(\mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}}^{\text{gr}}(\wedge^{\alpha} P \otimes \bar{\mathbb{Q}}_\ell[x], \bar{\mathcal{A}}_n)), \end{aligned}$$

where we have implicitly picked an isomorphism of S_n -representations $P \simeq P^\vee$. Moreover, the factor $(2\alpha, \alpha)$ in $\tilde{\varphi}_\alpha$ is to account for the degrees of θ in \mathcal{A}_n which are not there anymore in $\wedge^\alpha P$: these are precisely the (Q', A') -degrees of homogeneous wedges of θ of order α . Thus we have an equivalence of functors

$$\bar{\Gamma}^{\heartsuit_w} \simeq \bigoplus_{\alpha} \tilde{t}_{\alpha}(\mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w}}^{\text{gr}}(\wedge^{\alpha} P \otimes \bar{\mathbb{Q}}_\ell[x], -)) : \bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf},\heartsuit_w} \rightarrow \text{Vect}^{\heartsuit_t, \text{gr}_{Q'}, \text{gr}_{A'}}.$$

But now, the second functor has an extension to the whole of $\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}}$, which is given by

$$\bigoplus_{\alpha} \tilde{t}_{\alpha}(\mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}}}^{\text{gr}}(\wedge^{\alpha} P \otimes \bar{\mathbb{Q}}_\ell[x], -)) : \bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}} \rightarrow \text{Vect}^{\text{gr}_{Q'}, \text{gr}_{A'}}.$$

The proof thus concludes by Lemma 5.1.2. □

Definition 5.3.13 We define

$$\begin{aligned} \bar{\Gamma}_\alpha &:= \mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}}}^{\text{gr}}(\bar{\mathbb{Q}}_\ell[x] \otimes \wedge^{\alpha} P, -) : \bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}} \rightarrow \text{Vect}^{\text{gr}}, \\ \Gamma_\alpha &:= \bar{\Gamma}_\alpha \circ \text{wt} \simeq \bar{\Gamma}_\alpha \circ \widetilde{\text{inv}}_\theta : \mathcal{A}_n\text{-Mod}^{\text{gr,perf}} \rightarrow \text{Vect}^{\text{gr}}, \\ \widetilde{\text{HH}}_\alpha &:= \Gamma_\alpha \circ \text{tr} : \text{H}_n^{\text{gr}} \rightarrow \text{Vect}^{\text{gr}}. \end{aligned}$$

Here wt is the weight complex functor of $\text{Ch}_G^{\text{u,gr}} \simeq \mathcal{A}_n\text{-Mod}^{\text{gr,perf}}$, which is identified with $\widetilde{\text{inv}}_\theta$ by Proposition 5.3.4.

As a direct consequence of the direct sum decomposition of $\bar{\Gamma}$ in Proposition 5.3.12, we get the following direct sum decompositions of Γ and $\widetilde{\text{HH}}$:

Corollary 5.3.14 We have the following equivalences of functors:

$$\bar{\Gamma} \simeq \bigoplus_{\alpha} \tilde{t}_{\alpha} \circ \bar{\Gamma}_\alpha, \quad \Gamma \simeq \bigoplus_{\alpha} \tilde{t}_{\alpha} \circ \Gamma_\alpha \quad \text{and} \quad \widetilde{\text{HH}} \simeq \bigoplus_{\alpha} \tilde{t}_{\alpha} \circ \widetilde{\text{HH}}_\alpha.$$

5.3.15 Gradings By construction, for any integer α , \tilde{t}_α sends a graded vector space of graded degree X and cohomological degree C to an object of (Q', A', T') -degree $(X + 2\alpha, X + \alpha, C - X)$. Written multiplicatively (see also Section 5.2.4), this object has degree

$$Q'^{X+2\alpha} A'^{X+\alpha} T'^{C-X} = Q'^{2\alpha} A'^{\alpha} Q'^X A'^X T'^{C-X} = a^\alpha Q^X T^{C-X}.$$

In other words, the target of $\tilde{\iota}_\alpha$ has a -degree α . Consequently, $\tilde{\iota}_\alpha \circ \bar{\Gamma}_\alpha$, $\tilde{\iota}_\alpha \circ \Gamma_\alpha$, and $\tilde{\iota}_\alpha \circ \widetilde{\text{HH}}_\alpha$ all land in the a -degree α part. This justifies the notation given in Definition 5.3.13: $\bar{\Gamma}_\alpha$, Γ_α , and $\widetilde{\text{HH}}_\alpha$ are the a -degree α parts of $\bar{\Gamma}$, Γ , and $\widetilde{\text{HH}}$, respectively.

Observe that the factor $(2\alpha, \alpha)$ in $\tilde{\varphi}_\alpha$ is responsible precisely for the a -degree. In what follows, we will treat each a -degree separately, and for each a -degree we will ignore the a -factor and consider only the (Q, T) -degrees. Proposition 5.3.12 thus has the following reformulation:

Corollary 5.3.16 *The a -degree α part $\bar{\Gamma}_\alpha$ of $\bar{\Gamma}$ is given by $\mathcal{H}\text{Com}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}}}^{\text{gr}}(\bigwedge^\alpha P \otimes \bar{\mathbb{Q}}_\ell[x], -) \in \text{Vect}^{\text{gr}}$, where the degree (X, C) corresponds to $(X, C - X)$ in the (Q, T) -grading. Here X is the graded degree and C is the cohomological degree in Vect^{gr} .*

5.3.17 Identifying the functor Γ Our goal is now to show an analog of Corollary 5.3.16 for the functor Γ appearing in (5.2.10). We already saw in Corollary 5.3.14 above that Γ is decomposed into a direct sum of $\iota_\alpha \circ \Gamma_\alpha$. We will now show that the Γ_α are corepresentable, except that now, the corepresenting objects live in $\mathcal{A}_n\text{-Mod}^{\text{gr,coh}} \subseteq \mathcal{A}_n\text{-Mod}^{\text{gr}} \simeq \text{Ind}(\mathcal{A}_n\text{-Mod}^{\text{gr,perf}})$. In other words, they are corepresented by ind-objects (which happen to be coherent)!

Recall the functor

$$\widetilde{\text{triv}}_\theta := \text{Sym}^n V[-1] \otimes \text{triv}_\theta : \bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}} \rightarrow \mathcal{A}_n\text{-Mod}^{\text{gr,coh}},$$

where triv_θ sends a perfect $\bar{\mathcal{A}}_n$ complex to a coherent object where θ act by 0. This functor has a partially defined right adjoint

$$\widetilde{\text{inv}}_\theta := \text{Sym}^n V^\vee[1] \otimes \text{inv}_\theta : \mathcal{A}_n\text{-Mod}^{\text{gr,perf}} \rightarrow \bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}},$$

where $\mathcal{A}_n\text{-Mod}^{\text{gr,perf}}$ is naturally a full subcategory of $\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}$. This pair of partial adjoints comes from the standard adjunction pair $\text{triv}_\theta \dashv \text{inv}_\theta$ between the Ind-completed categories $\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr}} = \text{Ind}(\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}})$ and $\text{Ind}(\mathcal{A}_n\text{-Mod}^{\text{gr,coh}})$.

Remark 5.3.18 (abuse of notation) In what follows, we will abuse notation and implicitly view objects of $\mathcal{A}_n\text{-Mod}^{\text{gr,perf}}$ as living in $\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}$ whenever necessary. For example, when $c \in \mathcal{A}_n\text{-Mod}^{\text{gr,coh}}$ and $p \in \mathcal{A}_n\text{-Mod}^{\text{gr,perf}}$, we write

$$\mathcal{H}\text{Com}_{\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(c, p) := \mathcal{H}\text{Com}_{\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(c, \iota_{\text{perf} \hookrightarrow \text{coh}} p),$$

where $\iota_{\text{perf} \hookrightarrow \text{coh}} : \mathcal{A}_n\text{-Mod}^{\text{gr,perf}} \hookrightarrow \mathcal{A}_n\text{-Mod}^{\text{gr,coh}}$ is the natural inclusion.

In view of Corollaries 5.3.14 and 5.3.16, we have the following analog of Proposition 5.3.12 for Γ :

Proposition 5.3.19 *The a -degree α part Γ_α of Γ is given by*

$$\Gamma_\alpha \simeq \mathcal{H}\text{Com}_{\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(\bigwedge^\alpha P \otimes \text{Sym}^n V[-1] \otimes \bar{\mathbb{Q}}_\ell[x], -) : \mathcal{A}_n\text{-Mod}^{\text{gr,perf}} \rightarrow \text{Vect}^{\text{gr}},$$

where the degree (X, C) in Vect^{gr} corresponds to $(X, C - X)$ in the (Q, T) -grading. Here X is the graded degree and C is the cohomological degree in Vect^{gr} .

Proof Using the adjunction $\widetilde{\text{triv}}_\theta \dashv \widetilde{\text{inv}}_\theta$ discussed above, for any $M \in \mathcal{A}_n\text{-Mod}^{\text{gr,perf}}$, we have

$$\begin{aligned} \Gamma_\alpha(M) &\simeq \mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}}}^{\text{gr}}(\wedge^\alpha P \otimes \bar{\mathbb{Q}}_\ell[x], \widetilde{\text{inv}}_\theta M) \simeq \mathcal{H}\text{om}_{\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(\wedge^\alpha P \otimes \widetilde{\text{triv}}_\theta(\bar{\mathbb{Q}}_\ell[x]), M) \\ &\simeq \mathcal{H}\text{om}_{\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(\wedge^\alpha P \otimes \text{Sym}^n V[-1] \otimes \bar{\mathbb{Q}}_\ell[x], M). \end{aligned} \quad \square$$

We are now ready to state the main theorem of this subsection, which is a direct consequence of the results proved so far.

Theorem 5.3.20 *The a -degree α part $\widetilde{\text{HH}}_\alpha$ of $\widetilde{\text{HH}}$ is given by*

$$\widetilde{\text{HH}}_\alpha \simeq \mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}}}^{\text{gr}}(\wedge^\alpha P \otimes \bar{\mathbb{Q}}_\ell[x], \widetilde{\text{inv}}_\theta \text{tr}(-)) \simeq \mathcal{H}\text{om}_{\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(\wedge^\alpha P \otimes \text{Sym}^n V[-1] \otimes \bar{\mathbb{Q}}_\ell[x], \text{tr}(-))$$

as functors $\text{H}_n^{\text{gr}} \rightarrow \text{Vect}^{\text{gr}}$, where the degree (X, C) corresponds to $(X, C - X)$ in the (Q, T) -grading, where X is the graded degree and C is the cohomological degree in Vect^{gr} .

Proof This follows directly from the computation of Γ_α in Proposition 5.3.19 and the fact that $\widetilde{\text{HH}}_\alpha = \Gamma_\alpha \circ \text{tr}$ (see Definition 5.3.13 and Corollary 5.3.14). □

Example 5.3.21 The a -degree α part of the HOMFLY-PT link homology of the unlink of n components is given by

$$\mathcal{H}\text{om}_{\bar{\mathcal{A}}_n\text{-Mod}^{\text{gr,perf}}}^{\text{gr}}(\wedge^\alpha P \otimes \bar{\mathbb{Q}}_\ell[x], \widetilde{\text{inv}}_\theta \text{tr}(1)) \simeq \mathcal{H}\text{om}_{\mathcal{A}_n\text{-Mod}^{\text{gr,coh}}}^{\text{gr}}(\wedge^\alpha P \otimes \text{Sym}^n V[-1] \otimes \bar{\mathbb{Q}}_\ell[x], \text{tr}(1)),$$

where 1 is the monoidal unit of H_n^{gr} and where the Q and T gradings are given as in Theorem 5.3.20. Now, recall that $\widetilde{\text{inv}}_\theta \text{tr}(1) \simeq \bar{\mathcal{A}}_n$. The computation in the proof of Proposition 5.3.12 then shows that the left-hand side of the equivalence above is simply $\wedge^\alpha P \otimes \bar{\mathbb{Q}}_\ell[x]$. The associated three-variable polynomial is thus

$$\left(\sum_{\alpha=0}^n \binom{n}{\alpha} a^\alpha \right) \left(\sum_{l=0}^{\infty} Q^{2l} \right)^n = \left(\frac{1+a}{1-Q^2} \right)^n = \left(\frac{1+a}{1-q} \right)^n.$$

5.4 HOMFLY-PT homology via $\text{Hilb}(\mathbb{C}^2)$

We will now finally establish the realization of HOMFLY-PT homology via Hilbert schemes of points on \mathbb{C}^2 . The main point is to transport Theorem 5.3.20 to the Hilbert scheme side via Koszul duality and the result of Krug as encapsulated in Section 4.

5.4.1 Koszul duality We will now reformulate Theorem 5.3.20 using Koszul duality, which is an equivalence of categories (4.1.7). We start with the following matching of objects:

Lemma 5.4.2 *Under the equivalence of categories (4.1.7), we have the following matching of objects:*

$$\widetilde{\text{inv}}_\theta^{\text{enh}}(\wedge^\alpha P \otimes \text{Sym}^n V[-1] \otimes \bar{\mathbb{Q}}_\ell[x]) \simeq \wedge^\alpha P \otimes \bar{\mathbb{Q}}_\ell[x, \underline{y}] \simeq \wedge^\alpha P \otimes \text{Sym}(V_x[-2] \oplus V_y^\vee) \in \mathcal{B}_n\text{-Mod}^{\text{gr,perf}}.$$

Proof Directly from the construction (see also Section 4.1.5), we have

$$\text{inv}_\theta^{\text{enh}}(\overline{\mathbb{Q}}_\ell[x]) \simeq \overline{\mathbb{Q}}_\ell[x, \underline{y}].$$

Thus,

$$\widetilde{\text{inv}}_\theta^{\text{enh}}(\overline{\mathbb{Q}}_\ell[x]) := \text{Sym}^n V^\vee[1] \otimes \text{inv}_\theta^{\text{enh}}(\overline{\mathbb{Q}}_\ell[x]) \simeq \text{Sym}^n V^\vee[1] \otimes \overline{\mathbb{Q}}_\ell[x, \underline{y}].$$

Since all functors involved are linear over $\overline{\mathbb{Q}}_\ell[\mathbb{S}_n]\text{-Mod}^{\text{gr,perf}}$, we get

$$\widetilde{\text{inv}}_\theta^{\text{enh}}(\wedge^\alpha P \otimes \text{Sym}^n V[-1] \otimes \overline{\mathbb{Q}}_\ell[x]) \simeq \wedge^\alpha P \otimes \overline{\mathbb{Q}}_\ell[x, \underline{y}] \in \mathcal{B}_n\text{-Mod}^{\text{gr,perf}}. \quad \square$$

Corollary 5.4.3 *The a -degree α part $\widetilde{\text{HH}}_\alpha$ of $\widetilde{\text{HH}}$ is given by the cohomology of*

$$\widetilde{\text{HH}}_\alpha(R_\beta) \simeq \mathcal{H}\text{om}_{\mathcal{B}_n\text{-Mod}^{\text{gr,perf}}}^{\text{gr}}(\wedge^\alpha P \otimes \overline{\mathbb{Q}}_\ell[x, \underline{y}], \widetilde{\text{inv}}_\theta^{\text{enh}}(\text{tr}(R_\beta))) \in \text{Vect}^{\text{gr}},$$

where the degree (X, C) corresponds to $(X, C - X)$ in the (Q, T) -grading. Here X is the graded degree and C is the cohomological degree in Vect^{gr} .

Proof This is simply Theorem 5.3.20 transported to $\mathcal{B}_n\text{-Mod}^{\text{gr,perf}}$ using the equivalence of categories (4.1.7) and the matching of objects done in Lemma 5.4.2. □

5.4.4 HOMFLY-PT homology via 2-periodized Hilbert schemes We will now relate HOMFLY-PT homology and 2-periodized Hilbert schemes of points. Since there are now multiple gradings, we will include them in the notation. Recall that on the $\mathcal{B}_n\text{-Mod}^{\text{gr,perf}}$ side, there are two sets of gradings: (X, Y) and (\tilde{X}, \tilde{Y}) . Here X is the original grading coming from graded sheaves whereas Y is the extra grading (to be “canceled out” by 2-periodization). Moreover, the (\tilde{X}, \tilde{Y}) -grading is related to the (X, Y) -grading via a simple change of coordinates; see Sections 4.4.1 and 4.4.6.

Definition 5.4.5 We define

$$\widetilde{\text{HH}}_\alpha^{2\text{-per}} : \text{H}_n^{\text{gr}} \rightarrow \overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}]\text{-Mod}^{\text{gr}_X, \text{gr}_Y} =: \text{Vect}^{\text{gr}_X, \text{gr}_Y, 2\text{-per}} \simeq \text{Vect}^{\text{gr}_{\tilde{X}}, \text{gr}_{\tilde{Y}}, 2\text{-per}}$$

by $\overline{\mathbb{Q}}_\ell[\beta, \beta^{-1}] \otimes \widetilde{\text{HH}}_\alpha$, where β lives in cohomological degree -2 and (X, Y) -degree $(0, 1)$, and where $\widetilde{\text{HH}}_\alpha$ is viewed as a complex with two formal gradings by setting the Y -degree to 0 while keeping the X -degree the same. See also Section 4.4.8 for the various definitions and grading conventions for 2-periodization.

We note that $\widetilde{\text{HH}}_\alpha^{2\text{-per}}$ and $\widetilde{\text{HH}}_\alpha$ contain the same amount of information. The introduction of 2-periodization allows us to introduce a formal grading Y that is twice the cohomological degree. For example, the $Y = l$ part of $\text{H}^0(\widetilde{\text{HH}}_\alpha^{2\text{-per}}(-))$ is precisely $\text{H}^{2l}(\widetilde{\text{HH}}_\alpha(-))$. In this precise sense, we have $Y = C^2$ (in multiplicative notation) if we use C to denote the cohomological degree. All the cohomology groups of $\widetilde{\text{HH}}_\alpha$ can then be recovered by looking only at H^0 and H^1 of $\widetilde{\text{HH}}_\alpha^{2\text{-per}}$.

Theorem 5.4.6 [28, Conjecture 7.2(a)] *We have the equivalence of functors from $H_n^{\text{gr}} \rightarrow \text{Vect}^{\text{gr } \tilde{X}, \text{gr } \tilde{Y}, 2\text{-per}}$*

$$\widetilde{\text{HH}}_{\alpha}^{2\text{-per}} \simeq \mathcal{H}\text{om}_{\text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)^{2\text{-per}}}^{\text{gr } \tilde{X}, \text{gr } \tilde{Y}} \left((\wedge^{\alpha} \mathcal{T})^{2\text{-per}}, \Psi^{2\text{-per}}(\widetilde{\text{inv}}_{\theta}^{\text{enh}}(\text{tr}(-))^{2\text{-per}}) \right),$$

where \mathcal{T} is the tautological bundle on Hilb_n . The (\tilde{X}, \tilde{Y}) -grading matches with the (q, t) -grading as follows: $\tilde{X} = q$ and $\tilde{Y} = \sqrt{t}$. Moreover, the cohomological degree C on the right corresponds to \sqrt{qt} . Note also that for any $R \in H_n^{\text{gr}}$, $\Psi^{2\text{-per}}(\widetilde{\text{inv}}_{\theta}^{\text{enh}}(\text{tr}(R))^{2\text{-per}}) \in \text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{x}}}^{2\text{-per}}$, ie it is supported along the x -axis.

In particular, for any $R_{\beta} \in H_n^{\text{gr}}$ associated to a braid β , there exists a natural

$$\mathcal{F}_{\beta} := \Psi^{2\text{-per}}(\widetilde{\text{inv}}_{\theta}^{\text{enh}}(\text{tr}(R_{\beta}))^{2\text{-per}}) \in \text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{x}}}^{2\text{-per}}$$

such that the a -degree α component of the HOMFLY-PT homology of β is given by

$$\mathcal{H}\text{om}_{\text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)^{2\text{-per}}}^{\text{gr } \tilde{X}, \text{gr } \tilde{Y}} \left(\wedge^{\alpha} \mathcal{T}^{2\text{-per}}, \mathcal{F}_{\beta} \right).$$

Proof For any $R \in H_n^{\text{gr}}$, we have the natural equivalences

$$\begin{aligned} \widetilde{\text{HH}}_{\alpha}^{2\text{-per}}(R) &:= \overline{\mathbb{Q}}_{\ell}[\beta, \beta^{-1}] \otimes \widetilde{\text{HH}}_{\alpha}(R) \\ &\simeq \overline{\mathbb{Q}}_{\ell}[\beta, \beta^{-1}] \otimes \mathcal{H}\text{om}_{\mathcal{B}_n\text{-Mod}^{\text{gr}, \text{perf}}}^{\text{gr } X} \left(\wedge^{\alpha} P \otimes \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{y}], \widetilde{\text{inv}}_{\theta}^{\text{enh}}(\text{tr}(R)) \right) \\ &\simeq \mathcal{H}\text{om}_{\Rightarrow (\mathcal{B}_n\text{-Mod}^{\text{gr}, \text{perf}})}^{\text{gr } X, \text{gr } Y} \left(\wedge^{\alpha} P \otimes \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{y}], \widetilde{\text{inv}}_{\theta}^{\text{enh}}(\text{tr}(R)) \right) \\ &\simeq \mathcal{H}\text{om}_{\mathcal{B}_n^{\text{gr}}\text{-Mod}^{\text{gr } \tilde{X}, \text{gr } \tilde{Y}, 2\text{-per}, \text{perf}}}^{\text{gr } \tilde{X}, \text{gr } \tilde{Y}} \left((\wedge^{\alpha} T)^{2\text{-per}}, (\widetilde{\text{inv}}_{\theta}^{\text{enh}}(\text{tr}(R)))^{2\text{-per}} \right) \\ &\simeq \mathcal{H}\text{om}_{\text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)^{2\text{-per}}}^{\text{gr } \tilde{X}, \text{gr } \tilde{Y}} \left((\wedge^{\alpha} \mathcal{T})^{2\text{-per}}, \Psi^{2\text{-per}}(\widetilde{\text{inv}}_{\theta}^{\text{enh}}(\text{tr}(R))^{2\text{-per}}) \right), \end{aligned}$$

where the first follows from Corollary 5.4.3, the second from Lemma 4.4.2, the third from Corollary 4.4.5, and the last from Theorem 4.4.15 and Corollary 4.5.6. Note that as before, we implicitly passed from (X, Y) -grading to (\tilde{X}, \tilde{Y}) -grading.

For the matching of gradings, observe that $\tilde{X}^k \tilde{Y}^l = X^{2k} Y^k X^{-l} = X^{2k-l} C^{2k}$ corresponds to

$$Q^{2k-l} T^{2k-2k+l} = Q^{2k-l} T^l = Q^{2k} Q^{-l} T^l = q^k (\sqrt{t})^l$$

in the (q, t) -degree. In other words, (\tilde{X}, \tilde{Y}) -grading corresponds to (q, \sqrt{t}) -grading precisely. Moreover, cohomological degree 1, which is C , corresponds to

$$T = QQ^{-1}T = \sqrt{qt}. \quad \square$$

Remark 5.4.7 When decategorified, a term in cohomological degree 1 of $\widetilde{\text{HH}}_{\alpha}^{2\text{-per}}$ contributes a factor of $-\sqrt{qt}$. The sign is there because of the odd cohomological degree.

Example 5.4.8 We now rephrase Example 5.3.21 in terms of 2-periodic complexes. By Definition 5.4.5, the a -degree α part of the 2-periodized HOMFLY-PT homology is given by

$$\widetilde{\text{HH}}_{\alpha}^{2\text{-per}} = \overline{\mathbb{Q}}_{\ell}[\beta, \beta^{-1}] \otimes \wedge^{\alpha} P \otimes \overline{\mathbb{Q}}_{\ell}[\underline{x}].$$

Since there's no odd cohomological degree, the associated polynomial is extracted from

$$H^0(\widetilde{\text{HH}}_{\alpha}^{2\text{-per}}) \simeq \wedge^{\alpha} P \otimes \overline{\mathbb{Q}}_{\ell}[\beta \underline{x}],$$

where $\beta \underline{x}$ has cohomological degree 0 and (X, Y) -degree $(2, 1)$, or equivalently (\tilde{X}, \tilde{Y}) -degree $(1, 0)$, which is the same as (q, t) -degree $(1, 0)$. The associated polynomial is thus

$$\left(\sum_{\alpha=0}^n \binom{n}{\alpha} a^{\alpha} \right) \left(\sum_{l=0}^{\infty} q^l \right)^n = \left(\frac{1+a}{1-q} \right)^n.$$

5.5 Matching the actions of $\overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n}$

In [28, Section 5.1], an action of the variables \underline{x} (or equivalently, $\tilde{\underline{x}}$, depending on whether we are working with the sheared or the 2-periodized version) on HHH is constructed. In particular, symmetric functions on \underline{x} (or equivalently, $\tilde{\underline{x}}$) also act. On the other hand, by construction, the geometric realization of $\widetilde{\text{HH}}^{2\text{-per}}$ via Hilbert schemes in Theorem 5.4.6 automatically equips it with an action of $\overline{\mathbb{Q}}_{\ell}[\tilde{\underline{x}}, \tilde{\underline{y}}]^{S_n} = H^0(\text{Hilb}_n, \mathcal{O})$. In particular, it admits an action of $\overline{\mathbb{Q}}_{\ell}[\tilde{\underline{x}}]^{S_n} \subset \overline{\mathbb{Q}}_{\ell}[\tilde{\underline{x}}, \tilde{\underline{y}}]^{S_n}$.

This subsection is dedicated to showing the following result, whose proof will conclude in Section 5.5.16:

Theorem 5.5.1 (part of [28, Conjecture 7.2(b)]) *The two actions above coincide.*

The proof is a simple matter of chasing through the construction. To keep the main ideas evident, we will elide the difference between sheared and 2-periodic versions. For example, we will pass seamlessly between $\underline{x}, \underline{y}$ and $\tilde{\underline{x}}, \tilde{\underline{y}}$.

The argument can be most conceptually and conveniently phrased in terms of module categories. For example, instead of saying that a (graded) commutative ring R acts on objects of a $(\text{Vect}^{\text{gr}}\text{-module})$ category \mathcal{C} , we formulate everything in terms of the symmetric monoidal category $\mathcal{O} = R\text{-Mod}$ (or $\mathcal{O} = R\text{-Mod}^{\text{gr}}$) acting on \mathcal{C} itself. Then the core of the argument revolves around observing that all functors involved are linear over \mathcal{O} .

5.5.2 Module categories and actions of a commutative ring

We start with some generalities regarding module categories and actions of commutative rings.

Let \mathcal{O} and \mathcal{O}' be compactly generated rigid symmetric monoidal categories equipped with a symmetric monoidal functor $F: \mathcal{O} \rightarrow \mathcal{O}'$ and a right adjoint G , which is necessarily right-lax symmetric monoidal. The symmetric monoidal functor F equips \mathcal{O}' with the structure of an \mathcal{O} -module category. F can thus

be written as $F = - \otimes 1_{\mathcal{O}'}$. The functor G can thus be written as $G \simeq \mathcal{H}\text{om}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'}, -)$, where the superscript \mathcal{O} in $\mathcal{H}\text{om}$ denotes the \mathcal{O} -enriched Hom. See [31, Appendix A.2.6] for a quick review on enriched Hom-spaces.

Lemma 5.5.3 *Let \mathcal{O} and \mathcal{O}' be as above, and \mathcal{C} an \mathcal{O}' -module category. Then, for any $c_1, c_2 \in \mathcal{C}$, $\mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2)$ has a natural $\mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'})$ -module structure. Moreover, this module structure is compatible with*

- (i) *\mathcal{O}' -linear functors: if $F: \mathcal{C} \rightarrow \mathcal{D}$ is a morphism of \mathcal{O}' -module categories, then $\mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2) \rightarrow \mathcal{H}\text{om}_{\mathcal{D}}^{\mathcal{O}}(F(c_1), F(c_2))$ is a morphism of modules over $\mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'})$;*
- (ii) *the restriction of structure via a symmetric monoidal functor $F': \mathcal{O}' \rightarrow \mathcal{O}''$: if the \mathcal{O}' -module structure on \mathcal{C} comes from restricting an \mathcal{O}'' -module structure, then the $\mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'})$ -module structure on $\mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2)$ is obtained by restriction of scalars along the commutative algebra map $\mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'}) \rightarrow \mathcal{E}\text{nd}_{\mathcal{O}''}^{\mathcal{O}}(1_{\mathcal{O}''})$.*

Proof We have $\mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2) \in \mathcal{O}'$, and thus it naturally has a module structure over $1_{\mathcal{O}'}$. Since G is right-lax symmetric monoidal, $G \mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2)$ has a natural module structure over $G(1_{\mathcal{O}'}) \simeq \mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'})$. But, by adjunction, $G \mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2) \simeq \mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2)$ and we are done.

The second part regarding various compatibilities is an easy diagram chase. □

Lemma 5.5.4 *Let \mathcal{O} , \mathcal{O}' , and \mathcal{C} be as above. Then for any c there is a natural map of algebra objects in \mathcal{O}*

$$\varphi_c: \mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'}) \rightarrow \mathcal{E}\text{nd}_{\mathcal{C}}^{\mathcal{O}}(c).$$

Moreover, this map is compatible with \mathcal{O}' -linear functors and with restriction of structure via a symmetric monoidal functor $\mathcal{O}' \rightarrow \mathcal{O}''$.

Proof $\mathcal{E}\text{nd}_{\mathcal{C}}^{\mathcal{O}}(c)$ is an algebra object in \mathcal{O}' , and hence it receives a natural algebra map from $1_{\mathcal{O}'}$ since $1_{\mathcal{O}'}$ is the initial algebra object. Applying G to this map, we obtain the desired map of algebras in \mathcal{O} .

The second part follows from Lemma 5.5.3. □

Remark 5.5.5 In the case where $\mathcal{O} = \text{Vect}$ (resp. $\mathcal{O} = \text{Vect}^{\text{gr}}$), Lemma 5.5.4 states that any object $c \in \mathcal{C}$ admits a natural action (resp. a graded action) of $\mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'})$ (resp. $\mathcal{E}\text{nd}_{\mathcal{O}'}^{\text{gr}}(1_{\mathcal{O}'})$).

Corollary 5.5.6 *Let \mathcal{O} , \mathcal{O}' , and \mathcal{C} be as above. Then for any $c_1, c_2 \in \mathcal{C}$ the action of $\mathcal{E}\text{nd}_{\mathcal{O}'}^{\mathcal{O}}(1_{\mathcal{O}'})$ on $\mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2)$ factors through $\mathcal{E}\text{nd}_{\mathcal{C}}^{\mathcal{O}}(c_1)^{\text{rev}}$ and $\mathcal{E}\text{nd}_{\mathcal{C}}^{\mathcal{O}}(c_2)$. Here the superscript *rev* denotes the reverse multiplication.*

Proof Since $1_{\mathcal{O}'}$ is the initial algebra object in \mathcal{O}' , the action of $1_{\mathcal{O}'}$ on $\mathcal{H}\text{om}_{\mathcal{C}}^{\mathcal{O}}(c_1, c_2)$ factors through the natural actions of $\mathcal{E}\text{nd}_{\mathcal{C}}^{\mathcal{O}}(c_1)^{\text{rev}}$ and $\mathcal{E}\text{nd}_{\mathcal{C}}^{\mathcal{O}}(c_2)$. Applying G , we obtain the desired conclusion. □

In what follows, we will apply the statements above to the case where $\mathcal{O}' = R\text{-Mod}^{\text{gr}}$ for some commutative ring R and $\mathcal{O} = \text{Vect}^{\text{gr}}$. Moreover, F is the functor of taking free R -modules and G is the forgetful functor.

Remark 5.5.7 In the situations that we are interested in, the functor F and all the actions (in various module category structures) are compact preserving. Note, however, that G does not preserve compactness. Thus, when working with the small full subcategory spanned by compact objects, Lemmas 5.5.3 and 5.5.4 and Corollary 5.5.6 still apply, except that the enriched Homs are not necessarily compact.

5.5.8 The action of $\overline{\mathbb{Q}}_\ell[x]$ on HHH We will now recast the action of $\overline{\mathbb{Q}}_\ell[x]$ on HHH as described in [28, Section 5.1] in the categorical language above. To start, observe that $H_n^{\text{gr}} \simeq \text{Shv}_{\text{gr},c}(BB \times_{BG} BB)$ has an action of $\text{Shv}_{\text{gr},c}(BB \times BB) \simeq \overline{\mathbb{Q}}_\ell[x]^{\otimes 2}\text{-Mod}^{\text{gr,perf}}$ where the two factors act on the left and the right, respectively. The usual action of $\overline{\mathbb{Q}}_\ell[x]^{\otimes 2}$ on any $R \in H_n^{\text{gr}}$ (ie any Soergel bimodule) can thus be recovered from Lemma 5.5.3. Note also that the commutative diagram

$$\begin{array}{ccc} BB \times_{BG} BB & \longrightarrow & BB \\ \downarrow & & \downarrow \\ BB & \longrightarrow & BG \end{array}$$

implies that the induced actions of $\text{Shv}_{\text{gr},c}(BG) \simeq C_{\text{gr}}^*(BG)\text{-Mod}^{\text{gr,perf}} \simeq \overline{\mathbb{Q}}_\ell[x]^{S_n}\text{-Mod}^{\text{gr,perf}}$ via these two actions (on the left and right) agree. In particular, the two actions of $C_{\text{gr}}^*(BG)$ on any $R \in H_n^{\text{gr}}$ obtained from restricting the two actions of $C_{\text{gr}}^*(BB)$ along $C_{\text{gr}}^*(BG) \rightarrow C_{\text{gr}}^*(BB)$ agree; see also [28, Lemma 5.1].

5.5.9 Reducing to the weight heart By construction HHH is obtained by applying Hochschild homology *termwise* (via the weight complex functor) to an element $R \in H_n^{\text{gr}}$. This was formulated geometrically in Section 5.2.1. Since all the categorical actions are weight exact, for the purpose of describing the action of $\overline{\mathbb{Q}}_\ell[x]$ on HHH, we can assume that $R \in H_n^{\text{gr},\heartsuit w}$ and forget about the weight complex functor. The action of $\overline{\mathbb{Q}}_\ell[x] \simeq C_{\text{gr}}^*(BB)$ on $\text{HHH}(R)$ can thus be obtained by applying Lemma 5.5.3 where $\mathcal{O} = \text{Vect}^{\text{gr}}$, $\mathcal{O}' = \text{Shv}_{\text{gr}}(BB)^{\text{ren}}$, $\mathcal{C} = \text{Shv}_{\text{gr}}(G/B)^{\text{ren}}$, $c_1 = \overline{\mathbb{Q}}_{\ell,G/B}$, and $c_2 = p^*R$. Indeed, this is because $\text{HHH}(R)$ is given by (the cohomology of)

$$C_{\text{gr}}^*(BB, \pi_* p^* R) \simeq \mathcal{H}\text{om}_{\text{Shv}_{\text{gr},c}(G/B)}^{\text{gr}}(\overline{\mathbb{Q}}_{\ell,G/B}, p^* R).$$

5.5.10 The action of $\overline{\mathbb{Q}}_\ell[x]^{S_n}$ on HHH We can restrict the action of $\overline{\mathbb{Q}}_\ell[x]$ on HHH to $\overline{\mathbb{Q}}_\ell[x]^{S_n} \subseteq \overline{\mathbb{Q}}_\ell[x]$ and obtain an action of $\overline{\mathbb{Q}}_\ell[x]^{S_n}$, which is the action described in [28]. We will now realize this action more geometrically. We start with the following result:

Lemma 5.5.11 For $R \in H_n^{\text{gr},\heartsuit w}$, the action of $\overline{\mathbb{Q}}_\ell[x]^{S_n} = C_{\text{gr}}^*(BG)$ on

$$\text{HHH}(R) = H^*(\mathcal{H}\text{om}_{\text{Shv}_{\text{gr},c}(G/B)}^{\text{gr}}(\overline{\mathbb{Q}}_\ell, p^* R))$$

is given by the $\text{Shv}_{\text{gr},c}(BG)$ -module structure on $\text{Shv}_{\text{gr},c}(G/B)$ via Lemma 5.5.3, where the former acts via pulling back along $G/B \rightarrow BB \rightarrow BG$.

Proof By definition, the $C_{\text{gr}}^*(BG)$ -action is given by restricting along $C_{\text{gr}}^*(BG) \rightarrow C_{\text{gr}}^*(BB)$. The result thus follows by applying Lemma 5.5.3(ii) to the case where $\mathcal{O} = \text{Vect}^{\text{gr}}$, $\mathcal{O}' = \text{Shv}_{\text{gr}}(BG)^{\text{ren}}$, $\mathcal{O}'' = \text{Shv}_{\text{gr}}(BB)^{\text{ren}}$, $\mathcal{C} = \text{Shv}_{\text{gr}}(G/B)^{\text{ren}}$, $c_1 = \overline{\mathbb{Q}}_{\ell, G/B}$, and c_2 is p^*R . \square

Lemma 5.5.12 For $R \in H_n^{\text{gr}, \heartsuit w}$, the action of $C_{\text{gr}}^*(BG)$ on $\text{HHH}(R)$ described above agrees with the one coming from the $\text{Shv}_{\text{gr}, c}(BG)$ -module structure on $\text{Shv}_{\text{gr}, c}(G/G)$ via Lemma 5.5.3 and the equivalence

$$\text{HHH}(R) = H^*(\mathcal{H}\text{om}_{\text{Shv}_{\text{gr}, c}(G/G)}^{\text{gr}}(\overline{\mathbb{Q}}_{\ell, G/G}, q_*p^*R)) \simeq H^*(\mathcal{H}\text{om}_{\text{Ch}_G^{\text{u,gr}}}^{\text{gr}}(\overline{\mathbb{Q}}_{\ell, G/G}, \text{tr}(R))).$$

Proof Consider the (non-Cartesian) commutative square

$$\begin{array}{ccc} & G/B & \\ p \swarrow & & \searrow q \\ BB \times_{BG} BB & & G/G \\ & \searrow & \swarrow \\ & BG & \end{array}$$

where p and q form the horocycle correspondence. Note that p^* and q^* are $\text{Shv}_{\text{gr}, c}(BG)$ -linear. By rigidity of $\text{Shv}_{\text{gr}, c}(BG)$, p_* and q_* are also $\text{Shv}_{\text{gr}, c}(BG)$ -linear. But now, we have a natural equivalence

$$\begin{aligned} \mathcal{H}\text{om}_{\text{Shv}_{\text{gr}}(G/B)^{\text{ren}}}^{\text{Shv}_{\text{gr}}(BG)^{\text{ren}}}(\overline{\mathbb{Q}}_{\ell}, p^*R) &\simeq \mathcal{H}\text{om}_{\text{Shv}_{\text{gr}}(G/B)^{\text{ren}}}^{\text{Shv}_{\text{gr}}(BG)^{\text{ren}}}(q^*\overline{\mathbb{Q}}_{\ell}, p^*R) \\ &\simeq \mathcal{H}\text{om}_{\text{Shv}_{\text{gr}}(G/G)^{\text{ren}}}^{\text{Shv}_{\text{gr}}(BG)^{\text{ren}}}(\overline{\mathbb{Q}}_{\ell}, q_*p^*R) \in \text{Shv}_{\text{gr}, c}\left(\frac{G}{G}\right) \subseteq \text{Shv}_{\text{gr}}\left(\frac{G}{G}\right)^{\text{ren}}. \end{aligned}$$

The proof concludes by applying $\mathcal{H}\text{om}_{\text{Shv}_{\text{gr}, c}(G/G)}^{\text{gr}}(\overline{\mathbb{Q}}_{\ell}, -)$ as in Lemma 5.5.3 and using Lemma 5.5.11. \square

The action above has a more explicit description in terms of Theorem 3.3.4. We start with the following result:

Lemma 5.5.13 We have an equivalence of $\overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n}\text{-Mod}^{\text{gr}, \text{perf}}$ -module categories (or equivalently, of $\text{Shv}_{\text{gr}, c}(BG)$ -module categories)

$$\text{Hom}_{\text{Ch}_G^{\text{u,gr}}}^{\text{gr}}(\text{Spr}_G^{\text{gr}}, -) : \text{Ch}_G^{\text{u,gr}} \xrightarrow{\simeq} \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[S_n]\text{-Mod}^{\text{gr}, \text{perf}},$$

where the $\text{Shv}_{\text{gr}, c}(BG)$ -module structure on the left-hand side is inherited from the embedding $\text{Ch}_G^{\text{u,gr}} \hookrightarrow \text{Shv}_{\text{gr}, c}(G/G)$ and where the $C_{\text{gr}}^*(BG)\text{-Mod}^{\text{gr}, \text{perf}}$ -module structure on the right-hand side comes from the natural morphism of algebras

$$(5.5.14) \quad \overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n} \simeq C_{\text{gr}}^*(BG) \simeq \mathcal{E}\text{nd}_{\text{Shv}_{\text{gr}, c}(BG)}^{\text{gr}}(\overline{\mathbb{Q}}_{\ell}) \rightarrow \mathcal{E}\text{nd}_{\text{Ch}_G^{\text{u,gr}}}^{\text{gr}}(\text{Spr}_G^{\text{gr}}) \simeq \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[S_n]$$

given by Lemma 5.5.4. Moreover, this morphism of algebra is the obvious one, given by the embedding of symmetric polynomials on \underline{x} to all polynomials.

Proof Instead of running the proof of Theorem 3.3.4 relatively over Vect^{gr} , we could have run it relatively over $\text{Shv}_{\text{gr}}(BG)^{\text{ren}}$. Then we obtain an equivalence of $\text{Shv}_{\text{gr}}(BG)^{\text{ren}}$ -module categories (or equivalently, of $\overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n}$ -module categories)

$$\text{Ch}_G^{\text{u,gr,ren}} \xrightarrow{\cong} \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[\mathbb{S}_n]\text{-Mod}(\text{Shv}_{\text{gr}}(BG)^{\text{ren}}) \simeq \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[\mathbb{S}_n]\text{-Mod}(\overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n}\text{-Mod}),$$

where the category on the right-hand side is formed using the natural morphism of algebras (5.5.14), by Corollary 5.5.6. Note also that the equivalence of categories in Theorem 3.3.4 is obtained by further composing with the forgetful functor

$$\begin{aligned} \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[\mathbb{S}_n]\text{-Mod}(\text{Shv}_{\text{gr}}(BG)^{\text{ren}}) &\simeq \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[\mathbb{S}_n]\text{-Mod}(\overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n}\text{-Mod}) \\ &\xrightarrow{\text{forgetful}} \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[\mathbb{S}_n]\text{-Mod}(\text{Vect}^{\text{gr}}) \simeq \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[\mathbb{S}_n]\text{-Mod}^{\text{gr}}. \end{aligned}$$

Now, observe that the forgetful functor is in fact an equivalence of categories. The proof of the first part thus concludes.

We will compute the morphism (5.5.14) explicitly, as stated in the last sentence of the lemma. Consider the following (non-Cartesian) commutative diagram:

$$\begin{array}{ccc} B/B & \xrightarrow{q} & G/G \\ \downarrow h & & \downarrow h \\ BB & \xrightarrow{q} & BG \end{array}$$

Now, applying Lemma 5.5.4 and the fact that $\text{Spr}_G^{\text{gr}} \simeq q_* \overline{\mathbb{Q}}_{\ell}$, we see that the natural algebra morphism

$$C_{\text{gr}}^*(BG) \simeq \mathcal{E}nd_{\text{Shv}_{\text{gr},c}(BG)}^{\text{gr}}(\overline{\mathbb{Q}}_{\ell}) \rightarrow \mathcal{E}nd_{\text{Shv}_{\text{gr},c}(G/G)}^{\text{gr}}(\text{Spr}_G^{\text{gr}})$$

factors as

$$\begin{array}{ccccccc} C_{\text{gr}}^*(BG) & \longrightarrow & C_{\text{gr}}^*(BB) & \longrightarrow & \mathcal{E}nd_{\text{Shv}_{\text{gr},c}(B/B)}^{\text{gr}}(\overline{\mathbb{Q}}_{\ell}) & \longrightarrow & \mathcal{E}nd_{\text{Shv}_{\text{gr},c}(G/G)}^{\text{gr}}(\text{Spr}_G^{\text{gr}}) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n} & \longrightarrow & \overline{\mathbb{Q}}_{\ell}[\underline{x}] & \longrightarrow & \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] & \longrightarrow & \overline{\mathbb{Q}}_{\ell}[\underline{x}, \underline{\theta}] \boxtimes \overline{\mathbb{Q}}_{\ell}[\mathbb{S}_n] \end{array}$$

where the maps in the bottom row are the obvious ones. The commutativity of the third square is due to Theorem 3.2.1(iii). The commutativity of the other two squares is obvious. The desired statement then follows from the commutativity of the outer square in the diagram above. \square

We obtain the following refinement of Theorem 4.4.15:

Corollary 5.5.15 *We have an equivalence of $\overline{\mathbb{Q}}_{\ell}[\tilde{x}]^{S_n}\text{-Mod}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}, \text{perf}}$ -module categories*

$$\Rightarrow \text{Ch}_G^{\text{u,gr}} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\tilde{y}}}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}} \xrightarrow[\simeq]{\Psi^{2\text{-per}}} \text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2 / \text{Hilb}_{n, \tilde{x}}^{2\text{-per}}),$$

where the $\overline{\mathbb{Q}}_{\ell}[\underline{x}]^{S_n}\text{-Mod}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}, \text{perf}}$ -module category structure on the first two (resp. last) categories (resp. category) is given by Lemma 5.5.13 (resp. by the Hilbert–Chow map and forgetting the \tilde{y} -variables).

Similarly, we have the corresponding statement for the small category variant by passing to the full subcategories of compact objects.

Proof Lemma 5.5.13 above, combined with Koszul duality, implies that we have an equivalence of $\overline{\mathbb{Q}}_\ell[\tilde{x}]^{S_n}\text{-Mod}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}, \text{perf}}$ -module categories

$$\Rightarrow \text{Ch}_G^{\text{u,gr,ren}} \simeq \tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\tilde{y}}}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}},$$

where on the right, the module structure is induced by the algebra map

$$\overline{\mathbb{Q}}_\ell[\tilde{x}]^{S_n} \rightarrow \overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}]^{S_n} \rightarrow \overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}] \otimes \overline{\mathbb{Q}}_\ell[S_n] =: \tilde{\mathcal{B}}_n^{\text{gr}},$$

which is geometrically realized as coming from the natural morphism

$$\mathbb{A}^{2n}/S_n \rightarrow \mathbb{A}^{2n} // S_n \rightarrow \mathbb{A}^n // S_n.$$

The commutative diagram (4.2.2) implies that $\Psi^{2\text{-per}}$ is linear over $\overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}]^{S_n}\text{-Mod}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}}$ (see also [35, Remark 3.11]), where the $\overline{\mathbb{Q}}_\ell[\tilde{x}, \tilde{y}]^{S_n}\text{-Mod}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}}$ -module category structure is given by the Hilbert–Chow morphism $\text{Hilb}_n \rightarrow \mathbb{A}^{2n} // S_n$. In particular, $\Psi^{2\text{-per}}$ is linear over $\overline{\mathbb{Q}}_\ell[\tilde{x}]^{S_n}\text{-Mod}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}}$. Thus we obtain an equivalence of $\overline{\mathbb{Q}}_\ell[\tilde{x}]^{S_n}\text{-Mod}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}}$ -module categories

$$\tilde{\mathcal{B}}_n^{\text{gr}}\text{-Mod}_{\text{nilp}_{\tilde{y}}}^{\text{gr}_{\tilde{x}}, \text{gr}_{\tilde{y}}, 2\text{-per}} \xrightarrow[\simeq]{\Psi^{2\text{-per}}} \text{QCoh}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{x}}}^{2\text{-per}}. \quad \square$$

5.5.16 Completing the proof of Theorem 5.5.1 In Lemma 5.5.12, we show that the action of $\overline{\mathbb{Q}}_\ell[\underline{x}]^{S_n}$ on HHH as given in [28] comes from the $\text{Shv}_{\text{gr}, c}(BG)$ -module category structure on $\text{Ch}_G^{\text{u,gr}}$. In Lemma 5.5.13, we show that this module category structure is induced by an explicit map of algebras, which is then used in Corollary 5.5.15 to show that this module structure is compatible with the one on the Hilbert scheme side, where the structure is induced by the Hilbert–Chow morphism. But now, this structure is responsible for the geometric construction of the action of $\overline{\mathbb{Q}}_\ell[\underline{x}]^{S_n}$ on HHH and the proof concludes. \square

5.6 The support of \mathcal{F}_β on Hilb_n

We will now show that for a braid β , the support of \mathcal{F}_β on Hilb_n , where \mathcal{F}_β is as in Theorem 5.4.6, can be bounded above using the number of components of the link associated to β . This is the content of [28, Conjecture 7.2(b) and (c)]. Unlike what is implied over there, however, we do not deduce this as a consequence of Theorem 5.5.1. The obstacle is that Theorem 5.5.1 is a statement about global sections of a sheaf, whereas the statement we are after is one about the support *before* taking global sections, which is more local in nature. The route we take is thus via the theory of supports as developed in [3; 10], which is inherently local.

As in the previous subsection, to keep the exposition light, we will elide the difference between the 2-periodic and sheared versions of the various categories involved, passing seamlessly between, for instance, $\underline{x}, \underline{y}$ and \tilde{x}, \tilde{y} .

5.6.1 Subspaces of Hilb_n and supports of \mathcal{F}_β We start with some notation regarding various subspaces of Hilb_n which will appear as supports of \mathcal{F}_β .

Definition 5.6.2 (i) For $\beta \in \text{Br}_n$ a braid on n strands, we define $w_\beta \in S_n$ to be the corresponding permutation.

(ii) For each $w \in S_n$, we let $\mathbb{A}_w^{2n} \subseteq \mathbb{A}^{2n}$ denote the closed subscheme of \mathbb{A}^{2n} defined by the relations $x_i = x'_{w(i)}$, where $x_1, \dots, x_n, x'_1, \dots, x'_n$ are the coordinates of the \mathbb{A}^{2n} .²²

(iii) We let $\mathbb{A}_w^n := \mathbb{A}_w^{2n} \times_{\mathbb{A}^{2n}} \mathbb{A}^n$ be the closed subscheme of the diagonal \mathbb{A}^n , where $\mathbb{A}^n \rightarrow \mathbb{A}^{2n}$ is the diagonal map. Alternatively, if we let the x_i denote the coordinates of the diagonal \mathbb{A}^n , then \mathbb{A}_w^n is defined by the relations $x_i = x_{w(i)}$.

(iv) For each $w \in S_n$, we let \bar{w} denote its conjugacy class, $(\mathbb{A}_w^n) // S_n \subseteq \mathbb{A}^n // S_n$ the image of \mathbb{A}_w^n in $\mathbb{A}^n // S_n$, and $\mathbb{A}_{\bar{w}}^n \subseteq \mathbb{A}^n$ the preimage of $(\mathbb{A}_w^n) // S_n$ in \mathbb{A}^n . In other words, $\mathbb{A}_{\bar{w}}^n$ is the orbit of \mathbb{A}_w^n under the S_n action and $\mathbb{A}_{\bar{w}}^n // S_n$ its GIT quotient.

(v) For each $w \in S_n$, we let $\text{Hilb}_{n, \bar{w}} \subseteq \text{Hilb}_n$ denote the preimage of $\mathbb{A}_{\bar{w}}^n // S_n$ under $\text{Hilb}_n \rightarrow \mathbb{A}^{2n} // S_n \rightarrow \mathbb{A}^n // S_n$, where the first map is the Hilbert–Chow map and the second map is the projection onto the first factor.

(vi) For each $w \in S_n$, we let $\text{Hilb}_{n, \tilde{x}, \bar{w}}$ be the closed subscheme of Hilb_n fitting in the Cartesian square

$$\begin{array}{ccccc} \text{Hilb}_{n, \tilde{x}, \bar{w}} & \longrightarrow & \text{Hilb}_{n, \tilde{x}} & \longrightarrow & \text{Hilb}_n \\ \downarrow & & \downarrow & & \downarrow \\ \mathbb{A}_{\bar{w}}^n // S_n & \longrightarrow & \mathbb{A}^n // S_n & \longrightarrow & \mathbb{A}^{2n} // S_n \end{array}$$

where the bottom right horizontal map is induced by $\mathbb{A}^n \simeq \mathbb{A}^n \times \{0\} \rightarrow \mathbb{A}^{2n}$. Equivalently, $\text{Hilb}_{n, \tilde{x}, \bar{w}} = \text{Hilb}_{n, \tilde{x}} \cap \text{Hilb}_{n, \bar{w}}$.

Note that all of these subschemes are stable under the scaling action of \mathbb{G}_m . It thus makes sense to talk about \mathbb{G}_m -equivariant (quasi)coherent sheaves on these spaces.

Remark 5.6.3 Since these schemes are used for the sole purpose of making statements about set-theoretical supports of (quasi)coherent sheaves, any possible nonreduced or derived structure is not relevant to us. For definiteness, we take the underlying reduced classical schemes if any nonreduced or derived structure is present.

With the definitions in place, we are now ready to state the main result of this subsection, whose proof will conclude in Section 5.6.12 below.

Theorem 5.6.4 [28, Conjecture 7.2(b) and (c)] *Let β be a braid on n strands and $w_\beta \in S_n$ the associated permutation. Then $\mathcal{F}_\beta \in \text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{x}}}^{2\text{-per}}$ (see Theorem 5.4.6) is set-theoretically supported on $\text{Hilb}_{n, \tilde{x}, \bar{w}_\beta}$, ie $\mathcal{F}_\beta \in \text{Perf}(\text{Hilb}_n / \mathbb{G}_m^2)_{\text{Hilb}_{n, \tilde{x}, \bar{w}_\beta}}^{2\text{-per}}$.*

²²Note that \mathbb{A}_w^{2n} has dimension n ; the superscript is simply to indicate that it is a closed subscheme of \mathbb{A}^{2n} .

5.6.5 Support in DG-categories As in Section 5.5, the proof is most conceptually explained in terms of module category structures. These structures are naturally in contact with support conditions in the sense of Arinkin and Gaitsgory in [3].

More precisely, let $\mathcal{O} = \mathcal{A}\text{-Mod}^{\text{gr,perf}}$ be a symmetric monoidal category where \mathcal{A} is a graded commutative (DG) algebra and \mathcal{C} an \mathcal{O} -module category. Consider the commutative algebra $A = \bigoplus_n H^{2n}(A)$ which is doubly graded. Equivalently, $\text{Spec } A$ is equipped with an action of \mathbb{G}_m^2 . Then, for any \mathbb{G}_m^2 -invariant closed subset Y of $\text{Spec } A$, [3] defines the full subcategory \mathcal{C}_Y of \mathcal{C} consisting of objects supported along Y .

The algebras \mathcal{A} in our cases are of the forms $C_{\text{gr}}^*(BB \times BB) \simeq \overline{\mathbb{Q}}_\ell[x, x']$, $C_{\text{gr}}^*(BB) \simeq \overline{\mathbb{Q}}_\ell[x]$, and $C_{\text{gr}}^*(BG) \simeq \overline{\mathbb{Q}}_\ell[x]^{S_n}$, which are pure and concentrate in only even degrees. The two \mathbb{G}_m actions thus coincide, and moreover the algebra A is simply a shear of \mathcal{A} . We can therefore ignore one of the gradings and the closed subsets Y we will consider are simply \mathbb{G}_m -invariant closed subsets of \mathbb{A}^{2n} , \mathbb{A}^n , and $\mathbb{A}^n // S_n$. These are precisely the ones that appear in Definition 5.6.2.

Remark 5.6.6 Since these are the only cases that we are interested in, we will assume that all of our commutative graded DG-algebras \mathcal{A} used to study supports are pure and concentrated in even degrees.

Remark 5.6.7 Strictly speaking, [3] works with big categories whereas we formulate everything using small categories. This is not a problem, however, since all of our categories are compactly generated and all actions are compact preserving. Alternatively, we could formulate everything using big categories by working with, for example, $H_n^{\text{gr,ren}}$ and $\text{Shv}_{\text{gr},c}(BG)^{\text{ren}}$, etc. We chose not to do so since, for example, H_n^{gr} is a much more familiar object than $H_n^{\text{gr,ren}}$.

5.6.8 Supports of Rouquier complexes We will now formulate the support conditions satisfied by Rouquier complexes. Recall that for a braid β on n strands, we use $R_\beta \in H_n^{\text{gr}}$ to denote the associated Rouquier complex. As seen above, $H_n^{\text{gr}} := \text{Shv}_{\text{gr},c}(BB \times_{BG} BB)$ has a natural module category structure over $\text{Shv}_{\text{gr},c}(BB \times BB) \simeq \overline{\mathbb{Q}}_\ell[x, x']\text{-Mod}^{\text{gr,perf}}$. By [3, Section 3.5 and E.3.2], it makes sense to talk about the support of any element in H_n^{gr} relative to $\overline{\mathbb{Q}}_\ell[x, x']\text{-Mod}^{\text{gr,perf}}$. More precisely, given any conical closed subset $Y \subset \mathbb{A}^{2n}$ (with respect to the usual scaling action of \mathbb{G}_m), we can talk about the full subcategory $(H_n^{\text{gr}})_Y \simeq H^{\text{gr}} \otimes_{\text{Shv}_{\text{gr},c}(BB \times BB)} \text{Shv}_{\text{gr},c}(BB \times BB)_Y \simeq H^{\text{gr}} \otimes_{\overline{\mathbb{Q}}_\ell[x, x']\text{-Mod}^{\text{gr,perf}}} \overline{\mathbb{Q}}_\ell[x, x']\text{-Mod}_Y^{\text{gr,perf}} \hookrightarrow H_n^{\text{gr}}$ of objects supported on Y .

The main computational input is the following result:

Proposition 5.6.9 *Let $\beta \in \text{Br}_n$ be a braid on n strands. Then $R_\beta \in H_n^{\text{gr}}$ is supported on $\mathbb{A}_{w_\beta}^{2n}$ with respect to the action of $\text{Shv}_{\text{gr},c}(BB \times BB)$ on H_n^{gr} .*

Proof This is essentially [28, Theorem 5.2] but formulated in a more “local” way. For the sake of completeness, let us also sketch a more geometric proof, which follows the usual strategy. Namely, we decompose β into a product of simple crossings.

We thus start with the case of a single simple crossing β_i . Then w_{β_i} is a simple permutation. Consider the stack $B \setminus Bw_{\beta_i} B / B$. Observe that

$$C_{\text{gr}}^*(B \setminus Bw_{\beta_i} B / B) \simeq \overline{\mathbb{Q}}_\ell[x, x'] / (x_k - x_{w_{\beta_i}(k)}),$$

and hence,

$$\text{Shv}_{\text{gr},c}(B \setminus Bw_{\beta_i} B / B) \simeq \overline{\mathbb{Q}}_\ell[x, x'] / (x_k - x_{w_{\beta_i}(k)})\text{-Mod}^{\text{gr},\text{perf}}.$$

Consequently, any object in $\text{Shv}_{\text{gr},c}(B \setminus Bw_{\beta_i} B / B)$ is supported on $\mathbb{A}_{w_{\beta_i}}^{2n}$, relative to

$$C_{\text{gr}}^*(BB) \otimes C_{\text{gr}}^*(BB)\text{-Mod}^{\text{gr},\text{perf}} \simeq \text{Shv}_{\text{gr},c}(BB \times BB).$$

Now, let $j_{w_{\beta_i}} : B \setminus Bw_{\beta_i} B / B \rightarrow B \setminus G / B$. Then by rigidity of $\text{Shv}_{\text{gr},c}(BB \times BB)$,

$$j_{w_{\beta_i},*}, j_{w_{\beta_i},!} : \text{Shv}_{\text{gr},c}(B \setminus Bw_{\beta_i} B / B) \rightarrow \text{Shv}_{\text{gr},c}(B \setminus G / B)$$

are linear over $\text{Shv}_{\text{gr},c}(BB \times BB)$, and hence they both preserve supports. In particular, the Rouquier complexes R_{β_i} and $R_{\beta_i^{-1}}$ are both supported on $\mathbb{A}_{w_{\beta_i}}^{2n}$.

For a general braid β , we write $\beta = \beta_1 \dots \beta_m$. Then R_β is obtained using the usual convolution diagram for Hecke categories. The proof concludes by applying [3, Proposition 3.5.9] to the convolution diagram realizing R_β as a product of R_{β_i} . □

5.6.10 The support of $\text{tr}(R_\beta)$ The rest of the proof of Theorem 5.6.4 is straightforward and follows essentially the same strategy as the one used in Section 5.5. Namely, it amounts to transporting the support condition on R_β established in Proposition 5.6.9 around to $\text{Ch}_G^{\text{u,gr}}$ and then to Hilb_n . We will now consider the first part.

Corollary 5.6.11 *Let R_β be the Rouquier complex associated to a braid β . Then the support of $\text{tr}(R_\beta) \in \text{Ch}_G^{\text{u,gr}}$, relative to $\text{Shv}_{\text{gr},c}(BG) \simeq \overline{\mathbb{Q}}_\ell[x]^{S_n}\text{-Mod}^{\text{gr},\text{perf}}$, lies inside $\mathbb{A}_{\overline{w}}^n // S_n$.*

Proof Recall that $\text{tr} = q_! p^*$, where p and q are given in the following diagram:

$$\begin{array}{ccccc} BB \times_{BG} BB & \xleftarrow{p} & G/B & \xrightarrow{q} & G/G \\ \downarrow & & \downarrow & & \downarrow \\ BB \times BB & \longleftarrow & BB & \longrightarrow & BG \end{array}$$

Applying [3, Proposition 3.5.9] and the fact that supports are preserved under functors compatible with module category structures, we see that $p^* R_\beta$ has support in $\mathbb{A}_{w_\beta}^n := \mathbb{A}_{w_\beta}^{2n} \times_{\mathbb{A}^{2n}} \mathbb{A}^n$ relative to $\text{Shv}_{\text{gr},c}(BB)$, as a consequence of Proposition 5.6.9. Applying [3, Proposition 3.5.9] again, we see that $p^* R_\beta$ has support $\mathbb{A}_{\overline{w}_\beta}^n // S_n$ relative to $\text{Shv}_{\text{gr},c}(BG)$. Now, the proof concludes by using the fact that $q_!$ is $\text{Shv}_{\text{gr},c}(BG)$ -linear. □

5.6.12 Completing the proof of Theorem 5.6.4 First, recall that in geometric settings, ie when all categories involved are QCoh of Artin stacks of finite type and module category structures are given by pulling back along morphisms of stacks, categorical support in the sense above agrees with the usual notion of set-theoretic support. Thus it suffices to show that the sheaf \mathcal{F}_β is supported on $\mathbb{A}_{\overline{w}_\beta}^n // S_n$ relative

to $\mathbb{A}^n // S_n$. Indeed, this is equivalent to stating that \mathcal{F}_β is set-theoretically supported on $\text{Hilb}_{n, \bar{w}} \subseteq \text{Hilb}_n$, and hence also on $\text{Hilb}_{n, \tilde{x}, \bar{w}} = \text{Hilb}_{n, \tilde{x}} \cap \text{Hilb}_{n, \bar{w}}$ since \mathcal{F}_β is known to have support on $\text{Hilb}_{n, \tilde{x}}$ by Theorem 5.4.6 already.

But now, by Corollary 5.5.15 and, again, the fact that supports are preserved by functors that are compatible with the module category structures, this support condition is equivalent to the one proved in Corollary 5.6.11 above. \square

Acknowledgements

Ho would like to thank the organizers of the workshop *Categorification in quantum topology and beyond* at the Erwin Schrödinger International Institute for Mathematics and Physics (ESI) in 2019, where he learned about the conjecture in [29]. He would also like to thank the organizers of the American Institute of Mathematics's *Link homology research community* for providing a stimulating (online) research environment.

Ho would like to thank Minh-Tam Quang Trinh for many conversations over the years on HOMFLY-PT homology, Paul Wedrich for helpful email exchanges at the initial stage of this project, and Oscar Kivinen for many helpful email exchanges (especially regarding the various grading conventions for HHH) and for pointing out the reference [35]. Li would like to thank Peng Shan and Yin Tian for many helpful conversations on Soergel bimodules and Hilbert schemes. We thank the referee for carefully reading the manuscript and providing extensive feedback which greatly helped improve the clarity of the exposition.

Ho is partially supported by Hong Kong RGC ECS grant 26305322 and RGC GRF grant 16304923. Li is partially supported by National Natural Science Foundation of China grant 12101348.

References

- [1] **PN Achar, S Makisumi, S Riche, G Williamson**, *Koszul duality for Kac–Moody groups and characters of tilting modules*, J. Amer. Math. Soc. 32 (2019) 261–310 MR Zbl
- [2] **B Antieau, D Gepner, J Heller**, *K-theoretic obstructions to bounded t -structures*, Invent. Math. 216 (2019) 241–300 MR Zbl
- [3] **D Arinkin, D Gaitsgory**, *Singular support of coherent sheaves and the geometric Langlands conjecture*, Selecta Math. 21 (2015) 1–199 MR Zbl
- [4] **D Arinkin, D Gaitsgory, D Kazhdan, S Raskin, N Rozenblyum, Y Varshavsky**, *The stack of local systems with restricted variation and geometric Langlands theory with nilpotent singular support*, preprint (2020) arXiv 2010.01906
- [5] **A Beilinson, J Bernstein, P Deligne, O Gabber**, *Faisceaux pervers*, Astérisque 100 (2018) MR Zbl
- [6] **A Beilinson, V Ginzburg, W Soergel**, *Koszul duality patterns in representation theory*, J. Amer. Math. Soc. 9 (1996) 473–527 MR Zbl

- [7] **A Beliakova, K K Putyra, S M Wehrli**, *Quantum link homology via trace functor, I*, *Invent. Math.* 215 (2019) 383–492 MR Zbl
- [8] **D Ben-Zvi, J Francis, D Nadler**, *Integral transforms and Drinfeld centers in derived algebraic geometry*, *J. Amer. Math. Soc.* 23 (2010) 909–966 MR Zbl
- [9] **D Ben-Zvi, D Nadler**, *The character theory of a complex group*, preprint (2009) arXiv 0904.1247
- [10] **D Benson, S B Iyengar, H Krause**, *Local cohomology and support for triangulated categories*, *Ann. Sci. Éc. Norm. Supér.* 41 (2008) 573–619 MR Zbl
- [11] **R Bezrukavnikov**, *On two geometric realizations of an affine Hecke algebra*, *Publ. Math. Inst. Hautes Études Sci.* 123 (2016) 1–67 MR Zbl
- [12] **R Bezrukavnikov, M Finkelberg, V Ostrik**, *Character D -modules via Drinfeld center of Harish–Chandra bimodules*, *Invent. Math.* 188 (2012) 589–620 MR Zbl
- [13] **R Bezrukavnikov, K Tolmachov**, *Monodromic model for Khovanov–Rozansky homology*, *J. Reine Angew. Math.* 787 (2022) 79–124 MR Zbl
- [14] **R Bezrukavnikov, Z Yun**, *On Koszul duality for Kac–Moody groups*, *Represent. Theory* 17 (2013) 1–98 MR Zbl
- [15] **M V Bondarko**, *Weight structures vs. t -structures; weight filtrations, spectral sequences, and complexes (for motives and in general)*, *J. K-Theory* 6 (2010) 387–504 MR Zbl
- [16] **M V Bondarko, V A Sosnilo**, *On constructing weight structures and extending them to idempotent completions*, *Homology Homotopy Appl.* 20 (2018) 37–57 MR Zbl
- [17] **T Bridgeland, A King, M Reid**, *The McKay correspondence as an equivalence of derived categories*, *J. Amer. Math. Soc.* 14 (2001) 535–554 MR Zbl
- [18] **T-H Chen**, *On a conjecture of Braverman–Kazhdan*, *J. Amer. Math. Soc.* 35 (2022) 1171–1214 MR Zbl
- [19] **J N Eberhardt, J Scholbach**, *Integral motivic sheaves and geometric representation theory*, *Adv. Math.* 412 (2023) art. id. 108811 MR Zbl
- [20] **B Elias**, *Gaitsgory’s central sheaves via the diagrammatic Hecke category*, preprint (2018) arXiv 1811.06188
- [21] **B Elias**, *The Hecke category is bigraded*, preprint (2023) arXiv 2305.08278
- [22] **E Elmanto, V Sosnilo**, *On nilpotent extensions of ∞ -categories and the cyclotomic trace*, *Int. Math. Res. Not.* 2022 (2022) 16569–16633 MR Zbl
- [23] **D Gaitsgory**, *ind-coherent sheaves*, *Mosc. Math. J.* 13 (2013) 399–528 MR Zbl
- [24] **D Gaitsgory, D Kazhdan, N Rozenblyum, Y Varshavsky**, *A toy model for the Drinfeld–Lafforgue shtuka construction*, *Indag. Math.* 33 (2022) 39–189 MR Zbl
- [25] **D Gaitsgory, N Rozenblyum**, *A study in derived algebraic geometry, I: Correspondences and duality*, *Mathematical Surveys and Monographs* 221, Amer. Math. Soc., Providence, RI (2017) MR Zbl
- [26] **D Gaitsgory, N Rozenblyum**, *A study in derived algebraic geometry, II: Deformations, Lie theory and formal geometry*, *Mathematical Surveys and Monographs* 221, Amer. Math. Soc., Providence, RI (2017) MR Zbl
- [27] **E Gorsky, M Hogancamp, P Wedrich**, *Derived traces of Soergel categories*, *Int. Math. Res. Not.* 2022 (2022) 11304–11400 MR Zbl

- [28] **E Gorsky, O Kivinen, J Simental**, *Algebra and geometry of link homology: lecture notes from the IHES 2021 Summer School*, Bull. Lond. Math. Soc. 55 (2023) 537–591 MR Zbl
- [29] **E Gorsky, A Neğu, J Rasmussen**, *Flag Hilbert schemes, colored projectors and Khovanov–Rozansky homology*, Adv. Math. 378 (2021) art. id. 107542 MR Zbl
- [30] **Q P Ho, P Li**, *Eisenstein series via factorization homology of Hecke categories*, Adv. Math. 404 (2022) art. id. 108410 MR Zbl
- [31] **Q P Ho, P Li**, *Revisiting mixed geometry*, preprint (2022) arXiv 2202.04833
- [32] **M Khovanov**, *Triply-graded link homology and Hochschild homology of Soergel bimodules*, Internat. J. Math. 18 (2007) 869–885 MR Zbl
- [33] **R Kiehl, R Weissauer**, *Weil conjectures, perverse sheaves and l -adic Fourier transform*, Ergebnisse der Math. 42, Springer (2001) MR Zbl
- [34] **L Kong, H Zheng**, *The center functor is fully faithful*, Adv. Math. 339 (2018) 749–779 MR Zbl
- [35] **A Krug**, *Remarks on the derived McKay correspondence for Hilbert schemes of points and tautological bundles*, Math. Ann. 371 (2018) 461–486 MR Zbl
- [36] **P Li**, *Derived categories of character sheaves*, preprint (2018) arXiv 1803.04289
- [37] **P Li**, *Derived categories of character sheaves II: canonical induction/restriction functors*, preprint (2023) arXiv 2305.04444
- [38] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR Zbl
- [39] **J Lurie**, *Higher algebra* (2017) Available at <https://www.math.ias.edu/~lurie/papers/HA.pdf>
- [40] **G Lusztig**, *Intersection cohomology complexes on a reductive group*, Invent. Math. 75 (1984) 205–272 MR Zbl
- [41] **G Lusztig**, *Character sheaves, I*, Adv. in Math. 56 (1985) 193–237 MR Zbl
- [42] **G Lusztig**, *Character sheaves, II*, Adv. in Math. 57 (1985) 226–265 MR Zbl
- [43] **G Lusztig**, *Character sheaves, III*, Adv. in Math. 57 (1985) 266–315 MR Zbl
- [44] **G Lusztig**, *Character sheaves, IV*, Adv. in Math. 59 (1986) 1–63 MR Zbl
- [45] **G Lusztig**, *Character sheaves, V*, Adv. in Math. 61 (1986) 103–155 MR Zbl
- [46] **G Lusztig, Z Yun**, *Endoscopy for Hecke categories, character sheaves and representations*, Forum Math. Pi 8 (2020) art. id. e12 MR Zbl
- [47] **V Melani, M Rubi3**, *Formality criteria for algebras over operads*, J. Algebra 529 (2019) 65–88 MR Zbl
- [48] **A Oblomkov, L Rozansky**, *Knot homology and sheaves on the Hilbert scheme of points on the plane*, Selecta Math. 24 (2018) 2351–2454 MR Zbl
- [49] **A Oblomkov, L Rozansky**, *Soergel bimodules and matrix factorizations*, preprint (2020) arXiv 2010.14546
- [50] **A Oblomkov, L Rozansky**, *Categorical Chern character and braid groups*, Adv. Math. 437 (2024) art. id. 109436 MR Zbl
- [51] **H Queffelec, D E V Rose**, *Sutured annular Khovanov–Rozansky homology*, Trans. Amer. Math. Soc. 370 (2018) 1285–1319 MR Zbl
- [52] **H Queffelec, D E V Rose, A Sartori**, *Annular evaluation and link homology*, preprint (2018) arXiv 1802.04131

- [53] **L Rider**, *Formality for the nilpotent cone and a derived Springer correspondence*, Adv. Math. 235 (2013) 208–236 MR Zbl
- [54] **V Shende, D Treumann, E Zaslow**, *Legendrian knots and constructible sheaves*, Invent. Math. 207 (2017) 1031–1133 MR Zbl
- [55] **W Soergel, R Virk, M Wendt**, *Equivariant motives and geometric representation theory*, preprint (2018) arXiv 1809.05480
- [56] **V Sosnilo**, *Theorem of the heart in negative K-theory for weight structures*, Doc. Math. 24 (2019) 2137–2158 MR Zbl
- [57] **M-T Q Trinh**, *From the Hecke category to the unipotent locus*, preprint (2021) arXiv 2106.07444
- [58] **B Webster, G Williamson**, *The geometry of Markov traces*, Duke Math. J. 160 (2011) 401–419 MR Zbl
- [59] **B Webster, G Williamson**, *A geometric construction of colored HOMFLYPT homology*, Geom. Topol. 21 (2017) 2557–2600 MR Zbl

*Department of Mathematics, The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong*

*Yau Mathematical Sciences Center, Tsinghua University
Beijing, China*

quoc.ho@ust.hk, lipenghui@mail.tsinghua.edu.cn

Proposed: Ciprian Manolescu
Seconded: Benson Farb, David Fisher

Received: 28 August 2023
Revised: 10 September 2024

Asymptotically Calabi metrics and weak Fano manifolds

HANS-JOACHIM HEIN

SONG SUN

JEFFREY VIACLOVSKY

RUOBING ZHANG

We show that any asymptotically Calabi manifold which is Calabi–Yau can be compactified complex analytically to a weak Fano manifold. Furthermore, the Calabi–Yau structure arises from a generalized Tian–Yau construction on the compactification, and we prove a strong uniqueness theorem. We also give an application of this result to the surface case.

53C21, 53C25, 53C55; 14J32, 14J45

1 Introduction

This paper is concerned with the Tian–Yau construction [35] of complete Ricci-flat Kähler metrics on the complement of a smooth anticanonical divisor in a smooth Fano manifold. We begin by describing their geometry at infinity.

1.1 Asymptotically Calabi metrics

We begin by defining the asymptotic models for the metrics constructed in [35]. Let D be an $(n-1)$ -dimensional compact Kähler manifold with trivial canonical bundle and let $L \rightarrow D$ be an ample line bundle. Define

$$(1-1) \quad \deg(L) = \int_D c_1(L)^{n-1},$$

and fix a nowhere vanishing holomorphic $(n-1)$ -form Ω_D on D satisfying

$$(1-2) \quad \frac{1}{2} \int_D (\sqrt{-1})^{(n-1)^2} \Omega_D \wedge \bar{\Omega}_D = (2\pi c_1(L))^{n-1}.$$

Using Yau’s resolution [37] of the Calabi conjecture, there exists a unique Ricci-flat Kähler metric $\omega_D \in 2\pi c_1(L)$ with

$$(1-3) \quad \omega_D^{n-1} = \frac{1}{2} (\sqrt{-1})^{(n-1)^2} \Omega_D \wedge \bar{\Omega}_D.$$

There exists a unique hermitian metric h on L whose curvature form is $-\sqrt{-1}\omega_D$, up to scaling. Fixing a choice of h , the *Calabi model space* is the subset \mathcal{C} of L consisting of all elements ξ with $0 < |\xi|_h < 1$. We next define a nowhere vanishing holomorphic volume form $\Omega_{\mathcal{C}}$ and a Ricci-flat Kähler metric $\omega_{\mathcal{C}}$ which is incomplete as $|\xi|_h \rightarrow 1$ and complete as $|\xi|_h \rightarrow 0$. Let $p: \mathcal{C} \rightarrow D$ denote the bundle projection and Z be the holomorphic vector field generating the natural \mathbb{C}^* -action on the fibers of p . The holomorphic volume form $\Omega_{\mathcal{C}}$ is uniquely determined by the equation

$$(1-4) \quad Z \lrcorner \Omega_{\mathcal{C}} = p^* \Omega_D,$$

and the metric $\omega_{\mathcal{C}}$ is given by the *Calabi ansatz*

$$(1-5) \quad \omega_{\mathcal{C}} = \frac{n}{n+1} \sqrt{-1} \partial \bar{\partial} (-\log |\xi|_h^2)^{(n+1)/n},$$

which satisfies the complex Monge–Ampère equation

$$\omega_{\mathcal{C}}^n = \frac{1}{2} (\sqrt{-1})^{n^2} \Omega_{\mathcal{C}} \wedge \bar{\Omega}_{\mathcal{C}},$$

hence is Ricci-flat. The function $z = (-\log |\xi|_h^2)^{1/n}$ is the $\omega_{\mathcal{C}}$ -moment map for the natural S^1 -action on L . It is easily verified that the $\omega_{\mathcal{C}}$ -distance function r to a fixed point in \mathcal{C} satisfies

$$(1-6) \quad r^{-1} z^{(n+1)/2} = C + o(1) \quad \text{as } z \rightarrow \infty.$$

We will consider Kähler manifolds with trivial canonical bundle which will be denoted by (X, I, ω, Ω) , where X is a smooth manifold of real dimension $2n$, I is a complex structure, ω is a Kähler form, and Ω is a holomorphic volume form. We say that (X, I, ω, Ω) is a *Calabi–Yau structure* if it satisfies the complex Monge–Ampère equation

$$(1-7) \quad \omega_X^n = \tau^n \cdot \frac{1}{2} (\sqrt{-1})^{n^2} \cdot \Omega_X \wedge \bar{\Omega}_X$$

for some constant $\tau > 0$. We are interested in Calabi–Yau structures which are asymptotic to the above Calabi model spaces. More precisely, we make the following definition.

Definition 1.1 A Calabi–Yau structure $(X, I_X, \omega_X, \Omega_X)$ is *asymptotically Calabi* if there exists $\underline{\delta} > 0$, a Calabi model space \mathcal{C} , and a diffeomorphism $\Phi: \mathcal{C} \setminus K' \rightarrow X \setminus K$, where $K \subset X$ is compact and $K' = \{|\xi|_h \geq \frac{1}{2}\}$, such that for all $k \in \mathbb{N}_0$, the following hold uniformly as $z \rightarrow +\infty$:

$$(1-8) \quad |\nabla_{g_{\mathcal{C}}}^k (\Phi^* I_X - I_{\mathcal{C}})|_{g_{\mathcal{C}}} = O(e^{-\underline{\delta} z^{n/2}}),$$

$$(1-9) \quad |\nabla_{g_{\mathcal{C}}}^k (\Phi^* \Omega_X - \Omega_{\mathcal{C}})|_{g_{\mathcal{C}}} = O(e^{-\underline{\delta} z^{n/2}}),$$

$$(1-10) \quad |\nabla_{g_{\mathcal{C}}}^k (\Phi^* \omega_X - \omega_{\mathcal{C}})|_{g_{\mathcal{C}}} = O(e^{-\underline{\delta} z^{n/2}}).$$

Remark 1.2 The decay assumption on the error terms is the natural one since harmonic functions on the Calabi model space have this type of decay; see [19; 32].

Our main theorem is the following.

Theorem 1.3 *Any Calabi–Yau structure (X, I, ω, Ω) which is asymptotically Calabi can be compactified complex analytically to a weak Fano manifold \bar{X} . Furthermore, the Calabi–Yau structure arises from a generalized Tian–Yau construction on \bar{X} and ω is the unique Calabi–Yau metric with respect to (I, Ω) satisfying (1-10) and representing $[\omega] \in H^2(X)$.*

The above theorem solves a particular case of Yau’s conjecture in [38]. We give a precise description of this generalized Tian–Yau construction in Section 2. The main step in the proof is to show that for a certain choice of a parallel complex structure I , the underlying complex manifold of an Calabi–Yau asymptotically Calabi metric can be compactified to a weak Fano manifold. This involves producing holomorphic functions with controlled growth at infinity, which is typically done using weighted Fredholm theory. This strategy runs into trouble due to the fact that, in the Tian–Yau construction, the decay rate of the metric is much slower than the decay rate of the complex structure. To overcome this difficulty, our strategy is based on the L^2 -estimates in several complex variables pioneered by Hörmander [21]. The proof of Theorem 1.3 can be found in Section 2. We also have the following corollary. Recall the index of \bar{X} is the largest integer k such that $K_{\bar{X}}^{-1} = H^k$ for some line bundle H .

Corollary 1.4 *Let (X^n, I, ω, Ω) be a Calabi–Yau structure which is asymptotically Calabi. Then $\pi_1(X)$ is a cyclic group with order the index of \bar{X} . Furthermore, there exists a constant $C(n)$, depending only upon n , such that $\deg(L) \leq C(n)$.*

This degree bound is remarkable because in Definition 1.1, $\deg(L)$ could a priori be any integer, but if \mathcal{C} occurs at infinity for a Calabi–Yau asymptotically Calabi metric, then $\deg(L)$ must be bounded. The proof of this degree bound uses Theorem 1.3 and deep results in birational geometry.

1.2 The surface case

A gravitational instanton is by definition a complete noncompact hyperkähler 4-manifold (X, g, ω) with square-integrable curvature. By the results of the recent paper by Sun and Zhang [33], a gravitational instanton is always asymptotic to a model end. Accordingly, gravitational instantons can be classified into 6 families: ALE, ALF, ALG, ALH, ALG*, ALH*. There has been extensive work, much of it quite recent, on classifying the 6 families completely; see Chen and Chen [3; 2; 4], Chen and Viaclovsky [5], Kronheimer [25] and Minerbe [27]. The results in this paper are relevant to the ALH* family. This has the unique intriguing feature that its members have fractional asymptotic volume growth; indeed, the volume growth exponent of an ALH* model end is $\frac{4}{3}$. Gravitational instantons of type ALH* also appear as singularity models in polarized degenerations of K3 surfaces; see Hein, Sun, Viaclovsky and Zhang [19] and Sun and Zhang [31]. Their precise definition can be found in Section 3.

There are two known mathematical constructions of ALH* gravitational instantons. Both of them come with a preferred choice of complex structure, and are based on solving a complex Monge–Ampère equation on a quasiprojective surface with trivial canonical bundle. First, we have the Tian–Yau construction [35], which involves the complement of a smooth anticanonical divisor in a del Pezzo surface. Second, we

have the construction of Hein [18], which involves the complement of a singular fiber of Kodaira type I_b in a rational elliptic surface. Our next theorem relates the complex structures involved in these two constructions.

Theorem 1.5 *Let (X, g, ω) be an ALH* gravitational instanton.*

- (i) *Letting I denote the complex structure corresponding to ω_1 , then (X, I) is biholomorphic to weak del Pezzo surface \bar{X} minus a smooth anticanonical elliptic curve. Furthermore, the hyperkähler structure arises from a generalized Tian–Yau construction on this compactification and is the unique Calabi–Yau metric with respect to $(I, \Omega = \omega_2 + \sqrt{-1}\omega_3)$ satisfying (1-10) and representing $[\omega_1] \in H^2(X)$.*
- (ii) *Letting J denote the complex structure corresponding to I_2 , then (X, J) compactifies to a rational elliptic surface S with a global section by adding F , a Kodaira type I_b fiber of multiplicity 1. The 2-form $\Omega = \omega_2 + \sqrt{-1}\omega_3$ is a rational 2-form on S with a simple pole along F .*

The proof of Theorem 1.5 will be given in Section 3 and is our originally intended proof of some claims made in [19, Remark 2.4]. The main ingredients in the proof are Theorem 1.3, the decay estimates of [19, Section 3], and the analysis of harmonic functions on asymptotically Calabi spaces of [19, Section 4]. We also note that in the meantime Collins, Jacob and Lin [8, Theorem 1.3] have proved, using an entirely different method, that a Tian–Yau space can be compactified to a rational elliptic surface.

Acknowledgements The authors thank Gao Chen for valuable discussions on gravitational instantons. Hein was supported by NSF grant DMS-1745517 and by the DFG under Germany’s Excellence Strategy EXC 2044-390685587 *Mathematics Münster: Dynamics-Geometry-Structure* as well as by the CRC 1442 *Geometry: deformations and rigidity* of the DFG. Sun was supported by the Simons Collaboration on *Special holonomy in geometry, analysis and physics* (#488633), and NSF grant DMS-2004261. Viaclovsky was partially supported by NSF grants DMS-1811096 and DMS-2105478. Zhang was partially supported by NSF grants DMS-1906265 and DMS-2304818. Finally, the authors thank the referee for their careful reading of the paper, and for their questions and comments which improved the exposition of the paper.

2 Compactification to weak Fano manifold

In this section, we give the proof of Theorem 1.3. Let (X, I, ω, Ω) be a complete Calabi–Yau manifold that is asymptotically Calabi. We identify $X \setminus K$ smoothly with a Calabi model space $\mathcal{C} \setminus K'$, where $K' = \{z \leq z_0\}$ for some $z_0 \gg 0$, and assume that (1-8)–(1-10) are satisfied. Let $\phi_0 \equiv z^n - \delta z^{n/2}$, for some $\delta \in (0, \frac{1}{2})$ to be chosen later. Then

$$\begin{aligned}
 (2-1) \quad d d_{\mathcal{C}}^c \phi_0 &= -d J_{\mathcal{C}} d \phi_0 = -d \left(\left(n z^{n-1} - \frac{1}{2} n \delta z^{n/2-1} \right) J_{\mathcal{C}} dz \right) \\
 &= d \left(\left(n - \frac{1}{2} n \delta z^{-n/2} \right) \theta \right) \\
 &= \left(n - \frac{1}{2} n \delta z^{-n/2} \right) d\theta + \delta \left(\frac{1}{2} n \right)^2 z^{-n/2-1} dz \wedge \theta,
 \end{aligned}$$

where $\theta \equiv -z^{n-1} J_\epsilon dz$. Note also that

$$(2-2) \quad \omega_D = \sqrt{-1} \partial \bar{\partial} (-\log \|\xi\|_h^2) = \sqrt{-1} \partial \bar{\partial} (z^n) = \frac{1}{2} n d\theta,$$

$$(2-3) \quad \begin{aligned} \omega_\epsilon &= \frac{n}{n+1} \frac{1}{2} d d_\epsilon^c z^{n+1} = -\frac{n}{2(n+1)} d J_\epsilon dz^{n+1} \\ &= -\frac{1}{2} n d(z^n J dz) = \frac{1}{2} n d(z\theta) \\ &= \frac{1}{2} n (z d\theta + dz \wedge \theta). \end{aligned}$$

Then for $z_1 \gg z_0$,

$$(2-4) \quad \begin{aligned} d d_\epsilon \phi_0 &= z^{-n/2-1} \left((n z^{n/2+1} - \frac{1}{2} n \delta z) d\theta + \delta (\frac{1}{2} n)^2 dz \wedge \theta \right) \\ &= z^{-n/2-1} \left((n z^{n/2+1} - \frac{1}{2} n \delta z - (n/2)^2 \delta z) d\theta + \delta (\frac{1}{2} n)^2 (z d\theta + dz \wedge \theta) \right) \\ &\geq z^{-n/2-1} \frac{1}{2} n \delta \omega_\epsilon, \end{aligned}$$

since $\frac{1}{2} n d\theta = \omega_D$ is positive definite on D . Let $A = z_1^n - \delta z_1^{n/2}$, and choose a smooth increasing and convex function $u: \mathbb{R} \rightarrow \mathbb{R}$ such that $u(t) = 2A/3$ for $t \leq A/2$ and $u(t) = t$ for $t \geq A$. Write $\phi_1 \equiv u \circ \phi_0$. Then

$$(2-5) \quad \begin{aligned} d d_\epsilon^c \phi_1 &= -d J_\epsilon d(u \circ \phi_0) = -d J_\epsilon u'(\phi_0) d\phi_0 \\ &= -u''(\phi_0) d\phi_0 \wedge J_\epsilon d\phi_0 + u'(\phi_0) d d_\epsilon^c \phi_0. \end{aligned}$$

Since the form $-d\phi_0 \wedge J_\epsilon d\phi_0$ is positive semidefinite, we see that $d d_\epsilon^c \phi_1 \geq 0$ for all $z \geq z_0$ and $d d_\epsilon^c \phi_1 \geq C \delta z^{-n/2-1} \omega_\epsilon$ for $z > z_1$.

Notice that ϕ_1 can be naturally viewed as a smooth function on X , and it satisfies $d d_I^c \phi_1 \geq P(z)\omega$ for a nonnegative function $P(z)$ with $P(z) \geq C \delta z^{-n/2-1}$ when z is large. In particular, we know that (X, I) is 1-convex. So, by [13, Section 2], there is a Remmert reduction $\pi: X \rightarrow \tilde{X}$, where \tilde{X} is Stein and $\text{Sing}(\tilde{X})$ is a finite set contained in the region $\{z \leq z_1\}$. Then \tilde{X} admits an exhaustion function $\psi_{\tilde{X}}$ which is smooth on $\tilde{X} \setminus \text{Sing}(\tilde{X})$ and satisfies $d d_I^c \psi_{\tilde{X}} > 0$. Denote $\psi_X \equiv \pi^* \psi_{\tilde{X}}$. Then $d d_I^c \psi_X > 0$ on $X \setminus E$, where $E \equiv \pi^{-1}(\text{Sing}(\tilde{X}))$. Choose a cutoff function χ on X supported in $\{z \leq z_1 + 1\}$ with $\chi \equiv 1$ on $\{z \leq z_1\}$. Then let $\phi \equiv \epsilon \chi \psi_X + \phi_1$ for a fixed $0 < \epsilon \ll 1$. The above calculation shows that ϕ also satisfies $d d_I^c \phi \geq P(z)\omega$, where $P(z) \geq 0$ on X , $P(z) > 0$ on $X \setminus E$, and $P(z) \geq C \delta z^{-n/2-1}$ outside a compact set.

Any holomorphic section $s \in H^0(D, L^k)$ gives rise to a holomorphic function f_s on $L \setminus \mathbf{0}_L$ by defining

$$(2-6) \quad f_s(\xi) = s(\pi(\xi)) / \xi^{\otimes k},$$

where $\xi \in L \setminus \mathbf{0}_L$, and $\pi: L \rightarrow D$ is the bundle projection. In particular, f_s restricts to a holomorphic function on \mathcal{C} . Taking the logarithm of (2-6), and using that $z^n = -2 \log \|\xi\|_h$, we see that $f_s = \tilde{f}_s e^{(k/2)z^n}$, where \tilde{f}_s is a function on the unit circle bundle of L .

Lemma 2.1 For all $l \geq 0$ and any $\varepsilon > 0$, we have $|\nabla_{g_\varepsilon}^l f_s|_{g_\varepsilon} = O(e^{(k/2)z^n + \varepsilon z^{n/2}})$ as $z \rightarrow \infty$.

Proof The estimate for $l = 0$ holds with $\varepsilon = 0$, and follows from the previous remarks. Since f_s is holomorphic, it is harmonic. The curvature of the Calabi metric is in particular bounded at infinity; see [35, Lemma 4.3] and [19, Remark 3.2]. Given $p \in \mathcal{C}$, let $B_1(p)$ be a unit ball around p , and $B_1(\hat{p})$ be a unit ball in its universal cover. Then standard elliptic estimates for harmonic functions yield that

$$(2-7) \quad |\nabla^l f_s|(p) \leq \|\nabla^l f_s\|_{C^0(B_{1/2}(\hat{p}))} \leq C_l \|f_s\|_{C^0(B_1(\hat{p}))}.$$

Using (1-6), one can easily check that if $|r(q) - r(p)| < 1$, then

$$(2-8) \quad |z(q)^n - z(p)^n| \leq C z(p)^{(n-1)/2}.$$

Since $(n - 1)/2 < n/2$, the claim follows. □

Denote by $\mathcal{O}(X)$ the space of I -holomorphic functions on X .

Proposition 2.2 There is an injective linear map $\mathcal{L}: \bigoplus_{k=0}^\infty H^0(D, L^k) \rightarrow \mathcal{O}(X)$ such that for any nonzero section $s \in H^0(D, L^k)$, we have that

$$(2-9) \quad |\nabla_{\Phi^*g}^l (\mathcal{L}(s) \circ \Phi - f_s)|_{\Phi^*g} = O(e^{(k/2)z^n - (\delta_k/2)z^{n/2}})$$

for all $l \geq 0$ and for some $\delta_k > 0$ as $z \rightarrow \infty$.

Proof In the following proof, for simplicity of notation we will omit the diffeomorphism Φ . Fix a cutoff function χ on \mathcal{C} which is equal to 1 when $z \geq z_1 + 1$ and vanishes when $z \leq z_1$. Then for any $s \in H^0(D, L^k)$, the function χf_s naturally extends to a smooth function on X . Notice that for $z > z_0 + 1$ we have that

$$(2-10) \quad |\bar{\partial}_I(\chi f_s)|_g = |\bar{\partial}_I f_s|_g = |(\bar{\partial}_I - \bar{\partial}_{\mathcal{C}}) f_s|_g \leq e^{(k/2)z^n - \delta z^{n/2}}.$$

Set $\delta \equiv \min(\underline{\delta}/(4k), \frac{1}{4})$ in the definition of ϕ_0 above. Then we have that

$$(2-11) \quad \int_{X \setminus E} \frac{1}{P(z)} |\bar{\partial}_I(\chi f_s)|_g^2 e^{-k\phi} d\text{Vol}_g < \infty.$$

Notice that $X \setminus E \cong \tilde{X} \setminus \text{Sing}(\tilde{X})$ admits a complete Kähler metric; see [29, Proposition 4.1]. Also, by assumption K_X is trivial, so we can apply the standard L^2 -estimates for the $\bar{\partial}$ -operator on $X \setminus E$ — see for example [11, Chapter VIII.6, Theorem 6.1] — to find a unique solution u to the equation $\bar{\partial}_I u = \bar{\partial}_I(\chi f_s)$ with

$$(2-12) \quad \int_{X \setminus E} |u|^2 e^{-k\phi} d\text{Vol}_g \leq \int_{X \setminus E} \frac{1}{P(z)} |\bar{\partial}_I(\chi f_s)|_g^2 e^{-k\phi} d\text{Vol}_g$$

such that u is L^2 orthogonal to $\ker(\bar{\partial}_I)$. Notice that

$$(2-13) \quad \Delta_g u = \bar{\partial}_{I,g}^* \bar{\partial}_I(\chi f_s) = O(e^{(k/2)z^n - (\delta/2)z^{n/2}}).$$

Similar to the proof of Lemma 2.1, it follows from local L^2 elliptic estimates that for all $l \geq 0$,

$$(2-14) \quad |\nabla_g^l u|_g = O(e^{(k/2)z^n - \delta z^{n/2}}) \quad \text{as } z \rightarrow \infty.$$

Indeed, by straightforward computations, one can find a $C > 0$ such that $\text{Vol}_g(B_1(p)) \geq C \cdot z^{-(n-1)/2}(p)$. Therefore using (2-8) and (2-12), there exists a uniform constant $C > 0$ such that for any $z(p) \gg 1$,

$$(2-15) \quad \int_{B_1(p)} |u|^2 \, d\text{vol}_g \leq C e^{k \cdot z(p)^n - \delta' \cdot z(p)^{n/2}},$$

where $\delta' > 0$ is any smaller number than $\underline{\delta}$. By an obvious volume comparison argument, we obtain a similar estimate for the lifting of u to $B_{1/2}(\widehat{p})$ in the local universal cover $\widehat{B_1(p)}$ of $B_1(p)$. Since $\widehat{B_1(p)}$ is noncollapsed with bounded curvature, standard elliptic theory applied to the lifting of (2-13) yields (2-14).

To finish the proof, we let $\mathcal{L}(s) \equiv \chi f_s - u$. This function is holomorphic away from E , so by Hartogs's theorem (applied to \widetilde{X}) one can see that it is globally holomorphic on X . The conclusion then follows. \square

Fix k such that $L^l|_D$ is very ample for all $l \geq k$. Then we have a holomorphic embedding

$$(2-16) \quad F_k: L \rightarrow \mathbb{P}(H^0(D, L^k)^* \oplus H^0(D, L^{k+1})^*), \quad (x, \xi) \mapsto (\text{ev}_{x,k} / \xi^{\otimes k}, \text{ev}_{x,k+1} / \xi^{\otimes(k+1)}),$$

where $x \in D$, $\xi \in L_x$ and $\text{ev}_{x,l}: H^0(D, L^l) \rightarrow L^l|_x$ is the evaluation map. Alternatively, we can describe F_k as follows. By assumption, we have the embeddings

$$(2-17) \quad i_{L^k}: D \rightarrow \mathbb{P}(H^0(D, L^k)^*), \quad i_{L^{k+1}}: D \rightarrow \mathbb{P}(H^0(D, L^{k+1})^*).$$

We can view i_{L^k} and $i_{L^{k+1}}$ as mapping into $\mathbb{P}(H^0(D, L^k)^* \oplus H^0(D, L^{k+1})^*)$. Then for $x \in D$, F_k maps the fiber $\pi^{-1}(x) \subset L$ linearly to the line between $i_{L^k}(\pi(p))$ and $i_{L^{k+1}}(\pi(p))$, so it is clearly an embedding. Obviously, $F_k(\mathbf{0}_L)$ is isomorphic to D and is contained in the linear subspace $\mathbb{P}(H^0(D, L^{k+1})^*) \subset \mathbb{P}(H^0(D, L^k)^* \oplus H^0(D, L^{k+1})^*)$.

Now we define a holomorphic map $G_k: X \rightarrow \mathbb{P}(H^0(D, L^k)^* \oplus H^0(D, L^{k+1})^*)$ via

$$(2-18) \quad p \in X \mapsto (\widetilde{\text{ev}}_{p,k}, \widetilde{\text{ev}}_{p,k+1}),$$

where $\widetilde{\text{ev}}_{p,k}: H^0(D, L^k) \rightarrow \mathbb{C}$ is given by $\widetilde{\text{ev}}_{p,k}(s) = \mathcal{L}(s)(p)$.

We denote by \overline{X} the topological compactification of X by adding $F_k(D)$ to the end of $G_k(X)$. This is justified by the following.

Proposition 2.3 *There exists a compact set $K \subset X$ such that G_k is a holomorphic embedding on $X \setminus K$ with $G_k(X \setminus K) \cap F_k(D) = \emptyset$. Furthermore there is a neighborhood of $F_k(D)$ in \overline{X} which is homeomorphic to a neighborhood of $F_k(\mathbf{0}_L)$ in $F_k(L)$.*

Proof The key point is that we can compare G_k with F_k via the fixed embedding of the end of X into $\mathcal{C} \subset L$. Given any point $q_0 \in \mathbf{0}_L \cong D$, we can find sections $s_0, \dots, s_{n-1} \in H^0(D, L^k)$ and $s_n \in H^0(D, L^{k+1})$ such that $s_0(q_0) \neq 0$, $s_1(q_0) = \dots = s_{n-1}(q_0) = 0$, $ds_1(q_0), \dots, ds_{n-1}(q_0)$ are linearly independent, and $s_n(q_0) \neq 0$. Then $w_k \equiv s_k/s_0$ for $1 \leq k \leq n-1$ and $w_n \equiv s_0/s_n$ form local holomorphic coordinates in a neighborhood U of q_0 in L . To clarify this definition, note that w_n is

a local section of L^{-1} on D , but by duality we can view such a section as a local function on the total space of L which is linear on fibers. We therefore can think of w_1, \dots, w_{n-1} as coordinates on the divisor, and w_n as a fiber coordinate. Denote this coordinate system by $w : U \rightarrow \mathbb{C}^n$. Notice that $|\xi|_{h_L}^2 = |w_n|^2 e^{-\varphi(w_1, \dots, w_{n-1})}$ for a smooth function φ satisfying $\sqrt{-1} \partial \bar{\partial} \varphi = \omega_D$. The Calabi metric in the (w_1, \dots, w_n) coordinates is given by

$$(2-19) \quad \omega_{\mathbb{C}} = (-\log |\xi|_{h_L}^2)^{1/n} \omega_D + 1/n (-\log |\xi|_{h_L}^2)^{(1/n)-1} \cdot \sqrt{-1} \cdot \left(\frac{dw_n}{w_n} - \partial \varphi \right) \wedge \left(\frac{d\bar{w}_n}{\bar{w}_n} - \bar{\partial} \varphi \right).$$

We define a mapping Ψ from an open subset $V \subset P(H^0(D, L^k)^* \oplus H^0(D, L^{k+1})^*)$ containing $F_k(q_0)$ to \mathbb{C}^n by

$$(2-20) \quad [(\alpha, \beta)] \mapsto \left(\frac{\alpha(s_1)}{\alpha(s_0)}, \dots, \frac{\alpha(s_{n-1})}{\alpha(s_0)}, \frac{\alpha(s_0)}{\beta(s_n)} \right).$$

Then the restriction of Ψ to $\text{Image}(F_k)$ is a coordinate chart for $\text{Image}(F_k)$ near $F_k(q_0)$, and $\Psi \circ F_k = w$ in a neighborhood of $q_0 \in \mathbf{0}_L$. Next, let $\eta_j = \mathcal{L}(s_j)/\mathcal{L}(s_0)$ for $1 \leq j \leq n-1$, and $\eta_n = \mathcal{L}(s_0)/\mathcal{L}(s_n)$, and we denote this mapping $\eta : \Phi(U \setminus \mathbf{0}_L) \rightarrow \mathbb{C}^n$ (after possibly shrinking U). Note that for any p in the domain of η , we have

$$(2-21) \quad \begin{aligned} \Psi(G_k(p)) &= \Psi([\tilde{e}v_{p,k}, \tilde{e}v_{p,k+1}]) = \left(\frac{\tilde{e}v_{p,k}(s_1)}{\tilde{e}v_{p,k}(s_0)}, \dots, \frac{\tilde{e}v_{p,k}(s_0)}{\tilde{e}v_{p,k+1}(s_n)} \right) \\ &= \left(\frac{\mathcal{L}(s_1)}{\mathcal{L}(s_0)}(p), \dots, \frac{\mathcal{L}(s_0)}{\mathcal{L}(s_n)}(p) \right) = \eta(p). \end{aligned}$$

By Proposition 2.2, on $U \setminus \mathbf{0}_L$ we have that $\eta \circ \Phi = w(1 + \zeta)$ with $|\nabla_g^l \zeta|_g = O(e^{-\delta' z^{n/2}})$ for all $l \geq 0$ for some $\delta' > 0$. In the following, the constant δ' is allowed to change from line to line. Using (2-19), we have that

$$(2-22) \quad \left(\frac{\partial(\eta \circ \Phi)_\alpha}{\partial w_\beta} \right) = \begin{pmatrix} \mathbb{I}_{n-1} + O(e^{-\delta' z^{n/2}}) & & \vdots & \\ & O(e^{-\delta' z^{n/2}} |w_n|^{-1}) & & \\ \dots & & \vdots & \\ & O(e^{-\delta' z^{n/2}} |w_n|) & \dots & 1 + O(e^{-\delta' z^{n/2}}) \end{pmatrix}.$$

Thus, the Jacobian matrix $(\partial(\eta \circ \Phi)_\alpha / \partial w_\beta)$ is nondegenerate for $z \gg 1$, which implies that G_k is an immersion outside a compact set.

Next, we show that G_k is injective onto its image for $z \gg 1$. Suppose we have two points $p_1, p_2 \in X$ with $G_k(p_1) = G_k(p_2)$, and $z(q_2) \geq z(q_1) \gg 1$, where $\Phi(q_j) = p_j$. Let d_{FS} denote the Fubini–Study distance on $\mathbb{P}(H^0(D, L^k)^* \oplus H^0(D, L^{k+1})^*)$. We claim that there exists $\varepsilon > 0$ such that $d_{\text{FS}}(F_k(q), G_k(\Phi(q))) < \varepsilon$ if $q \in L$ is sufficiently near D . To see this, using the formula

$$(2-23) \quad d_{\text{FS}}(Z, W) = \arccos \sqrt{\frac{\langle Z, W \rangle^2}{\|Z\|^2 \|W\|^2}}$$

and Proposition 2.2, we obtain that

$$(2-24) \quad d_{\text{FS}}(F_k(q), G_k(\Phi(q))) = \arccos(1 + O(e^{-\delta'z(q)^{n/2}})) \quad \text{as } z \rightarrow \infty,$$

since L^{k+1} is very ample. Then

$$(2-25) \quad d_{\text{FS}}(F_k(q_1), F_k(q_2)) \leq d_{\text{FS}}(F_k(q_1), G_k(p_1)) + d_{\text{FS}}(G_k(p_2), F_k(q_2)) < 2\varepsilon.$$

This implies that q_1 and q_2 must be contained in the same w coordinate patch above, and therefore $\eta_\alpha(p_1) = \eta_\alpha(p_2)$ for $\alpha = 1, \dots, n$. Denote $\tau_\alpha \equiv |w_\alpha(q_2) - w_\alpha(q_1)|$. Then $\tau_\alpha = w_\alpha(q_1)O(e^{-\delta'z(q_1)^{n/2}})$. So we know in particular that $|w_n(q_1)| \leq C|w_n(q_2)|$. Notice that

$$(2-26) \quad |\tau_\alpha| = |w_\alpha(q_2)\zeta_\alpha(q_2) - w_\alpha(q_1)\zeta_\alpha(q_1)| \leq Ce^{-\delta'z(q_1)^{n/2}}|\tau_\alpha| + |w_\alpha(q_1)| \cdot |\zeta_\alpha(q_2) - \zeta_\alpha(q_1)|.$$

Let q_t with $t \in [0, 1]$ be the straight line connecting q_1 and q_2 in the w coordinates. Then

$$(2-27) \quad |\zeta_\alpha(q_2) - \zeta_\alpha(q_1)| \leq \sum_{\beta=1}^n \sup_{t \in [0,1]} \left| \frac{\partial \zeta_\alpha}{\partial w_\beta}(q_t) \right| \cdot |\tau_\beta| \leq Ce^{-\delta'z(q_1)^{n/2}} \left(\sum_{j=1}^{n-1} |\tau_j| + \frac{1}{|w_n(q_1)|} |\tau_n| \right).$$

Combining (2-26) and (2-27) with $\alpha = n$, we get that when $z_1 \gg 1$,

$$|\tau_n| \leq Ce^{-\delta'z(q_1)^{n/2}} |w_n(q_1)| \cdot \sum_{j=1}^{n-1} |\tau_j|.$$

If we now sum (2-26) for $\alpha = 1, \dots, n - 1$ and use (2-27), then we obtain that $\tau_\alpha = 0$ for $1 \leq \alpha \leq n - 1$ and hence that $\tau_n = 0$ as well.

Next, we claim that away from a sufficiently large compact subset, G_k maps no point into $F_k(\mathbf{0}_L)$. To see this, assume that there exists a point $p \in X$ with $G_k(p) = F_k(x)$, where $x \in \mathbf{0}_L$. Letting $q = \Phi^{-1}(p)$, if $z(q)$ is sufficiently large then q and x must be contained in a same w -coordinate chart as above. Taking the n^{th} component of $\Psi(G_k(p)) = \eta(p)$, yields

$$(2-28) \quad 0 = w_n(x) = \Psi_n(F_k(x)) = \eta_n(p) = \eta_n(\Phi(q)) = w_n(q)(1 + \zeta_n(q)).$$

But if $z(q) \gg 1$, we obtain a contradiction since $w_n(q) \neq 0$. The proposition then follows from the above. □

Remark 2.4 Recall the coordinate $(\eta \circ \Phi)_n = w_n(1 + \eta_n)$, and since $e^{-\delta z^{n/2}} \sim e^{-\delta' \sqrt{-\log |w_n|}}$, the complex structure I is not even Hölder continuous at the divisor. Thus, one cannot invoke any known weak version of the Newlander–Nirenberg theorem to construct a complex-analytic compactification; see for example [20]. This is the reason why we had to deviate from the known approaches to constructing such compactifications; compare for example the proofs of [16, Theorem 3.1] and [26, Theorem 1.6].

Denote by Z the image of \bar{G}_k . It follows from the Remmert–Stein theorem [11, Chapter II.8, Theorem 8.7] that Z is a complex analytic variety in a neighborhood of $F_k(\mathbf{0}_L) \cong D$. Since Z is topologically a manifold by Proposition 2.3, it follows that Z is locally irreducible. Thus, by [11, Chapter II.7, Corollary 7.13], the normalization map $Z' \rightarrow Z$ is a homeomorphism.

Denote by D' the copy of D in Z' . Notice that, in the proof of Proposition 2.3, the functions η_α for $1 \leq \alpha \leq n$ can be viewed as local holomorphic functions on a neighborhood of a point $q \in D'$ in Z' . Moreover, they define a local topological embedding of Z' into \mathbb{C}^n . So the inverse map is holomorphic, which implies that Z' must be smooth near D' . Now we may glue Z' to X and obtain a smooth complex-analytic compactification of X , which we denote by \bar{X}' . By construction, D' is a smooth divisor in \bar{X}' .

Write $\omega' \equiv \sqrt{-1}(\sum_{j=1}^{n-1} d\eta_j \wedge d\bar{\eta}_j + |\eta_n|^{-2} d\eta_n \wedge d\bar{\eta}_n)$. This is a locally defined Kähler metric in a punctured neighborhood of D' , with cylindrical behavior normal to D' . It is easy to check using the calculations in the proof of Proposition 2.3 that for all $l \geq 0$,

$$(2-29) \quad \sum_{j=1}^{n-1} |\nabla_{\omega'}^l(w_j)|_{\omega'} + |w_n^{-1} \nabla_{\omega'}^l(w_n)|_{\omega'} \leq C_l.$$

This is a crucial fact for us. One can also reinterpret this as saying that the metric ω' is C^∞ uniformly equivalent to the corresponding cylindrical Kähler metric defined using (w_1, \dots, w_n) . Note that $|\partial_{w_j} z| \leq Cz^{1-n}$ for $1 \leq j \leq n-1$ and $|\partial_{w_n} z| \leq Cz^{1-n}|w_n|^{-1}$. In particular, we have that

$$(2-30) \quad |\nabla_{\omega'}^l z|_{\omega'} \leq C_l z^{1-nl}.$$

Recall that Ω is the holomorphic volume form on (X, g) with respect to I , and $\Omega_\mathcal{C}$ is the corresponding holomorphic 2-form on the Calabi model space \mathcal{C} . Notice that $\Omega_\mathcal{C}$ is a meromorphic 2-form on L with a simple pole along the zero section $\mathbf{0}_L$, and Ω can be viewed as a holomorphic 2-form on $\bar{X}' \setminus D'$.

Lemma 2.5 Ω is a meromorphic volume form on \bar{X}' with a simple pole along D' . In particular, D' is an anticanonical divisor.

Proof We may locally, near a point $p \in D$, write $\Omega_\mathcal{C} = f(w_1, \dots, w_n)w_n^{-1}dw_1 \wedge \dots \wedge dw_n$, where f is a nowhere vanishing local holomorphic function. Then $\Omega = f(w_1, \dots, w_n)w_n^{-1}dw_1 \wedge \dots \wedge dw_n + \Gamma$ with $|\Gamma \wedge \bar{\Gamma} / \omega_\mathcal{C}^n| = O(e^{-\delta z^{n/2}})$. But $\omega_\mathcal{C}^n = \frac{1}{2}(\sqrt{-1})^{n^2} \Omega_\mathcal{C} \wedge \bar{\Omega}_\mathcal{C} \sim -|w_n|^{-2}dw_1 \wedge d\bar{w}_1 \wedge \dots \wedge dw_n \wedge d\bar{w}_n$ near D . It follows that $C^{-1} \leq |\Omega|_{\omega'} \leq C$, so the form $w_n \Omega$ extends holomorphically to D' locally. \square

Now choose a finite open cover $\{V_j\}$ of D such that each V_j has holomorphic coordinates $w_{1,j}, \dots, w_{n-1,j}$ given by the quotients $s_{k,j}/s_{0,j}$ for some sections $s_{0,j}, \dots, s_{n-1,j} \in H^0(D, L^k)$ for $1 \leq k \leq n-1$, and $L|_{V_j}$ has a trivializing section $e_j = s_{n,j}/s_{0,j}$ for some $s_{n,j} \in H^0(D, L^{k+1})$. Consequently we obtain an open cover $\{U_j\}$ of a neighborhood of $\mathbf{0}_L \simeq D$ in L , with holomorphic coordinates $\{w_{1,j}, \dots, w_{n,j}\}$, such that a point $\xi \in L$ is represented as $\xi = w_{n,j} \cdot e_j(w_{1,j}, \dots, w_{n-1,j})$. Write $e^{-\varphi_j(w_{1,j}, \dots, w_{n-1,j})} = |e_j(w_{1,j}, \dots, w_{n-1,j})|_h^2$.

Accordingly, we get an open cover $\{U'_j\}$ of a neighborhood of D' in \bar{X}' . In each U'_j we then have holomorphic coordinates $\eta_{k,j} = \mathcal{L}(s_{k,j})/\mathcal{L}(s_{0,j})$ and $\eta_{n,j} = \mathcal{L}(s_{0,j})/\mathcal{L}(s_{n,j})$. From the proof of Proposition 2.3 we know that $\eta_{k,j} = w_{k,j}$ on $U'_j \cap D'$ for $1 \leq k \leq n-1$, and the transition functions

satisfy $\partial\eta_{n,j}/\partial\eta_{n,i} = \partial w_{n,j}/\partial w_{n,i}$ on $U'_i \cap U'_j \cap D'$. This implies that the conormal bundle $N_{D'}^{-1}$ is isomorphic to L^{-1} , so in particular D' has ample normal bundle in \bar{X}' . Moreover, the local functions $\eta_{n,j}$ together give rise to a global section $S_{D'}$ of the line bundle $[D'] \simeq K_{\bar{X}'}^{-1}$ with a simple zero along D' .

Now the local functions φ_j define a hermitian metric on $N_{D'} \simeq [D']$. Fix an arbitrary extension of this to a hermitian metric h' on $[D']$ on a neighborhood of D' . On each chart U_j , this extension h' has a local representation given by $e^{-\phi_j}$, with $\phi_j = \varphi_j$ along D' . We consider a smooth closed $(1, 1)$ -form

$$(2-31) \quad \omega_m \equiv \frac{n}{n+1} \sqrt{-1} \partial\bar{\partial} (-\log |S_{D'}|_{h'}^2)^{(n+1)/n}.$$

It is well-defined and positive definite outside a compact subset of $\bar{X}' \setminus D'$.

One can check that ω_m and ω' satisfy

$$(2-32) \quad C^{-1} z^{1-n} \omega' \leq \omega_m \leq C z \omega'$$

for some constant C , and for each $k \geq 0$,

$$(2-33) \quad |\nabla_{\omega'}^k \omega_m|_{\omega'} = O(z^{m_k})$$

for some $m_k \in \mathbb{N}_0$, as $z \rightarrow \infty$. Therefore covariant derivative with respect to ω' and ω_m differ by at most polynomial error terms.

Proposition 2.6 *For all $l > 0$ and all $\delta < \min(\delta_k, \underline{\delta})$, we have that*

$$(2-34) \quad |\nabla_{\omega_m}^l (\omega - \omega_m)|_{\omega_m} = O(e^{-\delta z^{n/2}}).$$

Consequently, $[\omega] \in \text{im}(H_c^2(X) \rightarrow H^2(X))$.

Proof By assumption, $\omega = \omega_\epsilon + \beta_1$, where $|\nabla_{\omega_\epsilon}^l \beta_1|_{\omega_\epsilon} = O(e^{-\delta z^{n/2}})$ for all $l \geq 0$. Let

$$\tilde{\omega} \equiv \frac{n}{n+1} \sqrt{-1} \partial_I \bar{\partial}_I z^{n+1}.$$

Writing $\omega_\epsilon = \tilde{\omega} + \beta_2$, using (2-30), (2-32) and (2-33), we have that $|\nabla_{\tilde{\omega}}^l \beta_2|_{\tilde{\omega}} = O(e^{-\delta z^{n/2}})$ for all $l \geq 0$ and $\delta < \underline{\delta}$. In U_j , we have that

$$z^n = -\log |\xi|_{h_L}^2 = -\log |w_{n,j}|^2 + \varphi_j(w_{1,j}, \dots, w_{n-1,j}).$$

On the other hand, we have that

$$|S_{D'}|_{h'}^2 = -\log |\eta_{n,j}|^2 + \phi_j(\eta_{1,j}, \dots, \eta_{n,j}).$$

Notice that $\eta_{n,j} = w_{n,j}(1 + \beta_3)$, with $|\nabla_{\omega}^l \beta_3|_{\omega} = O(e^{-\delta_k z^{n/2}})$ for all $l \geq 0$, and $\phi_j(\eta_{1,j}, \dots, \eta_{n,j}) = \varphi_j(w_{1,j}, \dots, w_{n-1,j}) + \eta_{n,j} P(\eta_{1,j}, \dots, \eta_{n,j})$ for some smooth function P . Estimate (2-34) now follows from a straightforward computation, using the key property (2-29), (2-32) and (2-33).

By integrating from infinity as in [19, Lemma 3.7], away from a compact set we may write $\omega = \omega_m + d\sigma$, where σ is a smooth real-valued 1-form such that $|\nabla_{\omega_m}^l \sigma|_{\omega_m} = O(e^{-\delta z^{n/2}})$ for all $l \geq 0$. Then the smooth real-valued 2-form

$$(2-35) \quad \omega - d\left(\chi\left(\sigma + \frac{n}{2(n+1)}d_I^c(-\log|S_{D'}|_{h'}^2)^{(n+1)/n}\right)\right)$$

is cohomologous to ω and has compact support, where χ is a cutoff function which is 1 in a neighborhood of infinity and 0 on a compact set. \square

Now we claim that \bar{X}' is a *weak Fano manifold*, ie \bar{X}' is projective and $K_{\bar{X}'}^{-1}$ is big and nef. Assuming projectivity, the nef property is obvious because $K_{\bar{X}'}^{-1} = [D']$ for some smooth divisor D' with ample normal bundle, and it is also elementary to deduce the big property from this fact; see for example [36, Lemma 2.3]. Projectivity follows from a similar reasoning as in the proof of [10, Theorem 2.1]; see in particular [10, page 4, proof in the smooth case]. In short, one first proves, using the theory of the Remmert reduction (see the beginning of Section 2), that $K_{\bar{X}'}^{-1}$ is semiample. This also strongly relies on the fact that $K_{\bar{X}'}^{-1} = [D']$, where D' has ample normal bundle. It then follows that \bar{X}' is Moishezon, which implies that Hodge theory holds on \bar{X}' . Also, $h^{0,2}(\bar{X}') = 0$ thanks to the Grauert–Riemenschneider vanishing theorem [14, Satz 2.1]. Using the latter two properties, the fact that \bar{X}' admits compactly supported Kähler classes by Proposition 2.6, and another vanishing theorem on the open manifold X due to Grauert–Riemenschneider [14, page 278, Korollar], one can explicitly write down a Kähler form on \bar{X}' . The Kodaira embedding theorem now implies that \bar{X}' is projective because we already know that $h^{0,2}(\bar{X}') = 0$. We note that a similar gluing argument of Kähler forms will appear again in the proof of Lemma 2.7 below. The argument works here even though, unlike in [10], D' is not Fano, so that $H^{1,1}(\bar{X}')$ may not surject onto $H^2(X)$. The key point is that we are restricting ourselves to the case of compactly supported Kähler classes on X .

To finish the proof of Theorem 1.3, it remains to identify the Ricci-flat Kähler metric ω on X with a slight generalization of the construction given by Tian and Yau [35, Theorem 4.2]. Let M be an n -dimensional projective manifold containing a smooth divisor D with ample normal bundle such that $K_M^{-1} = [D]$. Then M is necessarily weakly Fano. Fix a defining section $S \in H^0(M, K_M^{-1})$ of the divisor D and denote $X \equiv M \setminus D$. Following the beginning of the argument from [10] sketched above, one first proves that K_M^{-1} is semiample. Here this also follows from the Kawamata basepoint-free theorem [22, Theorem 6.1] because we already know that M is projective. Let E be the nonample locus of K_M^{-1} . This is a union of some subvarieties of M that are disjoint from D . We fix a smooth hermitian metric \tilde{h} on K_M^{-1} such that its curvature form is nonnegative everywhere and is positive away from E . Write $\tilde{v} \equiv -\log|S|_{\tilde{h}}^2$.

We may view S^{-1} as a holomorphic volume form Ω_X on X with a simple pole along D . Let Ω_D be the holomorphic volume form on D given by the residue of Ω_X along D . Let h_D be a hermitian metric on $K_M^{-1}|_D$ such that its curvature form is a Calabi–Yau metric ω_D on D . Rescaling S if necessary, we may assume that $\omega_D^{n-1} = \frac{1}{2}(\sqrt{-1})^{(n-1)^2}\Omega_D \wedge \bar{\Omega}_D$.

One can extend h_D to a smooth hermitian metric h on K_M^{-1} such that its curvature form is positive definite in a neighborhood of D . Fix such an extension, denoted by h_M . For any $A \in \mathbb{R}$, we write $h_A \equiv h_M e^{-A}$. Then it is easy to see that outside a compact subset of X , $(-\log |S|_{h_A}^2)^{(n+1)/n}$ is strictly plurisubharmonic. As in the beginning of this section, by composing with a convex function we may find a global smooth function v_A on X with $dd^c v_A \geq 0$ such that

$$v_A = \frac{n}{n+1} (-\log |S|_{h_A}^2)^{(n+1)/n}$$

outside a compact set and v_A is constant in a neighborhood of E .

Abusing notation, we denote by $H_{c,+}^2(X)$ the subset of $\text{im}(H_c^2(X) \rightarrow H^2(X))$ consisting of classes \mathfrak{k} such that $\int_Y \mathfrak{k}^p > 0$ for any compact analytic subset Y of X of pure complex dimension $p > 0$. Notice that any such Y must be contained in E .

Lemma 2.7 *For all $\mathfrak{k} \in H_{c,+}^2(X)$ and all numbers $A \in \mathbb{R}$, $\tau > 0$, there exists a smooth Kähler form $\omega_{A,\tau}$ on X such that $[\omega_{A,\tau}] = \mathfrak{k}$ and $|\nabla_{\omega_{A,\tau}}^l (\omega_{A,\tau} - \tau dd^c v_A)|_{\omega_{A,\tau}} = O(|S|_{h_M}^{\delta_0})$ for some $\delta_0 > 0$ and all $l \geq 0$.*

Proof By hypothesis, \mathfrak{k} is represented by a closed 2-form β on X such that $\beta = 0$ away from some compact subset of X . In particular, β trivially extends to a smooth closed 2-form on M . By the Grauert–Riemenschneider vanishing theorem [14, Satz 2.1], or (because we are assuming that M is projective) by the Kawamata–Viehweg vanishing theorem (see for example [24, Theorem 2.64]), we have that $h^{0,2}(M) = 0$. Thus, there exist a smooth closed $(1, 1)$ -form η and a smooth 1-form γ on M such that $\beta = \eta + d\gamma$. Because β has compact support in $X \subset M$, it follows that η is d -exact on some open neighborhood of D in M . In particular, $\eta|_D = dd^c f$ for some smooth function f on D . We choose an arbitrary smooth extension of f to M and replace β by $\eta - dd^c f$. Thus, we are now assuming without loss of generality that β is a smooth closed $(1, 1)$ -form on M such that $\beta|_D = 0$ and $[\beta|_M] = \mathfrak{k}$. It is straightforward to check that such a form β satisfies the exponential decay estimate from the statement of the lemma with respect to any reference metric of the form $\tau dd^c v_A$ outside a sufficiently large compact subset of X . (If $n \geq 3$, then instead of applying the dd^c -lemma on D it is possible to apply a stronger dd^c -lemma on the complement of a large compact subset of X to arrange directly that β still has compact support in X ; see for example [7, Lemma 4.3].)

Because $\mathfrak{k} \in H_{c,+}^2(X)$, by the generalized Demailly–Păun criterion of [9, Theorem 1.1] there exists a smooth function u_0 on X such that $\beta + dd^c u_0$ is positive in a neighborhood U of E . Choose a cutoff function χ of compact support contained in U which equals 1 on a neighborhood of E , and let $\rho = \rho(t)$ be a cutoff function on $[0, \infty)$ which equals 1 when $t \leq 1$ and vanishes when $t \geq 2$. For fixed A and τ , we define

$$(2-36) \quad \omega_{A,\tau} \equiv \beta + dd^c \left(\chi u_0 + C_1 \left(\rho \circ \left(\frac{1}{C_2} \tilde{v} \right) \right) \cdot \tilde{v} + \tau v_A \right).$$

It is straightforward to verify that if we first choose C_1 large and then C_2 large (depending on C_1), then $\omega_{A,\tau}$ is globally positive on X . \square

One can also see that $\omega_{A,\tau}^n = e^{-f_{A,\tau}} \tau^n \cdot \frac{1}{2}(\sqrt{-1})^{n^2} \Omega_X \wedge \bar{\Omega}_X$ for some function $f_{A,\tau}$ which tends to zero at infinity.

Lemma 2.8 *Given any $\tau > 0$, there is a unique $A = A(\tau)$ such that*

$$(2-37) \quad \int_X (\omega_{A,\tau}^n - \tau^n \frac{1}{2}(\sqrt{-1})^{n^2} \Omega_X \wedge \bar{\Omega}_X) = 0.$$

Proof It is easy to check that the integral is finite for any A . For a given A , add and subtract $\omega_{0,\tau}^n$ under the integral sign and split up the integral accordingly. For $\varepsilon > 0$ sufficiently small, we have

$$\int_{|S|_h \leq \varepsilon} (\omega_{A,\tau}^n - \omega_{0,\tau}^n) = \tau \int_{|S|_h = \varepsilon} (d^c v_A \wedge \omega_{A,\tau}^{n-1} - d^c v_0 \wedge \omega_{0,\tau}^{n-1}).$$

By computing the boundary term explicitly and letting $\varepsilon \rightarrow 0$, one sees that

$$\int_X (\omega_{A,\tau}^n - \omega_{0,\tau}^n) = \lambda \cdot A \tau^n \int_D \omega_D^{n-1}$$

for some $\lambda = \lambda(n) \neq 0$. The key point is that we only care the limit as $\varepsilon \rightarrow 0$, and thus we only need to understand the leading term and ignore all the lower order terms. This argument goes back to [35, Lemma 4.6]. For a more detailed exposition, see also the very recent paper [6, page 14]. Then the desired condition becomes a linear equation for A . □

Using Lemmas 2.7 and 2.8 as ingredients, we now get the following existence theorem, which generalizes the classical existence result of Tian and Yau [35, Theorem 4.2].

Theorem 2.9 *Given any $\mathfrak{k} \in H_{c,+}^2(M)$ and $\tau > 0$, there is a complete Kähler metric $\omega_\tau = \omega_{A(\tau),\tau} + dd^c \phi$ on X such that $[\omega_\tau] = \mathfrak{k}$ and $\omega_\tau^n = \tau^n \frac{1}{2}(\sqrt{-1})^{n^2} \Omega_X \wedge \bar{\Omega}_X$, and such that for some $\delta_0 > 0$ and for all $l \geq 0$ we have that $|\nabla_{\omega_{A(\tau),\tau}}^l \phi|_{\omega_{A(\tau),\tau}} = O(e^{-\delta_0(-\log |S|_{h_M}^2)^{1/2}})$.*

This follows from [35, Theorem 1.1] and [18, Proposition 2.4, Proposition 2.9(ia) and (ii)]. The correct choice of $A = A(\tau)$ is crucial because otherwise the relevant Monge–Ampère equation cannot be solved by direct methods. Once the equation has been solved for $A = A(\tau)$, it follows that a solution exists for all A by adding $v_{A(\tau),\tau} - v_{A,\tau}$ to the potential, but $v_{A(\tau),\tau} - v_{A,\tau}$ is comparable to $(A(\tau) - A)\tau(-\log |S|_{h_M}^2)^{1/n}$ at infinity, which is not even uniformly bounded for $A \neq A(\tau)$.

Remark 2.10 The decay rate of the complex structure is of order $O(e^{-(\frac{1}{2}-\varepsilon)z^n})$ for all $\varepsilon > 0$. This is much faster than the decay rate of the metric, which is in general only of order $O(e^{-\delta z^{n/2}})$.

Theorem 2.11 *Suppose we have a Kähler metric ω on X such that $\omega^n = C \Omega_X \wedge \bar{\Omega}_X$ for some $C > 0$ and such that there exist $\tau > 0$ and $A \in \mathbb{R}$ such that $|\nabla_\omega^l (\omega - \tau dd^c v_A)|_\omega = O(e^{-\delta(-\log |S|_h^2)^{1/2}})$ for some $\delta > 0$ and for all $l \geq 0$. Then $\omega = \omega_\tau$, where ω_τ is the metric constructed in Theorem 2.9 in the class $\mathfrak{k} = [\omega] \in H_{c,+}^2(X)$.*

Proof By rescaling ω if necessary, we may assume that $C = 1$. As in the proof of Proposition 2.6, we can see that $\omega = \tau dd^c v_A + d\sigma$ outside a compact set, where σ is a smooth real-valued 1-form

such that $|\nabla_{\omega_m}^l \sigma|_{\omega_m} = O(e^{-\delta(-\log |S|_h^2)^{1/2}})$ for all $l \geq 0$. In particular, $[\omega] \in \text{im}(H_c^2(X) \rightarrow H^2(X))$. Thus, if we let \mathfrak{k} denote the de Rham cohomology class of ω in $H^2(X)$, then we have that $\mathfrak{k} \in H_{c,+}^2(X)$. Consequently, we can apply Theorem 2.9 to obtain a complete Calabi–Yau metric ω_τ on X with $[\omega_\tau] = \mathfrak{k}$. By construction, we have that $\omega^n = \omega_\tau^n$. Then, as in the proof of Lemma 2.8, we can see that $A = A(\tau)$, so we know that $\omega = \omega_\tau + d\tilde{\sigma}$, with $|\nabla_{\omega_\tau}^l \tilde{\sigma}|_{\omega_\tau} = O(e^{-\delta_1(-\log |S|_h^2)^{1/2}})$ for some $\delta_1 > 0$.

Next, we solve the Poisson equation $\bar{\partial}^* \bar{\partial} u = \bar{\partial}^* \tilde{\sigma}^{0,1}$ with respect to ω_τ . Note that $\int_X (\bar{\partial}^* \tilde{\sigma}^{0,1}) \omega_\tau^n = 0$ by definition, so we can apply [17, Theorem 1.5] to conclude the existence of a C^∞ solution u which is uniformly bounded and satisfies $\int_X |du|_{\omega_\tau}^2 \omega_\tau^n < \infty$. Moreover, by the integration argument in the proof of [18, Proposition 2.9(ia)], after subtracting a constant from u one actually has the asymptotics $|\nabla_{\omega_\tau}^l u|_{\omega_\tau} = O(e^{-\delta_2(-\log |S|_h^2)^{1/2}})$ for some $\delta_2 > 0$ and all $l \geq 0$. Let $\gamma \equiv \tilde{\sigma}^{0,1} - \bar{\partial} u$. By construction, $\bar{\partial} \gamma = \bar{\partial}^* \gamma = 0$ with respect to ω_τ . Thus, by the Bochner formula, $\Delta |\gamma|^2 \geq 0$. Since γ tends to zero at infinity, it must vanish identically. It follows that $\omega - \omega_\tau = dd^c(\sqrt{-1}(\bar{u} - u))$. Writing

$$(2-38) \quad 0 = \omega^n - \omega_\tau^n = (\omega - \omega_\tau)(\omega^{n-1} + \dots + \omega_\tau^{n-1}),$$

then multiplying (2-38) by $\bar{u} - u$ and integrating by parts, since $\bar{u} - u$ is decaying, we conclude that $u = \bar{u}$, hence $\omega = \omega_\tau$. \square

The second statement in Theorem 1.3 follows from Theorem 2.11 (applied to $M = (\bar{X}', \bar{I})$) and Proposition 2.6.

Proof of Corollary 1.4 By [34, Theorem 1.1], the weak Fano manifold $\bar{X} = X \cup D$ is simply connected. Since $-K_{\bar{X}}$ is big, D does not consist of a pencil, so by [28, Corollary 2.10], $\pi_1(X) \simeq \pi_1(-K_{\bar{X}}^\times)$, where $-K_{\bar{X}}^\times = -K_{\bar{X}} \setminus \mathbf{0}_{-K_{\bar{X}}}$. Letting N denote the unit circle bundle of $-K_{\bar{X}}$, the homotopy sequence of the fibration $S^1 \rightarrow N \rightarrow \bar{X}$ yields

$$(2-39) \quad \pi_2(\bar{X}) \rightarrow \mathbb{Z} \rightarrow \pi_1(X) \rightarrow 0.$$

Since \bar{X} is simply connected and $-K_{\bar{X}}^\times$ deformation retracts onto N , noticing that $K_{\bar{X}}$ is trivial over X , we have $\pi_1(N) \cong \pi_1(-K_{\bar{X}}^\times) \cong \pi_1(X)$. By [30, Proposition 2], the first mapping factors as the Hurewicz projection from $\pi_2(\bar{X}) \rightarrow H_2(\bar{X}, \mathbb{Z})$ composed with the mapping $\delta: H_2(\bar{X}, \mathbb{Z}) \rightarrow \mathbb{Z}$ given by the intersection pairing with $[D] \in H_{n-2}(\bar{X}, \mathbb{Z})$, that is,

$$(2-40) \quad \delta(\Sigma) = \int_{\Sigma} c_1([D]).$$

By Poincaré duality, this mapping is nontrivial, so the image is an infinite subgroup $k\mathbb{Z} \subset \mathbb{Z}$, and consequently $\pi_1(X) = \mathbb{Z}/k\mathbb{Z}$ is cyclic, with k equal to the index of \bar{X} .

The bound on the degree is a consequence of the following results. It is a classical result that the degree of a weak del Pezzo surface satisfies $1 \leq d \leq 9$; see for example [12, Chapter 8]. A bound on the degree of a weak Fano threefold is proved in [23, Corollary 1.3]. Finally, Birkar [1, Theorem 1.1] proved that in any dimension there is an a priori bound on the degree of a weak Fano manifold. \square

3 ALH* gravitational instantons

The 3-dimensional Heisenberg group is

$$(3-1) \quad H(1, \mathbb{R}) \equiv \left\{ \begin{bmatrix} 1 & x & t \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} : x, y, t \in \mathbb{R} \right\}.$$

Let $(\mathbb{T}^2, g_{c,\tau})$ be the flat 2-torus corresponding to the lattice $c(\mathbb{Z} \oplus \mathbb{Z}\tau) \subset \mathbb{C}$ with $\tau \in \mathbb{H}$ and $c > 0$, where \mathbb{H} is the upper half-plane in \mathbb{C} . Write $\tau_1 = \text{Re}(\tau)$ and $\tau_2 = \text{Im}(\tau)$. For $b \in \mathbb{Z}_+$, let $\text{Nil}_{b,c,\tau}^3$ be the nilmanifold corresponding to the quotient of the Heisenberg group by the subgroup generated by

$$(3-2) \quad (x, y, t) \mapsto (x + c, y, t + cy),$$

$$(3-3) \quad (x, y, t) \mapsto (x + c\tau_1, y + c\tau_2, t + c\tau_1 y),$$

$$(3-4) \quad (x, y, t) \mapsto (x, y, t + \tau_2 b^{-1} c^2).$$

Definition 3.1 For $b \in \mathbb{Z}_+, c > 0, \tau \in \mathbb{H}$, and $R > 0$, the ALH* model space is

$$(3-5) \quad \mathfrak{M}_{b,c,\tau}(R) \equiv (R, \infty) \times \text{Nil}_{b,\tau,c}^3$$

together with the hyperkähler Riemannian metric

$$(3-6) \quad g^{\mathfrak{M}} \equiv V(dx^2 + dy^2 + dz^2) + V^{-1}\theta^2,$$

where z is the coordinate on (R, ∞) , $V \equiv 2\pi bc^{-2}\tau_2^{-1}z$ and $\theta \equiv 2\pi bc^{-2}\tau_2^{-1}(dt - x dy)$.

Choose the following orthonormal frame for $T^*\mathfrak{M}$:

$$(3-7) \quad \{e_1, e_2, e_3, e_4\} = \{V^{\frac{1}{2}} dx, V^{\frac{1}{2}} dy, V^{\frac{1}{2}} dz, V^{-\frac{1}{2}} \theta\}.$$

Define three almost-complex structures on $T^*\mathfrak{M}$ by

$$(3-8) \quad I_{\mathfrak{M}}^*(e_1) = e_2, \quad I_{\mathfrak{M}}^*(e_3) = e_4,$$

$$(3-9) \quad J_{\mathfrak{M}}^*(e_1) = e_4, \quad J_{\mathfrak{M}}^*(e_2) = e_3,$$

$$(3-10) \quad K_{\mathfrak{M}}^*(e_2) = e_4, \quad K_{\mathfrak{M}}^*(e_3) = e_1,$$

which are dual to almost-complex structures $I_{\mathfrak{M}}, J_{\mathfrak{M}}, K_{\mathfrak{M}}$ on $T\mathfrak{M}$, respectively. Denote the Kähler forms associated to $I_{\mathfrak{M}}^*, J_{\mathfrak{M}}^*, K_{\mathfrak{M}}^*$ by $\omega_1^{\mathfrak{M}}, \omega_2^{\mathfrak{M}}, \omega_3^{\mathfrak{M}}$, respectively. These are explicitly given by

$$(3-11) \quad \omega_1^{\mathfrak{M}} = dz \wedge \theta + V dx \wedge dy,$$

$$(3-12) \quad \omega_2^{\mathfrak{M}} = dx \wedge \theta + V dy \wedge dz,$$

$$(3-13) \quad \omega_3^{\mathfrak{M}} = dy \wedge \theta + V dz \wedge dx.$$

Taken together, this data defines a hyperkähler structure on \mathfrak{M} , which we denote as $(\mathfrak{M}, g^{\mathfrak{M}}, \omega^{\mathfrak{M}})$. This structure can equivalently also be specified as $(\mathfrak{M}, g^{\mathfrak{M}}, I_{\mathfrak{M}}, J_{\mathfrak{M}}, K_{\mathfrak{M}})$.

Definition 3.2 (ALH* gravitational instanton) A hyperkähler structure (X, g, ω) on a 4-manifold X is called an ALH* gravitational instanton with parameters $b \in \mathbb{Z}_+$, $c > 0$ and $\tau \in \mathbb{H}$ if there exist $\delta, R > 0$, a compact subset $X_R \subset X$, an ALH* model space $(\mathfrak{M}_{b,c,\tau}(R), g^{\mathfrak{M}})$ and a diffeomorphism

$$(3-14) \quad \Phi: \mathfrak{M}_{b,c,\tau}(R) \rightarrow X \setminus X_R$$

such that for all $k \in \mathbb{N}_0$,

$$(3-15) \quad |\nabla_{g^{\mathfrak{M}}}^k (\Phi^* g - g^{\mathfrak{M}})|_{g^{\mathfrak{M}}} = O(e^{-\delta z}),$$

$$(3-16) \quad |\nabla_{g^{\mathfrak{M}}}^k (\Phi^* \omega_i - \omega_i^{\mathfrak{M}})|_{g^{\mathfrak{M}}} = O(e^{-\delta z}) \quad \text{for } i = 1, 2, 3,$$

as $z \rightarrow \infty$.

According to [19, Proposition 3.1], any ALH* model space \mathfrak{M} together with its complex structure $I_{\mathfrak{M}}$ is holomorphically isometric to a *Calabi model space* up to rescaling. This means the following. We can view the flat torus $(\mathbb{T}^2, g_{c,\tau})$ as an elliptic curve D . Then there exists an ample line bundle $L \rightarrow D$ of degree b , together with a hermitian metric h_L whose curvature form defines the flat metric

$$(3-17) \quad 2\pi b c^{-2} \tau_2^{-1} g_{c,\tau}$$

on D , such that the underlying complex manifold $(\mathfrak{M}_{b,c,\tau}(R), I_{\mathfrak{M}})$ can be identified with the open set

$$(3-18) \quad \mathcal{C} \equiv \{\xi \in L : 0 < |\xi|_{h_L} < e^{-\frac{1}{2}z_0^2}\}$$

for some $z_0 > 0$. Moreover, the Kähler form and the holomorphic 2-form on \mathfrak{M} are respectively given by $\omega_1 = \mu \omega_{\mathcal{C}}$ for some $\mu > 0$ and $\omega_2 + \sqrt{-1} \omega_3 = \nu \Omega_{\mathcal{C}}$ for some $\nu \in \mathbb{C}^*$, where

$$(3-19) \quad \omega_{\mathcal{C}} \equiv \frac{2}{3} \sqrt{-1} \partial_{\mathcal{C}} \bar{\partial}_{\mathcal{C}} (-\log |\xi|_{h_L}^2)^{\frac{3}{2}}$$

and $\Omega_{\mathcal{C}}$ is a holomorphic volume form which has a simple pole along the zero section $\mathbf{0}_L$ and is invariant under the natural \mathbb{C}^* -action on L . In addition, we have that

$$(3-20) \quad z = (-\log |\xi|_{h_L}^2)^{\frac{1}{2}},$$

and this is the $\omega_{\mathcal{C}}$ -moment map for the natural S^1 -action on L . We note here that h_L is unique only up to scaling, and the scale of h_L is an important free parameter; see Lemma 2.8.

Thanks to this identification, an ALH* gravitational instanton is the same as a complete hyperkähler 4-manifold which is asymptotically Calabi in the sense of [19, Definition 4.1]. The proof of part (i) of Theorem 1.5 follows from this equivalence and Theorem 1.3.

3.1 Compactification to rational elliptic surfaces

In this subsection, we prove part (ii) of Theorem 1.5. First, we recall some basic facts about the lowest nontrivial eigenvalue of the Laplacian on $(\mathbb{T}^2, g_{c,\tau})$. The \mathbb{Z}^2 -action on \mathbb{C} is generated by

$$(3-21) \quad (x, y) \mapsto (x + c, y), \quad (x, y) \mapsto (x + c\tau_1, y + c\tau_2).$$

The eigenfunctions of the Laplacian on $(\mathbb{T}^2, g_{c,\tau})$ are given by

$$(3-22) \quad \phi_{m,n}(x, y) = e^{2\pi imc^{-1}(x-\tau_1\tau_2^{-1}y)} e^{2\pi in c^{-1}\tau_2^{-1}y}$$

for $(m, n) \in \mathbb{Z}^2$. The eigenvalues are

$$(3-23) \quad \lambda_{m,n} = 4\pi^2 c^{-2} \{m^2 + \tau_2^{-2}(n - m\tau_1)^2\}.$$

Using the $\text{PSL}(2, \mathbb{Z})$ -action on \mathbb{H} , we can assume without loss of generality that $|\tau| \geq 1$. Then the lowest nontrivial eigenvalue is given by $\lambda_1 = \lambda_{0,1} = 4\pi^2 c^{-2} \tau_2^{-2}$, with eigenfunction $\phi_{0,1}$.

Next, we establish an almost optimal decay rate for any ALH* gravitational instanton.

Proposition 3.3 *Let (X, g, ω) be an ALH* gravitational instanton with parameters $b \in \mathbb{Z}_+, c > 0, \tau \in \mathbb{H}$. Then, in suitable ALH* coordinates, (3-15)–(3-16) hold with $\delta = \sqrt{\lambda_1} - \varepsilon$ for any $\varepsilon > 0$.*

Proof By Theorem 1.3, (X, g, I) arises as a Tian–Yau metric on a weak del Pezzo surface M minus an anticanonical divisor D . Thus, we can use the coordinate system Φ from [19, Proposition 3.4]. The Tian–Yau metric is of the form $\omega = \omega_0 + \sqrt{-1} \partial\bar{\partial}u$ with $|\nabla_{g_\varepsilon}^k u|_{g_\varepsilon} = O(e^{-\varepsilon z})$ as $z \rightarrow \infty$ for some $\varepsilon > 0$ and all $k \geq 0$. By [19, Proposition 3.4], the background Kähler form satisfies the asymptotics

$$(3-24) \quad |\nabla_{g_\varepsilon}^k (\omega_0 - \omega_\varepsilon)|_{g_\varepsilon} = O(e^{-\varepsilon z^2})$$

as $z \rightarrow \infty$, for some $\varepsilon > 0$ and all $k \geq 0$. Expanding the Calabi–Yau equation

$$(3-25) \quad (\omega_0 + \sqrt{-1} \partial\bar{\partial}u)^2 = \frac{1}{2} \Omega \wedge \bar{\Omega}$$

yields that $\Delta_{g_\varepsilon} u = O(e^{-2\varepsilon z})$ as $z \rightarrow \infty$. Assume that $2\varepsilon < \sqrt{\lambda_1}$. Then the arguments of [19, Section 4] allow us to conclude that $u = O(e^{-2\varepsilon z})$ for any $\underline{\varepsilon} \in (0, \varepsilon)$. Indeed, notice that [19, Proposition 4.12] holds with $\underline{\delta} = \sqrt{\lambda_1}$, as is clear from the proof. Using this, we can find a function $u_{-2\underline{\varepsilon}}$ defined on $\{z > R_1\}$ for some $R_1 > R$ such that $u_{-2\underline{\varepsilon}} = O(e^{-2\underline{\varepsilon} z})$ for any $\underline{\varepsilon} \in (0, \varepsilon)$ and $\Delta_{g_\varepsilon} u_{-2\underline{\varepsilon}} = \Delta_{g_\varepsilon} u$. The function $u - u_{-2\underline{\varepsilon}}$ is then harmonic with respect to g_ε and is also $o(1)$ as $z \rightarrow \infty$, so from [19, Proposition 4.10] (part (2) of which is easily seen to hold with $\underline{\delta}/2$ replaced by any number greater than $-\sqrt{\lambda_1}$), we have

$$(3-26) \quad u - u_{-2\underline{\varepsilon}} = O(e^{(-\sqrt{\lambda_1} + \varepsilon')z})$$

for any $\varepsilon' > 0$, which implies that $u = O(e^{-2\varepsilon z})$ for any $\underline{\varepsilon} \in (0, \varepsilon)$. We can iterate this argument together with standard local derivative estimates for the equation (3-25) to get

$$(3-27) \quad |\nabla_{g_\varepsilon}^k u|_{g_\varepsilon} = O(e^{(-\sqrt{\lambda_1} + \varepsilon)z})$$

as $z \rightarrow \infty$, for all $k \geq 0$ and $\varepsilon > 0$. This implies (3-16) for $i = 1$ with $\delta = \sqrt{\lambda_1} - \varepsilon$ for any $\varepsilon > 0$.

Also, from Theorem 1.3 and [19, Proposition 3.4(a)], we have that

$$(3-28) \quad |\nabla_{g_{\mathfrak{M}}}^k (\Phi^* I - I_{\mathfrak{M}})|_{g_{\mathfrak{M}}} = O(e^{-\varepsilon' z^2})$$

for some $\varepsilon' > 0$ and any $k \geq 0$ as $z \rightarrow \infty$. Since g is Kähler with respect to I , the metric condition (3-15) with $\delta = \sqrt{\lambda_1} - \varepsilon$ then follows from (3-28) and the above estimates on $\Phi^* \omega_1 - \omega_1^{\mathfrak{M}}$.

Finally, from [19, Proposition 3.4(b)], we have that

$$(3-29) \quad |\nabla_{g_{\mathfrak{M}}}^k (\Phi^* \Omega_X - \Omega_{\mathfrak{M}})|_{g_{\mathfrak{M}}} = O(e^{-\varepsilon' z^2})$$

for some $\varepsilon' > 0$ and any $k \geq 0$ as $z \rightarrow \infty$. Because $\Omega_X = \omega_2 + \sqrt{-1}\omega_3$ and $\Omega_{\mathfrak{M}} = \omega_2^{\mathfrak{M}} + \sqrt{-1}\omega_3^{\mathfrak{M}}$, the conditions (3-16) for $i = 2$ and $i = 3$ actually hold for any $\delta > 0$. \square

Remark 3.4 Proposition 3.3 could also be proved by following the arguments in [33, Section 6.5]. This way of reasoning is analytically more involved, but it does not require Theorem 1.3.

We now produce a J -holomorphic function on X asymptotic to a nice $J_{\mathfrak{M}}$ -holomorphic function on \mathfrak{M} that defines an elliptic fibration of \mathfrak{M} with the desired behavior at infinity.

Proposition 3.5 *Let (X, g, I, J, K) be an ALH* gravitational instanton. Then there exists a function $u: X \rightarrow \mathbb{C}$ which is holomorphic with respect to J and satisfies*

$$(3-30) \quad u = e^{\sqrt{\lambda_1}(z+iy)} + f,$$

where $|\nabla^k f|_g = O(e^{\varepsilon z})$ for all $k \geq 0$, as $z \rightarrow \infty$, for any $\varepsilon > 0$.

Proof The function $z + \sqrt{-1}y$ is a locally defined $J_{\mathfrak{M}}$ -holomorphic function on \mathfrak{M} because

$$(3-31) \quad J_{\mathfrak{M}}(dz + i dy) = -dy + i dz = i(dz + i dy).$$

Consequently, $h = e^{\sqrt{\lambda_1}(z+iy)}$ is a globally defined $J_{\mathfrak{M}}$ -holomorphic function on \mathfrak{M} . Identify X and \mathfrak{M} at infinity using the ALH* coordinate system of Proposition 3.3. Fix a cutoff function χ such that $\chi = 1$ for $z > 4R$ and $\chi = 0$ for $z < 2R$. Using the fact that $\Delta_{g_{\mathfrak{M}}} h = 0$ and Proposition 3.3, we have that

$$(3-32) \quad \Delta_g(\chi h) = O(e^{\varepsilon z}) \quad \text{as } z \rightarrow \infty, \text{ for all } \varepsilon > 0.$$

Claim 1 *There exists a function $f \in C^\infty(X)$ such that $\Delta_g f = -\Delta_g(\chi h)$ and*

$$(3-33) \quad f = O(e^{\varepsilon z}) \quad \text{as } z \rightarrow \infty, \text{ for all } \varepsilon > 0.$$

Proof of claim Using [19, Proposition 4.12], we can find a smooth function $f_{\mathfrak{M}}$ defined on \mathfrak{M} such that $f_{\mathfrak{M}} = O(e^{\varepsilon z})$ as $z \rightarrow \infty$ for any $\varepsilon > 0$ and such that $\Delta_{g_{\mathfrak{M}}} f_{\mathfrak{M}} = -\Delta_g h$. Thus, aiming to set $f = \chi f_{\mathfrak{M}} + f'$, we can reduce to solving the equation $\Delta_g f' = h'$ for $f' = O(e^{\varepsilon z})$ (for any $\varepsilon > 0$), where

$$(3-34) \quad h' \equiv -\Delta_g(\chi h) - \Delta_g(\chi f_{\mathfrak{M}}) = O(e^{(-\sqrt{\lambda_1} + \varepsilon)z})$$

for any $\varepsilon > 0$ thanks to Proposition 3.3. In fact, we will now show that for any $\delta > 0$ there exists a $\underline{\delta} > 0$ such that the equation $\Delta_g f' = h'$ with $h' = O(e^{-\delta z})$ is solvable with $f' = \alpha z + O(e^{-\delta z})$ for some $\alpha \in \mathbb{R}$. This implies what we want. To prove this new claim, we first observe that the function z is $g_{\mathfrak{M}}$ -harmonic. Thus, by Proposition 3.3, $\Delta_g(\chi z) = O(e^{(-\sqrt{\lambda_1} + \varepsilon)z})$ for any $\varepsilon > 0$, so in particular $\Delta_g(\chi z) \in L^1(X, d\text{Vol}_g)$. The divergence theorem now allows us to conclude that $\int_X \Delta_g(\chi z) d\text{Vol}_g > 0$.

Indeed, the corresponding boundary integrals in (X, g) and in $(\mathfrak{M}, g_{\mathfrak{M}})$ are asymptotic to each other because z grows more slowly than any exponential in z , and the boundary integral in $(\mathfrak{M}, g_{\mathfrak{M}})$ approaches a positive constant as $z \rightarrow \infty$ by direct computation. Thus, by replacing h' by $h' - \alpha \Delta_g(\chi z)$ for some suitable $\alpha \in \mathbb{R}$, we can assume without loss of generality that h' has mean value zero over X with respect to $d\text{Vol}_g$. It now follows from [17, Theorem 1.5] that the equation $\Delta_g f' = h'$ is solvable for some smooth and uniformly bounded f' with $\int_X |df'|_g^2 d\text{Vol}_g < \infty$. It then also follows from the integration argument in the proof of [18, Proposition 2.9(ia)] that any such f' satisfies the estimate $|\nabla^k(f' - C)|_g = O(e^{-\delta z})$ for some constant C and for any $k \geq 0$ as $z \rightarrow \infty$. After subtracting this constant, the (new) claim follows. \square

Since the function $u \equiv \chi h + f$ is harmonic, the 1-form $\alpha \equiv \bar{\partial}_J u$ satisfies

$$(3-35) \quad \bar{\partial}_J \alpha = 0, \quad \bar{\partial}_{J,g}^* \alpha = 0.$$

Since g is Kähler with respect to J , from Proposition 3.3, we have that

$$(3-36) \quad |J - J_{\mathfrak{M}}|_{g_{\mathfrak{M}}} = O(e^{(-\sqrt{\lambda_1} + \varepsilon)z}) \quad \text{as } z \rightarrow \infty.$$

Using the fact that $h = O(e^{\sqrt{\lambda_1}z})$ and $\bar{\partial}_{J_{\mathfrak{M}}} h = 0$, we can deduce from (3-36) that $\bar{\partial}_J h = O(e^{\varepsilon z})$ for any $\varepsilon > 0$ with respect to $g_{\mathfrak{M}}$. By (3-33) and standard local gradient estimates, the same bound holds for $\bar{\partial}_J f$. Thus, it holds for α as well. Since α has J -type $(0, 1)$, $\text{Re}(\alpha)$ is half-harmonic with respect to g thanks to (3-35). By [19, Theorem 5.1], we conclude that $\alpha \equiv 0$, which implies that u is J -holomorphic. \square

We now complete the proof of part (ii) of Theorem 1.5. Given our work so far, the proof is similar to the proof in [3, Section 4.7]. Let u be the holomorphic function from Proposition 3.5. From (3-30), it follows that all fibers of u near infinity are regular and are diffeomorphic to tori, hence all the regular fibers of u are diffeomorphic to tori, and so $u: X \rightarrow \mathbb{C}$ is an elliptic fibration. The ALH* coordinates define a submanifold near infinity by

$$(3-37) \quad \Sigma_0 = \{(x, y, t, z) : x = 0, t = 0\}.$$

This is clearly a smooth section of the model elliptic fibration over $U \equiv \mathbb{C} \setminus B_R(0)$ for some $R \gg 1$. It is moreover $J_{\mathfrak{M}}$ -holomorphic because for all $p \in \Sigma_0$,

$$(3-38) \quad T_p \Sigma_0 = \text{span}_p \left\{ \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\} = \{Y \in T_p \mathfrak{M} : e_1(Y) = e_4(Y) = 0\},$$

and $\text{span}\{e_1, e_4\}$ is invariant under $J_{\mathfrak{M}}$ by (3-9).

We view the section Σ_0 as a mapping $\sigma_0: U \rightarrow X \setminus X_R$. We next want to perturb σ_0 to a section σ which is holomorphic with respect to J . Given any smooth (not necessarily holomorphic) section σ over U , we can define $\bar{\partial} \sigma \in \Gamma(\Lambda^{0,1}(U) \otimes \sigma^* T^{1,0}(F))$, where $T^{1,0}(F)$ is the $(1, 0)$ part of the vertical tangent bundle, by restricting the differential $\sigma_* \otimes \mathbb{C}$ to $T^{0,1}(U)$, and then projecting to the $(1, 0)$ part. Next, we use the 2-form $\Omega = \omega_1 + \sqrt{-1}\omega_3$, which is holomorphic with respect to J . If we insert the $T^{1,0}(F)$

component of $\bar{\partial}\sigma$ into Ω , we can define $\Omega \odot \bar{\partial}\sigma \in C^\infty(U, \mathbb{C})$, since $\Lambda^{1,0}(U) \otimes \Lambda^{0,1}(U) \cong \Lambda^{1,1}(U)$ is a trivial bundle. Denoting $h(u, \bar{u}) = \Omega \odot \bar{\partial}\sigma_0$, from basic theory of the $\bar{\partial}$ -operator in U , we can solve the equation $(\partial/\partial\bar{u})H = h$ on U . Choose an arbitrary point $p \in U$ and an affine holomorphic fiber coordinate w over a small neighborhood U_p of p in U . Then $\{w = 0\}$ is a local holomorphic section over U_p , and the holomorphic 2-form can be written as $\Omega = f(u) du \wedge dw$, where $f(u)$ is nowhere vanishing. It is easy to see that the smooth local sections

$$(3-39) \quad \sigma_{h,p} \equiv \left(\frac{H(u, \bar{u})}{f(u)} \right) \frac{\partial}{\partial w}$$

over U_p patch up to a well-defined smooth section σ_h over U , independent of the choice of local w coordinate. Consequently, the section $\sigma \equiv \sigma_0 - \sigma_h$ is a holomorphic section defined over all of U .

After identifying U with a punctured disc Δ^* using $z = u^{-1}$ as a holomorphic coordinate, we can then identify the elliptic surface with $(\Delta_z^* \times \mathbb{C}_w)/(\mathbb{Z} \oplus \mathbb{Z})$, with the action given by

$$(3-40) \quad (m, n) \cdot (z, w) = (z, w + mt_1(z) + nt_2(z)), \quad \text{where } (m, n) \in \mathbb{Z} \oplus \mathbb{Z}$$

(t_1, t_2 are the periods), such that $\{w = 0\}$ defines σ ; see for example [18, pages 369–370]. Consequently, there exists a compactification of X to an elliptic surface S such that $X = S \setminus F$, where F is the fiber at infinity. Since the cross-sections are diffeomorphic to nilmanifolds of degree b , the only possibility is that the monodromy is of type I_b . It is easy to see that the form Ω is then a meromorphic 2-form on S with a pole of order 1 along F , which implies that F has multiplicity 1. Since $\text{div}(\Omega) = -F$, we have that $K_S = -[F]$. From Corollary 1.4, $b^1(X) = 0$. A Mayer–Vietoris argument similar to that in the proof of Corollary 1.4 above shows that $b^1(S) = 0$. Arguing exactly as in [4, Theorem 3.3], we see that S is a rational elliptic surface, with projection $u: S \rightarrow \mathbb{P}^1$, so is the blowup of \mathbb{P}^2 in 9 points. Consequently, there exists a (-1) -curve E (the exceptional divisor of the last blowup). The adjunction formula then implies that $K_S \cdot E = -1$, so the condition that $K_S = -[F]$ implies that there are no multiple fibers, and E is a global section; see [15, Proposition 4.1].

References

- [1] **C Birkar**, *Singularities of linear systems and boundedness of Fano varieties*, Ann. of Math. 193 (2021) 347–405 MR Zbl
- [2] **G Chen, X Chen**, *Gravitational instantons with faster than quadratic curvature decay, II*, J. Reine Angew. Math. 756 (2019) 259–284 MR Zbl
- [3] **G Chen, X Chen**, *Gravitational instantons with faster than quadratic curvature decay, I*, Acta Math. 227 (2021) 263–307 MR Zbl
- [4] **G Chen, X Chen**, *Gravitational instantons with faster than quadratic curvature decay, III*, Math. Ann. 380 (2021) 687–717 MR Zbl
- [5] **G Chen, J Viaclovsky**, *Gravitational instantons with quadratic volume growth*, J. Lond. Math. Soc. 109 (2024) art. id. e12886 MR Zbl

- [6] **Y Chen**, *Calabi–Yau metrics of Calabi type with polynomial rate of convergence*, preprint (2024) arXiv 2404.18070
- [7] **C van Coevering**, *A construction of complete Ricci-flat Kähler manifolds*, preprint (2008) arXiv 0803.0112
- [8] **T C Collins, A Jacob, Y-S Lin**, *Special Lagrangian submanifolds of log Calabi–Yau manifolds*, Duke Math. J. 170 (2021) 1291–1375 MR Zbl
- [9] **T C Collins, V Tosatti**, *A singular Demailly–Păun theorem*, C. R. Math. Acad. Sci. Paris 354 (2016) 91–95 MR Zbl
- [10] **R J Conlon, H-J Hein**, *Classification of asymptotically conical Calabi–Yau manifolds*, Duke Math. J. 173 (2024) 947–1015 MR Zbl
- [11] **J-P Demailly**, *Complex analytic and differential geometry*, book manuscript (2012) Available at <http://www-fourier.ujf-grenoble.fr/~demailly/manuscripts/agbook.pdf>
- [12] **I V Dolgachev**, *Classical algebraic geometry: a modern view*, Cambridge Univ. Press (2012) MR Zbl
- [13] **H Grauert**, *Über Modifikationen und exzeptionelle analytische Mengen*, Math. Ann. 146 (1962) 331–368 MR Zbl
- [14] **H Grauert, O Riemenschneider**, *Verschwindungssätze für analytische Kohomologiegruppen auf komplexen Räumen*, Invent. Math. 11 (1970) 263–292 MR Zbl
- [15] **B Harbourne, W E Lang**, *Multiple fibers on rational elliptic surfaces*, Trans. Amer. Math. Soc. 307 (1988) 205–223 MR Zbl
- [16] **M Haskins, H-J Hein, J Nordström**, *Asymptotically cylindrical Calabi–Yau manifolds*, J. Differential Geom. 101 (2015) 213–265 MR Zbl
- [17] **H-J Hein**, *Weighted Sobolev inequalities under lower Ricci curvature bounds*, Proc. Amer. Math. Soc. 139 (2011) 2943–2955 MR Zbl
- [18] **H-J Hein**, *Gravitational instantons from rational elliptic surfaces*, J. Amer. Math. Soc. 25 (2012) 355–393 MR Zbl
- [19] **H-J Hein, S Sun, J Viaclovsky, R Zhang**, *Nilpotent structures and collapsing Ricci-flat metrics on the $K3$ surface*, J. Amer. Math. Soc. 35 (2022) 123–209 MR Zbl
- [20] **C D Hill, M Taylor**, *Integrability of rough almost complex structures*, J. Geom. Anal. 13 (2003) 163–172 MR Zbl
- [21] **L Hörmander**, *An introduction to complex analysis in several variables*, 3rd edition, North-Holland Math. Libr. 7, North-Holland, Amsterdam (1990) MR Zbl
- [22] **Y Kawamata**, *Pluricanonical systems on minimal algebraic varieties*, Invent. Math. 79 (1985) 567–588 MR Zbl
- [23] **J Kollár, Y Miyaoka, S Mori, H Takagi**, *Boundedness of canonical \mathbb{Q} -Fano 3-folds*, Proc. Japan Acad. Ser. A Math. Sci. 76 (2000) 73–77 MR Zbl
- [24] **J Kollár, S Mori**, *Birational geometry of algebraic varieties*, Cambridge Tracts in Math. 134, Cambridge Univ. Press (1998) MR Zbl
- [25] **P B Kronheimer**, *A Torelli-type theorem for gravitational instantons*, J. Differential Geom. 29 (1989) 685–697 MR Zbl
- [26] **C Li**, *On sharp rates and analytic compactifications of asymptotically conical Kähler metrics*, Duke Math. J. 169 (2020) 1397–1483 MR Zbl

- [27] **V Minerbe**, *Rigidity for multi-Taub-NUT metrics*, J. Reine Angew. Math. 656 (2011) 47–58 MR Zbl
- [28] **M V Nori**, *Zariski’s conjecture and related problems*, Ann. Sci. École Norm. Sup. 16 (1983) 305–344 MR Zbl
- [29] **T Ohsawa**, *Vanishing theorems on complete Kähler manifolds*, Publ. Res. Inst. Math. Sci. 20 (1984) 21–38 MR Zbl
- [30] **I Shimada**, *Remarks on fundamental groups of complements of divisors on algebraic varieties*, Kodai Math. J. 17 (1994) 311–319 MR Zbl
- [31] **S Sun, R Zhang**, *Complex structure degenerations and collapsing of Calabi–Yau metrics*, preprint (2019) arXiv 1906.03368
- [32] **S Sun, R Zhang**, *A Liouville theorem on asymptotically Calabi spaces*, Calc. Var. Partial Differential Equations 60 (2021) art. id. 103 MR Zbl
- [33] **S Sun, R Zhang**, *Collapsing geometry of hyperkähler 4-manifolds and applications*, Acta Math. 232 (2024) 325–424 MR Zbl
- [34] **S Takayama**, *Simple connectedness of weak Fano varieties*, J. Algebraic Geom. 9 (2000) 403–407 MR Zbl
- [35] **G Tian, S-T Yau**, *Complete Kähler manifolds with zero Ricci curvature, I*, J. Amer. Math. Soc. 3 (1990) 579–609 MR Zbl
- [36] **C Voisin**, *Coniveau 2 complete intersections and effective cones*, Geom. Funct. Anal. 19 (2010) 1494–1513 MR Zbl
- [37] **S T Yau**, *On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equation, I*, Comm. Pure Appl. Math. 31 (1978) 339–411 MR Zbl
- [38] **S T Yau**, *The role of partial differential equations in differential geometry*, from “Proceedings of the International Congress of Mathematicians, I” (O Lehto, editor), Acad. Sci. Fennica, Helsinki (1980) 237–250 MR Zbl

H-JH: *Mathematisches Institut, Universität Münster
Münster, Germany*

SS: *Institute for Advanced Study in Mathematics, Zhejiang University
Hangzhou, China*

JV: *Department of Mathematics, University of California, Irvine
Irvine, CA, United States*

RZ: *Department of Mathematics, University of California
San Diego, CA, United States*

hhein@uni-muenster.de, songsun@zju.edu.cn, jviaclov@uci.edu, rzhang573@wisc.edu

Proposed: Simon Donaldson

Received: 12 September 2023

Seconded: Tobias H Colding, Mark Gross

Revised: 19 July 2024

The distribution of critical graphs of Jenkins–Strebel differentials

FRANCISCO ARANA-HERRERA

AARON CALDERON

By work of Jenkins and Strebel, given a Riemann surface X and a simple closed multicurve α on it, there exists a unique quadratic differential q on X whose horizontal foliation is measure equivalent to α . We study the distribution of the critical graphs of these differentials in the moduli space of metric ribbon graphs as the extremal length of the multicurves goes to infinity, showing they equidistribute to the Kontsevich measure regardless of the initial choice of X .

32G15, 57K20

1. Introduction	2571
2. Preliminaries	2576
3. Near the multiple zero locus	2581
4. Uniform geometric estimates	2585
5. Horoball measures	2589
6. Equidistribution of critical graphs	2597
References	2606

1 Introduction

Overview Famous independent works of Jenkins [1957] and Strebel [1966; 1975; 1976] show that, given a Riemann surface X and a simple closed multicurve α on it, there exists a unique quadratic differential q on X whose horizontal foliation represents each curve of α by a cylinder of height 1. To every such quadratic differential q one can associate its *critical graph*, a ribbon graph having the singularities of q as vertices and the horizontal saddle connections of q as edges. This ribbon graph inherits a metric from the singular flat metric induced by q on X . Roughly speaking, the critical graph encodes the conformal geometry of X away from α ; see Figure 1.

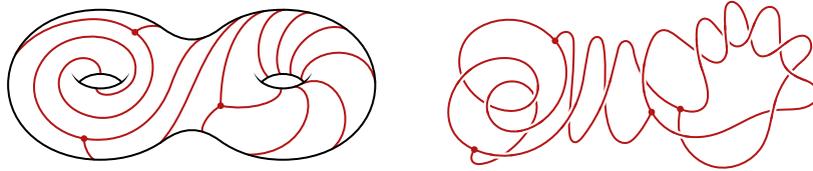


Figure 1: A Jenkins–Strebel differential and its critical graph.

By work of Mirzakhani¹ [2008b], given a closed connected oriented Riemann surface X , the number of simple closed (multi)curves on X of extremal length $\leq L^2$ is asymptotic to a polynomial of degree $6g - 6$. Each one of these curves gives rise to a critical graph as described above. We show these critical graphs, appropriately rescaled, equidistribute to the Kontsevich measure on the moduli space of metric ribbon graphs. In particular, the limiting distribution is independent of the initial choice of X .

The question at hand is heavily motivated by analogous results in the setting of hyperbolic surfaces. In [Arana-Herrera and Calderon 2022], the authors showed that, given a closed connected oriented hyperbolic surface X , the metric ribbon graph spines of complementary subsurfaces to simple closed multigeodesics also equidistribute to the Kontsevich measure on the corresponding moduli space as the lengths of the geodesics goes to infinity. This result in turn follows a line of investigation that can be traced back to several other authors [Mirzakhani 2016; Liu 2022; Arana-Herrera 2022; Erlandsson and Souto 2022]. Analogous questions have also been answered in the setting of homogeneous dynamics [Aka et al. 2016a; 2016b; Einsiedler et al. 2019].

Paralleling our approach in [Arana-Herrera and Calderon 2022], we reduce the problem to an equidistribution question concerning the dynamics of the Teichmüller horocycle flow on moduli spaces of quadratic differentials. These reductions combine Margulis’s well-known averaging and unfolding techniques [1970] with several recent developments in Teichmüller theory. Of particular importance are Delaunay triangulations of quadratic differentials following Masur and Smillie [1991], the AGY-metric on the moduli space of quadratic differentials developed by Avila, Gouëzel, and Yoccoz [Avila et al. 2006; Avila and Gouëzel 2013], and the study of the projection of the Masur–Veech measure to the moduli space of Riemann surfaces carried out by Athreya, Bufetov, Eskin, and Mirzakhani [Athreya et al. 2012].

Our main result also provides a new procedure for sampling random metric ribbon graphs. In particular, the geometry of any single conformal surface reflects the geometry of random metric ribbon graphs.

Main result To streamline our exposition, for the moment we only state our main result in the (representative) case of nonseparating simple closed curves. The statement for the general case appears as Theorem 6.2 below.

¹Although not explicitly stated, this follows directly from [Mirzakhani 2008b, Theorem 1.3]. See [Martínez-Granado and Thurston 2021, Corollary 5.13] for another proof from the viewpoint of geodesic currents.

For the rest of this discussion fix an integer $g \geq 2$ and denote by S_g a closed connected oriented surface of genus g . Let Mod_g be the mapping class group of S_g , \mathcal{T}_g be the Teichmüller space of marked conformal structures on S_g , and \mathcal{M}_g be the moduli space of conformal structures on S_g . Free homotopy classes of unoriented simple closed curves on S_g will be referred to as simple closed curves. Given a simple closed curve α on S_g and a marked conformal structure $X \in \mathcal{T}_g$, denote by $\text{Ext}_X(\alpha) > 0$ the extremal length of α with respect to X ; see Section 2 below for a discussion of the different definitions of extremal length.

Let γ be a nonseparating simple closed curve on S_g and $X \in \mathcal{T}_g$ be a marked conformal structure on S_g . For every $L > 0$ consider the counting function

$$s(X, \gamma, L) := \#\{\alpha \in \text{Mod}_g \cdot \gamma \mid \text{Ext}_X(\alpha) \leq L^2\}.$$

This function does not depend on the marking of $X \in \mathcal{T}_g$ but only on its underlying conformal structure $X \in \mathcal{M}_g$. Indeed, it is equal to the number of nonseparating simple closed curves on X of extremal length $\leq L^2$. By Mirzakhani’s seminal work [2008b], this counting function is asymptotically polynomial (see (6-1) below).

Given $X \in \mathcal{T}_g$ and a nonseparating simple closed curve α on S_g , denote by $\text{JS}(X, \alpha)$ the unique Jenkins–Strebel differential on X whose horizontal measured foliation is equivalent to α (where the curve is given weight 1); in particular, the horizontal foliation of $\text{JS}(X, \alpha)$ consists of a single cylinder of height 1. Consider the *critical graph* of this differential, that is, the union of all of the singular horizontal trajectories of $\text{JS}(X, \alpha)$ equipped with the restriction of the underlying singular flat metric and the ribbon structure coming from its embedding in S_g . Each of its boundaries has length $\text{Ext}_X(\alpha)$, so to understand the distribution of these graphs we need to rescale them (see Section 2).

Denote by $\mathcal{MRG}_{g-1,2}(1, 1)$ the moduli space of metric ribbon graphs of genus $g - 1$ with two boundary components, each of length 1. For X and α as above let

$$\Xi^1(X, \alpha) \in \mathcal{MRG}_{g-1,2}(1, 1)$$

denote the critical graph of $\text{JS}(X, \alpha)$ rescaled by $1/\text{Ext}_X(\alpha)$. We remark that $\Xi^1(X, \alpha)$ is *not* the same as the critical graph of the unit-area Jenkins–Strebel differential whose horizontal foliation is projectively equivalent to α ; this is because critical graphs scale linearly while extremal length scales quadratically.

On $\mathcal{MRG}_{g-1,2}(1, 1)$ consider the counting measure

$$\eta_{X,\gamma}^L := \sum_{\alpha \in \text{Mod}_g \cdot \gamma} \mathbb{1}_{[0,L^2]}(\text{Ext}_X(\alpha)) \delta_{\Xi^1(X,\alpha)}.$$

Just like the counting function $s(X, \gamma, L)$, this measure does not depend on the marking of $X \in \mathcal{T}_g$ but only on the underlying conformal structure. Denote by η_{Kon} the measure induced by the Kontsevich symplectic form on $\mathcal{MRG}_{g-1,2}(1, 1)$ and by $c_g > 0$ its total mass (see Section 5 below or [Arana-Herrera and Calderon 2022, Section 2]).

The following theorem, which shows that the rescaled critical graphs of Jenkins–Strebel differentials of nonseparating simple closed curves equidistribute over the moduli space $\mathcal{MRG}_{g-1,2}(1, 1)$, is an instance of our main result. For the general version see Theorem 6.2, as well as Theorem 6.12 for an even stronger version concerning simultaneous equidistribution.

Theorem 1.1 *Let γ be a nonseparating simple closed curve and $X \in \mathcal{M}_g$. Then*

$$\lim_{L \rightarrow \infty} \frac{\eta_{X,\gamma}^L}{s(X, \gamma, L)} = \frac{\eta_{\text{Kon}}}{c_g},$$

with respect to the weak- \star topology for measures on $\mathcal{MRG}_{g-1,2}(1, 1)$.

Main ideas of proof Our proof of Theorem 1.1 follows the same outline as the analogous result in the hyperbolic setting [Arana-Herrera and Calderon 2022, Theorem 1.1]. Namely, we reduce the desired equidistribution result on $\mathcal{MRG}_{g-1,2}(1, 1)$ to a curve counting result, which we then translate back into an equidistribution result over \mathcal{M}_g .

First, consider the following reformulation. Let $f : \mathcal{MRG}_{g-1,2}(1, 1) \rightarrow \mathbb{R}_{\geq 0}$ be a continuous compactly supported function. Then, for any nonseparating simple closed curve γ , any $X \in \mathcal{T}_g$, and any $L > 0$, define the f -weighted counting

$$c(X, \gamma, f, L) := \sum_{\alpha \in \text{Mod}_g \cdot \gamma} \mathbb{1}_{[0,L^2]}(\text{Ext}_X(\alpha)) f(\Xi^1(X, \alpha)) = \int_{\mathcal{MRG}_{g-1,2}(1,1)} f(\mathbf{x}) d\eta_{X,\gamma}^L(\mathbf{x}).$$

Theorem 1.1 is then equivalent to the following counting result:

Theorem 1.2 *Let γ be a nonseparating simple closed curve on S_g , $X \in \mathcal{M}_g$ be a conformal structure on S_g , and $f : \mathcal{MRG}_{g-1,2}(1, 1) \rightarrow \mathbb{R}_{\geq 0}$ be a continuous compactly supported function. Then*

$$\lim_{L \rightarrow \infty} \frac{c(X, \gamma, f, L)}{s(X, \gamma, L)} = \frac{1}{c_g} \int_{\mathcal{MRG}_{g-1,2}(1,1)} f(\mathbf{x}) d\eta_{\text{Kon}}(\mathbf{x}).$$

Once in this setting, we apply Margulis’s “averaging and unfolding” techniques [1970] to reduce the counting problem at hand to an equidistribution question over \mathcal{M}_g . Just as in [Arana-Herrera and Calderon 2022], during the averaging step one needs uniform control over how the metric ribbon graph $\Xi^1(X, \alpha)$ varies as $X \in \mathcal{T}_g$ does. The basic issue is that if $\text{JS}(X, \alpha)$ lies in (or near) a nonprincipal stratum, then small deformations of X can change the topology of the critical graph. To remedy this, we prove in Proposition 3.6 that for most α , the differential $\text{JS}(X, \alpha)$ lies far away from nonprincipal strata. Proposition 4.3 then shows that for such α we can achieve the desired uniform control on the geometry of the critical graph.

This control allows us to perform the averaging and unfolding step of our argument, after which our original problem reduces to a question regarding the equidistribution of certain subsets, which we call “critical-JS-horoballs”, in the moduli space \mathcal{M}_g . To tackle this question we use the rich dynamics of the

Teichmüller geodesic and horocycle flows. More concretely, we use work of Forni [2021] which in turn relies crucially on work of Eskin, Mirzakhani, and Mohammadi [Eskin and Mirzakhani 2018; Eskin et al. 2015].

Related results As mentioned above, our main result directly parallels the main theorem of [Arana-Herrera and Calderon 2022]. Indeed, as a consequence of our results, we find that the limiting distribution of complementary subsurfaces to hyperbolic geodesics and critical graphs of Jenkins–Strebel differentials are exactly the same. This is a reflection of the phenomenon that as the boundary lengths of hyperbolic surfaces go to infinity, they look more and more like ribbon graphs; compare to [Do 2010; Mondello 2009b]. There is a rich history of using both the conformal and hyperbolic incarnations of moduli space to probe its structure, often resulting in analogous theorems. Of particular relevance here are two different identifications of the moduli spaces of punctured/bordered Riemann surfaces with the space of metric ribbon graphs. The first, due to Harer, Mumford, Penner, and Thurston, uses Jenkins–Strebel differentials with specified poles and residues; this identification was used in the computation of the orbifold Euler characteristic of $\mathcal{M}_{g,n}$ [Harer and Zagier 1986] as well as Kontsevich’s proof [1992] of Witten’s conjecture. The second, due to Do [2010] and Luo [2007], uses the spine of a hyperbolic surface with boundary; see also [Mondello 2009b]. This can be used to unite Mirzakhani’s proof [2007] of Witten’s conjecture with Kontsevich’s; see [Do 2010]. Our previous work [Arana-Herrera and Calderon 2022] dealt with this second identification, while the paper at hand corresponds to the first paradigm.

Our theorem also echoes other equidistribution results on moduli space. The Teichmüller geodesic flow, represented by the action of the diagonal matrix

$$g_t := \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix},$$

expands the length of horizontal separatrices by a factor of e^t , so for every simple closed curve α the rescaled critical graph $\Xi^1(X, \alpha)$ is the same as the critical graph of $g_{-\log \text{Ext}_X(\alpha)}$ JS(X, α). Denote by $\|q\|$ the area of a quadratic differential q . With this identification, Theorem 1.1 states that point masses on the critical graphs of

$$(1-1) \quad \{g_{-\log\|q\|}q \mid q = \text{JS}(X, \alpha) \text{ for } \alpha \in \text{Mod}_g \cdot \gamma \text{ and } \|q\| \leq L\}$$

equidistribute in $\mathcal{MRG}_{g-1,2}(1, 1)$ as L tends to infinity.

This should be compared with work of [Athreya et al. 2012] and the first author [Arana-Herrera 2023], which give mean equidistribution theorems for expanding Teichmüller balls in moduli space. More precisely, for every $X \in \mathcal{T}_g$, one can consider the underlying Riemann surfaces of the expanding ball

$$(1-2) \quad \{g_{\log\|q\|}q \mid q \in Q(X) \text{ and } \|q\| \leq L\},$$

where $Q(X)$ is the vector space of holomorphic quadratic differentials on X ; this is the same as the set of points with Teichmüller distance at most $\log L$ from X . Theorem 1.2 of [Athreya et al. 2012] then states that on average, the images of these balls in \mathcal{M}_g equidistribute as L tends to infinity. The link between

these equidistribution results for (1-1) and (1-2) is the uniform distribution of simple closed curves on the space \mathcal{MF}_g of singular measured foliations on S_g [Mirzakhani 2008b, Theorem 1.3].

Lastly, it is interesting to contrast our theorem with the main result of [Dozier and Sapir 2021], where it is shown that the projections of some strata of $\mathcal{Q}^1\mathcal{M}_g$ to \mathcal{M}_g are not coarsely dense; in particular, the geometry of the critical graphs of horizontally periodic unit-area differentials may be strongly constrained. For example, if $X \in \mathcal{M}_g$ is not near the projection of the minimal stratum then it supports no horizontally periodic unit-area differential whose critical graph has a vertex of valence $4g - 2$, and supports no differential whose critical graph is close to such.

On the other hand, our main theorem implies that the set of critical graphs for a fixed X , rescaled to have unit length, is dense in the space of metric ribbon graphs. Since critical graphs are scaled under the Teichmüller geodesic flow, our result can also be interpreted as saying that any X supports a Jenkins–Strebel differential $q = \text{JS}(X, \alpha)$ such that $g_{-\log\|q\|}q$ is arbitrarily close to the minimal stratum.

Organization In Section 2 we collect preliminary material needed to understand the statements and proofs of our main results. In addition to standard background material, we focus on how taking the critical graph of a Jenkins–Strebel differential gives a map to the moduli space of metric ribbon graphs. In Section 3 we show that the Jenkins–Strebel differentials of most simple closed multicurves stay deep within the principal stratum; this allows us in the subsequent Section 4 to invoke results about the AGY metric to show that for most curves, the critical graphs of their Jenkins–Strebel differentials vary uniformly as the base surface varies. In Section 5 we define “critical-JS-horoballs”, compute their total mass, and show that they equidistribute over moduli space. The results in this section rely on transporting measures between \mathcal{T}_g and spaces of quadratic differentials, and leverage the ergodic theory of the $\text{SL}(2, \mathbb{R})$ action on the latter spaces. Finally, in Section 6 we apply Margulis’s averaging and unfolding strategy to prove Theorem 6.2, a generalization of Theorem 1.1 to arbitrary multicurves.

Acknowledgements The authors would like to thank Vincent Delecroix for suggesting the question answered in this paper and Curt McMullen for helpful comments. The authors would also like to thank Giovanni Forni for very enlightening conversations. Calderon acknowledges support from NSF grant DMS-2202703.

2 Preliminaries

Outline of this section We give of a brief overview of the basic objects and theorems that will be used throughout the rest of this paper. We begin with a discussion on the theory of Jenkins–Strebel differentials with a special emphasis on the work of Hubbard and Masur [1979]. We then briefly recall the moduli spaces of metric ribbon graphs; for a more detailed discussion of these spaces see [Arana-Herrera and Calderon 2022]. Lastly, we discuss in detail how the construction of critical graphs of Jenkins–Strebel differentials defines a map from (quotients of) Teichmüller space to appropriate moduli spaces of metric ribbon graphs.

Extremal length Given a Riemann surface X and a simple closed curve γ on it, the extremal length of γ with respect to X admits two equivalent definitions. First, it can be defined analytically as

$$(2-1) \quad \text{Ext}_X(\gamma) := \sup_{\rho} \frac{\ell_{\rho}(\gamma)^2}{\text{Area}(\rho)},$$

where the supremum ranges over all conformal metrics ρ on X of nonzero finite area and $\ell_{\rho}(\gamma)$ denotes the infimum of the ρ -lengths of simple closed curves isotopic to γ . Equivalently, it can be defined geometrically as

$$(2-2) \quad \text{Ext}_X(\gamma) := \inf_C \frac{1}{\text{mod}(C)},$$

where the infimum ranges over all embedded cylinders C on X with core curve isotopic to γ and $\text{mod}(C)$ denotes the modulus of the cylinder C .

In independent work, Jenkins [1957] and Strebel [1966; 1975; 1976] showed these two a priori different notions of extremal length are equivalent through the construction of so-called Jenkins–Strebel differentials; see Theorem 2.1 below.

Quadratic differentials A (holomorphic) quadratic differential q on a Riemann surface X is a differential which in local coordinates has the form $f(z) dz^2$ for some holomorphic function $f(z)$. Such a differential has a well defined notion of area,

$$\|q\| := \text{Area}(q) := \int_X |q|.$$

More precisely, the differential q induces a singular flat metric on X . If in local coordinates $z = x + iy$, then the metric is given by $dx^2 + dy^2$; the zeroes of the differential correspond to singularities of the metric. The area of q is the total area of this metric. We denote by $Q(X)$ the complex vector space of all quadratic differentials on a Riemann surface X , and by $S(X) \subseteq Q(X)$ the sphere of all unit-area quadratic differentials on X . We sometimes denote quadratic differentials by (X, q) to record the Riemann surface X on which they are defined.

The spaces $Q(X)$ and $S(X)$ for X ranging over the Teichmüller space \mathcal{T}_g can be arranged into bundles $Q\mathcal{T}_g$ and $Q^1\mathcal{T}_g$ over \mathcal{T}_g of marked (unit-area) quadratic differentials; note that “marking” here refers to a marking only of the underlying surface S_g and not the zeros of the quadratic differential. These bundles support natural Mod_g actions given by change of markings. The quotient $Q^1\mathcal{M}_g := Q^1\mathcal{T}_g/\text{Mod}_g$ is the bundle of unit-area quadratic differentials over moduli space.

The bundle $Q\mathcal{T}_g$ carries a natural Lebesgue-class measure called the *Masur–Veech measure*, which is induced from an integral lattice, corresponding to differentials tiled by unit-area squares. This measure induces a measure ν_{MV} on the hypersurface $Q^1\mathcal{T}_g$ also called the Masur–Veech measure. Both versions of this measure are Mod_g invariant, and the corresponding pushforward $\hat{\nu}_{\text{MV}}$ on $Q^1\mathcal{M}_g$ is a finite Lebesgue-class measure. Throughout the paper we will denote its mass by b_g .

Singular measured foliations Denote by \mathcal{MF}_g the space of singular measured foliations on S_g up to isotopy and Whitehead moves. The set of isotopy classes of weighted simple closed curves on S_g embeds densely into \mathcal{MF}_g . Furthermore, geometric intersection number extends continuously to a pairing on \mathcal{MF}_g . Train track coordinates (see Section 3) induce a natural integral piecewise-linear structure on \mathcal{MF}_g . In particular, \mathcal{MF}_g carries a natural Lebesgue class measure μ_{Thu} called the Thurston measure, which is invariant under the natural Mod_g -action on \mathcal{MF}_g . Denote by \mathcal{PMF}_g the projectivization of \mathcal{MF}_g under the action that scales transverse measures and by $[\lambda] \in \mathcal{PMF}_g$ the projective class of $\lambda \in \mathcal{MF}_g$.

Every quadratic differential q on a Riemann surface X gives rise to a pair of singular measured foliations $\mathfrak{R}(q)$ and $\mathfrak{S}(q)$ on X . If in local coordinates $z = x + iy$ the differential q corresponds to dz^2 , then $\mathfrak{R}(q)$ corresponds to the measured foliation induced by $|dx|$ while $\mathfrak{S}(q)$ corresponds to the measured foliation induced by $|dy|$; the zeroes of q correspond to the singularities of the foliations. We refer to $\mathfrak{R}(q)$ and $\mathfrak{S}(q)$ as the vertical and horizontal foliations of q . These constructions give rise to Mod_g -equivariant maps $\mathfrak{R}, \mathfrak{S}: \mathcal{QT}_g \rightarrow \mathcal{MF}_g$.

Jenkins–Strebel differentials It is natural to ask whether, given a marked Riemann surface $X \in \mathcal{T}_g$ and a simple closed curve γ on S_g , it is possible to find a marked quadratic differential $q \in Q(X)$ such that $\mathfrak{S}(q) = \gamma$. Jenkins [1957] and Strebel [1966; 1975; 1976] independently showed that this is always possible and, moreover, in a unique way. We refer to the corresponding quadratic differential $\text{JS}(X, \gamma) \in Q(X)$ as the *Jenkins–Strebel differential* of γ on X . Jenkins and Strebel’s motivation can be further understood through the following result:

Theorem 2.1 *Given a marked complex structure $X \in \mathcal{T}_g$ and a simple closed curve γ on S_g , the singular flat metric induced by $\text{JS}(X, \gamma) \in Q(X)$ realizes the supremum in (2-1). Furthermore, the complement of the critical leaves of vertical foliation of $\text{JS}(X, \gamma) \in Q(X)$ is a cylinder realizing the infimum in (2-2). In particular, the supremum and infimum in (2-1) and (2-2) are equal.*

The Hubbard–Masur theorem The question addressed by the works of Jenkins and Strebel can also be considered for more general singular measured foliations. Indeed, Hubbard and Masur [1979] proved the following structural result:

Theorem 2.2 *Given $X \in \mathcal{T}_g$ and a singular measured foliation $\lambda \in \mathcal{MF}_g$, there exists a unique quadratic differential $q = q(X, \lambda) \in Q(X)$ such that $\mathfrak{S}(q) = \lambda$. Furthermore, the map $q \in Q(X) \mapsto \mathfrak{S}(q) \in \mathcal{MF}_g$ is a homeomorphism.*

Said another way, the map associating to a simple closed curve its Jenkins–Strebel differential extends to a homeomorphism $\text{JS}: \mathcal{T}_g \times \mathcal{MF}_g \rightarrow \mathcal{QT}_g$.

Theorem 2.2 suggests the following extension of the notion of extremal length to singular measured foliations: the extremal length $\text{Ext}_X(\lambda)$ of $\lambda \in \mathcal{MF}_g$ with respect to $X \in \mathcal{T}_g$ is the area of $\text{JS}(X, \lambda) \in Q(X)$. With this definition, extremal length is 2-homogeneous with respect to the scaling action on transverse

measures; this prompts us to usually work with the square root of extremal length rather than with extremal length itself. For other methods of extending extremal lengths to singular measured foliations (and more general objects) see [Kerckhoff 1980; Martínez-Granado and Thurston 2021; Arana-Herrera 2024].

One can also ask which pairs of singular measured foliations can arise as the vertical and horizontal foliations of quadratic differentials. It turns out that as long as the pair of foliations appropriately fills the surface, this is always possible [Gardiner and Masur 1991]. A pair of singular measured foliations $(\lambda, \mu) \in \mathcal{MF}_g \times \mathcal{MF}_g$ is said to *fill* the surface S_g if their geometric intersection number with any singular measured foliation is positive. Denote by $\Delta \subseteq \mathcal{MF}_g \times \mathcal{MF}_g$ the set of pairs of nonfilling singular measured foliations.

Theorem 2.3 *Given a pair of filling singular measured foliations $\mu, \lambda \in \mathcal{MF}_g$, there exists a unique quadratic differential $q \in \mathcal{QT}_g$ such that $\Re(q) = \mu$ and $\Im(q) = \lambda$. Furthermore, we have that the map $(\Re, \Im): \mathcal{QT}_g \rightarrow \mathcal{MF}_g \times \mathcal{MF}_g \setminus \Delta$ is a homeomorphism.*

We remark that, under this isomorphism, the product measure $\mu_{\text{Thu}} \times \mu_{\text{Thu}}$ coincides with ν_{MV} up to a constant.

The Teichmüller metric The Teichmüller metric d_{Teich} on \mathcal{T}_g quantifies the minimal dilation among quasiconformal maps between marked complex structures on S_g . More precisely, for every $X, Y \in \mathcal{T}_g$,

$$d_{\text{Teich}}(X, Y) := \frac{1}{2} \log \left(\inf_{f: X \rightarrow Y} K(f) \right),$$

where the infimum runs over all quasiconformal maps $f: X \rightarrow Y$ in the homotopy class given by the markings of X and Y , and where $K(f)$ denotes the dilation of such maps. See [Farb and Margalit 2012, Chapter 11] for a more detailed definition. The action of the (extended) mapping class group on Teichmüller space is the isometry group of the Teichmüller metric [Royden 1971]. The Teichmüller metric is complete and its geodesics can be described explicitly in terms of the diagonal part of the natural $\text{SL}(2, \mathbb{R})$ -action on $Q^1\mathcal{T}_g$.

Kerckhoff [1980] proved the following formula for the Teichmüller metric; this formula provides control over the extremal lengths of singular measured foliations in terms of the Teichmüller distance between two marked complex structures.

Theorem 2.4 *For any pair of marked complex structures $X, Y \in \mathcal{T}_g$,*

$$(2-3) \quad d_{\text{Teich}}(X, Y) = \frac{1}{2} \max_{\lambda \in \mathcal{MF}_g} \log \left(\frac{\text{Ext}_Y(\lambda)}{\text{Ext}_X(\lambda)} \right).$$

Furthermore, if $q = q(X, Y) \in \mathcal{S}(X)$ is the unit-area quadratic differential corresponding to the unique Teichmüller geodesic from X to Y , then $\Im(q)$ realizes the maximum in (2-3).

The fact that $\mathfrak{S}(q)$ realizes the supremum (as opposed to $\mathfrak{R}(q)$) is due to the fact that Teichmüller mappings stretch the leaves of $\mathfrak{S}(q)$ while shrinking its measure. For example, suppose $q \in S(X)$; then $\lambda = \mathfrak{S}(q)$ has unit extremal length. Then its image $g_t q$ under the Teichmüller geodesic flow still has area 1 and $\mathfrak{S}(g_t q) = e^{-t} \lambda$. Thus λ has extremal length e^{2t} on $g_t q$.

Metric ribbon graphs A ribbon graph is the combinatorial data of a deformation retraction of a surface with boundary. More formally, it is a (simplicial) graph Γ equipped with a cyclic ordering of the edges at each vertex; this can be reconciled with the first notion by thickening each edge to a ribbon and gluing the edges of the ribbons according to the cyclic ordering. The genus and number of boundary components of Γ are the values for the resulting topological surface. One may equip a ribbon graph Γ with a metric \mathbf{x} assigning a length to each of its edges to obtain a so-called metric ribbon graph (Γ, \mathbf{x}) .

For $b \geq 0$ a nonnegative integer, denote by $\mathcal{MRG}_{g,b}$ the moduli space of all metric ribbon graphs with genus g and b distinctly labeled boundary components, all of whose vertices have valence at least three. This space is built out of cones corresponding to trivalent ribbon graphs, glued along faces corresponding to shared degenerations. See [Mondello 2009a] for a careful description of the topology of this space. For a given tuple $\mathbf{L} = (L_1, \dots, L_b)$ of positive numbers, define $\mathcal{MRG}_{g,b}(\mathbf{L}) \subseteq \mathcal{MRG}_{g,b}$ to be the subset of all ribbon graphs whose (labeled) boundary components have lengths L_1, \dots, L_b . Each slice is homeomorphic to the usual moduli space of b -pointed Riemann surfaces $\mathcal{M}_{g,b}$. The slices $\mathcal{MRG}_{g,b}(\mathbf{L})$ piece together to form a fibration $\mathcal{MRG}_{g,b} \rightarrow \mathbb{R}_{>0}^b$.

Critical graphs Given a marked complex structure $X \in \mathcal{T}_g$ and a simple closed curve γ on S_g , the critical graph of the Jenkins–Strebel differential $q = \text{JS}(X, \gamma) \in \mathcal{Q}(X)$ is the metric ribbon graph obtained as the union of the critical leaves of $\mathfrak{S}(q) \in \mathcal{MF}_g$ endowed with the restriction of the singular flat metric induced by q on X . A similar construction can be carried out for general multicurves, but to get a well defined map to a corresponding moduli space of metric ribbon graphs, care needs to be exercised with regards to the symmetries of the construction.

For the rest of this discussion let $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ be an ordered oriented simple closed multicurve on S_g , let $\text{Stab}_0(\vec{\gamma}) \subseteq \text{Mod}_g$ be its oriented stabilizer, ie the set of mapping classes that fix each component of $\vec{\gamma}$ together with its orientation, and let $\gamma \in \mathcal{MF}_g$ be its equivalence class as a singular measured foliation. Cutting S_g along the components of $\vec{\gamma}$ yields a (possibly disconnected) surface with boundary $S_g \setminus \vec{\gamma}$. Label the components of $S_g \setminus \vec{\gamma}$ by $(\Sigma_j)_{j=1}^c$; such a labeling is possible because the components of γ are labeled and oriented. For each $j \in \{1, \dots, c\}$ let $g_j, b_j \geq 0$ be nonnegative integers such that Σ_j is homeomorphic to S_{g_j, b_j} .

For any length vector $\mathbf{L} \in \mathbb{R}_{>0}^k$ we set

$$\mathcal{MRG}(S_g \setminus \vec{\gamma}; \mathbf{L}) := \prod_{j=1}^c \mathcal{MRG}_{g_j, b_j}(\mathbf{L}^{(j)}),$$

where $\mathbf{L}^{(j)}$ denotes the lengths corresponding to the boundary components of Σ_j .

These slices fit together into a larger moduli space $\mathcal{MRG}(S_g \setminus \vec{\gamma})$ which can be topologized through its natural embedding into a product of combinatorial moduli spaces with variable boundary lengths. Denote by $\Delta \subseteq \mathbb{R}_{>0}^k$ the open simplex

$$\Delta := \{\mathbf{L} \in \mathbb{R}_{>0}^k \mid L_1 + \cdots + L_k = 1\}$$

and let $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ denote the total space of the fibration over Δ whose fiber above $\mathbf{L} \in \Delta$ is $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \mathbf{L})$; this can be thought of as the “projectivization” of the full moduli space $\mathcal{MRG}(S_g \setminus \vec{\gamma})$ under the natural rescaling action.

For any marked complex structure $X \in \mathcal{T}_g$, the critical graph of $\text{JS}(X, \gamma)$ is the metric ribbon graph obtained as the union of the critical leaves of $\mathfrak{S}(q) \in \mathcal{MF}_g$ endowed with the restriction of the singular flat metric induced by q on X . Notice this graph has one connected component for each component of $S_g \setminus \vec{\gamma}$. Using the labeling of these components one can define the *critical graph map*

$$\Xi(\cdot, \vec{\gamma}): \mathcal{T}_g \rightarrow \mathcal{MRG}(S_g \setminus \vec{\gamma}).$$

Furthermore, after rescaling the metric of the critical graph $\Xi(X, \vec{\gamma}) \in \mathcal{MRG}(S_g \setminus \vec{\gamma})$ by $1/\text{Ext}_X(\gamma)$, one obtains the unit-length critical graph map

$$\Xi^1(\cdot, \vec{\gamma}): \mathcal{T}_g \rightarrow \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta).$$

Since $\text{Stab}_0(\vec{\gamma})$ preserves labelings of complementary subsurfaces, the maps $\Xi(\cdot, \vec{\gamma})$ and $\Xi^1(\cdot, \vec{\gamma})$ are $\text{Stab}_0(\vec{\gamma})$ -invariant.

3 Near the multiple zero locus

Outline of this section We show that the Jenkins–Strebel differentials of most simple closed multicurves lie deep within the principal stratum; see Proposition 3.6 for a precise statement. This control is crucial to apply the bounds on the variation of critical graphs proved in Section 4; compare to Proposition 4.3. The main tools used in this section are the theory of Delaunay triangulations of quadratic differentials developed in work of Masur and Smillie [1991] and dual train tracks to triangulations.

Triangulations of quadratic differentials For the rest of this paper fix $g \geq 2$ and let S_g be a compact connected oriented surface of genus $g \geq 2$. By a *triangulation* of a quadratic differential q we mean a triangulation of its underlying Riemann surface whose edges are saddle connections of q (and hence whose vertices are zeros of q). A triangulation of a quadratic differential q is said to be *L-bounded* for some $L > 0$ if its edges have flat length $\leq L$.

Given a marked quadratic differential $q \in \mathcal{Q}^1\mathcal{T}_g$ and a triangulation Δ' of q , one can pull back Δ' via the marking map to obtain an isotopy class of triangulation Δ on S_g . In particular, because we are not marking the zeros of q , the vertices of the triangulation Δ are not fixed. Recall that the mapping class

group Mod_g acts properly discontinuously on $\mathcal{Q}^1\mathcal{T}_g$ and \mathcal{T}_g by changing markings. It also acts on the set of isotopy classes of triangulations of S_g by applying the mapping class, and the association described above is equivariant.

Recall that $\pi: \mathcal{Q}^1\mathcal{T}_g \rightarrow \mathcal{T}_g$ denotes the forgetful map.

Lemma 3.1 *Let Δ be an isotopy class of triangulation of S_g , let $\mathcal{K} \subseteq \mathcal{T}_g$ be a compact subset, and $L > 0$. Then the subset of marked quadratic differentials $q \in \pi^{-1}(\text{Mod}_g \cdot \mathcal{K})$ having an L -bounded triangulation Δ' that pulls back to Δ via the marking of q has compact closure in $\mathcal{Q}^1\mathcal{T}_g$.*

Proof Let $q \in \pi^{-1}(\text{Mod}_g \cdot \mathcal{K})$ and Δ' be an L -bounded triangulation of q which pulls back to Δ via the marking of q . Fix a simple closed curve α on S_g . As Δ' is L -bounded and pulls back to Δ via the marking of q , one can bound the flat length $\ell_\alpha(q)$ of any geodesic representatives of α on q uniformly in terms of α , $\chi(S_g)$, and L . This together with the fact that $q \in \pi^{-1}(\text{Mod}_g \cdot \mathcal{K})$ implies the hyperbolic length $\ell_\alpha(\pi(q))$ of the unique geodesic representative of α with respect to the marked hyperbolic structure on S_g induced by $\pi(q) \in \mathcal{T}_g$ via uniformization can be bounded uniformly in terms of α , $\chi(S_g)$, L , and \mathcal{K} . As the bundle $\pi: \mathcal{Q}^1\mathcal{T}_g \rightarrow \mathcal{T}_g$ has compact fibers and as the only way of escaping to infinity in \mathcal{T}_g is to develop a simple closed curve of unbounded hyperbolic length, this finishes the proof. \square

Delaunay triangulations of quadratic differentials For every quadratic differential q denote by $\ell_{\min}(q)$ the length of its shortest saddle connections and by $\text{diam}(q)$ its diameter. Every quadratic differential (X, q) admits a triangulation by saddle connections which is *Delaunay* with respect to the singularities of q and the singular flat metric induced by q on X [Masur and Smillie 1991, Section 4]. We refer to any such triangulation as a *Delaunay triangulation* of q . For the purposes of this discussion we will not need to appeal to the explicit construction of these triangulations. Rather, it will suffice to know they exist and satisfy the following properties:

Lemma 3.2 [Athreya et al. 2012, Lemma 3.11] *Let $q \in \mathcal{Q}^1\mathcal{T}_g$ and Δ be a Delaunay triangulation of q . Let γ be a saddle connection of q .*

- (1) *If γ belongs to Δ , then $\ell_\gamma(q) \leq_g \text{diam}(q)$.*
- (2) *If $\ell_\gamma(q) \leq \sqrt{2}\ell_{\min}(q)$, then γ belongs to Δ .*

Directly from Lemmas 3.1 and 3.2 we deduce the following result:

Proposition 3.3 *For every compact subset $\mathcal{K} \subseteq \mathcal{T}_g$ there exists a constant $L = L(\mathcal{K}) > 0$ and a finite collection of isotopy classes of triangulations $\{\Delta_i\}_{i=1}^n$ on S_g with the following property. Let $q \in \pi^{-1}(\mathcal{K})$ and γ be a saddle connection of q attaining the minimal flat length among saddle connections of q . Then there exists $i \in \{1, \dots, n\}$ and an L -bounded triangulation Δ' of q having γ as one of its edges and which pulls back to Δ_i via the marking of q .*

Proof Fix a compact subset $\mathcal{K} \subseteq \mathcal{T}_g$. By Lemma 3.2.1, there exists a constant $L = L(\mathcal{K}) > 0$ such that every Delaunay triangulation of a quadratic differential in $\pi^{-1}(\mathcal{K})$ is L -bounded. Moreover, by Lemma 3.2.2, these triangulations always contain the shortest saddle connections of q . Thus it suffices to show these triangulations pull back to finitely many triangulations on S_g . This follows because triangulations of quadratic differentials in $\mathcal{Q}^1\mathcal{T}_g$ have at most $4g - 4$ vertices, so up to the action of the mapping class group there are finitely many combinatorial types of triangulations on S_g that can arise. Applying Lemma 3.1 and the proper discontinuity of the Mod_g action on $\mathcal{Q}^1\mathcal{T}_g$, we see that there are only finitely many isotopy classes of triangulations on S_g that are L -bounded on some $q \in \pi^{-1}(\mathcal{K})$. \square

Remark 3.4 This statement is false if we mark the zeros of q as well as the underlying surface. The zeros can braid around each other while q remains in a compact subset of $\mathcal{Q}^1\mathcal{T}_g$, yielding infinitely many distinct triangulations that differ by the surface braid group. This reflects the fact that the intersection of $\pi^{-1}(\mathcal{K})$ with any (nonminimal) stratum is not compact.

Train track coordinates We now discuss some aspects of the theory of train track coordinates. For more details we refer the reader to [Penner and Harer 1992]. A *train track* τ on S_g is an embedded 1-complex satisfying the following conditions:

- (1) Each edge of τ is a smooth path with a well-defined tangent vector at each endpoint. All edges at a given vertex are tangent.
- (2) For each component R of $S_g \setminus \tau$, the double of R along the smooth part, of the boundary ∂R has negative Euler characteristic; equivalently, R is not a bigon or a monogon.

The vertices of τ where three or more edges meet are called *switches*. By considering the inward-pointing tangent vectors of the edges incident to a switch, one can divide these edges into *incoming* and *outgoing* edges. A train track τ on S_g is said to be *maximal* if all the components of $S_g \setminus \tau$ are trigons, ie the interior of a disc with three nonsmooth points on its boundary.

A singular measured foliation $\lambda \in \mathcal{MF}_g$ is said to be *carried* by a train track τ on S_g if it can be obtained by collapsing the complementary regions in S_g of a measured foliation of a tubular neighborhood of τ whose leaves run parallel to the edges of τ . In this situation, the invariant transverse measure of λ corresponds to a counting measure ν on the edges of τ satisfying the *switch conditions*: at every switch of τ the sum of the measures of the incoming edges equals the sum of the measures of the outgoing edges. Every $\lambda \in \mathcal{MF}_g$ is carried by some maximal train track τ on S_g .

Given a maximal train track τ on S_g , denote by $V(\tau) \subseteq (\mathbb{R}_{\geq 0})^{18g-18}$ the $(6g-6)$ -dimensional closed cone of nonnegative counting measures on the edges of τ satisfying the switch conditions. The set $V(\tau)$ can be identified with the closed cone $U(\tau) \subseteq \mathcal{MF}_g$ of singular measured foliations carried by τ . These identifications give rise to coordinates on \mathcal{MF}_g called *train track coordinates*. The transition maps of these coordinates are piecewise integral linear. In particular, \mathcal{MF}_g can be endowed with a natural $(6g-6)$ -dimensional piecewise integral linear structure where the integral points $\mathcal{MF}_g(\mathbb{Z})$ correspond to integrally weighted simple closed multicurves.

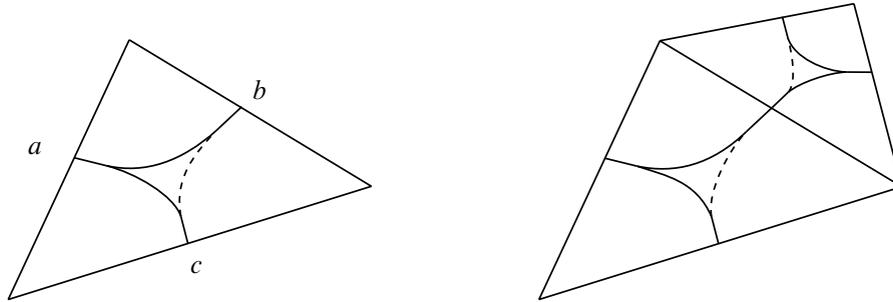


Figure 2: The 1-complex and the train track dual to a triangulation. The train track is obtained by removing the dashed edges from the 1-complex. Left: the dual 1-complex and dual train track in a triangle. Right: joining the dual 1-complexes and train tracks.

Train tracks dual to triangulations We now use our discussion of triangulations of $q \in \pi^{-1}(\mathcal{K})$ to control what the horizontal foliations can coarsely look like. We begin by recalling the construction of a train track dual to a triangulation; compare [Mirzakhani 2008a; Arana-Herrera 2025; Calderon and Farre 2024b; 2024a].

Let Δ be an isotopy class of triangulation on S_g . On each of the triangles of Δ consider a 1-complex as in Figure 2, left; the edges of this complex that do not intersect the sides of the triangle will be referred to as *inner edges*. Join these complexes along the edges of Δ as in Figure 2, right, to obtain a complex on S_g . We say that a train track τ on S_g is *dual* to Δ if it can be obtained from this complex by deleting one inner edge in each triangle of Δ . We remark that this operation is well defined even though we are only considering isotopy classes: two isotopic triangulations will give isotopic collections of dual train tracks.

Let $q \in \mathcal{Q}^1\mathcal{T}_g$ and Δ' be a triangulation of q . Denote by Δ the triangulation of S_g obtained by pulling back Δ' via the marking of q . The horizontal foliation $\mathfrak{F}(q) \in \mathcal{MF}_g$ is carried by a train track dual to Δ . Indeed, let T' be a triangle of Δ' . Label the edges of T' by a , b , and c so that

$$\int_a d\mathfrak{F}(q) = \int_b d\mathfrak{F}(q) + \int_c d\mathfrak{F}(q).$$

This labeling is unique unless one of the edges of T' is horizontal, in which case there exist two such labelings. On T' consider a 1-complex as in Figure 2, left, and delete the inner edge of this complex opposite to a . Consider the corresponding 1-complexes on all the triangles of Δ' . Joining these complexes along the edges of Δ' as in Figure 2, right, and pulling back the resulting complex to S_g yields a train track τ dual to Δ . This train track carries $\mathfrak{F}(q) \in \mathcal{MF}_g$ and, for each edge e of Δ' , the counting measure of the corresponding edge of τ is equal to $\int_e d\mathfrak{F}(q)$. If q is in the principal stratum of quadratic differentials, the train track τ obtained through this construction is maximal, and in general, the singularity structure of q is reflected in the combinatorics of τ .

Directly from the discussion above and Proposition 3.3 we deduce the following, which is analogous to both [Arana-Herrera 2025, Proposition 2.7] and [Calderon and Farre 2024b, Section 10]:

Proposition 3.5 For every compact subset $\mathcal{K} \subseteq \mathcal{T}_g$ there exists a finite collection of train tracks $\{\tau_i\}_{i=1}^n$ on S_g with the following property. Let $q \in \pi^{-1}(\mathcal{K})$ and γ be a saddle connection of q that attains the minimal flat length among saddle connections of q . Then there exists $i \in \{1, \dots, n\}$ such that τ_i carries $\mathfrak{S}(q)$ and such that the corresponding counting measure on the edges of τ_i gives weight $\int_\gamma d\mathfrak{S}(q)$ to one of the edges of τ_i .

The key estimate We are now ready to prove the main estimate of this section. This estimate will allow us control the equidistribution problem of interest near the multiple zero locus in a sufficiently strong way.

Let $\text{JS}: \mathcal{T}_g \times \mathcal{MF}_g \rightarrow \mathcal{Q}^1\mathcal{T}_g$ denote the Hubbard–Masur map from Theorem 2.2 which to every $(X, \lambda) \in \mathcal{T}_g \times \mathcal{MF}_g$ assigns the unique marked quadratic differential $q \in \pi^{-1}(X)$ such that $\mathfrak{S}(q) = \lambda$. This differential has area $\text{Ext}_X(\lambda)$, so rescaling it by the square root of extremal length results in a unit-area differential. Let $K_\sigma \subseteq \mathcal{Q}^1\mathcal{T}_g$ denote the set of marked unit-area quadratic differentials $q \in \mathcal{Q}^1\mathcal{T}_g$ belonging to the principal stratum with $\ell_{\min}(q) \geq \sigma$. Given $X \in \mathcal{T}_g$ consider the set

$$F(X, \sigma) := \left\{ \alpha \in \mathcal{MF}_g(\mathbb{Z}) \mid \frac{1}{\sqrt{\text{Ext}_X(\alpha)}} \text{JS}(X, \alpha) \notin K_\sigma \right\}$$

and for any compact set $\mathcal{K} \subset \mathcal{T}_g$ define $F(\mathcal{K}, \sigma) := \bigcup_{X \in \mathcal{K}} F(X, \sigma)$.

Proposition 3.6 Let $\mathcal{K} \subseteq \mathcal{T}_g$ be a compact subset. Then there exists a constant $C > 0$ such that for every $X \in \mathcal{K}$, every $\sigma > 0$, and every $L > 1$,

$$\#\{\alpha \in F(\mathcal{K}, \sigma) \mid \text{Ext}_\alpha(X) \leq L^2\} \leq CL^{6g-7} + C\sigma L^{6g-6}.$$

Proof Let $\{\tau_i\}_{i=1}^n$ be a finite collection of train tracks on S_g as provided by Proposition 3.5. Now consider $\alpha \in \mathcal{MF}_g(\mathbb{Z})$ such that $\text{Ext}_\alpha(X) \leq L^2$ and

$$q := \text{JS}(X, \alpha / \sqrt{\text{Ext}_\alpha(X)}) \in \mathcal{Q}^1\mathcal{T}_g \setminus K_\sigma.$$

Suppose first that q is not in the principal stratum. Then, by construction, any of the train tracks τ_i carrying $\alpha = \sqrt{\text{Ext}_\alpha(X)}\mathfrak{S}(q)$ is not maximal. Suppose now that q is in the principal stratum and that γ is a saddle connection of q attaining $\ell_{\min}(q)$. It follows by construction that, when carried by the corresponding τ_i , one of the resulting weights of α is at most $\sigma\sqrt{\text{Ext}_\alpha(X)} \leq \sigma L$. As the collection of train tracks $\{\tau_i\}_{i=1}^n$ is finite, the desired bound follows from a standard lattice-point counting argument, ie the fact that a ball of radius $R > 0$ in a d -dimensional Euclidean space contains approximately R^d integer points. □

4 Uniform geometric estimates

Outline of this section We show that the weights of the metric ribbon graph $\Xi^1(X, \vec{\alpha})$ vary uniformly over α as X varies in a suitably chosen neighborhood of moduli space; see Proposition 4.3 for a precise statement. The proof is based on a uniform continuity argument and uses the AGY metric on strata of

quadratic differentials introduced by Avila, Gouëzel, and Yoccoz [Avila et al. 2006] in a crucial way. Before proving Proposition 4.3 we discuss some of the basic properties of the AGY metric following [Avila et al. 2006; Avila and Gouëzel 2013].

The AGY metric Let $\mathcal{Q} \subseteq \mathcal{QT}_g$ be the principal stratum of marked quadratic differentials on S_g , that is, the subset of differentials with only simple zeros. Denote points in \mathcal{Q} by (X, q) , where X is a marked Riemann surface and q is a quadratic differential on X . Let $(Z, \omega) \rightarrow (X, q)$ be the holonomy double cover of (X, q) . The tangent space of \mathcal{Q} at (X, q) can be identified with the relative cohomology group $H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$, where $\Sigma \subseteq Z$ denotes the set of zeroes of ω , and the subscript odd denotes the -1 eigenspace of the canonical involution of $Z \rightarrow X$. For $v \in H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$ consider the norm

$$\|v\|_q := \sup_{s \in \mathcal{S}} \left| \frac{v(s)}{\text{hol}_\omega(s)} \right|,$$

where \mathcal{S} denotes the set of saddle connections of ω and $\text{hol}_\omega(s) \in \mathbb{C}$ denotes the holonomy of the saddle connection s with respect to ω ; we will sometimes denote this holonomy by $|s|_\omega$. By work of Avila, Gouëzel, and Yoccoz [Avila et al. 2006], this definition indeed gives rise to a norm on $H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$ and the restriction of the corresponding Finsler metric to any fixed-area slice of \mathcal{Q} is complete. We refer to this metric as the AGY metric of \mathcal{Q} and denote it by d_{AGY} . For any C^1 path $\kappa: [0, 1] \rightarrow \mathcal{Q}$, denote its AGY length by

$$\text{length}(\kappa) := \int_0^1 \|\kappa'(t)\|_{\kappa(t)} dt.$$

We observe that the rescaling action $q \mapsto rq$ for $r \in \mathbb{R}_{>0}$ is an isometry of the AGY metric. Indeed, given $v \in H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$ we have that $\|rv\|_{rq} = \|v\|_q$, and so given any C^1 path $\kappa(t)$ from q to q' we can compute the length of the corresponding path $r\kappa(t)$ from rq to rq' :

$$\text{length}(r\kappa) = \int_0^1 \|r\kappa'(t)\|_{r\kappa(t)} dt = \int_0^1 \|\kappa'(t)\|_{\kappa(t)} dt = \text{length}(\kappa).$$

The fundamental property Fix $(X, q) \in \mathcal{Q}$ and let $\Psi = \Psi_q$ be the local parametrization of \mathcal{Q} by its tangent plane $H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$ as above. More formally, define $\Psi(v)$ for $v \in H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$ as follows. Consider the path κ starting from (X, q) with $\kappa'(t) = v$ for all times t . For small t the path $\kappa(t)$ is well defined. It is possible that $\kappa(t)$ could be not defined for large t . If the path κ is well defined for all $t \in [0, 1]$ define $\Psi(v) := \kappa(1) \in \mathcal{Q}$.

Denote by $B(0, r)$ the ball of radius r centered at the origin of $H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$ with respect to the norm $\|\cdot\|_q$. The following results, which will be crucial in later sections, give control on the image of the “exponential map” Ψ :

Proposition 4.1 *The map Ψ is well defined on $B(0, \frac{1}{2})$. Moreover:*

- (1) For $v \in B(0, \frac{1}{2})$,

$$d_{\text{AGY}}(q, \Psi(v)) \leq 2\|v\|_q.$$

(2) For every $v \in B(0, \frac{1}{2})$ and every $w \in H_{\text{odd}}^1(Z, \Sigma; \mathbb{C})$,

$$\frac{1}{2} \leq \frac{\|w\|_q}{\|w\|_{\Psi(v)}} \leq 2.$$

(3) For $v \in B(0, \frac{1}{25})$,

$$d_{\text{AGY}}(q, \Psi(v)) \geq \frac{1}{2} \|v\|_q.$$

In particular, any $q' \in \mathcal{Q}$ with $d_{\text{AGY}}(q, q') < \frac{1}{50}$ is equal to $\Psi(v)$ for some $v \in B(0, \frac{1}{25})$.

Apart from the last claim of (3), this is just the statement of [Avila and Gouëzel 2013, Proposition 5.3] generalized to arbitrary vectors; we provide an outline of the final claim.

Proof sketch Consider the image of $B(0, \frac{1}{25})$ under Ψ . By invariance of domain, this is open; in particular, it contains an open AGY ball B_{AGY} of some maximal radius $r > 0$. Now by the first claim of (3), we have that $\Psi^{-1}(B_{\text{AGY}})$ is contained in $B(0, 2r)$. In particular, if $r < \frac{1}{50}$ then we can find a slightly larger ball $B(0, 2r + \varepsilon)$ whose Ψ image strictly contains B_{AGY} (by the first claim of (3) again); hence r was not maximal to begin with. So $r \geq \frac{1}{50}$. \square

A proof of the remaining statements can be obtained by closely following the arguments of [loc. cit.]. Indeed, the key input in the proof of Proposition 4.1 is [loc. cit., Proposition 5.5], which gives a bound on the growth of the parallel translate of a vector in terms of the length of the path along which it is translated, and is true for any v , not just those lying in the unstable distribution.

We record for future use a similar result on how the length of saddle connections changes along paths; compare with [loc. cit., Lemma 5.6]:

Lemma 4.2 Suppose that $\kappa: [0, 1] \rightarrow \mathcal{Q}$ is a C^1 path and that s is a saddle connection that survives in $\kappa(t)$ for all $t \in [0, 1]$. Then

$$e^{-\text{length}(\kappa)} \leq \frac{|s|_{\kappa(0)}}{|s|_{\kappa(1)}} \leq e^{\text{length}(\kappa)}.$$

Variation of weights With these preliminaries taken care of, we can now show that the geometry of the horizontal ribbon graph is controlled uniformly over most simple closed multicurves α as X varies in a small ball in \mathcal{T}_g ; compare with [Arana-Herrera and Calderon 2022, Lemma 4.1 and Proposition 4.2].

Recall that if $\vec{\alpha}$ is an ordered oriented simple closed multicurve on S_g , then $\alpha \in \mathcal{MF}_g(\mathbb{Z})$ denotes its equivalence class as a singular measured foliation. Recall that for a compact set $\mathcal{K} \subseteq \mathcal{T}_g$ and $\sigma > 0$, the set $F(\mathcal{K}, \sigma)$ denotes those integral simple closed multicurves $\alpha \in \mathcal{MF}_g(\mathbb{Z})$ so that, for some $X \in \mathcal{K}$, the differential $\text{JS}(X, \alpha)$ is either not in \mathcal{Q} or has a saddle connection shorter than $\sigma \sqrt{\text{Ext}_X(\alpha)}$.

Proposition 4.3 Fix a compact subset $\mathcal{K} \subseteq \mathcal{T}_g$. Then, for every $\varepsilon > 0$ and every $\sigma > 0$, there exists $\delta = \delta(\mathcal{K}, \varepsilon, \sigma) > 0$ such that for any ordered oriented simple closed multicurve $\vec{\alpha}$ on S_g such that $\alpha \notin F(\mathcal{K}, \sigma)$ and any two $X, X' \in \mathcal{K}$ with $d_{\text{Teich}}(X, X') \leq \delta$, the following hold:

- (1) The horizontal ribbon graphs $\Xi(X, \vec{\alpha})$ and $\Xi(X', \vec{\alpha})$ have the same topological type, ie live in the same facet of $\mathcal{MRG}(S_g \setminus \vec{\alpha})$.
- (2) For every edge e of $\Xi(X, \vec{\alpha})$ and/or $\Xi(X', \vec{\alpha})$, we have that

$$e^{-\varepsilon} \leq \frac{|e|_{\Xi^1(X, \vec{\alpha})}}{|e|_{\Xi^1(X', \vec{\alpha})}} \leq e^\varepsilon.$$

The proof of this proposition has two steps: we first show that the hypotheses imply that the differentials $JS(X, \alpha)$ and $JS(X', \alpha)$ are close in the AGY metric, then use properties of the metric to conclude the desired statements.

The first step is accomplished using the uniform continuity of the Hubbard–Masur map. To get this, we need to restrict to a compact set. Restricting to K_σ , as defined before Proposition 3.6, allows us to avoid a neighborhood of the multiple zero locus, but we must also impose a bound on the areas of differentials under consideration. As such, we first record an a priori bound on the extremal length of foliations as the base surface ranges over a compact set. The following is a direct consequence of Theorem 2.4:

Lemma 4.4 *For every compact $\mathcal{K} \subseteq \mathcal{T}_g$, there is a $c_{\mathcal{K}} > 1$ such that for every $\lambda \in \mathcal{MF}_g$ and every $X, X' \in \mathcal{K}$,*

$$c_{\mathcal{K}}^{-1} \leq \frac{\sqrt{\text{Ext}_{X'}(\lambda)}}{\sqrt{\text{Ext}_X(\lambda)}} \leq c_{\mathcal{K}}.$$

We now show that if X and X' are close in the Teichmüller metric, then most of their Jenkins–Strebel differentials are close in the AGY metric.

Lemma 4.5 *Fix \mathcal{K} and σ as in Proposition 4.3. For any $\zeta > 0$, there is a $\delta > 0$ such that for any $X, X' \in \mathcal{K}$ with $d_{\text{Teich}}(X, X') < \delta$ and any $\alpha \notin F(\mathcal{K}, \sigma)$,*

$$d_{\text{AGY}}(JS(X, \alpha), JS(X', \alpha)) < \zeta.$$

Proof Fix any metric $d_{\mathcal{MF}_g}$ on \mathcal{MF}_g equipping it with the standard topology. By Theorem 2.2, the map $JS: \mathcal{T}_g \times \mathcal{MF}_g \rightarrow \mathcal{QT}_g$ is a homeomorphism, and is hence uniformly continuous on the compact set

$$J := (\mathcal{K} \times \mathcal{MF}_g) \cap \text{Ext}^{-1}([c_{\mathcal{K}}^{-2}, c_{\mathcal{K}}^2]) \cap JS^{-1}(\mathbb{R}_{>0} \cdot K_\sigma).$$

That is, for any $\zeta > 0$ there is a $\delta > 0$ such that for any $(X, \lambda), (X', \lambda') \in J$ with

$$d_{\text{Teich}}(X, X') < \delta \quad \text{and} \quad d_{\mathcal{MF}_g}(\lambda, \lambda') < \delta,$$

we have that

$$d_{\text{AGY}}(JS(X, \lambda), JS(X', \lambda')) < \zeta.$$

In particular, Lemma 4.4 (alternatively, Theorem 2.4) implies that

$$\sqrt{\text{Ext}_{X'}}\left(\frac{\alpha}{\sqrt{\text{Ext}_X(\alpha)}}\right) = \frac{\sqrt{\text{Ext}_{X'}(\alpha)}}{\sqrt{\text{Ext}_X(\alpha)}} \in [c_{\mathcal{K}}^{-1}, c_{\mathcal{K}}],$$

so setting $\lambda = \lambda' = \alpha / \sqrt{\text{Ext}_X(\alpha)}$ we have that $\text{JS}(X, \lambda)$ and $\text{JS}(X', \lambda)$ are ζ apart. The statement for α itself follows from the fact that rescaling is an isometry of the AGY metric. \square

Now since $\text{JS}(X, \alpha)$ and $\text{JS}(X', \alpha)$ are close and live in the same leaf of the unstable foliation, we can use the exponential map Ψ to connect them via a path completely contained in the unstable leaf $\{\text{JS}(X, \alpha) \mid X \in \mathcal{T}_g\}$. Analyzing this map yields the proof of the main result of this section.

Proof of Proposition 4.3 We begin by observing that, by definition of the Ξ^1 map,

$$\frac{|e|_{\Xi^1(X, \vec{\alpha})}}{|e|_{\Xi^1(X', \vec{\alpha})}} = \frac{|e|_{\Xi(X, \vec{\alpha})} \text{Ext}_{X'}(\alpha)}{|e|_{\Xi(X', \vec{\alpha})} \text{Ext}_X(\alpha)}.$$

By Theorem 2.4, the ratio of extremal lengths is bounded arbitrarily close to 1 (so long as we take δ small enough), so it suffices to compare the geometry of the nonnormalized critical graphs. For ease of notation, throughout the rest of the proof let us define

$$q := \text{JS}(X, \alpha) \quad \text{and} \quad q' := \text{JS}(X', \alpha).$$

Fix $0 < \varepsilon < \frac{2}{25}$. Lemma 4.5 tells us that by taking δ small enough we can ensure q and q' are $\frac{1}{4}\varepsilon < \frac{1}{50}$ close in the AGY metric, so by Proposition 4.1(3) we have that $q' = \Psi(v)$ for some $v \in B(0, \frac{1}{2}\varepsilon)$. But now we know that q and q' have the same horizontal foliation, and hence the same vertical periods. Therefore v is real, ie $v \in H_{\text{odd}}^1(Z, \Sigma; \mathbb{R})$ with respect to the natural splitting

$$H_{\text{odd}}^1(Z, \Sigma; \mathbb{C}) \cong H_{\text{odd}}^1(Z, \Sigma; \mathbb{R}) \oplus H_{\text{odd}}^1(Z, \Sigma; i\mathbb{R}).$$

In particular, this implies that the period of every horizontal saddle remains real along $\kappa(t)$. Since the exponential map Ψ is well defined (and proper), every horizontal saddle must persist along the entire path $\{\kappa(t)\}_{t=0}^1$ (otherwise some saddle would shrink to zero, but doing so leaves the stratum). Hence the topological type of the horizontal saddle connection graphs of $\kappa(0) = q$ and $\kappa(1) = q'$ are the same, establishing the first part of the proposition.

For the second part, $\|v\|_{\kappa(t)}$ grows by at most a factor of 2 along the entire path (Proposition 4.1.(2)), so the length of κ is at most ε . Since each horizontal saddle persists along the entire path, Lemma 4.2 implies that the length of each can change by a factor of at most $e^{\text{length}(\kappa)}$. Thus the lengths on q and q' of every edge of $\Xi(X, \vec{\alpha})$ have ratio bounded by $e^{\pm\varepsilon}$, proving the second claim. \square

5 Horoball measures

Outline of this section We introduce “critical-JS-horoballs” and show that, as they expand over moduli space, they equidistribute with respect to the Masur–Veech measure. See Proposition 5.7 for a precise statement. Our preliminary discussion relies on work of Athreya, Bufetov, Eskin, and Mirzakhani [Athreya et al. 2012] as well as an expression for critical-JS-horoballs in terms of the fibered Kontsevich measure.

The equidistribution result is a consequence of work of Forni [2021], which in turn relies on breakthroughs of Eskin, Mirzakhani, and Mohammadi [Eskin and Mirzakhani 2018; Eskin et al. 2015]. Throughout we use the normalizations in [Athreya et al. 2012; Arana-Herrera 2023; Arana-Herrera and Calderon 2022] for all the measures considered.

The Masur–Veech measure Let $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ be an ordered oriented simple closed multicurve on S_g with underlying multicurve $\gamma \in \mathcal{MF}_g$. Recall that $\text{Stab}_0(\vec{\gamma}) \subseteq \text{Mod}_g$ denotes the oriented stabilizer of $\vec{\gamma}$, ie the set of mapping classes that fix each component of $\vec{\gamma}$ together with its orientation.

Recall that ν_{MV} denotes the Masur–Veech measure on $Q^1\mathcal{T}_g$. The forgetful map $\pi : Q^1\mathcal{T}_g \rightarrow \mathcal{T}_g$ pushes the Masur–Veech measure down to a Lebesgue-class measure $\mathbf{m} := \pi_*\nu_{\text{MV}}$ on \mathcal{T}_g , which we also refer to as the Masur–Veech measure. Both ν_{MV} and \mathbf{m} are Mod_g -invariant. Let $\tilde{\nu}_{\text{MV}}$ and $\tilde{\mathbf{m}}$ denote the corresponding local pushforwards to $Q^1\mathcal{T}_g/\text{Stab}_0(\vec{\gamma})$ and $\mathcal{T}_g/\text{Stab}_0(\vec{\gamma})$, and, similarly, let $\hat{\nu}_{\text{MV}}$ and $\hat{\mathbf{m}}$ be the pushforwards of $\tilde{\nu}_{\text{MV}}$ and $\tilde{\mathbf{m}}$ to $Q^1\mathcal{M}_g$ and \mathcal{M}_g . One could of course also define $\tilde{\mathbf{m}}$ and $\hat{\mathbf{m}}$ by pushing forward $\tilde{\nu}_{\text{MV}}$ and $\hat{\nu}_{\text{MV}}$ under the corresponding forgetful maps. Denote the total mass of $\hat{\nu}_{\text{MV}}$, or, equivalently, $\hat{\mathbf{m}}$, by $b_g > 0$.

Recall that for any Riemann surface X , we denote by $S(X)$ the sphere of unit-area quadratic differentials on X . Let s_X be the conditional probability measure on $S(X)$ induced by \mathbf{m} on \mathcal{T}_g . It is characterized by the disintegration formula

$$(5-1) \quad d\nu_{\text{MV}}(X, q) = ds_X(q) d\mathbf{m}(X),$$

together with similar expressions for the measures $\tilde{\nu}_{\text{MV}}$ and $\hat{\nu}_{\text{MV}}$ on quotients.²

The Hubbard–Masur function Recall that the singular measured foliation $\mathfrak{S}(q) \in \mathcal{MF}_g$ denotes the horizontal foliation of $q \in Q^1\mathcal{T}_g$, and that $[\mathfrak{S}(q)] \in \mathcal{PMF}_g$ denotes its projective class. As observed in [Athreya et al. 2012], every leaf of the unstable foliation \mathcal{F}^u of $Q^1\mathcal{T}_g$ carries a conditional measure that is uniformly expanded by the Teichmüller geodesic flow. This measure can be explicitly obtained as follows. By definition, any leaf of \mathcal{F}^u is of the form

$$\mathcal{F}^u_{[\beta]} = \{q \in Q^1\mathcal{T}_g \mid [\mathfrak{S}(q)] = [\beta] \in \mathcal{PMF}_g\}$$

for some $\beta \in \mathcal{MF}_g$. By Theorem 2.2, this set can in turn be identified with the open set $\mathcal{MF}_g(\beta) \subseteq \mathcal{MF}_g$ of singular measured foliations on S_g that together with β fill the surface. The Thurston measure μ_{Thu} on \mathcal{MF}_g therefore restricts to a nontrivial measure on $\mathcal{MF}_g(\beta)$ and hence gives rise to a measure on $\mathcal{F}^u_{[\beta]}$ which we denote by $\nu_{u,\beta}$.

Recall that $\{g_t : Q^1\mathcal{T}_g \rightarrow Q^1\mathcal{T}_g\}_{t \in \mathbb{R}}$ is the Teichmüller geodesic flow on $Q^1\mathcal{T}_g$. The fundamental scaling property described by the formula

$$(5-2) \quad (g_t)_*\nu_{u,\beta} = e^{-(6g-6)t} \nu_{u,\beta}$$

²We note that we do not need to worry about sizes of stabilizers when recording disintegration formulas at the level of moduli space. For $g \geq 3$, we have that \mathbf{m} -almost every $X \in \mathcal{M}_g$ has no nontrivial automorphisms and so the fiber of $Q^1\mathcal{M}_g$ over X is the entire sphere $S(X)$. For $g = 2$ every surface is hyperelliptic, but so is every quadratic differential.

is then a consequence of the fact that the Teichmüller geodesic flow stretches the horizontal direction, expanding the measure on the vertical foliation, and thus acting by multiplication by e^t on $\mathcal{MF}_g(\beta)$.

By Theorems 2.2 and 2.3, the forgetful map $\pi : \mathcal{Q}^1\mathcal{T}_g \rightarrow \mathcal{T}_g$ restricts to a homeomorphism $\pi_{[\beta]}$ between $\mathcal{F}_{[\beta]}^u \cong \mathcal{MF}(\beta)$ and \mathcal{T}_g . The pushforward of $\nu_{u,\beta}$ by this map is in the Lebesgue measure class. Following [loc. cit.], we denote by

$$\lambda^+(q) := \frac{d\mathbf{m}}{d(\pi_{[\beta]})_*\nu_{u,\beta}}(X)$$

the corresponding Radon–Nikodym derivative, where $q = \pi_{[\beta]}^{-1}(X)$, that is, $q = \text{JS}(X, \beta) / \sqrt{\text{Ext}_X(\beta)}$. We recall from [loc. cit.] that the *Hubbard–Masur function* is defined to be the following integral:

$$\Lambda(X) := \int_{\mathcal{Q}^1(X)} \lambda^+(q) ds_X.$$

Directly from the definitions, we observe that both $\lambda^+(q)$ and $\Lambda(X)$ are Mod_g -invariant; see also [loc. cit., top of page 1063].

Tracing through the definitions, one can arrive at the following formulation:

Lemma 5.1 [Athreya et al. 2012, Proposition 2.3(iii)] *For any $X \in \mathcal{T}_g$,*

$$\Lambda(X) = \mu_{\text{Thu}}(\{\beta \in \mathcal{MF} \mid \text{Ext}_X(\beta) \leq 1\}).$$

In fact, Mirzakhani proved the following strong result:

Theorem 5.2 [Dumas 2015, Theorem 5.10] *The function $X \in \mathcal{T}_g \mapsto \Lambda(X)$ is constant.*

As such, we refer to this value as the *Hubbard–Masur constant* $\Lambda_g > 0$.

Horoballs For every $L > 0$ we define the (total) extremal length horoball measure $\mathbf{m}_{\vec{\gamma}}^L$ on \mathcal{T}_g by restricting \mathbf{m} to the set of marked Riemann surfaces $X \in \mathcal{T}_g$ on which $\sqrt{\text{Ext}_X(\gamma)} \leq L$. We similarly define the (total) unstable horoball measure $\nu_{u,\vec{\gamma}}^L$ on $\mathcal{Q}^1\mathcal{T}_g$ by restricting $\nu_{u,\vec{\gamma}}$ to the preimage of this set under the forgetful map $\pi : \mathcal{Q}^1\mathcal{T}_g \rightarrow \mathcal{T}_g$. The extremal length of γ on X is the same as the area of the quadratic differential $q = \text{JS}(X, \gamma)$, which in turn is equal to the geometric intersection number of the horizontal and vertical foliations of q , so $\nu_{u,\vec{\gamma}}^L$ can equivalently be defined by restricting the Thurston measure on $\mathcal{MF}_g(\gamma)$ to the set

$$\{\beta \in \mathcal{MF}_g(\gamma) \mid i(\beta, \gamma) \leq L\}$$

and pushing this “Thurston horoball measure” μ_{Thu}^L forward to $\mathcal{F}_{[\gamma]}^u$. We then take local pushforwards to get measures

$$\tilde{\mathbf{m}}_{\vec{\gamma}}^L \text{ on } \mathcal{T}_g / \text{Stab}_0(\vec{\gamma}), \quad \tilde{\nu}_{u,\vec{\gamma}}^L \text{ on } \mathcal{Q}^1\mathcal{T}_g / \text{Stab}_0(\vec{\gamma}), \quad \tilde{\mu}_{\text{Thu}}^L \text{ on } \mathcal{MF}(\gamma) / \text{Stab}_0(\vec{\gamma}).$$

One can check that $\tilde{\mu}_{\text{Thu}}^L$ is finite, as the usual Thurston measure is locally finite on \mathcal{MF}_g and the closure (inside \mathcal{MF}_g , including the 0 foliation) of a piecewise-linear fundamental domain for the action of

$\text{Stab}_0(\vec{\gamma})$ on the support of μ_{Thu}^L is compact. This implies that $\tilde{\nu}_{u,\gamma}^L$ and $\tilde{\mathbf{m}}_{\vec{\gamma}}^L$ are also finite, so we can take the (global) pushforwards of $\tilde{\nu}_{u,\gamma}^L$ and $\tilde{\mathbf{m}}_{\vec{\gamma}}^L$ to the moduli spaces $\mathcal{Q}^1\mathcal{M}_g$ and \mathcal{M}_g ; denote the resulting measures by $\hat{\nu}_{u,\gamma}^L$ and $\hat{\mathbf{m}}_{\vec{\gamma}}^L$.

Recall from Section 2 that $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ denotes the moduli spaces of ribbon graphs of complementary subsurfaces to $\vec{\gamma}$ of total boundary length 2 and matching boundary lengths along the components of $\vec{\gamma}$. Recall also that $\Xi^1(X, \vec{\gamma}) \in \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ denotes the critical graph of the Jenkins–Strebel differential $\text{JS}(X, \gamma)$ rescaled so that the boundaries have total length 2.

We also want to consider subsets of the horoballs above by conditioning on the shape of the horizontal separatrices of the corresponding Jenkins–Strebel differentials. To this end, for any $L > 0$ and any nonzero continuous compactly supported function $h : \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta) \rightarrow \mathbb{R}$, define the Ξ^1 -horoball measure on \mathcal{T}_g by

$$(5-3) \quad d\mathbf{m}_{\vec{\gamma},h}^L(X) := \mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_X(\gamma)})h(\Xi^1(X, \vec{\gamma}))d\mathbf{m}(X).$$

We similarly define a version supported on the unstable leaf corresponding to γ :

$$(5-4) \quad d\nu_{u,\vec{\gamma},h}^L(X, q) := \mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_X(\gamma)})h(\Xi^1(X, \vec{\gamma}))d\nu_{u,\gamma}(q).$$

We informally refer to the measures $\mathbf{m}_{\vec{\gamma},h}^L$ as “critical-JS-horoballs”. Compare with the definition of “RSC-horoballs” from [Arana-Herrera and Calderon 2022, (5.1)]. Notice that the measures $\mathbf{m}_{\vec{\gamma},h}^L$ are not equal to the pushforwards of the measures $\nu_{u,\vec{\gamma},h}^L$ under the Hubbard–Masur map: they differ by the Hubbard–Masur function.

The measures $\mathbf{m}_{\vec{\gamma},h}^L(X)$ and $\nu_{u,\vec{\gamma},h}^L(q)$ are $\text{Stab}_0(\vec{\gamma})$ -invariant, and so as in the case of the total horoball measures we can take their local pushforwards $\tilde{\mathbf{m}}_{\vec{\gamma},h}^L$ and $\tilde{\nu}_{u,\vec{\gamma},h}^L$ after quotienting by $\text{Stab}_0(\vec{\gamma})$. We can then further push down to finite measures $\hat{\mathbf{m}}_{\vec{\gamma},h}^L$ and $\hat{\nu}_{u,\vec{\gamma},h}^L$ on \mathcal{M}_g and $\mathcal{Q}^1\mathcal{M}_g$, respectively.

The fibered Kontsevich measure In the next two subsections, we discuss how the “fibered Kontsevich measure” describes critical-JS-horoballs in terms of the combinatorial data of metric ribbon graphs; see Proposition 5.5. We begin by quickly recalling the definition of this measure and refer the reader to [Arana-Herrera and Calderon 2022, Sections 2 and 7] for a more detailed overview.

Kontsevich [1992] defined a piecewise 2-form ω_{Kon} on $\mathcal{MRG}_{g,b}$ that computes intersection numbers on moduli space. Restricting to a slice $\mathcal{MRG}_{g,b}(\mathbf{L})$ with fixed boundary lengths, this form is seen to be symplectic on every maximal facet. Thus the Kontsevich form gives rise to volume forms

$$\frac{1}{(3g-3+b)!} \wedge^{3g-3+b} \omega_{\text{Kon}}$$

on each maximal facet, which can be glued together into a volume form on the entire slice $\mathcal{MRG}_{g,b}(\mathbf{L})$. We will use $\eta_{\text{Kon}}^{\mathbf{L}}$ to denote the measure associated to this volume form and refer to it as the *Kontsevich measure*³ on $\mathcal{MRG}_{g,b}(\mathbf{L})$.

³See [Arana-Herrera and Calderon 2022, Remark 2.1] for a discussion of how to deal with this measure in the presence of nontrivial automorphism groups.

When $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ is an ordered oriented simple closed multicurve on S_g with complementary subsurfaces $(\Sigma_j)_{j=1}^c$, we recall that we set

$$\mathcal{MRG}(S_g \setminus \vec{\gamma}; \mathbf{L}) := \prod_{j=1}^c \mathcal{MRG}_{g_j, b_j}(\mathbf{L}^{(j)})$$

for any length vector $\mathbf{L} \in \mathbb{R}_{>0}^k$. As this is a product of moduli spaces with fixed boundary lengths, it has a (product) Kontsevich measure $\eta_{\text{Kon}}^{\vec{\gamma}, \mathbf{L}}$. Integrating against boundary lengths, the Kontsevich measures on each slice also fit together into a canonical measure on the total space, defined for any measurable subset $A \subseteq \mathcal{MRG}(S_g \setminus \vec{\gamma})$ by the formula

$$\eta_{\text{Kon}}^{\vec{\gamma}}(A) := \int_{\mathbb{R}_{>0}^k} \eta_{\text{Kon}}^{\vec{\gamma}, \mathbf{L}}(A \cap \mathcal{MRG}(S_g \setminus \vec{\gamma}; \mathbf{L})) \, dL_1 \cdots dL_k.$$

We now use this to induce a measure on $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$, the total space of the fibration over the standard simplex $\Delta \subset \mathbb{R}^k$. For $A \subseteq \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ set

$$\text{cone}(A) := \{(\Gamma, t\mathbf{x}) \mid (\Gamma, \mathbf{x}) \in A, t \in (0, 1]\} \subseteq \mathcal{MRG}(S_g \setminus \vec{\gamma}),$$

where Γ is the underlying ribbon graph and \mathbf{x} corresponds to its metric structure.

Denote by $\rho_g(\vec{\gamma}) \in \mathbb{N}$ the number of components of $\vec{\gamma}$ that bound a torus with one boundary component. Let $\sigma_g(\vec{\gamma}) > 0$ be the rational number given by

$$\sigma_g(\vec{\gamma}) := \frac{\prod_{j=1}^c |K_{g_j, b_j}|}{|\text{Stab}_0(\vec{\gamma}) \cap K_g|},$$

where $K_{g_j, b_j} \triangleleft \text{Mod}_{g_j, b_j}$ is the kernel of the mapping class group action on \mathcal{T}_{g_j, b_j} and $K_g \triangleleft \text{Mod}_g$ is the kernel of the mapping class group action on \mathcal{T}_g . These factors arise from special symmetries of moduli spaces of low-complexity surfaces; for a more extended discussion see [Arana-Herrera 2021; Arana-Herrera and Calderon 2022].

Definition 5.3 The *fibred Kontsevich measure* $\overset{\circ}{\eta}_{\text{Kon}}^\Delta$ on $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ is the measure which to every Borel measurable subset A assigns the value

$$\overset{\circ}{\eta}_{\text{Kon}}^\Delta(A) := \frac{\sigma_g(\vec{\gamma})}{2^{\rho_g(\vec{\gamma})}} \int_{\text{cone}(A)} L_1 \cdots L_k \, d\eta_{\text{Kon}}^{\vec{\gamma}}(\Gamma, \mathbf{x}).$$

Denote the total $\overset{\circ}{\eta}_{\text{Kon}}^\Delta$ -mass of $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ by $m_{\vec{\gamma}}$.

Integral points and horoball masses Let $m_{\vec{\gamma}, h}^L$ denote the total mass of $\hat{v}_{u, \vec{\gamma}, h}^L$ (equivalently, of $\tilde{v}_{u, \vec{\gamma}, h}^L$). Note that, since intersection numbers and square roots of extremal lengths scale homogeneously, (5-2) implies that

$$(5-5) \quad m_{\vec{\gamma}, h}^L = L^{6g-6} m_{\vec{\gamma}, h}^1.$$

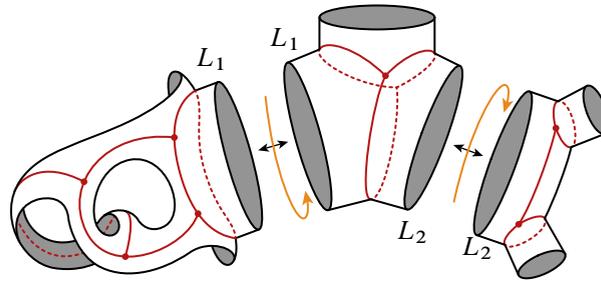


Figure 3: Thickening a ribbon graph and gluing boundaries to recover a Jenkins–Strebel differential with specified critical graph.

To compute $m_{\vec{\gamma}, h}^1$ and, in particular, to relate it to the (fibered) Kontsevich measure, we will need a better structural understanding of the unstable leaf $\mathcal{F}_{[\gamma]}^u$. The following statement gives us the desired control; compare to the discussion of moderately slanted cylinder diagrams in [Arana-Herrera 2020, Section 3] and to the discussion of shear-shape coordinates for quadratic differentials in [Calderon and Farre 2024b].

Lemma 5.4 *The critical graph map $\Xi: \mathcal{F}_{[\gamma]}^u / \text{Stab}_0(\vec{\gamma}) \rightarrow \mathcal{MRG}(S_g \setminus \vec{\gamma})$ demonstrates the quotient $\mathcal{F}_{[\gamma]}^u / \text{Stab}_0(\vec{\gamma})$ as a torus bundle over $\mathcal{MRG}(S_g \setminus \vec{\gamma})$.*

Proof By definition, every $q \in \mathcal{F}_{[\gamma]}^u$ is a unit-area quadratic differential whose horizontal foliation (ie imaginary part) is in the projective class of $[\gamma]$. In particular, this implies that its horizontal cylinders all have equal heights. Cutting along the core curves of these cylinder, we are left with a flat cone structure with totally geodesic boundary on $S_g \setminus \vec{\gamma}$. Collapsing the vertical leaves then defines a deformation retract onto the critical graph $\Xi(q) \in \mathcal{MRG}(S_g \setminus \vec{\gamma})$.

Conversely, given a tuple of metric ribbon graphs $(\Gamma, \mathbf{x}) \in \mathcal{MRG}(S_g \setminus \vec{\gamma})$, there exists a unique choice of height such that gluing together these metric ribbon graphs along cylinders of that height results in a unit-area quadratic differential with the given horizontal separatrices; if $(\Gamma, \mathbf{x}) \in \mathcal{MRG}(S_g \setminus \vec{\gamma}; \mathbf{L})$, then the corresponding height is $1 / \sum_{i=1}^k L_i$. Compare to Figure 3. The only ambiguity in this construction arises in choosing how much to shear along the cylinders of γ . Thus, for each $\mathbf{x} \in \mathcal{MRG}(S_g \setminus \vec{\gamma})$, there is a torus’s worth of ways to construct a quadratic differential $q \in \mathcal{F}_{[\gamma]}^u$ with critical graph $\Xi(q) = (\Gamma, \mathbf{x})$. \square

The fibers of the critical graph map $\Xi: \mathcal{F}_{[\gamma]}^u / \text{Stab}_0(\vec{\gamma}) \rightarrow \mathcal{MRG}(S_g \setminus \vec{\gamma})$ are tori of real dimension k equal to the number of components of $\vec{\gamma}$. Each dimension represents twisting about one of the components. Because of this, the fibers are naturally equipped with a notion of size coming from the circumferences of the corresponding cylinders (representing the possible amounts of twisting, plus a correction factor for extra symmetries). See also the discussion of the “cut-and-glue fibration” in [Arana-Herrera and Calderon 2022, Section 11].

This discussion allows us to express the pushforward of the Thurston measure by Ξ in terms of the Kontsevich measure on the base of the fibration.

Proposition 5.5 *The following identity of measures on $\mathcal{MRG}(S_g \setminus \vec{\gamma})$ holds:*

$$(5-6) \quad d(\Xi_* \tilde{\nu}_{u, \vec{\gamma}}) = \frac{\sigma_g(\vec{\gamma})}{2^{\rho_g(\vec{\gamma})}} L_1 \cdots L_k d\eta_{\text{Kon}}^{\vec{\gamma}, L} dL_1 \cdots dL_k.$$

That is, for every measurable subset $A \subseteq \mathcal{MRG}(S_g \setminus \vec{\gamma})$,

$$\tilde{\nu}_{u, \vec{\gamma}}(\Xi^{-1} A) = \frac{\sigma_g(\vec{\gamma})}{2^{\rho_g(\vec{\gamma})}} \int_{\mathbb{R}_{>0}^k} \eta_{\text{Kon}}^{\vec{\gamma}, L}(A \cap \mathcal{MRG}(S_g \setminus \vec{\gamma}; L)) L_1 \cdots L_k dL_1 \cdots dL_k.$$

Proof There are natural notions of integer points for both $\mathcal{F}_{[\gamma]}^u / \text{Stab}_0(\vec{\gamma})$ and $\mathcal{MRG}(S \setminus \vec{\gamma})$. In the first space, these are square-tiled surfaces. In the second space, these are integral metric ribbon graphs. The measure μ_{Thu} can be defined as a weak- \star limit of counting measures of integrally weighted simple closed multicurves, and the corresponding measure $\nu_{u, \vec{\gamma}}$ can hence be interpreted as a weak- \star limit of counting measures of square-tiled surfaces.

Observe that the critical graph map $\Xi : \mathcal{F}_{[\gamma]}^u / \text{Stab}_0(\vec{\gamma}) \rightarrow \mathcal{MRG}(S_g \setminus \vec{\gamma})$ takes integer points to integer points. Furthermore, over any integral metric ribbon graph in $(\Gamma, \mathbf{x}) \in \mathcal{MRG}(S \setminus \vec{\gamma}; L)$, there are exactly $L_1 \cdots L_k$ square-tiled surfaces in $\Xi^{-1}(\Gamma, \mathbf{x})$ (corresponding to integral amounts of twisting). Thus, up to getting the correct normalization factor, it suffices to show that we can reinterpret the right-hand side of (5-6) in terms of counting integer points. This statement was first observed by Norbury [2010], but is a consequence of the fact that the Kontsevich measure is the Lebesgue measure in the lengths of edges. Compare with the discussion on the fibered Kontsevich measure in [Arana-Herrera and Calderon 2022].

To get the normalizing constant, we must be careful to count integer points weighted by the size of their automorphism group. Equivalently, we must ensure that the pushforwards of measures to orbifolds are weighted by their symmetries. The symmetries of low-complexity moduli spaces (ie the kernel of the mapping class group action on these Teichmüller spaces) hence contribute a factor of $\sigma_g(\vec{\gamma})$. The $2^{-\rho_g(\vec{\gamma})}$ factor comes from the elliptic involution on $S_{1,1}$: its existence implies that there is only half as much twisting in the toral fibers as one might expect. □

Together with (5-5) and the definition of the fibered Kontsevich measure (Definition 5.3), Proposition 5.5 immediately implies the following:

Corollary 5.6 *For any $L > 0$ and any nonzero continuous compactly supported function*

$$h : \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta) \rightarrow \mathbb{R},$$

the total mass $m_{\vec{\gamma}, h}^L$ of $\hat{\nu}_{u, \vec{\gamma}, h}^L$ is equal to

$$m_{\vec{\gamma}, h}^L = L^{6g-6} \int_{\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)} h(\mathbf{x}) d\eta_{\text{Kon}}^{\circ \Delta}(\mathbf{x}).$$

Equidistribution of critical-JS-horoballs We now show that the expanding pushforwards of critical-JS-horoballs equidistribute with respect to the Masur–Veech measure; this is the analogue of [Arana-Herrera and Calderon 2022, Theorem 5.2]. As opposed to that paper, here we will be able to invoke strong results from Teichmüller dynamics to deduce equidistribution relatively quickly; the approach presented here is one of several possible ones. We first consider the horoball measures on unstable leaves.

Proposition 5.7 *For any nonzero continuous compactly supported function $h : \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta) \rightarrow \mathbb{R}$, the following convergence holds with respect to the weak- \star topology for measures on $\mathcal{Q}^1 \mathcal{M}_g$:*

$$\lim_{L \rightarrow \infty} \frac{\hat{v}_{u, \vec{\gamma}, h}^L}{m_{\vec{\gamma}, h}^L} = \frac{\hat{v}_{\text{MV}}}{b_g}.$$

Proof This result follows directly from [Forni 2021, Theorem 1.6].⁴ In fact, the cited theorem is stronger in the sense that it guarantees equidistribution of “horospheres.” The desired result for horoballs can be recovered by integrating horospheres along the direction of the Teichmüller geodesic flow. \square

We now use Proposition 5.7 to deduce the equidistribution of critical-JS-horoballs. Unlike in the hyperbolic setting (see [Arana-Herrera and Calderon 2022, Theorem 5.2 and Corollary 5.3]) we cannot simply push Proposition 5.7 down to \mathcal{M}_g to arrive at the following result; this is related to the fact that ν_{MV} is Mod_g -invariant while $\nu_{u, \vec{\gamma}}$ is not.

Corollary 5.8 *For any nonzero continuous compactly supported function $h : \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta) \rightarrow \mathbb{R}$, the following convergence holds with respect to the weak- \star topology for measures on \mathcal{M}_g :*

$$\lim_{L \rightarrow \infty} \frac{\hat{m}_{\vec{\gamma}, h}^L}{m_{\vec{\gamma}, h}^L} = \frac{\Lambda_g}{b_g} \hat{m}.$$

Proof Let $f : \mathcal{M}_g \rightarrow \mathbb{R}$ be a continuous compactly supported function and set $\tilde{f} : \mathcal{T}_g / \text{Stab}_0(\vec{\gamma}) \rightarrow \mathbb{R}$ to be its pullback to $\mathcal{T}_g / \text{Stab}_0(\vec{\gamma})$. As a direct consequence of the definitions and the Hubbard–Masur theorem, for every $L > 0$ we can rewrite

$$\begin{aligned} \int_{\mathcal{M}_g} f(X) d\hat{m}_{\vec{\gamma}, h}^L(X) &= \int_{\mathcal{T}_g / \text{Stab}_0(\vec{\gamma})} \tilde{f}(X) d\tilde{m}_{\vec{\gamma}, h}^L(X) \\ &= \int_{\mathcal{T}_g / \text{Stab}_0(\vec{\gamma})} \tilde{f}(X) \lambda^+ \left(\frac{\text{JS}(X, \gamma)}{\sqrt{\text{Ext}_X(\gamma)}} \right) d(\pi_{[\gamma]})_* \tilde{\nu}_{u, \vec{\gamma}, h}^L(X) \\ &= \int_{\mathcal{Q}^1 \mathcal{T}_g / \text{Stab}_0(\vec{\gamma})} \tilde{f}(\pi(q)) \lambda^+(q) d\tilde{\nu}_{u, \vec{\gamma}, h}^L(q). \end{aligned}$$

⁴An alternative proof of the desired equidistribution statement can be obtained using the mixing property of the Teichmüller geodesic flow; see for instance [Eskin et al. 2022, Proposition 3.2]. A proof using only the ergodicity of the Teichmüller horocycle flow should also follow from the methods discussed in [Arana-Herrera 2021].

Pushing back down to moduli space and dividing by the total mass $m_{\vec{\gamma},h}^L$ we get

$$\begin{aligned} \frac{1}{m_{\vec{\gamma},h}^L} \int_{\mathcal{M}_g} f(X) d\hat{m}_{\vec{\gamma},h}^L(X) &= \frac{1}{m_{\vec{\gamma},h}^L} \int_{\mathcal{Q}^1\mathcal{M}_g} f(\pi(q))\lambda^+(q) d\hat{\nu}_{u,\vec{\gamma},h}^L(q) \\ &\rightarrow \frac{1}{b_g} \int_{\mathcal{Q}^1\mathcal{M}_g} f(\pi(q))\lambda^+(q) d\hat{\nu}_{\text{MV}}(q), \end{aligned}$$

where the convergence as $L \rightarrow \infty$ follows from Proposition 5.7. We can then integrate off the s_X factor using the product formula (5-1) and invoke Theorem 5.2 to deduce the desired result. \square

6 Equidistribution of critical graphs

Outline of this section We state and prove our main result in the case of a general multicurve. Theorem 1.1 follows as a special case. The proof is obtained by putting the results of the previous sections into the outline discussed in the introduction. Namely, after reducing the equidistribution problem at hand to a counting problem for curves whose Jenkins–Strebel differentials have constrained critical graphs, averaging and unfolding techniques allow us to further reduce to an equidistribution question for critical horoball measures. Propositions 3.6 and 4.3 will play an important role at this stage of the proof. The results of Section 5, which rely on the ergodicity of the Teichmüller horocycle flow, guarantee these measures equidistribute. The relationship between the total mass of critical horoball measures and the Kontsevich measure, explained in Corollary 5.6, then allows us to relate the asymptotics of our curve counting problem to the Kontsevich measure.

From equidistribution to counting Fix an ordered oriented simple closed multicurve $\vec{\gamma}$. For every $L > 0$, consider the extremal length counting function

$$s(X, \vec{\gamma}, L) := \#\{\alpha \in \text{Mod}_g \cdot \vec{\gamma} \mid \sqrt{\text{Ext}_X(\vec{\alpha})} \leq L\}.$$

This does not depend on the marking of $X \in \mathcal{T}_g$ but only on its underlying conformal structure. Work of Mirzakhani [2008b] gives sharp asymptotics for this count:

$$(6-1) \quad \lim_{L \rightarrow \infty} \frac{s(X, \vec{\gamma}, L)}{L^{6g-6}} = \frac{m_{\vec{\gamma}}\Lambda_g}{b_g}.$$

Here $\Lambda_g > 0$ is the Hubbard–Masur constant, $b_g > 0$ is the total Masur–Veech volume of $\mathcal{Q}^1\mathcal{M}_g$, and $m_{\vec{\gamma}}$ is the total mass of the fibered Kontsevich measure $\overset{\circ}{\eta}_{\text{Kon}}^\Delta$ on the moduli space $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$.

Remark 6.1 As discussed in [Arana-Herrera and Calderon 2022, Remark 7.6], the constant $m_{\vec{\gamma}}$ can be reconciled with the usual statement of (6-1) involving the “frequency” $c(\gamma)$ by observing that we are counting ordered oriented multicurves (introducing a factor of $[\text{Stab}(\gamma) : \text{Stab}_0(\vec{\gamma})]$) and that the coefficient of the top-degree part of the Weil–Peterson volume polynomial of $S_g \setminus \vec{\gamma}$ is exactly $m_{\vec{\gamma}}$.

Recall that our goal is to study the asymptotic distribution of the counting measures on $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ given by

$$\eta_{X, \vec{\gamma}}^L := \sum_{\vec{\alpha} \in \text{Mod}_g \cdot \vec{\gamma}} \mathbb{1}_{[0, L]}(\sqrt{\text{Ext}_X(\alpha)}) \delta_{\Xi^1(X, \vec{\alpha})}.$$

The following is the general version of Theorem 1.1; its proof will occupy the rest of this section. Compare with [Arana-Herrera and Calderon 2022, Theorem 7.7].

Theorem 6.2 *Let $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ be an ordered oriented simple closed multicurve on S_g and $X \in \mathcal{M}_g$ be a complex structure on S_g . Then*

$$\lim_{L \rightarrow \infty} \frac{\eta_{X, \vec{\gamma}}^L}{s(X, \vec{\gamma}, L)} = \frac{\overset{\circ}{\eta}_{\text{Kon}}^\Delta}{m_{\vec{\gamma}}}$$

with respect to the weak- \star topology for measures on $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$.

Remark 6.3 For the sake of brevity and readability, throughout the rest of this section we will shorten $\mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ to \mathcal{MRG} . We will not consider any other spaces of ribbon graphs in the sequel.

As explained in Section 1, Theorem 6.2 is equivalent to a counting problem for metric ribbon graphs. More concretely, it is enough to show that for every continuous compactly supported $f : \mathcal{MRG} \rightarrow \mathbb{R}_{\geq 0}$,

$$\lim_{L \rightarrow \infty} \frac{1}{s(X, \vec{\gamma}, L)} \int_{\mathcal{MRG}} f(\mathbf{x}) d\eta_{X, \vec{\gamma}}^L(\mathbf{x}) = \frac{1}{m_{\vec{\gamma}}} \int_{\mathcal{MRG}} f(\mathbf{x}) d\overset{\circ}{\eta}_{\text{Kon}}^\Delta(\mathbf{x}).$$

For every $L > 0$ consider the f -weighted counting function

$$(6-2) \quad c(X, \vec{\gamma}, f, L) := \int_{\mathcal{MRG}} f(\mathbf{x}) d\eta_{X, \vec{\gamma}}^L(\mathbf{x}) = \sum_{\vec{\alpha} \in \text{Mod}_g \cdot \vec{\gamma}} \mathbb{1}_{[0, L]}(\sqrt{\text{Ext}_X(\alpha)}) f(\Xi^1(X, \vec{\alpha})).$$

Observe that, although it is not compactly supported, one can follow the same argument with $f \equiv 1$ and recover the usual counting function $s(X, \vec{\gamma}, L)$.

The rest of this section is devoted to proving the ensuing result, from which Theorem 6.2 follows directly by the above discussion. This is a generalized version of Theorem 1.2 from the introduction in the case of a general multicurve.

Theorem 6.4 *Let $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ be an ordered oriented simple closed multicurve on S_g and $X \in \mathcal{M}_g$ be a complex structure on S_g . Then, for every continuous compactly supported $f : \mathcal{MRG} \rightarrow \mathbb{R}_{\geq 0}$,*

$$\lim_{L \rightarrow \infty} \frac{c(X, \vec{\gamma}, f, L)}{s(X, \vec{\gamma}, L)} = \frac{1}{m_{\vec{\gamma}}} \int_{\mathcal{MRG}} f(\mathbf{x}) d\overset{\circ}{\eta}_{\text{Kon}}^\Delta(\mathbf{x}).$$

For the rest of this section we fix a marked conformal structure $X \in \mathcal{T}_g$, an ordered oriented simple closed multicurve $\vec{\gamma}$ on S_g , and a nonzero nonnegative continuous compactly supported function $f : \mathcal{MRG} \rightarrow \mathbb{R}_{\geq 0}$.

Averaging counts Our next goal is to average the counting functions introduced in (6-2) over small neighborhoods of moduli space. Using the results from Section 4, we first study how these counting functions vary in such neighborhoods.

Recall that Proposition 4.3 states that for every pair of conformal structures $X, Y \in \mathcal{T}_g$ that are sufficiently close in the Teichmüller metric and most ordered oriented simple closed multicurves $\vec{\alpha}$ on S_g , the critical graphs $\Xi^1(X, \vec{\alpha})$ and $\Xi^1(Y, \vec{\alpha})$ belong to the same facet of \mathcal{MRG} and the corresponding edges have length differing by a multiplicative constant. We now define analogous neighborhoods in the moduli space of ribbon graphs.

Given $\mathbf{x} \in \mathcal{MRG}$ and a positive constant $\varepsilon > 0$, denote by $N_\varepsilon(\mathbf{x})$ the set of all $\mathbf{y} \in \mathcal{MRG}$ in the same facet as \mathbf{x} , ie with the same topological type of underlying ribbon graph as \mathbf{x} , such that for every edge e of \mathbf{x} and \mathbf{y} ,

$$e^{-\varepsilon}|e|_{\mathbf{x}} \leq |e|_{\mathbf{y}} \leq e^\varepsilon|e|_{\mathbf{x}}.$$

Here we have implicitly fixed a local marking so we can compare the weights of specific edges. For every $\varepsilon > 0$ consider the averaged functions

$$f_\varepsilon^{\min}, f_\varepsilon^{\max}: \mathcal{MRG} \rightarrow \mathbb{R}_{\geq 0}$$

given by

$$f_\varepsilon^{\min}(\mathbf{x}) := \min_{\mathbf{y} \in N_\varepsilon(\mathbf{x})} f(\mathbf{y}), \quad f_\varepsilon^{\max}(\mathbf{x}) := \max_{\mathbf{y} \in N_\varepsilon(\mathbf{x})} f(\mathbf{y}).$$

We begin our proof of Theorem 6.4 with the following estimate, which allows us to compare counts of curves on X with counts on nearby surfaces. Propositions 3.6 and 4.3 play a crucial role in the proof of this result. For $X \in \mathcal{M}_g$, a function f on \mathcal{MRG} , and functions F and G on \mathbb{R} , we say that $F = O_{X,f}(G)$ if $F = O(G)$ where the implicit constants depend only on X and f .

Proposition 6.5 *For every $\varepsilon > 0$ there exists $\delta := \delta(X, \varepsilon) \in (0, \varepsilon)$ such that for every $Y \in \mathcal{M}_g$ with $d_{\text{Teich}}(X, Y) < \delta$, the following estimates hold:*

$$(6-3) \quad c(Y, \vec{\gamma}, f_\varepsilon^{\min}, e^{-\delta}L) + O_{X,f}(L^{6g-7} + \varepsilon L^{6g-6}) \leq c(X, \vec{\gamma}, f, L),$$

$$(6-4) \quad c(X, \vec{\gamma}, f, L) \leq c(Y, \vec{\gamma}, f_\varepsilon^{\max}, e^\delta L) + O_{X,f}(L^{6g-7} + \varepsilon L^{6g-6}).$$

Proof We prove (6-4). Similar arguments yield a proof of (6-3).

We begin by applying Proposition 3.6 to focus our attention on curves in the mapping class group orbit of $\vec{\gamma}$ whose corresponding Jenkins–Strebel differentials have no short saddle connections. Recall that for any compact set $\mathcal{K} \subseteq \mathcal{T}_g$ and any $\sigma > 0$, we use $F(\mathcal{K}, \sigma) \subseteq \mathcal{MF}_g(\mathbb{Z})$ to denote the set of integrally weighted simple closed multicurves α on S_g such that $\text{JS}(Y, \alpha)$, after rescaling to have unit area, has a σ -short saddle connection for some $Y \in \mathcal{K}$. By abuse of notation we also use $F(\mathcal{K}, \sigma)$ to denote the set of ordered oriented simple closed multicurves $\vec{\alpha}$ on S_g whose underlying multicurve belongs to $F(\mathcal{K}, \sigma)$.

Now take $\mathcal{K} \subseteq \mathcal{T}_g$ to be the closed unit ball in the Teichmüller metric centered at X and $\sigma = \varepsilon > 0$ arbitrary. Consider the truncated counting function

$$c^\dagger(X, \vec{\gamma}, f, L) := \sum_{\vec{\alpha} \in \text{Mod}_g \cdot \vec{\gamma} \setminus F(\mathcal{K}, \varepsilon)} \mathbb{1}_{[0, L]}(\sqrt{\text{Ext}_X(\alpha)}) f(\Xi^1(X, \vec{\alpha})).$$

By Proposition 3.6 it follows that

$$(6-5) \quad c(X, \vec{\gamma}, f, L) = c^\dagger(X, \vec{\gamma}, f, L) + \|f\|_\infty O_{\mathcal{K}}(L^{6g-7} + \varepsilon L^{6g-6}).$$

We can now invoke the geometric comparison results of Section 4. Consider $\delta = \delta(\mathcal{K}, \varepsilon, \varepsilon) > 0$ as in Proposition 4.3 and set

$$\delta' := \min\{1, \delta, \varepsilon\}.$$

Now for any $Y \in \mathcal{T}_g$ such that $d_{\text{Teich}}(X, Y) < \delta'$ and any $\vec{\alpha} \in \text{Mod}_g \cdot \vec{\gamma} \setminus F(\mathcal{K}, \varepsilon)$, it follows from Proposition 4.3 that $\Xi^1(X, \vec{\alpha})$ and $\Xi^1(Y, \vec{\alpha})$ are in the same facet of \mathcal{MRG} and that for every edge e of such metric ribbon graphs,

$$e^{-\varepsilon} |e|_{\Xi^1(Y, \vec{\alpha})} \leq |e|_{\Xi^1(X, \vec{\alpha})} \leq e^\varepsilon |e|_{\Xi^1(Y, \vec{\alpha})}.$$

It follows that $\Xi^1(X, \vec{\alpha}) \in N_\varepsilon(\Xi^1(Y, \vec{\alpha}))$, so by definition

$$(6-6) \quad f(\Xi^1(X, \vec{\alpha})) \leq f_\varepsilon^{\max}(\Xi^1(Y, \vec{\alpha})).$$

By Kerckhoff’s characterization of the Teichmüller metric (Theorem 2.4),

$$(6-7) \quad \sqrt{\text{Ext}_Y(\alpha)} \leq e^{\delta'} \sqrt{\text{Ext}_X(\alpha)}.$$

From (6-6) and (6-7) we deduce

$$(6-8) \quad c^\dagger(X, \vec{\gamma}, f, L) \leq c(Y, \vec{\gamma}, f_\varepsilon^{\max}, e^{\delta'} L).$$

Putting together (6-5) and (6-8) we conclude

$$c(X, \vec{\gamma}, f, L) \leq c(Y, \vec{\gamma}, f_\varepsilon^{\max}, e^{\delta'} L) + \|f\|_\infty O_{\mathcal{K}}(L^{6g-7} + \varepsilon L^{6g-6}). \quad \square$$

We now integrate Proposition 6.5 against the pushforward \hat{m} of the Masur–Veech measure. For every $\delta \in (0, 1)$ denote by $U_X(\delta) \subseteq \mathcal{M}_g$ the open ball of radius δ centered at $X \in \mathcal{M}_g$ with respect to the Teichmüller metric and let $\beta_{X, \delta} : \mathcal{M}_g \rightarrow \mathbb{R}_{\geq 0}$ be any bump function supported on $U_X(\delta)$ of total m mass 1.

Corollary 6.6 *Let all notation be as in Proposition 6.5. Then $c(X, \vec{\gamma}, f, L)$ is bounded below by*

$$(6-9) \quad \int_{\mathcal{M}_g} \beta_{X, \delta}(Y) c(Y, \vec{\gamma}, f_\varepsilon^{\min}, e^{-\delta} L) d\hat{m}(Y) + O_{X, f}(L^{6g-7} + \varepsilon L^{6g-6})$$

and above by

$$(6-10) \quad \int_{\mathcal{M}_g} \beta_{X, \delta}(Y) c(Y, \vec{\gamma}, f_\varepsilon^{\max}, e^\delta L) d\hat{m}(Y) + O_{X, f}(L^{6g-7} + \varepsilon L^{6g-6}).$$

Unfolding averaged counts Unfolding the integrals in (6-9) and (6-10) over $\mathcal{T}_g/\text{Stab}_0(\vec{\gamma})$ and pushing them back down to \mathcal{M}_g in a suitable way will reduce the proof of Theorem 6.4 to an application of Corollary 5.8. The following proposition describes this principle; the reader should also compare to [Arana-Herrera 2022, Proposition 3.3; Arana-Herrera and Calderon 2022, Proposition 6.6].

Proposition 6.7 Fix a continuous compactly supported function $h: \mathcal{MRG} \rightarrow \mathbb{R}_{\geq 0}$. Then for every $\delta > 0$ and every $L > 0$,

$$(6-11) \quad \int_{\mathcal{M}_g} \beta_{X,\delta}(Y)c(Y, \vec{\gamma}, h, L) d\hat{m}(Y) = \int_{\mathcal{M}_g} \beta_{X,\delta}(Y) d\hat{m}_{\vec{\gamma},h}^L(Y).$$

Remark 6.8 Our weight function has changed names; this is because we eventually apply Proposition 6.7 with h equal to the functions f_ε^{\max} and f_ε^{\min} .

Proof Let $\delta > 0$ and $L > 0$ be arbitrary. For every $Y \in \mathcal{M}_g$ one can rewrite the counting function $c(Y, \vec{\gamma}, h, L)$ as follows:

$$\begin{aligned} c(Y, \vec{\gamma}, h, L) &= \sum_{\vec{\alpha} \in \text{Mod}_g \vec{\gamma}} \mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_Y(\alpha)})h(\Xi^1(Y, \vec{\alpha})) \\ &= \sum_{[\phi] \in \text{Mod}_g/\text{Stab}_0(\vec{\gamma})} \mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_Y(\phi \cdot \vec{\gamma})})h(\Xi^1(Y, \phi \cdot \vec{\gamma})) \\ &= \sum_{[\phi] \in \text{Mod}_g/\text{Stab}_0(\vec{\gamma})} \mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_{\phi^{-1} \cdot Y}(\vec{\gamma})})h(\Xi^1(\phi^{-1} \cdot Y, \vec{\gamma})) \\ &= \sum_{[\phi] \in \text{Stab}_0(\vec{\gamma}) \setminus \text{Mod}_g} \mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_{\phi \cdot Y}(\vec{\gamma})})h(\Xi^1(\phi \cdot Y, \vec{\gamma})). \end{aligned}$$

Let us record this fact as

$$(6-12) \quad c(Y, \vec{\gamma}, h, L) = \sum_{[\phi] \in \text{Stab}_0(\vec{\gamma}) \setminus \text{Mod}_g} \mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_{\phi \cdot Y}(\vec{\gamma})})h(\Xi^1(\phi \cdot Y, \vec{\gamma})).$$

Denote by $p_{\vec{\gamma}}: \mathcal{T}_g/\text{Stab}_0(\vec{\gamma}) \rightarrow \mathcal{M}_g$ the quotient map and let $\tilde{\beta}_{X,\delta} = \beta_{X,\delta} \circ p_{\vec{\gamma}}$ be the lift of $\beta_{X,\delta}$ to this cover. Unfolding the integral on the left hand side of (6-11) using (6-12) it follows that

$$\begin{aligned} \int_{\mathcal{M}_g} \beta_{X,\delta}(Y)c(Y, \vec{\gamma}, h, L) d\hat{m}(Y) &= \int_{\mathcal{T}_g/\text{Stab}_0(\vec{\gamma})} \tilde{\beta}_{X,\delta}(Y)\mathbb{1}_{[0,L]}(\sqrt{\text{Ext}_Y(\vec{\gamma})})h(\Xi^1(Y, \vec{\gamma})) d\tilde{m}(Y) \\ &= \int_{\mathcal{T}_g/\text{Stab}_0(\vec{\gamma})} \tilde{\beta}_{X,\delta}(Y) d\tilde{m}_{\vec{\gamma},h}^L(Y) = \int_{\mathcal{M}_g} \beta_{X,\delta}(Y) d\hat{m}_{\vec{\gamma},h}^L(Y), \end{aligned}$$

where the second equality follows from the definition of the horoball measure $\tilde{m}_{\vec{\gamma},h}^L$ appearing in (5-3) and the third equality follows by taking the pushforward. \square

Reducing counting to equidistribution We can now finish the proof of Theorem 6.4 by applying our equidistribution results from Section 5. Proposition 6.7 relates averages of the f -weighted counting function $c(X, \vec{\gamma}, f, L)$ to horoball measures; our strategy now is to relate the original counting function and the mass of these measures, which we can then compare with the count $s(X, \vec{\gamma}, L)$ of all ordered oriented simple closed multicurves in the Mod_g -orbit of $\vec{\gamma}$.

Proof of Theorem 6.4 Recall that we are aiming to prove that

$$\lim_{L \rightarrow \infty} \frac{c(X, \vec{\gamma}, f, L)}{s(X, \vec{\gamma}, L)} = \frac{1}{m_{\vec{\gamma}}} \int_{\mathcal{MRG}} f(\mathbf{x}) d\hat{\eta}_{\text{Kon}}^\Delta(\mathbf{x}).$$

By Corollary 5.6, for any continuous compactly supported function $h: \mathcal{MRG} \rightarrow \mathbb{R}_{\geq 0}$, the total mass $m_{\vec{\gamma}, h}^L$ of the unstable horoball measure $\hat{m}_{\vec{\gamma}, h}^L$ on \mathcal{M}_g is L^{6g-6} times the integral

$$r(\vec{\gamma}, h) := \int_{\mathcal{MRG}} h(\mathbf{x}) d\hat{\eta}_{\text{Kon}}^\Delta(\mathbf{x}).$$

Proving Theorem 6.4 is then equivalent to showing that both of the following inequalities hold:

$$(6-13) \quad \frac{r(\vec{\gamma}, f)}{m_{\vec{\gamma}}} \leq \liminf_{L \rightarrow \infty} \frac{c(X, \vec{\gamma}, f, L)}{s(X, \vec{\gamma}, L)},$$

$$(6-14) \quad \limsup_{L \rightarrow \infty} \frac{c(X, \vec{\gamma}, f, L)}{s(X, \vec{\gamma}, L)} \leq \frac{r(\vec{\gamma}, f)}{m_{\vec{\gamma}}}.$$

We verify (6-14) by averaging and unfolding; a proof of (6-13) can be obtained following the same argument. Let $\varepsilon \in (0, 1)$ be arbitrary and $\delta = \delta(X, \varepsilon)$ as in Corollary 6.6. Shrinking δ as necessary, we can also assume that $e^{(6g-6)\delta} \leq 2$. Set $h := f_\varepsilon^{\max}$; Corollary 6.6 then implies we can average our counting function to get

$$c(X, \vec{\gamma}, f, L) \leq \int_{\mathcal{M}_g} \beta_{X, \delta}(Y) c(Y, \vec{\gamma}, h, e^\delta L) d\hat{m}(Y) + O_{X, f}(L^{6g-7} + \varepsilon L^{6g-6}).$$

Set $L' := e^\delta L$. Unfolding the integral, ie using Proposition 6.7, we deduce

$$c(X, \vec{\gamma}, f, L) \leq \int_{\mathcal{M}_g} \beta_{X, \delta}(Y) d\hat{m}_{\vec{\gamma}, h}^{L'}(Y) + O_{X, f}(L^{6g-7} + \varepsilon L^{6g-6}).$$

Dividing this inequality by $m_{\vec{\gamma}, h}^{L'}$ (which is nonzero so long as $f \neq 0$) we get

$$\frac{c(X, \vec{\gamma}, f, L)}{m_{\vec{\gamma}, h}^{L'}} \leq \int_{\mathcal{M}_g} \beta_{X, \delta}(Y) d\frac{\hat{m}_{\vec{\gamma}, h}^{L'}}{m_{\vec{\gamma}, h}^{L'}}(Y) + \frac{O_{X, f}(\varepsilon)}{r(\vec{\gamma}, h)},$$

where we have invoked Corollary 5.6 and our assumption on δ to simplify the bound on the far right. Since $h \geq f$, we know that $r(\vec{\gamma}, h) \geq r(\vec{\gamma}, f)$, so we can also absorb this term into our big O estimate.

Taking the \limsup as $L \rightarrow \infty$ and applying Corollary 5.8, we deduce that

$$(6-15) \quad \limsup_{L \rightarrow \infty} \frac{c(X, \vec{\gamma}, f, L)}{m_{\vec{\gamma}, h}^{L'}} \leq \frac{\Lambda_g}{b_g} \int_{\mathcal{M}_g} \beta_{X, \delta}(Y) d\hat{m}(Y) + O_{X, f}(\varepsilon) = \frac{\Lambda_g}{b_g} + O_{X, f}(\varepsilon).$$

Combining this with Mirzakhani’s asymptotic count (6-1) and our expression for horoball masses (Corollary 5.6), we arrive at the following estimate:

$$(6-16) \quad \limsup_{L \rightarrow \infty} \frac{c(X, \vec{\gamma}, f, L)}{s(X, \vec{\gamma}, L)} \leq \frac{r(\vec{\gamma}, h)e^{(6g-6)\delta}}{m_{\vec{\gamma}}} + O_{X,f}(\varepsilon).$$

We now shrink our approximating neighborhoods. By definition, $h := f_\varepsilon^{\max} \searrow f$ pointwise as $\varepsilon \searrow 0$. In particular, by the monotone convergence theorem,

$$\lim_{\varepsilon \searrow 0} r(\vec{\gamma}, f_\varepsilon^{\max}) = \lim_{\varepsilon \searrow 0} \int_{\mathcal{MRG}} f_\varepsilon^{\max}(\mathbf{x}) d\eta_{\text{Kon}}^\circ(\mathbf{x}) = \int_{\mathcal{MRG}} f(\mathbf{x}) d\eta_{\text{Kon}}^\circ(\mathbf{x}) = r(\vec{\gamma}, f).$$

Sending ε , and hence δ , to 0 we get that the right-hand side of (6-16) converges to $r(\vec{\gamma}, f)/m_{\vec{\gamma}}$, as desired. □

This completes the proof of the counting result (Theorem 6.4), and hence also the proof of our main equidistribution result (Theorem 6.2).

Simultaneous equidistribution Drawing inspiration from [Aka et al. 2016a; 2016b; Einsiedler et al. 2019; Arana-Herrera and Calderon 2022], we now discuss the issue of simultaneous equidistribution. More concretely, we show that the placement of simple closed multicurves in the space of singular measured foliations is asymptotically independent from the critical graph of the Jenkins–Strebel differential they define on a given Riemann surface.

Recall that \mathcal{PMF}_g denotes the space of projective singular measured foliations on S_g and that $[\lambda] \in \mathcal{PML}_g$ denotes the projective class of $\lambda \in \mathcal{MF}_g$. Given $X \in \mathcal{T}_g$, consider the coned-off Thurston measure μ_{Thu}^X which to every measurable subset $A \subseteq \mathcal{PML}_g$ assigns the value

$$\mu_{\text{Thu}}^X(A) := \mu_{\text{Thu}}(\{\lambda \in \mathcal{ML}_g \mid \text{Ext}_X(\lambda) \leq 1, [\lambda] \in A\}).$$

By Lemma 5.1 and Theorem 5.2, the total mass of this measure is precisely $\Lambda(X) = \Lambda_g > 0$, the Hubbard–Masur constant. Furthermore, as discussed in [Athreya et al. 2012, Proposition 2.3], this measure can be related to the fiberwise measures s_X and the Hubbard–Masur function λ^+ as follows:

Proposition 6.9 *Let $X \in \mathcal{T}_g$ be a marked complex structure on S_g . Consider the homeomorphism $[\mathfrak{S}]: S(X) \rightarrow \mathcal{PMF}_g$. Then*

$$\lambda^+(q) = \frac{d[\mathfrak{S}]_* \mu_{\text{Thu}}^X(q)}{ds_X}.$$

Fix an ordered oriented simple closed multicurve $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ on S_g and a marked hyperbolic structure $X \in \mathcal{T}_g$. Recall that $\gamma \in \mathcal{MF}_g$ denotes the equivalence class of $\vec{\gamma}$ as a singular measured foliation on S_g . For every $L > 0$ consider the counting measure on \mathcal{PML}_g given by

$$\zeta_{\vec{\gamma}, X}^L := \sum_{\vec{\alpha} \in \text{Mod}_g \cdot \vec{\gamma}} \mathbb{1}_{[0, L]}(\sqrt{\text{Ext}_X(\alpha)}) \delta_{[\alpha]}.$$

Observe that we weight a given projective class $[\alpha]$ by the index $[\text{Stab}(\gamma) : \text{Stab}_0(\vec{\gamma})]$; this allows us to use the oriented count $s(X, \vec{\gamma}, L)$ below. The following result can be deduced directly from Mirzakhani’s work [2008b, Theorem 6.4]:

Theorem 6.10 *In the weak- \star topology for measures on \mathcal{PML}_g ,*

$$\lim_{L \rightarrow \infty} \frac{\zeta_{\vec{\gamma}, X}^L}{s(X, \vec{\gamma}, L)} = \frac{\mu_{\text{Thu}}^X}{\Lambda_g}.$$

It is natural to consider the question of simultaneous equidistribution for the limits in Theorems 6.2 and 6.10. More precisely, fix an ordered oriented simple closed multicurve $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ on S_g and $X \in \mathcal{T}_g$. For every $L > 0$ consider the counting measure on $\mathcal{PMF}_g \times \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$ given by

$$\xi_{\vec{\gamma}, X}^L := \sum_{\vec{\alpha} \in \text{Mod}_g \cdot \vec{\gamma}} \mathbb{1}_{[0, L]}(\sqrt{\text{Ext}_X(\alpha)}) \delta_{[\alpha]} \otimes \delta_{\Xi^1(X, \vec{\alpha})}.$$

As always, this measure depends only on the underlying hyperbolic structure of $X \in \mathcal{T}_g$ and not on its marking. The question of equidistribution as $L \rightarrow \infty$ of these measures can be tackled by the same methods used in the proof of Theorem 6.2 subject to some important modifications we now discuss.

The most important difference in the proof comes from the equidistribution result that one must use instead of Corollary 5.8. Denote by $\mathcal{P}^1\mathcal{T}_g = \mathcal{T}_g \times \mathcal{PMF}_g$ the bundle projective singular measured foliations over Teichmüller space. This bundle carries a natural measure ν given by the following disintegration formula:

$$d\mathbf{n}(X, [\lambda]) = d\mu_{\text{Thu}}^X([\lambda]) d\mathbf{m}(X).$$

The quotient $\mathcal{P}^1\mathcal{M}_g = \mathcal{P}^1\mathcal{T}_g / \text{Mod}_g$ by the diagonal action of the mapping class group is the bundle of projective singular measured foliations over moduli space. As this action preserves the measure \mathbf{n} , we obtain a measure $\hat{\mathbf{n}}$ on $\mathcal{P}^1\mathcal{M}_g$ satisfying the following disintegration formula:

$$d\hat{\mathbf{n}}(X, [\lambda]) = d\mu_{\text{Thu}}^X([\lambda]) d\hat{\mathbf{m}}(X).$$

To define the relevant horoballs we want to consider over $\mathcal{P}^1\mathcal{M}_g$ we proceed as follows. For any $L > 0$ and any nonzero continuous compactly supported function $h : \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta) \rightarrow \mathbb{R}$, define the Ξ^1 -horoball measure on $\mathcal{P}^1\mathcal{T}_g$ by

$$d\mathbf{n}_{\vec{\gamma}, h}^L(X, [\lambda]) := d\delta_{[\gamma]}([\lambda]) d\mathbf{m}_{\vec{\gamma}, h}^L(X),$$

where $\delta_{[\gamma]}$ denotes the delta mass at $[\gamma] \in \mathcal{PMF}_g$. Directly from the definitions one can check that this measure is $\text{Stab}_0(\vec{\gamma})$ -invariant. It follows that one can locally push this measure forward to $\mathcal{P}^1\mathcal{T}_g / \text{Stab}_0(\vec{\gamma})$ to get a measure $\hat{\mathbf{n}}_{\vec{\gamma}, h}^L$. Denote by $\hat{\mathbf{n}}_{\vec{\gamma}, h}^L$ the measure on $\mathcal{P}^1\mathcal{M}_g$ obtained by pushing forward this measure. Notice that the total mass of this measure is exactly $m_{\vec{\gamma}, h}^L > 0$.

The following is the main equidistribution result needed to address the simultaneous equidistribution question alluded to above:

Proposition 6.11 For any nonzero continuous compactly supported function $h : \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta) \rightarrow \mathbb{R}$, the following convergence holds with respect to the weak- \star topology for measures on $\mathcal{P}^1 \mathcal{M}_g$:

$$\lim_{L \rightarrow \infty} \frac{\hat{n}_{\vec{\gamma}, h}^L}{m_{\vec{\gamma}, h}^L} = \frac{\hat{n}}{b_g}.$$

Proof Let $f : \mathcal{P}^1 \mathcal{M}_g \rightarrow \mathbb{R}$ be a continuous compactly supported function and set $\tilde{f} : \mathcal{P}^1 \mathcal{T}_g / \text{Stab}_0(\vec{\gamma}) \rightarrow \mathbb{R}$ to be its pullback. As a direct consequence of the definitions and the Hubbard–Masur theorem, for every $L > 0$ we can rewrite

$$\begin{aligned} \int_{\mathcal{P}^1 \mathcal{M}_g} f(X, [\lambda]) d\hat{n}_{\vec{\gamma}, h}^L(X, [\lambda]) &= \int_{\mathcal{P}^1 \mathcal{T}_g / \text{Stab}_0(\vec{\gamma})} \tilde{f}(X, [\lambda]) d\tilde{n}_{\vec{\gamma}, h}^L(X, [\lambda]) \\ &= \int_{\mathcal{T}_g / \text{Stab}_0(\vec{\gamma})} \tilde{f}(X, [\gamma]) \lambda^+ \left(\frac{\text{JS}(X, \gamma)}{\sqrt{\text{Ext}_X(\gamma)}} \right) d(\pi[\gamma])_* \tilde{v}_{u, \vec{\gamma}, h}^L(X) \\ &= \int_{\mathcal{Q}^1 \mathcal{T}_g / \text{Stab}_0(\vec{\gamma})} \tilde{f}(\pi(q), [\mathfrak{S}(q)]) \lambda^+(q) d\tilde{v}_{u, \vec{\gamma}, h}^L(q). \end{aligned}$$

Pushing back down to moduli space and dividing by the total mass $m_{\vec{\gamma}, h}^L$ we get

$$\begin{aligned} \frac{1}{m_{\vec{\gamma}, h}^L} \int_{\mathcal{P}^1 \mathcal{M}_g} f(X, [\lambda]) d\hat{n}_{\vec{\gamma}, h}^L(X, [\lambda]) &= \frac{1}{m_{\vec{\gamma}, h}^L} \int_{\mathcal{Q}^1 \mathcal{M}_g} f(\pi(q), [\mathfrak{S}(q)]) \lambda^+(q) d\hat{v}_{u, \vec{\gamma}, h}^L(q) \\ &\rightarrow \frac{1}{b_g} \int_{\mathcal{Q}^1 \mathcal{M}_g} f(\pi(q), [\mathfrak{S}(q)]) \lambda^+(q) d\hat{v}_{\text{MV}}(q), \end{aligned}$$

where the convergence as $L \rightarrow \infty$ follows from Proposition 5.7. Disintegrating the Masur–Veech measure fiberwise and using Proposition 6.9 we deduce

$$\begin{aligned} \frac{1}{b_g} \int_{\mathcal{Q}^1 \mathcal{M}_g} f(\pi(q), [\mathfrak{S}(q)]) \lambda^+(q) d\hat{v}_{\text{MV}}(q) &= \frac{1}{b_g} \int_{\mathcal{M}_g} \int_{\mathcal{S}(X)} f(X, [\mathfrak{S}(q)]) \lambda^+(q) ds_X(q) d\hat{m}(X) \\ &= \frac{1}{b_g} \int_{\mathcal{P}^1 \mathcal{M}_g} f(X, [\lambda]) d\hat{n}(X, [\lambda]). \end{aligned}$$

Putting the identities above together finishes the proof. □

The following simultaneous equidistribution result can be proved by using similar arguments as in the proof of Theorem 6.2 but working over the bundle $\mathcal{P}^1 \mathcal{M}_g$ instead of \mathcal{M}_g and using Proposition 6.11 in place of Corollary 5.8; compare to [Arana-Herrera 2022, Proof of Theorem 3.5].

Theorem 6.12 Let $\vec{\gamma} := (\vec{\gamma}_1, \dots, \vec{\gamma}_k)$ be an ordered oriented simple closed multicurve on S_g and $X \in \mathcal{M}_g$ be a complex structure on S_g . Then, with respect to the weak- \star topology for measures on $\mathcal{PMF}_g \times \mathcal{MRG}(S_g \setminus \vec{\gamma}; \Delta)$,

$$\lim_{L \rightarrow \infty} \frac{\xi_{\vec{\gamma}, X}^L}{s(X, \vec{\gamma}, L)} = \frac{\mu_{\text{Thu}}^X}{\Lambda_g} \otimes \frac{\hat{\eta}_{\text{Kon}}^\Delta}{m_{\vec{\gamma}}}.$$

As a consequence, we see that even when prescribing how a set of curves coarsely wraps around S_g (for example, by fixing a maximal train track chart for \mathcal{MF}_g), the critical graphs defined by those curves remain uniformly distributed.

References

- [Aka et al. 2016a] **M Aka, M Einsiedler, U Shapira**, *Integer points on spheres and their orthogonal grids*, J. Lond. Math. Soc. 93 (2016) 143–158 MR Zbl
- [Aka et al. 2016b] **M Aka, M Einsiedler, U Shapira**, *Integer points on spheres and their orthogonal lattices*, Invent. Math. 206 (2016) 379–396 MR Zbl
- [Arana-Herrera 2020] **F Arana-Herrera**, *Counting square-tiled surfaces with prescribed real and imaginary foliations and connections to Mirzakhani’s asymptotics for simple closed hyperbolic geodesics*, J. Mod. Dyn. 16 (2020) 81–107 MR Zbl
- [Arana-Herrera 2021] **F Arana-Herrera**, *Equidistribution of families of expanding horospheres on moduli spaces of hyperbolic surfaces*, Geom. Dedicata 210 (2021) 65–102 MR Zbl
- [Arana-Herrera 2022] **F Arana-Herrera**, *Counting hyperbolic multigeodesics with respect to the lengths of individual components and asymptotics of Weil–Petersson volumes*, Geom. Topol. 26 (2022) 1291–1347 MR Zbl
- [Arana-Herrera 2023] **F Arana-Herrera**, *Effective mapping class group dynamics, I: Counting lattice points in Teichmüller space*, Duke Math. J. 172 (2023) 1437–1529 MR Zbl
- [Arana-Herrera 2024] **F Arana-Herrera**, *Effective mapping class group dynamics, III: Counting filling closed curves on surfaces*, Camb. J. Math. 12 (2024) 563–622 MR Zbl
- [Arana-Herrera 2025] **F Arana-Herrera**, *Effective count of square-tiled surfaces with prescribed real and imaginary foliations in connected components of strata*, Ergodic Theory Dynam. Systems 45 (2025) 649–662 MR
- [Arana-Herrera and Calderon 2022] **F Arana-Herrera, A Calderon**, *The shapes of complementary subsurfaces to simple closed hyperbolic multi-geodesics*, preprint (2022) arXiv 2208.04339
- [Athreya et al. 2012] **J Athreya, A Bufetov, A Eskin, M Mirzakhani**, *Lattice point asymptotics and volume growth on Teichmüller space*, Duke Math. J. 161 (2012) 1055–1111 MR Zbl
- [Avila and Gouëzel 2013] **A Avila, S Gouëzel**, *Small eigenvalues of the Laplacian for algebraic measures in moduli space, and mixing properties of the Teichmüller flow*, Ann. of Math. 178 (2013) 385–442 MR Zbl
- [Avila et al. 2006] **A Avila, S Gouëzel, J-C Yoccoz**, *Exponential mixing for the Teichmüller flow*, Publ. Math. Inst. Hautes Études Sci. 104 (2006) 143–211 MR Zbl
- [Calderon and Farre 2024a] **A Calderon, J Farre**, *Continuity of the orthogeodesic foliation and ergodic theory of the earthquake flow* (2024) arXiv 2401.12299
- [Calderon and Farre 2024b] **A Calderon, J Farre**, *Shear-shape cocycles for measured laminations and ergodic theory of the earthquake flow*, Geom. Topol. 28 (2024) 1995–2124 MR Zbl
- [Do 2010] **N Do**, *The asymptotic Weil–Petersson form and intersection theory on $\mathcal{M}_{g,n}$* , preprint (2010) arXiv 1010.4126
- [Dozier and Sapir 2021] **B Dozier, J Sapir**, *Coarse density of subsets of moduli space*, Ann. Inst. Fourier (Grenoble) 71 (2021) 1121–1134 MR Zbl
- [Dumas 2015] **D Dumas**, *Skinning maps are finite-to-one*, Acta Math. 215 (2015) 55–126 MR Zbl

- [Einsiedler et al. 2019] **M Einsiedler, R Rühr, P Wirth**, *Distribution of shapes of orthogonal lattices*, Ergodic Theory Dynam. Systems 39 (2019) 1531–1607 MR Zbl
- [Erlandsson and Souto 2022] **V Erlandsson, J Souto**, *Mirzakhani’s curve counting and geodesic currents*, Progr. Math. 345, Birkhäuser, Cham (2022) MR Zbl
- [Eskin and Mirzakhani 2018] **A Eskin, M Mirzakhani**, *Invariant and stationary measures for the $SL(2, \mathbb{R})$ action on moduli space*, Publ. Math. Inst. Hautes Études Sci. 127 (2018) 95–324 MR Zbl
- [Eskin et al. 2015] **A Eskin, M Mirzakhani, A Mohammadi**, *Isolation, equidistribution, and orbit closures for the $SL(2, \mathbb{R})$ action on moduli space*, Ann. of Math. 182 (2015) 673–721 MR Zbl
- [Eskin et al. 2022] **A Eskin, M Mirzakhani, A Mohammadi**, *Effective counting of simple closed geodesics on hyperbolic surfaces*, J. Eur. Math. Soc. 24 (2022) 3059–3108 MR Zbl
- [Farb and Margalit 2012] **B Farb, D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl
- [Forni 2021] **G Forni**, *Limits of geodesic push-forwards of horocycle invariant measures*, Ergodic Theory Dynam. Systems 41 (2021) 2782–2804 MR Zbl
- [Gardiner and Masur 1991] **F P Gardiner, H Masur**, *Extremal length geometry of Teichmüller space*, Complex Variables Theory Appl. 16 (1991) 209–237 MR Zbl
- [Harer and Zagier 1986] **J Harer, D Zagier**, *The Euler characteristic of the moduli space of curves*, Invent. Math. 85 (1986) 457–485 MR Zbl
- [Hubbard and Masur 1979] **J Hubbard, H Masur**, *Quadratic differentials and foliations*, Acta Math. 142 (1979) 221–274 MR Zbl
- [Jenkins 1957] **J A Jenkins**, *On the existence of certain general extremal metrics*, Ann. of Math. 66 (1957) 440–453 MR Zbl
- [Kerckhoff 1980] **S P Kerckhoff**, *The asymptotic geometry of Teichmüller space*, Topology 19 (1980) 23–41 MR Zbl
- [Kontsevich 1992] **M Kontsevich**, *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. 147 (1992) 1–23 MR Zbl
- [Liu 2022] **M Liu**, *Length statistics of random multicurves on closed hyperbolic surfaces*, Groups Geom. Dyn. 16 (2022) 437–459 MR Zbl
- [Luo 2007] **F Luo**, *On Teichmüller spaces of surfaces with boundary*, Duke Math. J. 139 (2007) 463–482 MR Zbl
- [Margulis 1970] **G A Margulis**, *On some aspects of the theory of Anosov systems*, PhD thesis, Moscow State University (1970) In Russian; translation published by Springer (2003)
- [Martínez-Granado and Thurston 2021] **D Martínez-Granado, D P Thurston**, *From curves to currents*, Forum Math. Sigma 9 (2021) art. id. e77 MR Zbl
- [Masur and Smillie 1991] **H Masur, J Smillie**, *Hausdorff dimension of sets of nonergodic measured foliations*, Ann. of Math. 134 (1991) 455–543 MR
- [Mirzakhani 2007] **M Mirzakhani**, *Weil–Petersson volumes and intersection theory on the moduli space of curves*, J. Amer. Math. Soc. 20 (2007) 1–23 MR Zbl
- [Mirzakhani 2008a] **M Mirzakhani**, *Ergodic theory of the earthquake flow*, Int. Math. Res. Not. 2008 (2008) art. id. rnm116 MR Zbl

- [Mirzakhani 2008b] **M Mirzakhani**, *Growth of the number of simple closed geodesics on hyperbolic surfaces*, Ann. of Math. 168 (2008) 97–125 MR Zbl
- [Mirzakhani 2016] **M Mirzakhani**, *Counting mapping class group orbits on hyperbolic surfaces*, preprint (2016) arXiv 1601.03342
- [Mondello 2009a] **G Mondello**, *Riemann surfaces, ribbon graphs and combinatorial classes*, from “Handbook of Teichmüller theory, II”, IRMA Lect. Math. Theor. Phys. 13, Eur. Math. Soc., Zürich (2009) 151–215 MR Zbl
- [Mondello 2009b] **G Mondello**, *Triangulated Riemann surfaces with boundary and the Weil–Petersson Poisson structure*, J. Differential Geom. 81 (2009) 391–436 MR Zbl
- [Norbury 2010] **P Norbury**, *Counting lattice points in the moduli space of curves*, Math. Res. Lett. 17 (2010) 467–481 MR Zbl
- [Penner and Harer 1992] **R C Penner**, **J L Harer**, *Combinatorics of train tracks*, Ann. of Math. Stud. 125, Princeton Univ. Press (1992) MR Zbl
- [Royden 1971] **H L Royden**, *Automorphisms and isometries of Teichmüller space*, from “Advances in the theory of Riemann surfaces”, Ann. of Math. Stud. 66, Princeton Univ. Press (1971) 369–383 MR Zbl
- [Strebel 1966] **K Strebel**, *Über quadratische Differentiale mit geschlossenen Trajektorien und extremale quasikonforme Abbildungen*, from “Festband zum 70 Geburtstag von Rolf Nevanlinna”, Springer (1966) 105–127 MR Zbl
- [Strebel 1975] **K Strebel**, *On quadratic differentials and extremal quasi-conformal mappings*, from “Proceedings of the International Congress of Mathematicians, II” (R D James, editor), Canad. Math. Congr., Montreal (1975) 223–227 MR Zbl
- [Strebel 1976] **K Strebel**, *On the existence of extremal Teichmüller mappings*, J. Anal. Math. 30 (1976) 464–480 MR Zbl

*Department of Mathematics, University of Maryland
College Park, MD, United States*

*Department of Mathematics, University of Chicago
Chicago, IL, United States*

farana@umd.edu, aaroncalderon@uchicago.edu

Proposed: Benson Farb

Seconded: David Fisher, Anna Wienhard

Received: 21 November 2023

Revised: 3 September 2024

Tian’s stabilization problem for toric Fanos

CHENZI JIN

YANIR A RUBINSTEIN

In 1988, Tian posed the stabilization problem for equivariant global log canonical thresholds. We solve it in the case of toric Fano manifolds. This is the first general result on Tian’s problem. A key new estimate involves expressing complex singularity exponents associated to orbits of a group action in terms of support and gauge functions from convex geometry. These techniques also yield a resolution of another conjecture of Tian from 2012 on more general thresholds associated to Grassmannians of plurianticanonical series.

14C20, 14J45, 14J50, 14M25; 14L30, 32Q20, 32U05, 52A40

In memory of Eugenio Calabi

1. Introduction	2609
2. Toric and convex analysis setup	2616
3. A natural equivariant Hermitian metric and volume form	2621
4. An algebraic $\alpha_{k,G}$ -invariant and a Demailly type result	2623
5. The case of divisible k	2630
6. Estimating singularities associated to orbits	2632
7. Grassmannian Tian invariants	2638
8. Examples	2643
References	2650

1 Introduction

This article is the first in a series in which we study asymptotics of invariants related to existence of canonical metrics on Kähler manifolds.

Here we focus on Tian’s $\alpha_{k,G}$ - and $\alpha_{k,m,G}$ -invariants that were defined in 1988 and 1991, and to a large extent are still rather mysterious. The presence of the compact symmetry group G is a major source of difficulty and new ideas are needed here as these invariants have not been previously systematically studied or computed. In particular, we provide a formula for such invariants, valid for all toric Fano manifolds, leading to a resolution of Tian’s stabilization conjecture in this setting, which is also the first general result on Tian’s conjecture.

A sequel to this article deals with the δ_k -invariants of Fujita and Odaka (also called k^{th} stability thresholds) on toric Fano manifolds. It turns out that for these invariants there is a quantitative dichotomy regarding stabilization, and when stabilization fails we derive their complete asymptotic expansion.

1.1 Tian's stabilization problem

Let (X, L, ω) be a polarized Kähler manifold of dimension n with L a very ample line bundle over X , and ω a Kähler form representing $c_1(L)$. The space

$$(1) \quad \mathcal{H}_L := \{\varphi : \omega_\varphi := \omega + \sqrt{-1}\partial\bar{\partial}\varphi > 0\} \subset C^\infty(X)$$

of Kähler potentials of metrics cohomologous to ω was introduced by Calabi in a short visionary talk in the joint AMS–MAA annual meetings held at Johns Hopkins University in December 1953 [7]. In a groundbreaking article some 35 years later, Tian proved (motivated by a question of Yau [41, page 139]) that \mathcal{H}_L is approximated (or “quantized”) in the C^2 sense by the finite-dimensional spaces \mathcal{H}_k consisting of pullbacks of Fubini–Study metrics on $\mathbb{P}(H^0(X, L^k)^*)$ under all possible Kodaira embeddings induced by $H^0(X, L^k)$ [36]. A decade later this was improved to a complete asymptotic expansion (see Catlin [10] and Zelditch [42]) and so the \mathcal{H}_k can be considered as the Taylor (or Fourier, depending on the point of view) expansion of \mathcal{H}_L . The theme that holomorphic and other invariants associated to X and \mathcal{H} may be quantized using the spaces \mathcal{H}_k has dominated Kähler geometry for the last 35 years.

In the 1980s, Futaki's invariant was a new obstruction for the existence of Kähler–Einstein metrics, but there were no invariants that guaranteed existence. At best, there were constructions that utilized symmetry to reduce the Kähler–Einstein equation to a simpler equation that could be solved and lead to specific examples (another theme pioneered by Calabi, which makes a surprise appearance to close this paper; see Example 8.12). Given a maximal compact subgroup G of the automorphism group $\text{Aut } X$, Tian introduced the invariant

$$(2) \quad \alpha_G := \sup \left\{ c > 0 : \sup_{\varphi \in \mathcal{H}^G} \int_X e^{-c(\varphi - \sup \varphi)} \omega^n < \infty \right\}$$

(where \mathcal{H}^G denotes the G -invariant elements of \mathcal{H}_L), and its quantized version (Definition 4.1)

$$\alpha_{k,G},$$

computed over the G -invariant elements of \mathcal{H}_k , and obtained the sufficient condition

$$(3) \quad \alpha_G > \frac{n}{n+1}$$

for the existence of a Kähler–Einstein metric when $L = -K_X$ (which we will henceforth assume unless otherwise stated); see Tian [34, Theorem 4.1]. Initially, the main interest in the invariants α_G was as the first systematic tool for constructing Kähler–Einstein metrics on Fano manifolds, but later it was also conjectured by I Cheltsov and established by Demailly that α_G actually coincides with the G -equivariant global log canonical threshold from algebraic geometry; see Cheltsov and Shramov [13]. Since $\mathcal{H}_k \subset \mathcal{H}_L$, it follows that $\alpha_G \leq \inf_k \alpha_{k,G}$ [36, page 128], yet this does not help obtain (3). Instead, Tian posed

the following difficult question that would reduce the computation of the invariant α_G from the infinite-dimensional space \mathcal{H}_L to a finite-dimensional one \mathcal{H}_k , and establish a highly nontrivial relation between the different \mathcal{H}_k ; see Tian [35; 36, Question 1]:

Problem 1.1 *Let X be Fano, $L = -K_X$, and let G be a maximal compact subgroup $G \subset \text{Aut } X$. Is $\alpha_{k,G} = \alpha_G$ for all sufficiently large $k \in \mathbb{N}$?*

It is interesting to note that a slight variation of Tian's α - and α_k -invariants turned out to lead about three decades later [43; 30] to the very closely related δ - and δ_k -invariants of Fujita and Odaka [19] that are in turn a slight (and ingenious) variation on global log canonical thresholds, and turn out to essentially characterize the existence of Kähler–Einstein metrics. We return to these invariants in a sequel [25].

1.2 A Demailly type identity in the presence of symmetry

Another (easier) problem motivating this article concerns the by-now classical relation between Tian's (holomorphic) invariants and the (algebraic) global log canonical thresholds. The relationship was first conjectured by Cheltsov (see Cheltsov and Shramov [13]) and proved by Demailly and Shi; see Shi [31]. However, so far, this relationship has only been shown for the α - and α_k -invariants, or for the α_G -invariant (see [13, Theorem A.3], [31, Proposition 2.1], and [13, (A.1)], respectively), and not for the more subtle invariants $\alpha_{k,G}$. Inspired by Demailly, we introduce (Definition 4.4) the k^{th} G -equivariant global log canonical threshold

$$\text{glct}_{k,G}$$

as an algebraic counterpart of Tian's $\alpha_{k,G}$ (Definition 4.1). A natural question is:

Problem 1.2 *Let X be Fano and $L = -K_X$, and let G be a compact subgroup $G \subset \text{Aut } X$. Is $\text{glct}_{k,G} = \alpha_{k,G}$?*

Note that the proof of Demailly in the equivariant case [13, (A.2)] works for the limit as $k \rightarrow \infty$ but not on a fixed level k .

1.3 Results

We resolve both Problems 1.1 and 1.2 in the toric setting. It is perhaps not well known, but Calabi was interested in toric geometry and computed certain geodesics in \mathcal{H}_L in the toric setting, although he never published the result [9].

As standard, we allow the slightly more flexible situation of any (and not just a maximal) compact subgroup of the normalizer

$$N((\mathbb{C}^*)^n)$$

of the complex torus $(\mathbb{C}^*)^n$ in $\text{Aut } X$. We refer the reader to Section 2, where these and other toric notation is set up carefully. Denote by

$$(4) \quad \text{Aut } P \subseteq \text{GL}(M) \cong \text{GL}(n, \mathbb{Z})$$

the subgroup of the automorphism group of the lattice M that leaves the polytope P (27) invariant. It is necessarily a finite group (see Section 2). In fact, $\text{Aut } P$ is the quotient of the normalizer $N((\mathbb{C}^*)^n)$ of the complex torus $(\mathbb{C}^*)^n$ in $\text{Aut } X$ by $(\mathbb{C}^*)^n$, so that $N((\mathbb{C}^*)^n)$ consists of finitely many components, each isomorphic to a complex torus [3, Proposition 3.1]. For $H \subset \text{Aut } P$, let

$$(5) \quad G(H) := H \ltimes (S^1)^n \subset N((\mathbb{C}^*)^n) \subset \text{Aut } X$$

denote the compact group generated by H and $(S^1)^n$ (the latter is the maximal compact subgroup of the complex torus $(\mathbb{C}^*)^n$).

Our first result resolves Problem 1.2 in this generality:

Proposition 1.3 *Let X be toric Fano and $L = -K_X$. Let $P \subset M_{\mathbb{R}}$ (see (18) and (27)) be the polytope associated to $(X, -K_X)$, let $H \subset \text{Aut } P$, and let $G(H)$ be as in (5). Then $\text{glct}_{k,G(H)} = \alpha_{k,G(H)}$.*

The proof of Proposition 1.3 relies on the structure of the ring of holomorphic sections on a toric line bundle established in Section 4.3. For general nontoric cases, such a proposition will hold once a similar result on the space of sections can be established.

Using this result, and several new estimates, we can resolve Tian’s Problem 1.1 in the toric setting in a surprisingly strong sense, showing that equality holds for all $k \in \mathbb{N}$. We also allow for all groups $G(H)$ (and not just the maximal toric one $G(\text{Aut } P)$).

To state the precise result we introduce some more notation. For $H \subset \text{Aut } P$, denote by

$$(6) \quad P^H := \{y \in P : h \cdot y = y \text{ for all } h \in H\} \subset P \subset M_{\mathbb{R}}$$

the fixed-point set of H in P , and let

$$(7) \quad \pi_H := \frac{1}{|H|} \sum_{\eta \in H} \eta \in \text{End}(M_{\mathbb{Q}})$$

be the map that takes a point in $M_{\mathbb{R}}$ to the average of its H -orbit. Note that π_H is a projection map (see Section 6.3 for details).

Theorem 1.4 *Let X be toric Fano associated to a fan Δ whose rays are generated by primitive elements v_i in the lattice N dual to M . Let $P \subset M_{\mathbb{R}}$ (see (18) and (27)) be the polytope associated to $(X, -K_X)$, let $H \subset \text{Aut } P$, and let $G(H)$ be as in (5). Then for any $k \in \mathbb{N}$,*

$$(8) \quad \alpha_{k,G(H)} = \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} P^H \subset P \right\} = \min_{u \in \text{Ver } P^H} \frac{1}{\max_i \langle u, v_i \rangle + 1} \\ = \min_{u \in \pi_H(\text{Ver } P)} \frac{1}{\max_i \langle u, v_i \rangle + 1},$$

where P^H and π_H are defined in (6)–(7) and $\text{Ver}(\cdot)$ denotes the vertex set of a polytope. In particular, $\alpha_{k,G(H)}$ is independent of $k \in \mathbb{N}$ and is equal to $\alpha_{G(H)}$.

There are a few new ingredients in the proof of Theorem 1.4. The first is a useful formula for the spaces $\mathcal{H}_k^{G(H)}$ in terms of the H -orbits of the finite group action (Lemma 4.5). This together with a trick that amounts to estimating the singularities associated to a basis of sections in terms of the finite group action orbit of a section yields a useful formula for $\alpha_{k,G(H)}$ (Proposition 4.6), as well as the equality $\alpha_{k,G(H)} = \text{glct}_{k,G(H)}$, ie a solution to Problem 1.2 (Proposition 1.3). These then yield a corresponding useful formula for $\alpha_{G(H)}$ (Corollary 4.7). One may prove, using the results of Sections 3–4, that $\alpha_{k,G(H)} \geq \alpha_{k\ell,G(H)}$ for any fixed k and all $\ell \in \mathbb{N}$ (Proposition 5.1). In Proposition 5.3 it is shown that there is a special k_0 for which $\alpha_{k_0\ell,G(H)} = \alpha_{G(H)}$ for all $\ell \in \mathbb{N}$, as well as observed that this does not seem to imply Tian's conjecture (Remark 5.6). Finally, key new estimates occur in Section 6. First, we show that rather general complex singularity exponents associated to collections of toric monomials are independent of k (Proposition 6.2). Our proof uses a new observation about the relation between the support functions of collections of lattice points associated to the toric monomials and complex singularity exponents. We then apply this to our $G(H)$ -invariant setting, using the aforementioned expression of $\mathcal{H}_k^{G(H)}$ and a reduction lemma to the H -invariant subspace (Lemma 6.10), to conclude the proof of Theorem 1.4.

It is perhaps of some interest to include here a rather immediate application of this circle of ideas to a slightly more technical set of invariants, also introduced by Tian, which we call Tian's Grassmannian α -invariants. These invariants are defined a little differently, algebraically, and are denoted by $\alpha_{k,m}$ or $\alpha_{k,m,G}$ (Definition 7.1). At least in the nonequivariant setting (as well as in the torus-equivariant setting, see Remark 7.2) these can be considered as generalizations of the α_k as $\alpha_k = \alpha_{k,1}$; see Cheltsov and Shramov [13] and Shi [31]. The invariants $\alpha_{k,2}$ were used by Tian implicitly in his proof of Calabi's conjecture for del Pezzo surfaces [36, Appendix A] (cf [37, Theorem 6.1]) to overcome the most difficult case (of a cubic surface with an Eckardt point) where equality holds in (3), and this was improved by Shi to $\alpha_{k,2} > 2/3 = \alpha_{k,1}$ in that case; see Cheltsov [11] and Shi [31, Theorem 1.3].

Tian also posed a stabilization conjecture for these invariants in 2012 [38, Conjecture 5.3]:

Conjecture 1.5 *Let X be Fano. Fix $m \in \mathbb{N}$. For sufficiently large k , $\alpha_{k,m,G}$ is constant.*

We completely resolve Conjecture 1.5 in the toric setting. Theorem 1.4 resolved Problem 1.1 in the affirmative (corresponding to the case $m = 1$ of Conjecture 1.5). For $m \geq 2$, Conjecture 1.5 turn out to be only partially true as determined by a novel convex geometric obstruction we introduce:

$$(*_P) \quad \|\cdot\|_{-P}|_{P \setminus \text{Ver } P} < \max_P \|\cdot\|_{-P}.$$

Note that this condition depends only on P (and not on m or k). The condition $(*_P)$ means that the function $\|\cdot\|_{-P}$ on P achieves its maximum only at the vertices of P , ie

$$(*_P) \quad \text{argmax}_P \|\cdot\|_{-P} \subset \text{Ver } P.$$

When $(*_P)$ fails, the maximum is achieved also at some point that is not a vertex of P .

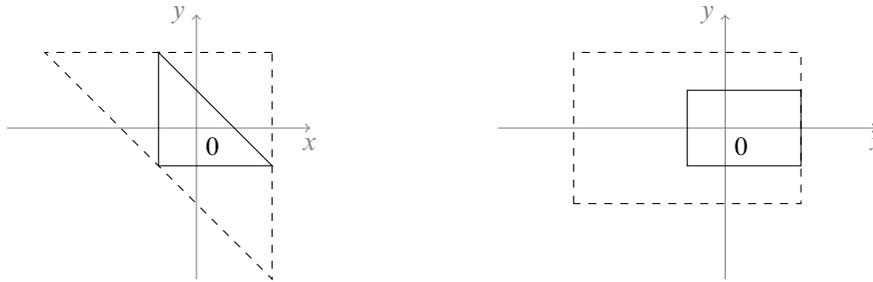


Figure 1: The polytope P (solid line) and the level set $\{\|\cdot\|_{-P} = \max_P \|\cdot\|_{-P}\}$ (dashed line). For $P = \text{co}\{(-1, -1), (2, -1), (-1, 2)\}$, the maximum is only attained at the vertices of P . In particular, $(*_P)$ holds. For $P = [-1, 2] \times [-1, 1]$, the maximum is attained on the line segment $\{2\} \times [-1, 1]$. In particular, $(*_P)$ does not hold.

Theorem 1.6 *Let X be toric Fano with associated polytope P (28) and let $\alpha := \alpha_{(S^1)^n}$ be as in (2) with $L = -K_X$. Conjecture 1.5 holds if and only if $(*_P)$ fails. More precisely, if $(*_P)$ holds,*

$$(9) \quad \alpha_{k,m,(S^1)^n} > \alpha \quad \text{for } k \in \mathbb{N} \text{ and } m \in \mathbb{N} \setminus \{1\},$$

otherwise

$$(10) \quad \alpha_{k,m,(S^1)^n} = \alpha \quad \text{for } m \in \mathbb{N} \text{ and for sufficiently large } k \in \mathbb{N}.$$

Theorem 1.6 is proven in Section 7.1, where we also explain the intuition behind it (see also Examples 8.6 and 8.9). For now, let us elucidate the condition $(*_P)$ a bit. The level set $\{\|\cdot\|_{-P} = \lambda\}$ is the dilation $\lambda\partial(-P)$, and $\{\|\cdot\|_{-P} = \max_P \|\cdot\|_{-P}\}$ is the largest dilation that intersects P (by Lemma 7.4). Thus condition $(*_P)$ states that P intersects $\max_P \|\cdot\|_{-P}\partial(-P)$ only at vertices. When $(*_P)$ fails, convexity arguments show the intersection will contain a positive-dimensional face of P . See Figure 1 for two examples.

Relation to earlier works Theorem 1.4 strengthens and clarifies work of Song [33] and Li and Zhu [28]. Song obtained a formula for α_G but not for $\alpha_{k,G}$. In particular, there seems to be a gap in the proof of [33, Theorem 1.2] that claims that $\alpha_G = \alpha_{k,G}$ for all sufficiently large k . This claim relies on proving that for some $k_0 \in \mathbb{N}$ and all $\ell \in \mathbb{N}$, $\alpha_G = \alpha_{k_0\ell,G}$, and then invoking that $\alpha_{k,G}$ is eventually monotone in k and hence must be independent of k for sufficiently large k . Unfortunately, the proof of monotonicity is omitted from [33, page 1257, line 7], and it appears to be difficult to reproduce. It seems that such monotonicity is not currently known (cf Remark 5.6). Indeed, there is no obvious relationship between the various \mathcal{H}_k^G coming from different Kodaira embeddings. As noted above, Song showed (for $H = \text{Aut } P$) that $\alpha_{G(H)} = \alpha_{k_0\ell,G(H)}$ for some $k_0 \in \mathbb{N}$ and all $\ell \in \mathbb{N}$. Li and Zhu showed the same identity (essentially for $H = \{\text{id}\}$) for a group compactification of a reductive complex Lie group and also claimed, similarly to Song, that this implies eventual constancy in k in that setting [28, Theorem 1.3, page 233]. Unfortunately, they also do not provide a proof of the needed eventual monotonicity or constancy. In the nonequivariant

setting, and for rather general Fano varieties for which $\alpha \leq 1$, Birkar showed the deep result that $\alpha = \alpha_{k_0 \ell}$ for some $k_0 \in \mathbb{N}$ and all $\ell \in \mathbb{N}$ [5, Theorem 1.7]. However, this result also does not imply Tian stabilization due to the aforementioned unknown monotonicity. Thus Theorem 1.4 seems to be the first general result on Tian's stabilization Problem 1.1.

Similarly, Theorem 1.6 seems to be the first general result on Conjecture 1.5. Indeed, Li and Zhu showed the same type of result Song obtained in the $m = 1$ setting, ie that $\alpha_{k_0 \ell, m, (S^1)^n} = \alpha_{(S^1)^n}$ under a condition depending on k_0 and m , and hence different from our $(*P)$ (with the minor caveat that their statement as written [28, Theorem 1.4] is incorrect, though can be easily fixed by replacing “facet” by “face”, see Section 7.1). However, again, due to the lack of monotonicity, they do not obtain a resolution of Conjecture 1.5 though they do obtain the first counterexamples to it when $m \geq 2$.

Combining Theorem 1.4 and Demailly's theorem [13, (A.1)] also recovers Song's formula for the $\alpha_{G(\text{Aut } P)}$ (that itself generalized Batyrev and Selivanova's formula that $\alpha_{G(\text{Aut } P)} = 1$ whenever $P^{\text{Aut } P} = \{0\}$ (recall (6)) [3, Theorem 1.1, page 233]). Cheltsov and Shramov claimed a more general formula for α_H (ie without the real torus symmetry included in $G(H)$, recall (5)) however (as kindly pointed out to us by Cheltsov) there is an error in the proof of [13, Lemma 5.1] as the toric degeneration used there need not respect the H -invariance. Further generalizations of Song's formula for α_G to general polarizations and group compactifications are due to Delcroix [14; 15] and Li, Shi, and Yao [27], and our methods should generalize to those settings as well as to the setting of log toric Fano pairs and edge singularities [12, Sections 6–7]. Although this is not the topic of our article, for completeness we mention in passing that the stabilization of the $\alpha_k(L)$ -invariants for arbitrary toric varieties (ie not necessarily Fano) but for the *nonequivariant* setting (ie $H = \{\text{id}\}$) is true, ie $\alpha_{k, (S^1)^n}(L) = \alpha(L)$ for all $k \in \mathbb{N}$, yet is much-simpler; the crux of our work is to address the equivariant setting that presents new challenges.

Theorem 1.7 *Let X be a toric variety and L ample, with associated lattice polytope $P = \bigcap_{i=1}^d \{y \in M_{\mathbb{R}} : \langle y, v_i \rangle \geq -b_i\}$ with $v_i \in N$ and $b_i \in \mathbb{Z}$. For $k \in \mathbb{N}$,*

$$\alpha_{k, (S^1)^n}(L) = \min_{u \in P \cap M/k} \frac{1}{\max_i \langle u, v_i \rangle + b_i} = \min_{u \in \text{Ver } P} \frac{1}{\max_i \langle u, v_i \rangle + b_i} = \alpha(L).$$

The same proof for Theorem 1.4 works in this case, with the additional assumption that $H = \{\text{id}\}$, except that the coefficient 1 is replaced by b_i . Proposition 4.6, for instance, gives the left-hand side. Blum and Jonsson's formula [6, (7.2)] gives the right-hand side. And the equality again relies on $\text{Ver } P \subset M$ and the fact that a convex function on a convex polytope attains its maximum on a vertex.

Finally, it is also worth mentioning that Tian also posed more general conjectures [38, Conjecture 5.4] for general polarizations (ie L not being $-K_X$) for which there are already some counterexamples [1].

Organization Section 2 sets up the necessary notation concerning toric varieties and convex analysis. Section 3 constructs natural equivariant reference Hermitian metrics and volume forms. Proposition 1.3

is proved in Section 4.4. Section 5 explains a trick that allows us to deal with divisible $k \in \mathbb{N}$, but also highlights the difficulties in dealing with general k . Theorem 1.4 is proved in Section 6. Theorem 1.6 is proved in Section 7. We conclude with examples in Section 8.

Acknowledgments Eugenio Calabi has had a profound influence on Rubinstein, who first met him as a graduate student in 2007. This article is dedicated to his memory and to the celebration of his mathematical culture, curiosity, and generosity. Calabi’s contributions pop up in several places in this article, starting from the space of Kähler metrics, the space of toric metrics, and the Calabi–Hirzebruch manifold at the very end.

The authors are grateful to the referees for a careful reading. Thanks to C Birkar, H Blum, I Cheltsov, A Litvak, L Rotem, and K Zhang for helpful references. Research supported in part by NSF grants DMS-1906370 and 2204347, BSF grant 2020329, and an Ann G Wylie dissertation fellowship.

2 Toric and convex analysis setup

2.1 Notions from convexity

Consider an n -dimensional real vector space $V \cong \mathbb{R}^n$ and let $V^* \cong \mathbb{R}^n$ denote its dual with the pairing denoted by $\langle \cdot, \cdot \rangle$. Given a set $A \subset V$, denote by

$$A^\circ = \{y \in V^* : \langle x, y \rangle \leq 1 \text{ for all } x \in A\}$$

the polar of A [29, page 125], and by

$$(11) \quad \text{co } A$$

the convex hull of A [29, page 12]. For a finite set [29, Theorem 2.3],

$$(12) \quad \text{co}\{p_1, \dots, p_\ell\} = \left\{ \sum_{i=1}^{\ell} \lambda_i p_i : \sum_{i=1}^{\ell} \lambda_i = 1, \lambda_i \in [0, 1] \right\}.$$

Also, set

$$-K := \{-x : x \in K\}.$$

Note $(-K)^\circ = -K^\circ$. Also, $K^\circ = (\text{co } K)^\circ \subset V^*$ whenever $K \subset V$. The polar can also be described via the support function $h_K : V^* \rightarrow \mathbb{R}$,

$$(13) \quad h_K(y) := \sup_{x \in K} \langle x, y \rangle \quad \text{for } y \in V^* \cong \mathbb{R}^n,$$

by $K^\circ = \{h_K \leq 1\}$.

A dual notion to the support function is the near-norm function

$$(14) \quad \|x\|_K := \inf\{t \geq 0 : x \in tK\}$$

associated to any compact convex set K with $0 \in \text{int } K$. Note that [29, Corollary 14.5]

$$(15) \quad \|\cdot\|_K = h_{K^\circ}.$$

Also note that $\|\cdot\|_K$ is a norm when K is centrally symmetric (ie $K = -K$), otherwise it is only a near-norm in the sense that it satisfies all the properties of a norm but is only \mathbb{R}_+ -homogeneous: $\|\lambda x\|_K = \|\lambda\|_K \|x\|_K$ for $\lambda \in \mathbb{R}_+$ (and not fully \mathbb{R} -homogeneous). The infimum in (14) is achieved: for a minimizing sequence $\{t_i\}$, $x/t_i \in K$ for any i , so

$$(16) \quad x \in \|x\|_K K$$

since K is closed.

Lemma 2.1 *Let V be an \mathbb{R} -vector space. Consider the polytope*

$$A = \bigcap_{j=1}^d \{x \in V : \langle x, v_j \rangle \leq 1\},$$

where $v_j \in V^*$. Then

$$\|x\|_A = \max_{1 \leq j \leq d} \langle x, v_j \rangle.$$

Proof Notice that $x \in tA$ if and only if for any $1 \leq j \leq d$, $\langle x, v_j \rangle \leq t$. Thus

$$\|x\|_A = \inf\{t \geq 0 : \max_{1 \leq j \leq d} \langle x, v_j \rangle \leq t\} = \max_{1 \leq j \leq d} \langle x, v_j \rangle.$$

Alternatively, observe $A = \{v_1, \dots, v_d\}^\circ$ and $A^\circ = (\{v_1, \dots, v_d\}^\circ)^\circ = \text{co}\{v_1, \dots, v_d\}$ [29, Theorem 14.5] and then use (57) and (15). □

2.2 Toric algebra

Consider a lattice of rank n and its dual lattice,

$$(17) \quad N \quad \text{and} \quad M := N^* := \text{Hom}(N, \mathbb{Z}).$$

Both N and M are isomorphic to \mathbb{Z}^n , but we do not specify the isomorphism (see eg the proof of Lemma 2.3 for this point). The notation is useful as it serves to distinguish between objects living in one lattice and its dual (although of course in computations we simply work on \mathbb{Z}^n ; see Section 8). Denote the corresponding \mathbb{R} -vector space and its dual (both isomorphic to \mathbb{R}^n) by

$$(18) \quad N_{\mathbb{R}} := N \otimes_{\mathbb{Z}} \mathbb{R} \quad \text{and} \quad M_{\mathbb{R}} := M \otimes_{\mathbb{Z}} \mathbb{R} = N_{\mathbb{R}}^*.$$

A rational convex polyhedral cone in $N_{\mathbb{R}}$ takes the form

$$(19) \quad \sigma = \sigma(v_1, \dots, v_d) := \left\{ \sum_{i=1}^d a_i v_i : a_i \geq 0, v_i \in N \right\}.$$

The rays $\mathbb{R}_+ v_i$ for $i \in \{1, \dots, d\}$ are called the generators of the cone [20, page 9]. They are (1-dimensional) cones themselves, of course. Our convention will be that the

$$(20) \quad v_i \text{ for } i \in \{1, \dots, d\} \text{ are primitive elements of the lattice } N,$$

which means there is no $m \in \mathbb{N} \setminus \{1\}$ such that $v_i/m \in N$. A cone is called strongly convex if $\sigma \cap -\sigma = \{0\}$ [20, page 14]. A face of σ is any intersection of σ with a supporting hyperplane.

Definition 2.2 A fan $\Delta = \{\sigma_i\}_{i=1}^\delta$ in N is a finite set of rational strongly convex polyhedral cones σ_i in $N_{\mathbb{R}}$ such that

- (i) each face of a cone in Δ is also (a cone) in Δ ,
- (ii) the intersection of two cones in Δ is a face of each.

Such a fan gives rise to a toric variety $X(\Delta)$: each cone σ_i in Δ gives rise to an affine toric variety [20, Section 1.3] that serves as (a Zariski open) chart in $X(\Delta)$ with the transition between the charts constructed by (i) and (ii) above [20, page 21]. For instance, the zero cone corresponds to the open dense orbit $(\mathbb{C}^*)^n$ [32, page 64], and more generally there is a bijection between the cones $\{\sigma_i\}_{i=1}^\delta$ and the orbits of the complex torus $(\mathbb{C}^*)^n$ in $X(\Delta)$ [32, Proposition 5.6.2], with the nonzero cones corresponding precisely to all the toric subvarieties of $X(\Delta)$ of positive codimension.

When $X(\Delta)$ is a smooth toric Fano variety (as we always assume), the fan Δ must arise from an integral polytope as follows (but in general, ie for singular toric varieties, this need not be the case [20, page 25]). Let Δ be a fan such that $X(\Delta)$ is smooth Fano and let $\sigma_1, \dots, \sigma_d$ be its 1-dimensional cones (ie rays) generated by primitive generators

$$(21) \quad \Delta_1 := \{v_1, \dots, v_d\} \subset N,$$

so $\sigma_i = \mathbb{R}_+ v_i$, and set (recall (11))

$$(22) \quad Q := \text{co } \Delta_1 = \text{co}\{v_1, \dots, v_d\} = \text{co } \Delta_1 \subset N_{\mathbb{R}}.$$

Then Δ is equal to the collection of cones over each face of Q plus the zero cone [20, page 25]. In other words if $F \subset Q$ is a face, then

$$(23) \quad \sigma_F := \{rx \in N_{\mathbb{R}} : r \geq 0, x \in F\}$$

is the union of all rays through F and the origin, and

$$\Delta = \{\sigma_F\}_{F \subset Q} \cup \{0\},$$

where 0 denotes the zero cone. Denote by

$$(24) \quad \text{Ver } A$$

the vertices of a polytope A . Note that $\text{Ver } F \subset \text{Ver } Q = \Delta_1$, and by (19)–(20),

$$(25) \quad \sigma_F = \sigma(\text{Ver } F).$$

Smoothness of X means that the generators of σ_F form a \mathbb{Z} -basis for N [20, page 29]. By (25) this means

$$(26) \quad \text{the vertices of } F \text{ form a } \mathbb{Z}\text{-basis for } N \text{ (for any facet } F \subset Q).$$

Thus each facet F of Q is an $(n-1)$ -simplex whose vertices form a \mathbb{Z} -basis of N . In Lemma 2.3 we show this means the vertices of the polar polytope belong to the dual lattice M .

When $L = -K_X$, there is an $\text{Aut } X$ action on $H^0(X, -kK_X)$ for every $k \in \mathbb{N}$. To get an induced linear action on $M_{\mathbb{Q}}$ we must restrict to the normalizer $N((\mathbb{C}^*)^n)$ of the complex torus $(\mathbb{C}^*)^n$ in $\text{Aut } X$. The representation of $(\mathbb{C}^*)^n$ on $H^0(X, -kK_X)$ splits into 1-dimensional spaces, whose generators are called the monomial basis. There is a one-to-one correspondence between the monomial basis of $H^0(X, -kK_X)$ and points in $kP \cap M$, and the quotient $N((\mathbb{C}^*)^n)/(\mathbb{C}^*)^n$ is a linear group that can be identified with $\text{Aut } P \subset \text{GL}(M) \cong \text{GL}(n, \mathbb{Z})$; see (4). Since P is defined as the convex hull of vertices in M , it follows that $\text{Aut } P$ is finite. Alternatively, this can be seen by observing that $N_{\mathbb{R}}$ is canonically isomorphic to the quotient of $(\mathbb{C}^*)^n$ by its maximal compact subgroup $(S^1)^n$ [3, page 229] and the induced action on $M_{\mathbb{R}}$ is then defined by transposing via the pairing. Conversely, all compact subgroups of $N((\mathbb{C}^*)^n)$ that contain $(S^1)^n$ are generated by $(S^1)^n$ and a finite subgroup H of $\text{Aut } P$ [3, Proposition 3.1], and we denote such a group by $G(H) \subset \text{Aut } X$ as in (5). We describe in the proof of Lemma 3.1 concretely how the action of $\text{Aut } P$ is expressed in coordinates. For a finite group H or finite set \mathcal{A} we denote by

$$|H|, \text{ respectively, } |\mathcal{A}|,$$

its order or cardinality.

Oftentimes we will work with

$$(27) \quad P := -Q^\circ = \{-v_1, \dots, -v_d\}^\circ = -\{v_1, \dots, v_d\}^\circ \subset M_{\mathbb{R}},$$

as it has the nice geometric property of faces of (real) dimension k corresponding to toric subvarieties of dimension k , and since the metric properties (eg volume) of P correspond to those of X . (Moreover, P can also be realized as the Delzant (moment) polytope associated to any $(S^1)^n$ -invariant Kähler metric representing the anticanonical class.) In particular,

$$(28) \quad P = \bigcap_{i=1}^d \{y \in M_{\mathbb{R}} : \langle y, -v_i \rangle \leq 1\} = \{h_{-\Delta_1} \leq 1\} = \{y \in M_{\mathbb{R}} : \max_j \langle -v_j, y \rangle \leq 1\},$$

and irreducible toric divisors correspond to facets of P

$$(29) \quad \{y \in P : \langle y, -v_i \rangle = 1\}.$$

Note that P contains the origin in its interior. Also note that (27) is the standard convention since then lattice points of P correspond to monomials via (34). Batyrev and Selivanova use $-P$ instead.

Lemma 2.3 *Let $P \subset M_{\mathbb{R}}$ be the polytope associated to a smooth toric variety X . Then P is an integral lattice polytope, ie $\text{Ver } P \subset M$.*

Proof By duality, if $u \in M_{\mathbb{R}}$ is a vertex of P then

$$(30) \quad F = \{v \in Q : \langle u, -v \rangle = 1\} \subset N_{\mathbb{R}}$$

is a facet of $Q = -P^\circ \subset N_{\mathbb{R}}$. Since X is smooth, the vertices of F form a \mathbb{Z} -basis for N by (26). Choose coordinates on N associated to this \mathbb{Z} -basis, ie the vertices of F are the standard basis vectors

e_1, \dots, e_n . Thus the facet $F = \text{co}\{e_1, \dots, e_n\}$ is the standard $(n-1)$ -simplex in \mathbb{R}^n cut out by the equation $\langle u, -v \rangle = 1$ where $u = (-1, \dots, -1) \in M$, as desired.

Alternatively, if one does not wish to choose coordinates but rather work invariantly, let

$$\text{Ver } F := \{f_1, \dots, f_n\} \subset N,$$

and note $\text{span}_{\mathbb{Z}} \text{Ver } F = N$. Thus any $v \in N$ can be written uniquely as $v = \sum_{i=1}^n a_i f_i$ (with $a_i \in \mathbb{Z}$), and so $u \in M_{\mathbb{R}} = \text{Hom}(N, \mathbb{R})$ can be identified (recall (17)) with the map

$$N \ni v \mapsto - \sum_{i=1}^n a_i.$$

As $\sum_{i=1}^n a_i \in \mathbb{Z}$, this map actually belongs to $\text{Hom}(N, \mathbb{Z}) = M$. □

Another useful fact is a sort of maximum principle for convex polytopes, saying essentially that a convex function on a polytope achieves its maximum at some vertex (regardless of continuity).

Lemma 2.4 *Let A be a convex polytope and $f : A \rightarrow \mathbb{R} \cup \{\infty\}$ a convex function. Then:*

- (i) $\sup_A f = \sup_{\text{Ver } A} f$.
- (ii) *If f is bounded on $\text{Ver } A$ it is bounded on A and its maximum is achieved in a vertex.*
- (iii) *If f attains its finite maximum on $\text{int } A$ it is constant.*
- (iv) *If f attains its finite maximum on the relative interior of a face $F \subset A$ it is constant on F .*

Proof Write $\text{Ver } A = \{p_1, \dots, p_\ell\}$ and assume $f(p_1) \leq \dots \leq f(p_\ell)$. By convexity, $A = \text{co } \text{Ver } A$, and (12) implies that any $x \in A$ can be expressed as

$$(31) \quad x = \sum_{i=1}^{\ell} \lambda_i p_i \quad \text{where} \quad \sum_{i=1}^{\ell} \lambda_i = 1 \quad \text{and} \quad \lambda \in [0, 1]^\ell.$$

Then $f(x) \leq \sum_{i=1}^{\ell} \lambda_i f(p_i) \leq \sum_{i=1}^{\ell} \lambda_i f(p_\ell) = f(p_\ell)$, proving (i) and (ii).

To see (iii), again express any $x \in A$ using (31). Suppose that $x_{\text{int}} \in \text{int } A$ achieves the (finite) maximum of f . Note that $x_{\text{int}} \in \text{int } A$ means one has the representation (31) for x_{int} for some $\lambda_{\text{int}} \in (0, 1)^\ell$. Choose $\delta = \delta(x) \in (0, 1)$ so that $\lambda_{\text{int}} - \delta\lambda \in \mathbb{R}_+^\ell$. Define

$$\lambda' := \frac{1}{1-\delta}(\lambda_{\text{int}} - \delta\lambda) \in \mathbb{R}_+^\ell.$$

Note that λ' satisfies $\sum_{i=1}^{\ell} \lambda'_i = 1$, so letting $x' = \sum_{i=1}^{\ell} \lambda'_i p_i$ we have $x' \in A$ by (31). Also, $\delta x + (1-\delta)x' = x_{\text{int}}$. Thus $\max_A f = f(x_{\text{int}}) \leq \delta f(x) + (1-\delta)f(x') \leq \max_A f$, forcing equality, ie $f(x) = \max_A f$ (since $\delta > 0$ and $\max_A f < \infty$), proving (iii). The proof of (iv) is identical by working on the polytope F . □

Finally, we recall Ehrhart’s theorem on the polynomiality of the number of lattice points in dilations of lattice polytopes [16; 17; 23, Theorem 19.1]. Set

$$(32) \quad E_P(k) := |kP \cap M| \quad \text{for } k \in \mathbb{N}.$$

Proposition 2.5 *Let M be a lattice and $P \subset M_{\mathbb{R}}$ be a lattice polytope of dimension n . Then*

$$E_P(k) = \sum_{i=0}^n a_i k^i \quad \text{for any } k \in \mathbb{N},$$

with $a_n = \text{Vol}(P)$, and

$$(33) \quad k \leq k' \implies E_P(k) \leq E_P(k').$$

3 A natural equivariant Hermitian metric and volume form

In light of Lemma 4.2 below it makes sense to choose a convenient pair (μ, h) of a volume form and a Hermitian metric. In fact, we are free to choose such a pair for each k . The special feature of working with $L = -K_X$ is that in fact a volume form essentially doubles as a Hermitian metric, which is sometimes a bit confusing to keep track of in terms of notation, but is quite convenient for computations. This section serves to explain this choice $(\mu_k = h_k^{1/k}, h_k)$; see (40) and (42), originally due to Song [33, Lemma 4.3]. We emphasize that the Hermitian metric must additionally be chosen G -invariant in Definition 4.1, and this is confirmed for h_k in Lemma 3.1.

Let X be a toric Fano manifold and P its associated polytope. There is a natural basis of the space of holomorphic sections $H^0(X, -kK_X)$ defined by the monomials z^{ku} where $u \in P \cap (1/k)M$. That is, there exists an invariant frame e over the open orbit such that

$$(34) \quad s_{k,u}(z) = z^{ku} e.$$

What does e actually look like? This is most naturally expressed in terms of the *monomial basis*. When $k = 1$ and $u \in P \cap M$ [20, Section 4.3],

$$s_{1,u} = z^u \prod_{i=1}^n z_i \partial_{z_1} \wedge \cdots \wedge \partial_{z_n}.$$

In general, for any $k \in \mathbb{N}$ and $u \in P \cap (1/k)M$,

$$(35) \quad s_{k,u} = z^{ku} \left(\prod_{i=1}^n z_i \right)^k (\partial_{z_1} \wedge \cdots \wedge \partial_{z_n})^{\otimes k}.$$

In other words,

$$(36) \quad e = \left(\prod_{i=1}^n z_i \right)^k (\partial_{z_1} \wedge \cdots \wedge \partial_{z_n})^{\otimes k}.$$

Next, let us construct a canonical Hermitian metric h_k on $-kK_X$. Since $-kK_X$ is very ample [20, page 70], it is natural to pull back the Fubini–Study Hermitian metric via the Kodaira embedding. It turns out that choosing the Kodaira embedding given by the monomial basis

$$(37) \quad \iota_k : X \ni z \mapsto [s_{k,u}(z)/e(z)]_{u \in P \cap (1/k)M} = [z^{ku}]_{u \in P \cap (1/k)M} \in \mathbb{P}^{E_P(k)-1}$$

will yield the desired h_k ; importantly, the resulting h_k will be *torus-invariant*, smooth, and essentially transform computations on X to P . To wit, the Fubini–Study metric on $\mathcal{O}(1) \rightarrow \mathbb{P}^{E_P(k)-1}$ is (where $E_P(k)$ is the number of lattice points in kP)

$$h_{\text{FS}}(Z_i, Z_j) := \frac{Z_i \bar{Z}_j}{\sum_{\ell} |Z_{\ell}|^2},$$

and we define

$$(38) \quad h_k := \iota_k^* h_{\text{FS}}.$$

Note that each homogeneous coordinate $Z_i \in H^0(\mathbb{P}^{E_P(k)-1}, \mathcal{O}(1))$ pulls back via ι_k to one of the monomial sections $s_{k,u}$ (which one depends on the ordering for the elements of $P \cap (1/k)M$ chosen in (37)). Thus to express h_k it suffices to compute it on the monomial basis of $H^0(X, -kK_X)$:

$$(39) \quad h_k(s_{k,u_1}, s_{k,u_2})(z) = h_{\text{FS}}(s_{k,u_1}, s_{k,u_2})(\iota_k(z)) = \frac{z^{ku_1} \bar{z}^{ku_2}}{\sum_{u \in P \cap (1/k)M} |z^{ku}|^2}.$$

Comparing (35) and (39) means that h_k can be written as

$$(40) \quad h_k = \frac{(dz^1 \wedge \bar{d}z^1 \wedge \dots \wedge dz^n \wedge \bar{d}z^n)^{\otimes k}}{(\prod_{i=1}^n |z_i|^2)^k \sum_{u \in P \cap (1/k)M} |z^{ku}|^2}.$$

In conclusion, $h_k^{1/k}$ is a smooth metric on $-K_X$ (h_k being obtained as a pullback of a smooth metric under the Kodaira *embedding*), and hence it is a smooth volume form on X . To express this volume form, on the open orbit $(\mathbb{C}^*)^n = \mathbb{R}^n \times (S^1)^n$ consider the holomorphic coordinates

$$(41) \quad w_i := \frac{1}{2}x_i + \sqrt{-1}\theta_i = \log z_i \in \mathbb{C}^n.$$

In these coordinates this volume form, on the open orbit, is

$$(42) \quad \mu_k := h_k^{1/k} = \frac{dx_1 \wedge \dots \wedge dx_n \wedge d\theta_1 \wedge \dots \wedge d\theta_n}{(\sum_{u \in P \cap M/k} e^{(ku,x)})^{1/k}}.$$

Lemma 3.1 *Let $G(H) \subseteq \text{Aut } X$ (5) be a subgroup generated by $(S^1)^n$ and a subgroup H of $\text{Aut } P$. Then h_k (40) is $G(H)$ -invariant.*

Proof From (42) it is evident that h_k is independent of $(\theta_1, \dots, \theta_n)$, ie it is $(S^1)^n$ -invariant. An automorphism $\sigma \in \text{Aut } P \subseteq \text{GL}(M)$ can be represented (via choosing a basis for the lattice M) by a matrix in $\text{GL}(n, \mathbb{Z}) \cong \text{GL}(M)$. Since σ preserves the polytope P , $\det \sigma \in \{\pm 1\}$ (σ could be orientation reversing, eg in the case of a reflection). The induced action of σ on the dual space $N_{\mathbb{R}}$ is naturally represented (via the pairing between M and N) by the transpose matrix, which we denote by σ^T , and this action actually comes from the \mathbb{C} -linear action of σ^T on $N_{\mathbb{C}} \cong \mathbb{C}^n$ (41). Thus

$$\sigma.(dx_1 \wedge \dots \wedge dx_n) = d(x_1 \circ \sigma) \wedge \dots \wedge d(x_n \circ \sigma) = \det(\sigma^T) dx_1 \wedge \dots \wedge dx_n$$

(here $\sigma \cdot$ denotes the action of σ on forms, ie by pullback), and $\sigma.(d\theta_1 \wedge \dots \wedge d\theta_n) = \det(\sigma^T) d\theta_1 \wedge \dots \wedge d\theta_n$. Since $(\det \sigma^T)^2 = 1$ it remains to consider the denominator of (42),

$$\begin{aligned} \sigma \cdot \sum_{u \in P \cap M/k} e^{k\langle u, x \rangle} &= \sum_{u \in P \cap M/k} e^{k\langle u, \sigma^T \cdot x \rangle} = \sum_{u \in P \cap M/k} e^{k\langle \sigma \cdot u, x \rangle} = \sum_{u \in \sigma(P \cap M/k)} e^{k\langle u, x \rangle} \\ &= \sum_{u \in P \cap M/k} e^{k\langle u, x \rangle}, \end{aligned}$$

since σ preserves both P and M/k . In particular, h_k is invariant under σ , concluding the proof. \square

Remark 3.2 An alternative, more invariant, proof of Lemma 3.1 is as follows. By (38), $\sigma.h_k = (\iota_k \circ \sigma)^* h_{FS}$. Now $\iota_k \circ \sigma$ induces the exact same Kodaira embedding if $\sigma \in (S^1)^n < G(H)$. So it suffices to consider $\sigma \in H$; then, since H preserves $P \cap M/k$, one obtains the same Kodaira embedding up to permutation of the coordinates in $\mathbb{P}^{E_P(k)-1}$. Either way, one obtains the same pullback of the Fubini–Study metric as can be from the definition of the Fubini–Study metric or directly from (40), ie $\sigma.h_k = h_k$.

4 An algebraic $\alpha_{k,G}$ -invariant and a Demailly type result

4.1 Analytic definition

Analogous to the classical α_G -invariant, Tian [36, page 128] defined the $\alpha_{k,G}$ -invariant. For this, one restricts to a G -invariant subset of \mathcal{H}_k . To write the subset explicitly in terms of global Kähler potentials it is necessary to choose a continuous G -invariant Hermitian metric h on $-kK_X$:

$$(43) \quad \mathcal{H}_k^G(h) := \left\{ \varphi = \frac{1}{k} \log \sum_i |s_i|_h^2 : \varphi \text{ is } G\text{-invariant and } \{s_i\} \text{ is a basis of } H^0(X, -kK_X) \right\} \subset C^\infty(X).$$

Definition 4.1 Let $G \subseteq \text{Aut } X$ be a compact subgroup. Let h be a fixed continuous G -invariant Hermitian metric on $-kK_X$, and μ a fixed continuous volume form on X . Then

$$\alpha_{k,G}(h, \mu) := \sup \left\{ c > 0 : \sup_{\varphi \in \mathcal{H}_k^G} \int_X e^{-c(\varphi - \sup \varphi)} d\mu < \infty \right\}.$$

Lemma 4.2 Definition 4.1 does not depend on the choice of h or μ .

For this reason we will simply denote the invariants by $\alpha_{k,G}$ from now on.

Proof Since X is compact, any two continuous volume forms are uniformly bounded and hence define the same L^1 spaces. Next, given two continuous G -invariant Hermitian metrics h and \tilde{h} on $-kK_X$ there is an isomorphism from $\mathcal{H}_k^G(h)$ to $\mathcal{H}_k^G(\tilde{h})$ given by $\varphi \mapsto \varphi + (1/k) \log(\tilde{h}/h)$. Observe that $(1/k) \log(\tilde{h}/h)$ is (again by compactness of X) a uniformly bounded function on X . Hence, for a fixed $c > 0$,

$$\sup_{\varphi \in \mathcal{H}_k^G(h)} \int_X e^{-c(\varphi - \sup \varphi)} d\mu < \infty \iff \sup_{\varphi \in \mathcal{H}_k^G(\tilde{h})} \int_X e^{-c(\varphi - \sup \varphi)} d\mu < \infty,$$

as desired. \square

4.2 Algebraic definition

Consider a (complex) nonzero vector subspace V of $H^0(X, kL)$. Associated to it is the (not necessarily complete) linear system $|V| := \mathbb{P}V \subset |kL| := \mathbb{P}H^0(X, kL)$ [22, page 137].

Lemma 4.3 *Let V be vector subspace of $H^0(X, kL)$ of dimension $p > 0$. For any basis $v_1, \dots, v_p \in H^0(X, kL)$ of V , the number*

$$\sup \left\{ c > 0 : \left(\sum_{j=1}^p |v_j(z)|^2 \right)^{-c} \text{ is locally integrable on } X \right\}$$

is the same.

Proof Let $\{v_1^{(\ell)}, \dots, v_p^{(i)}\}$ for $\ell \in \{1, 2\}$, be two bases for $V \in H^0(X, kL)$. Let $A \in \text{GL}(p, \mathbb{C})$ be the change-of-basis matrix, ie $v_j^{(2)} = A_j^i v_i^{(1)}$. Observe that $A^H A$ is a positive Hermitian matrix-valued on X , and denote its eigenvalues by $0 < \lambda_1 \leq \dots \leq \lambda_p$. Define $v^{(\ell)}(z) := (v_1^{(\ell)}(z), \dots, v_p^{(\ell)}(z)) \in \mathbb{C}^p$ and $|v^{(\ell)}(z)|^2 = \sum_{i=1}^p |v_i^{(\ell)}(z)|^2$. Then

$$\frac{|v^{(2)}(z)|^2}{|v^{(1)}(z)|^2} = \frac{|Av^{(1)}(z)|^2}{|v^{(1)}(z)|^2} \in [\lambda_1, \lambda_p].$$

So $|v^{(2)}|^2$ is locally integrable if and only if $|v^{(1)}|^2$ is. □

Thus define the log canonical threshold of the linear system $|V|$ by

$$(44) \quad \text{lct}|V| := \sup \left\{ c > 0 : \left(\sum_j |v_j(z)|^2 \right)^{-c} \text{ is locally integrable on } X \right\}.$$

When $L = -K_X$ there is an $\text{Aut } X$ action on $H^0(X, -kK_X)$ for every $k \in \mathbb{N}$. In [13, Theorem A.3, (A.1)] Demilly noted that then α_G -invariants (for compact subgroup $G \subset \text{Aut } X$) can be algebraically computed as

$$(45) \quad \alpha_G = \inf_{k \in \mathbb{N}} k \inf_{\substack{|V| \subset |-kK_X| \\ V^G = V \neq 0}} \text{lct}|V|,$$

where

$$V^G := \{v \in V : g.v \in V \text{ for all } g \in G\}.$$

Note that in (45) $V \neq 0$ ranges over all G -invariant vector subspaces of $H^0(X, -kK_X)$ (ie of any positive dimension). For example, $\text{lct}|-kK_X| = \infty$. From the definition, if $V_1 \subset V_2$ are two such subspaces it suffices to compute $\text{lct}|V_1|$ since

$$(46) \quad \text{lct}|V_1| \leq \text{lct}|V_2|.$$

It is thus natural to define the algebraic counterpart of the $\alpha_{k,G}$ -invariant as follows:

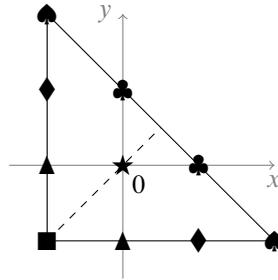


Figure 2: The six orbits $O_1^{(1)}, \dots, O_6^{(1)}$ of the action of the group generated by the reflection about $y = x$ on the polytope corresponding to \mathbb{P}^2 with $k = 1$.

Definition 4.4 Let X be a Fano manifold and $G \subset \text{Aut } X$ be a compact subgroup of the automorphism group. Define

$$(47) \quad \text{glct}_{k,G} := k \inf_{\substack{|V| \subset [-kK_X] \\ V^G = V}} \text{lct}|V|.$$

4.3 Characterization of the equivariant Bergman spaces

First, we show that in the toric setting, the space $\mathcal{H}_k^{G(H)}$ consists of Kähler potentials induced by Kodaira embeddings of multiples of monomials sections, with the norming constants constant along orbits of H .

Denote by

$$(48) \quad O_1^{(k)}, \dots, O_N^{(k)},$$

the orbits of H in $k^{-1}M \cap P$ (see Figure 2 for an example).

Lemma 4.5 Let $\{s_{k,u}\}_{u \in k^{-1}M \cap P}$ be the monomial basis (34) of $H^0(X, -kL)$. Let $G(H) \subseteq \text{Aut } X$ be a subgroup generated by $(S^1)^n$ and a subgroup H of $\text{Aut } P$. Then (recall (43) and (48))

$$\mathcal{H}_k^{G(H)} := \mathcal{H}_k^{G(H)}(h_k) = \left\{ k^{-1} \log \sum_{i=1}^N \lambda_i \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 : \lambda_i > 0 \right\}.$$

Proof Recall the natural basis $\{s_{k,u}\}_{u \in P \cap M/k}$ defined by the lattice monomials (34). Given any other basis $\{s_i\}$ for $H^0(X, -kK_X)$, let $A \in \text{GL}(E_P(k), \mathbb{C})$ be the change-of-basis matrix, so

$$s_i = A_i^u s_{k,u} \quad \text{for } i \in \{1, \dots, E_P(k)\}$$

(we use the Einstein summation convention).

Let $\varphi \in \mathcal{H}_k^{G(H)}$ and let $\{s_i\}$ be the associated basis (43). Expanding φ in terms of the monomials,

$$e^{k\varphi} = \sum_i |s_i|_{h_k}^2 = \sum_{u,u' \in P \cap M/k} \sum_{i,j=1}^{E_P(k)} A_i^u \overline{A_j^{u'}} \langle s_{k,u}, s_{k,u'} \rangle_{h_k} =: |s_{k,0}|_{h_k}^2 \sum_{u,u' \in P \cap M/k} c_{u,u'} z^{ku} \bar{z}^{ku'},$$

for some coefficients $\{c_{u,u'}\}_{u,u' \in P \cap M/k}$. We claim that $\{z^u \bar{z}^{u'}\}_{u,u' \in M}$ are linearly independent. To see that, suppose

$$f(z) = \sum_{u,u' \in M} c_{u,u'} z^u \bar{z}^{u'} = 0$$

with all but finitely many coefficients being 0. We may assume for any $c_{u,u'} \neq 0$ that u and u' lie in the positive orthant, by multiplying by $\prod_i |z_i|^2$ to some sufficiently large power. Then

$$c_{u,u'} = \frac{1}{u!u'!} \left. \frac{\partial^{|u|+|u'|}}{\partial z^u \partial \bar{z}^{u'}} \right|_{z=0} f(z) = 0,$$

proving the claim. Now the subgroup $(S^1)^n < G(H)$ acts on the open orbit $(\mathbb{C}^*)^n$ by

$$(49) \quad (\beta_1, \dots, \beta_n) \cdot (z_1, \dots, z_n) = (e^{\sqrt{-1}\beta_1} z_1, \dots, e^{\sqrt{-1}\beta_n} z_n).$$

So by $(S^1)^n$ -invariance and Lemma 3.1,

$$\beta \cdot e^{k\varphi} = |s_{k,0}|_{h_k}^2 \sum_{u,u' \in P \cap M/k} c_{u,u'} e^{\sqrt{-1}(\beta, k(u-u'))} z^{ku} \bar{z}^{ku'} = |s_{k,0}|_{h_k}^2 \sum_{u,u' \in P \cap M/k} c_{u,u'} z^{ku} \bar{z}^{ku'} = e^{k\varphi},$$

for any $\beta \in (S^1)^n = (\mathbb{R}/2\pi\mathbb{Z})^n$. By comparing the coefficients and using the claim we just demonstrated, it follows that for $e^{k\varphi}$ to be $(S^1)^n$ -invariant we must have $c_{u,u'} = 0$ whenever $u \neq u'$ (the converse is also true, of course).

Moreover, we also require $e^{k\varphi}$ to be invariant under the action of H , ie $c_{u,u} = c_{\sigma u, \sigma u}$ for any $\sigma \in H$. In conclusion, let $O_1^{(k)}, \dots, O_N^{(k)}$ be the orbits of the action of H on $P \cap M/k$. Then

$$e^{k\varphi} = \sum_{i=1}^N \lambda_i \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2,$$

for some coefficients $\{\lambda_i\}_{i=1}^N$. Thus (43) simplifies to

$$\mathcal{H}_k^{G(H)}(h) = \left\{ \varphi = \frac{1}{k} \log \sum_{i=1}^N \lambda_i \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 : \lambda_i > 0 \right\},$$

as claimed. □

4.4 Replacing a basis of sections by an orbit of a section

The next result generalizes [31, Proposition 2.1] to the equivariant setting using Lemma 4.5.

Proposition 4.6 *For a subgroup $G(H) \subseteq \text{Aut } X$ (5) generated by $(S^1)^n$ and a subgroup H of $\text{Aut } P$ (recall (42) and Definition 4.1),*

$$(50) \quad \alpha_{k,G(H)} = \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \text{ for all } u \in P \cap M/k \right\}.$$

Proof Step 1 (estimate a basis-type element using the worst orbit present in its expansion) Comparing (38) and (39) gives a useful expression for h_k (which is not strictly needed for the computation below, but makes it slightly easier to follow):

$$(51) \quad h_k = \left(\sum_{u \in P \cap M/k} |s_{k,u}|^2 \right)^{-1}.$$

Now, assuming $\lambda_N = \max_{i=1, \dots, N} \lambda_i$, following the notation of Lemma 4.5,

$$\begin{aligned} \sup \varphi &\leq \sup \left(\frac{1}{k} \log \sum_{i=1}^N \lambda_N \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right) = \sup \left(\frac{1}{k} \log \lambda_N \sum_{i=1}^N \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right) \\ &= \sup \left(\frac{1}{k} \log \lambda_N \sum_{u \in P \cap M/k} |s_{k,u}|_{h_k}^2 \right) = \frac{1}{k} \log \lambda_N, \end{aligned}$$

since by (51),

$$|s|_{h_k}^2 = \frac{|s|^2}{\sum_{u \in P \cap M/k} |s_{k,u}|^2} \quad \text{for any } s \in H^0(X, -kK_X).$$

If $c > 0$ is such that for each $i \in \{1, \dots, N\}$

$$\int_X \left(\sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k \leq C_c,$$

then for $\varphi \in \mathcal{H}_k^{G(H)}$ (using Lemma 4.5),

$$\begin{aligned} \int_X e^{-c(\varphi - \sup \varphi)} d\mu_k &= e^{c \sup \varphi} \int_X \left(\sum_{i=1}^N \lambda_i \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k \\ &\leq e^{(c/k) \log \lambda_N} \int_X \left(\lambda_N \sum_{u \in O_N^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k \\ &= \lambda_N^{c/k} \lambda_N^{-c/k} \int_X \left(\sum_{u \in O_N^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k \leq C_c. \end{aligned}$$

Step 2 (estimate an orbit-type element using an approximation by degenerating basis-type elements) Conversely, assume $c > 0$ is such that for any $\varphi \in \mathcal{H}_k^{G(H)}$,

$$\int_X e^{-c(\varphi - \sup \varphi)} d\mu_k \leq C_c.$$

Since for any $\ell \in \{1, \dots, N\}$

$$\left(\sum_{u \in O_\ell^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} = \lim_{\lambda_i \rightarrow 0, i \neq \ell} \left(\sum_{i=1}^N \lambda_i \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}},$$

where $\lambda_\ell = 1$, by Fatou’s lemma [18, Lemma 2.18],

$$\begin{aligned} \int_X \left(\sum_{u \in O_\ell^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k &\leq \liminf_{\lambda_i \rightarrow 0, i \neq \ell} e^{-c \sup \varphi_\lambda} e^{c \sup \varphi_\lambda} \int_X \left(\sum_{i=1}^N \lambda_i \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k \\ &= \left(\sup_{u \in O_\ell^{(k)}} \sum |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} \liminf_{\lambda_i \rightarrow 0, i \neq \ell} \int_X e^{-c(\varphi_\lambda - \sup \varphi_\lambda)} d\mu_k \\ &\leq \left(\sup_{u \in O_\ell^{(k)}} \sum |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} C_c, \end{aligned}$$

where

$$\varphi_\lambda := \frac{1}{k} \log \sum_{i=1}^N \lambda_i \sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2.$$

Thus we have shown that

$$\begin{aligned} (52) \quad \alpha_{k,G(H)} &= \sup \left\{ c > 0 : \int_X \left(\sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \text{ for all } i \in \{1, \dots, N\} \right\} \\ &= \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \text{ for all } u \in P \cap M/k \right\}, \end{aligned}$$

proving (50). □

From (52):

Corollary 4.7 Fix $k \in \mathbb{N}$ and let $O_1^{(k)}, \dots, O_N^{(k)}$ be the orbits of the action of H on $k^{-1}M \cap P$. Then

$$\alpha_{k,G(H)} = \min_{1 \leq i \leq N} \sup \left\{ c > 0 : \int_X \left(\sum_{u \in O_i^{(k)}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \right\}.$$

We can now answer affirmatively Problem 1.2 in our setting.

Proof of Proposition 1.3 Let $|V_{O_i^{(k)}}|$ be the linear system generated by $\{(s_{k,u}) : u \in O_i^{(k)}\}$ (recall (48)). By (44) and Corollary 4.7,

$$\alpha_{k,G(H)} = k \min_{i \in \{1, \dots, N\}} \text{lct} |V_{O_i^{(k)}}|.$$

It remains to show

$$\text{glct}_{k,G(H)} = k \min_{i \in \{1, \dots, N\}} \text{lct} |V_{O_i^{(k)}}|.$$

By (46), it suffices in (47) to restrict to irreducible $G(H)$ -invariant linear systems; see [27, page 146]. Now, any $(S^1)^n$ -invariant linear system $|V|$ is spanned by monomials (see Lemma 4.8 below). By irreducibility, this means $|V|$ is spanned by the monomials coming from some H -orbit O_i , ie $|V| = |V_{O_i^{(k)}}|$. □

Lemma 4.8 Let $V \subseteq H^0(X, -kK_X)$ be a complex vector subspace invariant under $(S^1)^n$. Then there exists a unique subset $\mathcal{F} \subset P \cap M/k$ such that

$$V = \text{span}\{s_{k,u}\}_{u \in \mathcal{F}}.$$

Proof It suffices to prove that any section in V is generated by monomials in V . That is, given any section

$$s = \sum_{u \in P \cap M/k} a_u s_{k,u} \in V,$$

if $a_u \neq 0$, then $s_{k,u} \in V$.

By $(S^1)^n$ -invariance, for any $\beta \in (S^1)^n$ and $s \in V$, also $\beta \cdot s \in V$ (recall (49)), that is

$$\beta \cdot s = \sum_{u \in P \cap M/k} a_u \beta \cdot s_{k,u} = \sum_{u \in P \cap M/k} a_u e^{\sqrt{-1}\langle \beta, ku \rangle} s_{k,u} \in V.$$

Thus for any $u_0 \in P \cap M/k$,

$$\begin{aligned} \int_{(S^1)^n} e^{-\sqrt{-1}\langle \beta, ku_0 \rangle} \beta \cdot s \, d\beta &= \sum_{u \in P \cap M/k} a_u s_{k,u} \int_{(S^1)^n} e^{\sqrt{-1}\langle \beta, ku - ku_0 \rangle} \, d\beta \\ &= \sum_{u \in P \cap M/k} a_u s_{k,u} (2\pi)^n \delta_{u-u_0} = (2\pi)^n a_{u_0} s_{k,u_0} \in V. \end{aligned}$$

If $a_{u_0} \neq 0$, then $s_{k,u_0} \in V$. □

Corollary 4.9

$$(53) \quad \alpha_{G(H)} = \inf_k \alpha_{k,G(H)} = \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \text{ for all } u \in P \cap M/k, k \in \mathbb{N} \right\}.$$

Proof By Demailly's theorem (45) [13, (A.1)], (47), and Proposition 1.3,

$$\alpha_{G(H)} = \inf_{k \in \mathbb{N}} \text{glct}_{k,G(H)} = \inf_{k \in \mathbb{N}} \alpha_{k,G(H)}.$$

By (50),

$$\begin{aligned} \alpha_{G(H)} &= \inf_{k \in \mathbb{N}} \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \text{ for all } u \in P \cap M/k \right\} \\ &= \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \text{ for all } u \in P \cap M/k, k \in \mathbb{N} \right\}. \quad \square \end{aligned}$$

Remark 4.10 Here it would not have been enough to invoke Tian's theorem [36, Proposition 6.1], and it was necessary to invoke Demailly's theorem [13, (A.1)]. The reason is that the $\alpha_{k,G}$ -invariants are defined by requiring certain integrals to be merely finite (ie bounded), but with a constant possibly depending on k . Tian's theorem says that \mathcal{H}^G is approximated by \mathcal{H}_k^G , but in approximating $\varphi \in \mathcal{H}^G$ by a sequence $\varphi_k \in \mathcal{H}_k^G$ it might happen that the integrals $\int_X e^{-c(\varphi_k - \sup \varphi_k)} \omega^n$ blow up as k tends to infinity. Demailly precludes that from happening by using the Demailly–Kollár lower semicontinuity of complex singularity exponents.

5 The case of divisible k

We ultimately improve on the results of this section, but we include them since they serve to emphasize the difficulties that still need to be dealt with (see Remark 5.6).

Proposition 5.1 *Let $G(H) \subseteq \text{Aut } X$ (5) be a subgroup generated by $(S^1)^n$ and a subgroup H of $\text{Aut } P$. For $k, \ell \in \mathbb{N}$ we have $\alpha_{k,G(H)} \geq \alpha_{k\ell,G(H)}$.*

Proof Since $h_{k\ell}$ and h_k^ℓ (39) are both smooth metrics on $-k\ell K_X$, by compactness of X they are equivalent. We also use the following lemma:

Lemma 5.2 *Let $a_1, \dots, a_N \geq 0$. Then for $\ell \in \mathbb{N}$,*

$$\sum_{i=1}^N a_i^\ell \leq \left(\sum_{i=1}^N a_i \right)^\ell \leq N^{\ell-1} \sum_{i=1}^N a_i^\ell.$$

Proof The first inequality follows by expanding $(\sum_{i=1}^N a_i)^\ell$. For the second inequality, consider the function $f(x) := x^\ell$ on $[0, +\infty)$. Since f is convex, by Jensen’s inequality

$$f\left(\frac{1}{N} \sum_{i=1}^N a_i\right) \leq \frac{1}{N} \sum_{i=1}^N f(a_i),$$

ie $(\sum_{i=1}^N a_i)^\ell \leq N^{\ell-1} \sum_{i=1}^N a_i^\ell$. □

By Proposition 4.6, and since $M/k\ell \subset M/k$,

$$\begin{aligned} \alpha_{k\ell,G(H)} &= \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k\ell,\sigma u}|_{h_{k\ell}}^2 \right)^{-\frac{c}{k\ell}} d\mu_k < \infty \text{ for all } u \in P \cap M/k\ell \right\} \\ &\leq \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k\ell,\sigma u}|_{h_{k\ell}}^2 \right)^{-\frac{c}{k\ell}} d\mu_k < \infty \text{ for all } u \in P \cap M/k \right\} \\ &= \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}^{\otimes \ell}|_{h_k^\ell}^2 \right)^{-\frac{c}{k\ell}} d\mu_k < \infty \text{ for all } u \in P \cap M/k \right\} \\ &= \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}|_{h_k}^{2\ell} \right)^{-\frac{c}{k\ell}} d\mu_k < \infty \text{ for all } u \in P \cap M/k \right\} \\ &= \sup \left\{ c > 0 : \int_X \left(\sum_{\sigma \in H} |s_{k,\sigma u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \text{ for all } u \in P \cap M/k \right\} = \alpha_{k,G(H)}, \end{aligned}$$

where Lemma 5.2 was invoked in the penultimate equality. □

Proposition 5.3 *Let $G(H) \subseteq \text{Aut } X$ (5) be a subgroup generated by $(S^1)^n$ and a subgroup H of $\text{Aut } P$. There exists $k_0 \in \mathbb{N}$ such that $\alpha_{K,G(H)} = \alpha_{G(H)}$ for all K divisible by k_0 .*

Remark 5.4 The proof will show that k_0 is determined by the fan as follows: let $u_0 \in P^H$ attain $\sup_{u \in P^H} \max_i \langle u, v_i \rangle$. Since P^H is a convex polytope, u_0 will be a vertex of P^H , ie cut out by P^H and the supporting hyperplanes of P containing u_0 . The equations defining P^H are determined by $H \subset GL(M)$, and hence are linear equations with integer coefficients. So are the equations cutting out ∂P . It follows that u_0 is a rational point, ie $u_0 \in M/k_0$ for some k_0 . Let k_0 be the smallest such positive integer. The fact that u_0 is a rational point is originally due to Song [33, page 1257], who proved that $\alpha_{k_0, G(\text{Aut } P)} = \alpha_{G(\text{Aut } P)}$.

Proof We can further simplify (53). Recall (7). By the geometric–arithmetic mean inequality,

$$\begin{aligned} \left(\frac{1}{|H|} \sum_{\sigma \in H} |s_{k, \sigma u}|_{h_k}^2 \right)^{-\frac{\alpha}{k}} &\leq \prod_{\sigma \in H} |s_{k, \sigma u}|_{h_k}^{-2\alpha/(|H|k)} = |s_{|H|k, \pi_H(u)}|_{h_{|H|k}}^{-2\alpha/(|H|k)} \\ &= \left(\frac{1}{|H|} \sum_{\sigma \in H} |s_{|H|k, \sigma \pi_H(u)}|_{h_{|H|k}}^2 \right)^{-\alpha/(|H|k)}, \end{aligned}$$

with the last equality since $\pi_H(u)$ is fixed by H (recall (6)–(7)) so each term in the sum is identical. Therefore the supremum in (53) is unchanged if restricted to those u fixed by H . Thus, using Lemma 5.5 (proven below), (53) simplifies to (recall (6))

$$\begin{aligned} \alpha_{G(H)} &= \sup \left\{ c > 0 : \int_X |s_{k, u}|_{h_k}^{-2c/k} d\mu_k < \infty \text{ for all } u \in P^H \cap \frac{1}{k}M, k \in \mathbb{N} \right\} \\ &= \inf \left\{ k \text{ lct}(s_{k, u}) : u \in P^H \cap \frac{1}{k}M, k \in \mathbb{N} \right\} = \inf_{k \in \mathbb{N}} \inf_{u \in P^H \cap (1/k)M} \frac{1}{\max_i \langle u, v_i \rangle + 1} \\ &= \inf_{u \in P^H} \frac{1}{\max_i \langle u, v_i \rangle + 1} = \frac{1}{\max_i \langle u_0, v_i \rangle + 1}, \end{aligned}$$

where we used the notation of Remark 5.4. By that same remark, $u_0 \in P^H \cap (1/K)M$ whenever $K \in \mathbb{N}$ is divisible by k_0 . In particular, by (50), and using Lemma 5.5 again,

$$\alpha_{K, G(H)} \leq \sup \left\{ c > 0 : \int_X |s_{K, u_0}|_{h_K}^{-2c/K} d\mu_K < \infty \right\} = K \text{ lct}(s_{K, u_0}) = \min_i \frac{1}{\langle u_0, v_i \rangle + 1} = \alpha_{G(H)}.$$

Since by definition $\alpha_{G(H)} \leq \alpha_{K, G(H)}$, equality is achieved and $\alpha_{K, G(H)} = \alpha_{G(H)}$ for all K divisible by k_0 . □

Lemma 5.5 For $u \in P \cap k^{-1}M$, $\text{lct}(s_{k, u}) = (1/k)(1/(1 + \max_i \langle u, v_i \rangle))$ (recall (21) and (34)).

Proof This is well known; see eg [6, Corollary 7.4; 27, Theorem 4.1]. It is also a consequence of Proposition 6.8 proven below (put $\mathcal{F} = \{u\}$). □

Remark 5.6 According to Proposition 5.3, the sequence $\{\alpha_{k, G(H)}\}_{k \in \mathbb{N}}$ is constant (equal to $\alpha_{G(H)}$) along the subsequence $k_0, 2k_0, \dots$. Proposition 5.3 does not yield information for all $k \in \mathbb{N}$ unless $k_0 = 1$, and examples show (see Section 8) that oftentimes $k_0 > 1$. Moreover, even though the sequence

also satisfies $\alpha_{k,G(H)} \geq \alpha_{k\ell,G(H)} \geq \alpha_{G(H)}$ for any $k, \ell \in \mathbb{N}$ (Proposition 5.1), these facts combined are still not enough to conclude Tian’s conjecture without further work. For instance, the sequence

$$a_k := \begin{cases} \alpha_{G(H)} & \text{if } k \text{ is even,} \\ \alpha_{G(H)} + k^{-1} & \text{if } k \text{ is odd,} \end{cases}$$

satisfies these requirements for $k_0 = 2$ (and also satisfy $\lim_k a_k = \alpha_{G(H)}$). Perhaps a more natural sequence that satisfies all the requirements and even for the k_0 of Remark 5.4 is

$$a_k := \inf \left\{ k \operatorname{lct}(s_{k,u}) : u \in P^H \cap \frac{1}{k} M \right\},$$

with the proof of Proposition 5.3 showing that $a_{k_0\ell} = a_{k_0} = \alpha_{G(H)}$ for $\ell \in \mathbb{N}$ but possibly $a_k > \alpha_{G(H)}$ for k not divisible by k_0 . Thus we are led to develop more refined estimates that are the topic of the next section.

6 Estimating singularities associated to orbits

6.1 Real singularity exponents and support functions

In this subsection we develop a key new technical estimate that expresses real singularity exponents associated to collections of toric monomials in terms of support functions.

Definition 6.1 For nonempty finite sets $\mathcal{F}, \mathcal{U} \subseteq \mathbb{R}^n$ and $k \in \mathbb{N}$,

$$c_k(\mathcal{F}, \mathcal{U}) := \sup \left\{ c \in (0, 1) : \int_{\mathbb{R}^n} \frac{(\sum_{u \in \mathcal{F}} e^{\langle ku, x \rangle})^{-c/k}}{(\sum_{u \in \mathcal{U}} e^{\langle ku, x \rangle})^{(1-c)/k}} dx < \infty \right\}.$$

Proposition 6.2 For $\mathcal{F}, \mathcal{U} \subseteq \mathbb{R}^n$ nonempty finite sets with $0 \in \operatorname{int} \operatorname{co} \mathcal{U}$ (recall (11)),

$$c_k(\mathcal{F}, \mathcal{U}) = \sup \{ c \in (0, 1) : 0 \in (1-c) \operatorname{co} \mathcal{U} + c \operatorname{co} \mathcal{F} \} = \sup \left\{ c \in (0, 1) : \left(-\frac{c}{1-c} \operatorname{co} \mathcal{F} \right) \cap \operatorname{co} \mathcal{U} \neq \emptyset \right\} > 0.$$

In particular, $c_k(\mathcal{F}, \mathcal{U})$ is independent of k .

Proof Using polar coordinates $x = rv$ and $dx = r^{n-1} dr \wedge dS(v)$ with $r \in \mathbb{R}_+$ and $v \in S^{n-1}(1) := \{x \in \mathbb{R}^n : |x| = 1\}$,

$$(54) \quad \int_{\mathbb{R}^n} \frac{(\sum_{u \in \mathcal{F}} e^{\langle ku, x \rangle})^{-c/k}}{(\sum_{u \in \mathcal{U}} e^{\langle ku, x \rangle})^{(1-c)/k}} dx = \int_{S^{n-1}(1)} f_c(v) dS(v),$$

where $f_c : S^{n-1}(1) \rightarrow \mathbb{R}_+$ is defined by

$$f_c(v) := \int_0^\infty \frac{(\sum_{u \in \mathcal{F}} e^{kr\langle u, v \rangle})^{-c/k}}{(\sum_{u \in \mathcal{U}} e^{kr\langle u, v \rangle})^{(1-c)/k}} r^{n-1} dr.$$

Notice that for any finite set $\mathcal{A} \subset \mathbb{R}^n$ and $r \in \mathbb{R}_+$ (recall (13)),

$$e^{krh_{\mathcal{A}}(v)} \leq \sum_{u \in \mathcal{A}} e^{kr\langle u, v \rangle} \leq |\mathcal{A}| e^{krh_{\mathcal{A}}(v)}.$$

Hence for $c \in (0, 1)$,

$$(55) \quad \frac{|\mathcal{F}|^{-c/k}}{|\mathcal{U}|^{(1-c)/k}} e^{-kr g_c(v)} \leq \frac{(\sum_{u \in \mathcal{F}} e^{kr \langle u, v \rangle})^{-c/k}}{(\sum_{u \in \mathcal{U}} e^{kr \langle u, v \rangle})^{(1-c)/k}} \leq e^{-kr g_c(v)},$$

where

$$(56) \quad g_c(x) := c \max_{u \in \mathcal{F}} \langle u, x \rangle + (1-c) \max_{u \in \mathcal{U}} \langle u, x \rangle = h_{c\mathcal{F} + (1-c)\mathcal{U}}(x) = c \max_{u \in \text{co}\mathcal{F}} \langle u, x \rangle + (1-c) \max_{u \in \text{co}\mathcal{U}} \langle u, x \rangle \\ = ch_{\text{co}\mathcal{F}}(x) + (1-c)h_{\text{co}\mathcal{U}}(x) = h_{c\text{co}\mathcal{F} + (1-c)\text{co}\mathcal{U}}(x),$$

and where $A + B := \{x + y : x \in A, y \in B\}$ is the Minkowski sum.

We need the following property of support functions:

Claim 6.3 *Let $\mathcal{A} \subset \mathbb{R}^n$ be a nonempty finite set. Then $h_{\mathcal{A}}|_{S^{n-1}(1)} > 0$ if and only if $0 \in \text{int co } \mathcal{A}$.*

Proof Note first that

$$(57) \quad h_{\mathcal{A}} = h_{\text{co}\mathcal{A}};$$

$h_{\mathcal{A}} \leq h_{\text{co}\mathcal{A}}$ since $\mathcal{A} \subset \text{co}\mathcal{A}$, while if $h_{\text{co}\mathcal{A}}(y) = \langle a(y), y \rangle$ for some $a(y) \in \text{co}\mathcal{A}$, and writing $a(y) = \sum_{i=1}^{|\mathcal{A}|} \lambda_i a_i$ with $\sum_{i=1}^{|\mathcal{A}|} \lambda_i = 1$ and $\lambda_i \geq 0$ where $\mathcal{A} = \{a_i\}$, then

$$h_{\text{co}\mathcal{A}}(y) = \sum_{i=1}^{|\mathcal{A}|} \lambda_i \langle a_i, y \rangle \leq \sum_{i=1}^{|\mathcal{A}|} \lambda_i h_{\mathcal{A}}(y) = h_{\mathcal{A}}(y).$$

Next, if $0 \in \text{int co } \mathcal{A}$ then $\epsilon B_2^n \subset \text{co } \mathcal{A}$ where $B_2^n := \{x \in \mathbb{R}^n : |x| \leq 1\}$ and $h_{\text{co}\mathcal{A}} \geq h_{\epsilon B_2^n} = \epsilon h_{B_2^n} = \epsilon$ when restricted to $S^{n-1}(1)$.

Conversely, if $0 \notin \text{int co } \mathcal{A}$ then by convexity of $\text{co } \mathcal{A}$ there exists a hyperplane H passing through a boundary point of $\text{co } \mathcal{A}$ so that $\text{co } \mathcal{A}$ lies on one side of it and 0 lies on the other side of it. If $\nu \in S^{n-1}(1)$ is normal to H and points toward the side not containing $\text{co } \mathcal{A}$ then $h_{\text{co}\mathcal{A}}(\nu) \leq 0$. □

Claim 6.4 *Let $\mathcal{A} \subset \mathbb{R}^n$ be a nonempty finite set. Then $h_{\mathcal{A}}|_{S^{n-1}(1)}$ is somewhere negative if and only if $0 \in \text{int}(\mathbb{R}^n \setminus \text{co } \mathcal{A}) = \mathbb{R}^n \setminus \text{co } \mathcal{A}$.*

Proof Suppose $0 \notin \text{int}(\mathbb{R}^n \setminus \text{co } \mathcal{A}) = \mathbb{R}^n \setminus \text{co } \mathcal{A}$. Equivalently $0 \in \text{co } \mathcal{A}$. Then $h_{\mathcal{A}} \geq h_{\{0\}} = 0$, so $h_{\mathcal{A}}$ is nowhere negative.

Conversely, if $0 \notin \text{co } \mathcal{A}$ then by convexity of $\text{co } \mathcal{A}$ there exists a hyperplane H passing through the origin so that $\text{co } \mathcal{A}$ lies *strictly* on one side. Let $\nu \in S^{n-1}(1)$ be a unit normal to H that points toward the side not containing $\text{co } \mathcal{A}$. By the strictness mentioned above and compactness of $\text{co } \mathcal{A}$, for some small $\epsilon > 0$, $\text{co } \mathcal{A} + \epsilon\nu$ still lies on the same side of H as $\text{co } \mathcal{A}$. Hence, as in the proof of Claim 6.3, $h_{\text{co}\mathcal{A} + \epsilon\nu}(\nu) \leq 0$. In other words,

$$h_{\text{co}\mathcal{A}}(\nu) = h_{\text{co}\mathcal{A} + \epsilon\nu}(\nu) - \epsilon|\nu|^2 \leq -\epsilon,$$

as claimed. □

Corollary 6.5 For $\mathcal{F}, \mathcal{U} \subseteq \mathbb{R}^n$ nonempty finite sets with $0 \in \text{int co } \mathcal{U}$ (recall (11)),

$$(58) \quad \sup \left\{ c \in (0, 1) : \left(-\frac{c}{1-c} \text{co } \mathcal{F} \right) \cap \text{co } \mathcal{U} \neq \emptyset \right\} = \begin{cases} 1 & \text{if } 0 \in \text{co } \mathcal{F}, \\ [1 - \min_{S^{n-1}(1)} (h_{\mathcal{F}}/h_{\mathcal{U}})]^{-1} & \text{if } 0 \notin \text{co } \mathcal{F}. \end{cases}$$

Proof By assumption $0 \in \text{int co } \mathcal{U}$, so by Claim 6.3, $h_{\mathcal{U}}|_{S^{n-1}(1)} > 0$. By (56),

$$(59) \quad g_c = h_{\text{co } \mathcal{U}} \left(1 - c \left(1 - \frac{h_{\text{co } \mathcal{F}}}{h_{\text{co } \mathcal{U}}} \right) \right).$$

Case 1 ($0 \in \text{co } \mathcal{F}$) If $0 \in \text{co } \mathcal{F}$, ie $h_{\text{co } \mathcal{F}} \geq 0$, then $g_c \geq h_{\text{co } \mathcal{U}}(1 - c) > 0$ for any $c \in (0, 1)$. By (59) and Claim 6.3, $0 \in \text{int}(c \text{co } \mathcal{F} + (1 - c) \text{co } \mathcal{U}) \subseteq c \text{co } \mathcal{F} + (1 - c) \text{co } \mathcal{U}$ and so (58) holds in this case.

Case 2 ($0 \notin \text{co } \mathcal{F}$) By assumption and Claim 6.4, $h_{\text{co } \mathcal{F}} < 0$ somewhere; by continuity/compactness let v_0 be a minimizer of the function

$$\frac{h_{\text{co } \mathcal{F}}}{h_{\text{co } \mathcal{U}}}$$

on $S^{n-1}(1)$. In particular,

$$\frac{h_{\text{co } \mathcal{F}}(v_0)}{h_{\text{co } \mathcal{U}}(v_0)} = \min_{S^{n-1}(1)} \frac{h_{\mathcal{F}}}{h_{\mathcal{U}}} < 0.$$

Set

$$c(\mathcal{F}, \mathcal{U}) := \left[1 - \frac{h_{\text{co } \mathcal{F}}(v_0)}{h_{\text{co } \mathcal{U}}(v_0)} \right]^{-1} = \left[1 - \min_{S^{n-1}(1)} \frac{h_{\mathcal{F}}}{h_{\mathcal{U}}} \right]^{-1}.$$

First, if $c \in (0, c(\mathcal{F}, \mathcal{U}))$, for all $v \in S^{n-1}(1)$,

$$g_c(v) = h_{\text{co } \mathcal{U}}(v) \left(1 - c \left(1 - \frac{h_{\text{co } \mathcal{F}}(v)}{h_{\text{co } \mathcal{U}}(v)} \right) \right) \geq h_{\text{co } \mathcal{U}}(v) \left(1 - c \left(1 - \frac{h_{\text{co } \mathcal{F}}(v_0)}{h_{\text{co } \mathcal{U}}(v_0)} \right) \right) = h_{\text{co } \mathcal{U}}(v) \left(1 - \frac{c}{c(\mathcal{F}, \mathcal{U})} \right) > 0.$$

Thus, by (56) and Claim 6.3, $0 \in \text{int}(c \text{co } \mathcal{F} + (1 - c) \text{co } \mathcal{U}) \subseteq c \text{co } \mathcal{F} + (1 - c) \text{co } \mathcal{U}$; in particular,

$$\left(-\frac{c}{1-c} \text{co } \mathcal{F} \right) \cap \text{co } \mathcal{U} \neq \emptyset.$$

So

$$\sup \left\{ c \in (0, 1) : \left(-\frac{c}{1-c} \text{co } \mathcal{F} \right) \cap \text{co } \mathcal{U} \neq \emptyset \right\} \geq c(\mathcal{F}, \mathcal{U}).$$

Second, if $c \in (c(\mathcal{F}, \mathcal{U}), 1)$,

$$g_c(v_0) = h_{\text{co } \mathcal{U}}(v_0) \left(1 - c \left(1 - \frac{h_{\text{co } \mathcal{F}}(v_0)}{h_{\text{co } \mathcal{U}}(v_0)} \right) \right) = h_{\text{co } \mathcal{U}}(v_0) \left(1 - \frac{c}{c(\mathcal{F}, \mathcal{U})} \right) < 0.$$

Thus, by Claim 6.4, $0 \in \mathbb{R}^n \setminus (c \text{co } \mathcal{F} + (1 - c) \text{co } \mathcal{U})$, equivalently,

$$\left(-\frac{c}{1-c} \text{co } \mathcal{F} \right) \cap \text{co } \mathcal{U} = \emptyset,$$

and

$$\sup \left\{ c \in (0, 1) : \left(-\frac{c}{1-c} \text{co } \mathcal{F} \right) \cap \text{co } \mathcal{U} \neq \emptyset \right\} \leq c(\mathcal{F}, \mathcal{U}).$$

Thus also in this case (58) holds. □

The next corollary follows from the proof of Corollary 6.5:

Corollary 6.6 For $\mathcal{F}, \mathcal{U} \subseteq \mathbb{R}^n$ nonempty finite sets with $0 \in \text{int co } \mathcal{U}$ (recall (11)), let

$$c(\mathcal{F}, \mathcal{U}) := \sup \left\{ c \in (0, 1) : \left(-\frac{c}{1-c} \text{co } \mathcal{F} \right) \cap \text{co } \mathcal{U} \neq \emptyset \right\} = \sup \{ c \in (0, 1) : 0 \in (1-c) \text{co } \mathcal{U} + c \text{co } \mathcal{F} \}.$$

Recall (56). If $c < c(\mathcal{F}, \mathcal{U})$, then $g_c > 0$; if $c > c(\mathcal{F}, \mathcal{U})$, then $g_c < 0$ somewhere.

We can now complete the proof of Proposition 6.2. Set

$$c(\mathcal{F}, \mathcal{U}) := \sup \left\{ c \in (0, 1) : \left(-\frac{c}{1-c} \text{co } \mathcal{F} \right) \cap \text{co } \mathcal{U} \neq \emptyset \right\}.$$

By assumption $0 \in \text{int co } \mathcal{U}$, thus $c(\mathcal{F}, \mathcal{U}) > 0$. We consider two cases.

Case 1 ($c \in (0, c(\mathcal{F}, \mathcal{U}))$) By the proof of Corollary 6.5, $g_c|_{S^{n-1}(1)} > 0$. By continuity/compactness it attains its minimum $g_c(\hat{v}) > 0$. Therefore

$$(60) \quad f_c(v) \leq \int_0^\infty e^{-kr g_c(v)} r^{n-1} dr \leq \int_0^\infty e^{-kr g_c(\hat{v})} r^{n-1} dr < \infty,$$

so by (54), $c_k(\mathcal{F}, \mathcal{U}) \geq c$, ie $c_k(\mathcal{F}, \mathcal{U}) \geq c(\mathcal{F}, \mathcal{U})$.

Case 2 ($c \in (c(\mathcal{F}, \mathcal{U}), 1)$) By the proof of Corollary 6.5, $g_c(v_0) < 0$ for some open neighborhood $U \subset S^{n-1}(1)$. Thus, by (55), for some constant $C = C(c, k, \mathcal{F}, \mathcal{U}) > 0$,

$$f_c(v) \geq C \int_0^\infty e^{-kr g_c(v)} r^{n-1} dr = \infty \quad \text{for } v \in U,$$

so $c_k(\mathcal{F}, \mathcal{U}) \leq c$, ie $c_k(\mathcal{F}, \mathcal{U}) \leq c(\mathcal{F}, \mathcal{U})$.

In conclusion, $c_k(\mathcal{F}, \mathcal{U}) = c(\mathcal{F}, \mathcal{U})$ and Proposition 6.2 is proved. □

6.2 Complex singularity exponents

Now we go back to our setting of estimating complex singularity exponents, where P was a reflexive polytope coming from a fan. For any $k \in \mathbb{N}$ and a nonempty subset $\mathcal{F} \subseteq P \cap M/k$ set

$$(61) \quad c_k(\mathcal{F}) := \sup \left\{ c \in (0, 1) : \int_X \left(\sum_{u \in \mathcal{F}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \right\}.$$

Remark 6.7 Consider

$$\tilde{c}_k(\mathcal{F}) := \sup \left\{ c \in (0, \infty) : \int_X \left(\sum_{u \in \mathcal{F}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k < \infty \right\}.$$

Then, by definition (44), $\tilde{c}_k(\mathcal{F}) = k \text{lct}|\text{span}\{s_{k,u}\}_{u \in \mathcal{F}}|$. However, Proposition 6.2 would not be applicable then. The point is that for the proof of Tian's conjecture we need to compute equivariant *global* log canonical thresholds and there will always be "admissible" invariant linear series for which k times the log canonical threshold is at most 1 (ie the "worst" ones will not have $\tilde{c}_k > 1$), ie for which $c_k = \tilde{c}_k$. This is essentially because our H are linear groups so $\{0\}$ is always an H -orbit and $\text{lct}(s_{k,0}) = 1/k = c_k(\{0\}) = \tilde{c}_k(\{0\})$ (by Lemma 5.5).

Proposition 6.8 For any $k \in \mathbb{N}$ and a nonempty subset $\mathcal{F} \subseteq P \cap M/k$,

$$c_k(\mathcal{F}) = \sup \left\{ c \in (0, 1) : P \cap \left(-\frac{c}{1-c} \operatorname{co} \mathcal{F} \right) \neq \emptyset \right\} > 0.$$

In particular, it is independent of k .

Proof By (42),

$$\int_X \left(\sum_{u \in \mathcal{F}} |s_{k,u}|_{h_k}^2 \right)^{-\frac{c}{k}} d\mu_k = (2\pi)^n \int_{\mathbb{R}^n} \frac{(\sum_{u \in \mathcal{F}} e^{(ku,x)})^{-c/k}}{(\sum_{u \in P \cap M/k} e^{(ku,x)})^{(1-c)/k}} dx.$$

By Proposition 6.2 with $\mathcal{U} = P \cap M/k$,

$$c_k(\mathcal{F}) = \sup \left\{ c \in (0, 1) : P \cap \left(-\frac{c}{1-c} \operatorname{co} \mathcal{F} \right) \neq \emptyset \right\} > 0,$$

since $\operatorname{co} \mathcal{U} = P$ (as the vertices of P belong to M/k for all $k \in \mathbb{N}$). □

Remark 6.9 Although not needed for our analysis, it might be of independent interest to understand whether the supremum in (61)—and in the definition of $\alpha_{k,G(H)}$ (50)—is attained. In other words, is (recall the notation of Corollary 6.6)

$$(62) \quad \int_{S^{n-1}(1)} f_c(v) dS(v)$$

finite for $c = c(\mathcal{F}, \mathcal{U})$? By (54)–(55) it is equivalent to consider this question for the integral

$$(63) \quad \int_{S^{n-1}(1)} \int_0^\infty e^{-kr g_c(v)} r^{n-1} dr dS(v) = \int_{\mathbb{R}^n} e^{-kh_{(1-c)\operatorname{co} \mathcal{U} + c \operatorname{co} \mathcal{F}}(x)} dx.$$

A classical formula in convex geometry says that whenever $K \subset \mathbb{R}^n$ is a compact convex set with the origin in its interior, then $n!|K^\circ| = \int_{\mathbb{R}^n} e^{-h_K(y)} dy$; for a detailed proof see [4, (4.2)]. In fact, it is possible to show that when $0 \in \partial K$ the formula still holds in the sense that both sides are equal to ∞ (this is related to, but stronger than, the classical fact that $0 \in \operatorname{int} K$ if and only if K° is bounded [29, Corollary 14.5.1]). To summarize, both (62) and (63) are infinite when $c = c(\mathcal{F}, \mathcal{U})$. We remark that one can also use this point of view to give an alternative, but equivalent, proof of Corollary 6.6.

6.3 Group orbits and singularities

Consider the orbit-averaging map π_H (7), taking a point to the average of its image under H . In particular, points on the same orbit have the same image. Let $M_{\mathbb{R}}^H$ be the subspace of fixed points of H in $M_{\mathbb{R}}$. Then $\pi_H|_{M_{\mathbb{R}}^H} = \operatorname{id}_{M_{\mathbb{R}}^H}$, and $M_{\mathbb{R}}^H$ is the image of π_H . Note that unless H is trivial, this is a proper subspace of $\mathbb{M}_{\mathbb{R}}$, so π_H is a projection from $M_{\mathbb{R}}$ to the invariant subspace $M_{\mathbb{R}}^H$, justifying the notation.

Recalling (6),

$$P^H := P \cap M_{\mathbb{R}}^H.$$

Since P is convex, $\pi_H(P) \subseteq P$. Hence $\pi_H(P) \subseteq P^H$. On the other hand, $P^H = \pi_H(P^H) \subseteq \pi_H(P)$. This implies

$$(64) \quad P^H = \pi_H(P).$$

Then (recall (24))

$$P^H = \pi_H(P) = \pi_H(\text{co}(\text{Ver } P)) = \text{co}(\pi_H(\text{Ver } P)).$$

In particular,

$$(65) \quad \text{Ver } P^H \subseteq \pi_H(\text{Ver } P).$$

Note that equality does not hold in general (for instance, consider $P = [-1, 1]^2$ and H the reflection group about a diagonal).

Lemma 6.10 Fix $u_0 \in M_{\mathbb{R}}^H$, and let \mathcal{F} be a nonempty H -invariant convex subset of some fiber $\pi_H^{-1}(u_0)$ of π_H . Then $\mathcal{F} \cap P \neq \emptyset$ if and only if $u_0 \in P$.

Proof If $\mathcal{F} \cap P \neq \emptyset$, then $\pi_H(\mathcal{F} \cap P) \neq \emptyset$. By (64), $\pi_H(\mathcal{F} \cap P) \subset \pi_H(P) = P^H \subset P$, and since $\emptyset \neq \mathcal{F} \subset \pi_H^{-1}(u_0)$, we have $\pi_H(\mathcal{F} \cap P) \subset \pi_H(\mathcal{F}) = \{u_0\}$. Thus $\pi_H(\mathcal{F} \cap P) \subset \{u_0\} \cap P$, and for this to be nonempty we must have $u_0 \in P$.

For the converse, since \mathcal{F} is H -invariant and convex, $\pi_H(\mathcal{F}) \subset \mathcal{F}$. Additionally, as before $\emptyset \neq \mathcal{F} \subset \pi_H^{-1}(u_0)$ implies $\pi_H(\mathcal{F}) = \{u_0\}$. Thus $u_0 \in \mathcal{F}$. Therefore if $u_0 \in P$ then actually $u_0 \in \mathcal{F} \cap P$. \square

6.4 Proof of Tian's conjecture

We can now prove our main result:

Proof of Theorem 1.4 Fix $k \in \mathbb{N}$ and let $O_1^{(k)}, \dots, O_N^{(k)}$ be the orbits of the action of H on $P \cap M/k$. Recall that π_H maps an orbit to a singleton:

$$\{o_i^{(k)}\} := \pi_H(O_i^{(k)}).$$

Note that $O_i^{(k)} \subset \pi_H^{-1}(o_i^{(k)})$, and hence (as the action of H and hence also π_H is linear) also $\text{co } O_i^{(k)} \subset \pi_H^{-1}(o_i^{(k)})$. Moreover, $\text{co } O_i^{(k)}$ is convex and H -invariant. By Corollary 4.7, Proposition 6.8, and Lemma 6.10,

$$\begin{aligned} \alpha_{k,G(H)} &= \min_{1 \leq i \leq N} c(O_i^{(k)}) = \min_{1 \leq i \leq N} \sup \left\{ c \in (0, 1) : \left(-\frac{c}{1-c} \text{co } O_i^{(k)} \right) \cap P \neq \emptyset \right\} \\ &= \min_{1 \leq i \leq N} \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} o_i^{(k)} \in P \right\} = \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \{o_1^{(k)}, \dots, o_N^{(k)}\} \subset P \right\}. \end{aligned}$$

Since

$$\{o_1^{(k)}, \dots, o_N^{(k)}\} = \pi_H \left(\bigcup_{i=1}^N O_i^{(k)} \right) = \pi_H(k^{-1}M \cap P),$$

we have

$$\begin{aligned}
 (66) \quad \alpha_{k,G(H)} &= \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \pi_H(k^{-1}M \cap P) \subset P \right\} \\
 &= \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \text{co}(\pi_H(k^{-1}M \cap P)) \subset P \right\} \\
 &= \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \pi_H(\text{co}(k^{-1}M \cap P)) \subset P \right\} = \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \pi_H(P) \subset P \right\} \\
 &= \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} P^H \subset P \right\} = \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \text{Ver } P^H \subset P \right\} \\
 &= \min_{u \in \text{Ver } P^H} \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} u \in P \right\}.
 \end{aligned}$$

Also, by (65),

$$\begin{aligned}
 \alpha_{k,G(H)} &= \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \text{Ver } P^H \subset P \right\} \geq \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \pi_H(\text{Ver } P) \subset P \right\} \\
 &\geq \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \pi_H(P) \subset P \right\} = \alpha_{k,G(H)}.
 \end{aligned}$$

Therefore

$$(67) \quad \alpha_{k,G(H)} = \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} \pi_H(\text{Ver } P) \subset P \right\} = \min_{u \in \pi_H(\text{Ver } P)} \sup \left\{ c \in (0, 1) : -\frac{c}{1-c} u \in P \right\}.$$

Recall (28)

$$P = \bigcap_{j=1}^d \{x \in M_{\mathbb{R}} : \langle x, -v_j \rangle \leq 1\}.$$

For any point u and $c \in (0, 1)$,

$$-\frac{c}{1-c} u \in P \quad \text{if and only if} \quad c \leq \min_j \frac{1}{\langle u, v_j \rangle + 1} \quad \text{for all } j \in \{1, \dots, d\}.$$

In sum, combining (66) and (67) implies (8). In particular, $\alpha_{k,G(H)}$ is independent of $k \in \mathbb{N}$ and, by Corollary 4.9, is equal to $\alpha_{G(H)}$. □

7 Grassmannian Tian invariants

The following definition is due to Tian [37, (6.1)]:

Definition 7.1 Let X be a Fano manifold and $G \subseteq \text{Aut } X$ a compact subgroup. For $k, m \in \mathbb{N}$, define (recall (44))

$$\alpha_{k,m,G} := k \inf_{\substack{|V| \leq -kK_X \\ \dim V = m \\ V^G = V}} \text{lct}|V|,$$

with the convention $\inf \emptyset = \infty$.

Equivalently, these invariants are the minimum of the function $V \mapsto k \text{lct}|V|$ over the Grassmannian restricted to the subset of G -invariant m -dimensional subspaces of $H^0(X, -kK_X)$. This subset can be, in some situations, even a finite set (see Remark 7.9). By Remark 7.2 below, these invariants generalize the

invariants $\alpha_{k,G}$. Note also that Definition 7.1 is an *algebraic* one. Also, $\alpha_{k,E_P(k),G} = \infty$ (recall (32)) so it is *not* equal to $\alpha_{k,G}$ from Definition 4.1.

In the toric setting it is natural to consider Conjecture 1.5 for the invariants $\alpha_{k,m,(S^1)^n}$. For these invariants we show that Conjecture 1.5 does not quite hold, and we give a precise breakdown of when it does in terms of a natural convex geometric criterion. In contrast with previous sections, we do not work with the additional symmetry corresponding to the groups $G(H)$ (5). The reason for that is that even in simple situations $\alpha_{k,m,G(H)}$ will not be well-behaved; see Example 8.6.

Remark 7.2 Putting $H = \{\text{id}\}$ in Proposition 4.6, $\alpha_{k,(S^1)^n} = \alpha_{k,1,(S^1)^n}$. In this sense, $\alpha_{k,m,(S^1)^n}$ is a generalization of the $\alpha_{k,(S^1)^n}$ -invariants.

First we show, as a rather immediate consequence of our work, an explicit formula for the invariants of Definition 7.1.

Corollary 7.3 *Let X be toric Fano with associated polytope P (28). For $k, m \in \mathbb{N}$ (recall (61)),*

$$\alpha_{k,m,(S^1)^n} = \min_{\substack{\mathcal{F} \subseteq P \cap M/k \\ |\mathcal{F}|=m}} c_k(\mathcal{F}).$$

Proof Fix $k, m \in \mathbb{N}$. Let $V \subseteq H^0(X, -kK_X) = \text{span}\{s_u\}_{u \in P \cap M/k}$ be a subspace invariant under $(S^1)^n$. By Lemma 4.8, there exists a unique $\mathcal{F} \subseteq P \cap M/k$ such that

$$V = \text{span}\{s_u\}_{u \in \mathcal{F}}.$$

Note that $|\mathcal{F}| = \dim V$. By (44) and (61),

$$\alpha_{k,m,(S^1)^n} = k \min_{\substack{\mathcal{F} \subseteq P \cap M/k \\ |\mathcal{F}|=m}} \text{lct}|\text{span}\{s_u\}_{u \in \mathcal{F}}| = \min_{\substack{\mathcal{F} \subseteq P \cap M/k \\ |\mathcal{F}|=m}} c_k(\mathcal{F}). \quad \square$$

Lemma 7.4 *Let $K \subset \mathbb{R}^n$ be a compact convex set with $0 \in \text{int } K$. For any set $S \subset \mathbb{R}^n$,*

$$\inf_S \|\cdot\|_K = \inf\{\lambda \geq 0 : S \cap \lambda K \neq \emptyset\}.$$

Proof If $S \cap \lambda K \neq \emptyset$ for some $\lambda \geq 0$, pick $x \in S \cap \lambda K$. Since $x \in \lambda K$, by (14), $\|x\|_K \leq \lambda$. Therefore $\inf_{x \in S} \|x\|_K \leq \inf\{\lambda \geq 0 : S \cap \lambda K \neq \emptyset\}$.

On the other hand, for any $x \in S$ we have $S \cap \|x\|_K K \supseteq \{x\} \neq \emptyset$ by (16). Therefore $\inf_{x \in S} \|x\|_K \geq \inf\{\lambda \geq 0 : S \cap \lambda K \neq \emptyset\}$. □

Proposition 6.8 can be reformulated in terms of near-norms:

Corollary 7.5 *For $k \in \mathbb{N}$ and a nonempty set $\mathcal{F} \subseteq P \cap M/k$ (recall (61)),*

$$c_k(\mathcal{F}) = \frac{1}{1 + \min_{\text{co } \mathcal{F}} \|\cdot\|_{-P}}.$$

Proof By Proposition 6.8 and Lemma 7.4,

$$\begin{aligned} c_k(\mathcal{F}) &= \sup\left\{c \in (0, 1) : P \cap \left(-\frac{c}{1-c} \operatorname{co} \mathcal{F}\right) \neq \emptyset\right\} = \sup\left\{c \in (0, 1) : \operatorname{co} \mathcal{F} \cap \frac{1-c}{c}(-P) \neq \emptyset\right\} \\ &= \sup\left\{\frac{1}{1+a} : a \in (0, \infty), \operatorname{co} \mathcal{F} \cap (-aP) \neq \emptyset\right\} = [1 + \inf\{a \in (0, \infty) : \operatorname{co} \mathcal{F} \cap (-aP) \neq \emptyset\}]^{-1} \\ &= [1 + \inf_{\operatorname{co} \mathcal{F}} \|\cdot\|_{-P}]^{-1} = [1 + \min_{\operatorname{co} \mathcal{F}} \|\cdot\|_{-P}]^{-1}, \end{aligned}$$

using compactness of $\operatorname{co} \mathcal{F}$. □

Combining Corollaries 7.3 and 7.5 yields an explicit formula for $\alpha_{k,m,(S^1)^n}$.

Corollary 7.6 *Let X be toric Fano with associated polytope P (28). For $k, m \in \mathbb{N}$,*

$$\alpha_{k,m,(S^1)^n} = \left[1 + \max_{\substack{\mathcal{F} \subseteq P \cap M/k \\ |\mathcal{F}|=m}} \min_{\operatorname{co} \mathcal{F}} \|\cdot\|_{-P}\right]^{-1}.$$

Remark 7.7 For $m = 1$, Corollary 7.6 reads

$$\alpha_{k,(S^1)^n} = \frac{1}{1 + \max_{u \in P \cap M/k} \|u\|_{-P}}.$$

Being a near-norm, the function $\|\cdot\|_{-P}$ satisfies the triangle inequality and is positively 1-homogeneous; hence it is convex. By Lemma 2.3, the vertex set of P is in the integral lattice M , and hence also contained in M/k for all $k \in \mathbb{N}$. This combined with Lemma 2.4 gives

$$(68) \quad \alpha_{k,(S^1)^n} = \frac{1}{1 + \max_{\operatorname{Ver} P} \|\cdot\|_{-P}} = \frac{1}{1 + \max_P \|\cdot\|_{-P}}.$$

This is, of course, a special case of Theorem 1.4 (proved in Section 6), derived here in slightly different notation (using the near-norm instead of the support function). Thus

$$(69) \quad \alpha_{k,(S^1)^n} = \alpha_{(S^1)^n} = \alpha,$$

where the first equality follows from (68) and Demailly’s result [13, (A.1)], and the last equality follows by comparing Theorem 1.4 with Blum and Jonsson’s formula [6, (7.2)].

Proposition 7.8 *Let X be toric Fano with associated polytope P (28). For $k \in \mathbb{N}$ and $m' > m$,*

$$(70) \quad \alpha_{k,m',(S^1)^n} \geq \alpha_{k,m,(S^1)^n}.$$

In particular,

$$\alpha_{k,m,(S^1)^n} \geq \alpha_{k,1,(S^1)^n} = \alpha_{k,(S^1)^n} = \alpha.$$

Equality holds if and only if there is $\mathcal{F} \subseteq P \cap M/k$ with $|\mathcal{F}| = m$ satisfying

$$(71) \quad \|\cdot\|_{-P}|_{\operatorname{co} \mathcal{F}} = \max_P \|\cdot\|_{-P}.$$

Remark 7.9 For a general Fano X , it does not follow from Definition 7.1 that $\alpha_{k,m',G} \geq \alpha_{k,m,G}$, ie there may be no monotonicity in m . For instance, there might be no m -dimensional G -invariant

subspaces. Thanks to Lemma 4.8 the situation for $G = (S^1)^n$ is particularly simple: there are finitely many m -dimensional $(S^1)^n$ -invariant subspaces (in fact, the Grassmannian of m -dimensional subspaces has exactly

$$\binom{E_P(k)}{m}$$

fixed points by the $(S^1)^n$ -action), and moreover every $(S^1)^n$ -invariant subspace consists of 1-dimensional blocks.

Proof Pick $\mathcal{F}' \subseteq P \cap M/k$ that computes $\alpha_{k,m',(S^1)^n}$ (such a subset exists since $P \cap M/k$ is a finite set). That is, $|\mathcal{F}'| = m'$ and $\alpha_{k,m',(S^1)^n} = [1 + \min_{\text{co } \mathcal{F}'} \|\cdot\|_{-P}]^{-1}$ (recall Corollary 7.6). For \mathcal{F} a subset of \mathcal{F}' with $|\mathcal{F}| = m$, Corollary 7.6 implies

$$\alpha_{k,m,(S^1)^n} = \frac{1}{1 + \max_{\substack{\mathcal{A} \subseteq P \cap M/k \\ |\mathcal{A}|=m}} \min_{\text{co } \mathcal{A}} \|\cdot\|_{-P}} \leq \frac{1}{1 + \min_{\text{co } \mathcal{F}} \|\cdot\|_{-P}} \leq \frac{1}{1 + \min_{\text{co } \mathcal{F}'} \|\cdot\|_{-P}} = \alpha_{k,m',(S^1)^n},$$

proving (70).

Next, suppose $\alpha_{k,m,(S^1)^n} = \alpha$, ie (recall (68)–(69))

$$(72) \quad \max_{\substack{\mathcal{F} \subseteq P \cap M/k \\ |\mathcal{F}|=m}} \min_{\text{co } \mathcal{F}} \|\cdot\|_{-P} = \max_P \|\cdot\|_{-P}.$$

For any $\mathcal{F} \subseteq P$, convexity of P (recall (28)) implies $\text{co } \mathcal{F} \subseteq P$. Hence

$$\min_{\text{co } \mathcal{F}} \|\cdot\|_{-P} \leq \max_P \|\cdot\|_{-P}.$$

The equality (72) holds if and only if there is $\mathcal{F} \subseteq P \cap M/k$ such that $|\mathcal{F}| = m$ and

$$\min_{\text{co } \mathcal{F}} \|\cdot\|_{-P} = \max_P \|\cdot\|_{-P},$$

which forces $\|\cdot\|_{-P}$ to be the constant $\max_P \|\cdot\|_{-P}$ on $\text{co } \mathcal{F}$. □

7.1 Proof of Theorem 1.6 and the intuition behind it

Theorem 1.4 resolved Tian's classical stabilization Problem 1.1 in the affirmative. This corresponds to the case $m = 1$ of Conjecture 1.5. Next we show that Conjecture 1.5 is only partially true for $m \geq 2$.

Proof of Theorem 1.6 Let $\mathcal{F} \subset P \cap M/k$ with $|\mathcal{F}| = m$. By convexity of P and as $m \geq 2$, $\text{co } \mathcal{F}$ contains an interval in P . In particular, \mathcal{F} contains a point of $P \setminus \text{Ver } P$. Thus if $(*_P)$ holds then (71) does not hold (for any such \mathcal{F}), and Proposition 7.8 implies (9).

Next, suppose $(*_P)$ fails (and now we relax to $m \geq 1$). Let $\hat{x} \in P \setminus \text{Ver } P$ achieve the maximum of $\|\cdot\|_{-P}$ on P . Express \hat{x} as a convex combination of a subset of vertices of P , $\{p_1, \dots, p_\ell\} \subset \text{Ver } P$, namely

$$(73) \quad \hat{x} = \sum_{i=1}^{\ell} \hat{\lambda}_i p_i \quad \text{where} \quad \sum_{i=1}^{\ell} \hat{\lambda}_i = 1 \quad \text{and} \quad \hat{\lambda} \in (0, 1)^\ell.$$

Note that necessarily

$$(74) \quad \ell > 1.$$

We claim that

$$\|x\|_{-P} = \max_{x \in P} \|\cdot\|_{-P} \quad \text{for any } x \in \text{co}\{p_i\}_{i=1}^\ell.$$

Indeed, letting

$$P' := \text{co}\{p_i\}_{i=1}^\ell \subset P,$$

the claim follows from $\hat{x} \in \text{int } P' \subset P$ and Lemma 2.4(iii).

Now the lattice polytope $P' := \text{co}\{p_i\}_{i=1}^\ell$ has positive dimension by (74). Equivalently, the Ehrhart polynomial of P' has degree at least 1. Hence, by Proposition 2.5, $|P' \cap M/k| \geq m$ for all sufficiently large k . Pick $\mathcal{F} \subseteq P' \cap M/k \subseteq P \cap M/k$ such that $|\mathcal{F}| = m$. Then $\text{co } \mathcal{F} \subseteq P'$. By Proposition 7.8, (10) follows. The case $m = 1$ was already treated in Theorem 1.4, but also from this proof we see that any $k \in \mathbb{N}$ works for $m = 1$. □

The proof of Theorem 1.6 and condition $(*_P)$ have a pleasing geometric interpretation. Geometrically, by Proposition 7.8, $\alpha_{k,m,(S^1)^n} = \alpha$ if and only if there is a subset $\mathcal{F} \subseteq P \cap M/k$ such that $|\mathcal{F}| = m$ and $\text{co } \mathcal{F}$ lies entirely in the level set

$$(75) \quad \{\|\cdot\|_{-P} = \max_P \|\cdot\|_{-P}\} \subseteq \partial P.$$

Recall that any convex subset of the boundary of a polytope must lie in a single face. Hence \mathcal{F} must be a subset of a face F of P . We may assume F is the minimal face that contains \mathcal{F} , ie \mathcal{F} intersects the relative interior of F . In particular, $\|\cdot\|_{-P}|_F$ attains its maximum in the interior of F and so F has to lie entirely in the level set (75) by Lemma 2.4. Therefore $\alpha_{k,m,(S^1)^n} = \alpha$ if and only if there is a subset $\mathcal{F} \subseteq P \cap M/k$ such that $|\mathcal{F}| = m$ and \mathcal{F} is a subset of a face that lies in the level set (75), or equivalently, there is a face F that lies in the level set (75), and $|F \cap M/k| \geq m$.

Notice that the level set $\{\|\cdot\|_{-P} = \lambda\}$ is the dilation $\lambda\partial(-P)$, and (75) is the largest dilation that intersects P (by Lemma 7.4); see Figure 1. Combining all the above, $(*_P)$ states that

$$P \text{ intersects } \max_P \|\cdot\|_{-P} \partial(-P) \text{ only at vertices.}$$

If $(*_P)$ holds, then any such face F consists of a single lattice point. In particular, $|F \cap M/k| = 1$ for any k , and $\alpha_{k,m,(S^1)^n}$ stabilizes if and only if $m = 1$. On the other hand, if $(*_P)$ fails, then there is a face F of positive dimension that lies in the level set (75). Since $|F \cap M/k|$ increases to infinity, for any m , we can find k_0 such that $|F \cap M/k| \geq m$ for $k \geq k_0$, ie $\alpha_{k,m,(S^1)^n}$ stabilizes in k .

7.2 A Demailly result for Grassmannian invariants

The next result was obtained independently by Li and Zhu [28, Proposition 4.1].

Proposition 7.10 Let X be toric Fano with associated polytope P (28). For $m \in \mathbb{N}$,

$$\lim_{k \rightarrow \infty} \alpha_{k,m,(S^1)^n} = \alpha.$$

Proof By continuity, given $\epsilon > 0$ there exists an open set $U_\epsilon \subset P$ such that

$$(76) \quad \|\cdot\|_{-P}|_{U_\epsilon} > \max_P \|\cdot\|_{-P} - \epsilon.$$

Fix $\delta > 0$ sufficiently small so that there is a closed cube $C_\delta \subset U_\epsilon$ with edge length δ . Now, for $k \geq K$, $|C_\delta \cap M/k| \geq \lfloor K\delta \rfloor^n$. Given $m \in \mathbb{N}$, choose K so $\lfloor K\delta \rfloor^n \geq m$. Then for each $k \geq K$ there exists $\mathcal{F} \subset C_\delta \cap M/k \subset P \cap M/k$ such that $|\mathcal{F}| = m$. Since $\mathcal{F} \subset C_\delta$ and C_δ is convex, $\text{co } \mathcal{F} \subset C_\delta \subset U_\epsilon$. By (76),

$$\min_{\text{co } \mathcal{F}} \|\cdot\|_{-P} \geq \min_{C_\delta} \|\cdot\|_{-P} > \max_P \|\cdot\|_{-P} - \epsilon.$$

So by Corollary 7.6 and (68)–(69),

$$\alpha_{k,m,(S^1)^n} \leq \frac{1}{1 + \min_{\text{co } \mathcal{F}} \|\cdot\|_{-P}} < \frac{1}{1 + \max_P \|\cdot\|_{-P} - \epsilon} = \frac{\alpha}{1 - \epsilon\alpha},$$

for any k such that $\lfloor k\delta \rfloor \geq \sqrt[n]{m}$. Hence $\limsup_{k \rightarrow \infty} \alpha_{k,m,(S^1)^n} \leq \alpha/(1 - \epsilon\alpha)$. On the other hand, by Proposition 7.8, $\liminf_{k \rightarrow \infty} \alpha_{k,m,(S^1)^n} \geq \alpha$. Since $\epsilon > 0$ is arbitrary the proof is complete. \square

8 Examples

In Section 8.1 we carefully illustrate Theorems 1.4 and 1.6 in dimension 2 and for all toric del Pezzo surfaces. In Section 8.2 we discuss central symmetry and its characterization in dimension 2 in terms of the α -invariant. In Section 8.3 we draw the connection between $(*_P)$ and work in the convex geometry literature on the asymmetry constant. We illustrate this with computations related to a remarkable toric 3-fold that uniquely minimizes α among all toric (and possibly even nontoric) 3-folds.

8.1 Del Pezzo surfaces

Example 8.1 Let $v_1 = (1, 0)$, $v_2 = (0, 1)$, and $v_3 = (-1, -1)$. Then $\mathbb{P}^2 = X(\Delta)$ for

$$\Delta = \{\{0\}, \mathbb{R}_+v_1, \mathbb{R}_+v_2, \mathbb{R}_+v_3, \mathbb{R}_+v_1 + \mathbb{R}_+v_2, \mathbb{R}_+v_1 + \mathbb{R}_+v_3, \mathbb{R}_+v_2 + \mathbb{R}_+v_3\}$$

and $P = \{-v_1, -v_2, -v_3\}^\circ = \{y \in M_{\mathbb{R}} : \langle y, -v_i \rangle \leq 1 \text{ for } i \in \{1, 2, 3\}\} \subset M_{\mathbb{R}}$ is depicted in Figure 3. Then $\text{Aut } P \cong D_6 = S_3$, the dihedral/cyclic group of order 6 consisting of the permutations of the 3 vertices, or the 3 homogeneous coordinates \mathbb{P}^2 , as in Figure 2. For $H = \text{Aut } P$, P^H is the origin, so by Theorem 1.4, $\alpha_{k,G(H)} = \alpha_{G(H)} = 1$. When H is the trivial group $P^H = P$ and Theorem 1.4 gives $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{3}$, with minimum achieved at any vertex of P . When H is the subgroup \mathbb{Z}_2 given by reflection about, say, $y = x$ (ie generated by the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$) associated to switching two of the homogeneous coordinates of \mathbb{P}^2), P^H is the line segment $\text{co}\{(-1, -1), (\frac{1}{2}, \frac{1}{2})\}$ and $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{3}$, with minimum achieved at $(-1, -1)$. When H is the subgroup of order 3, $P^H = \{0\}$ and $\alpha_{k,G(H)} = \alpha_{G(H)} = 1$,

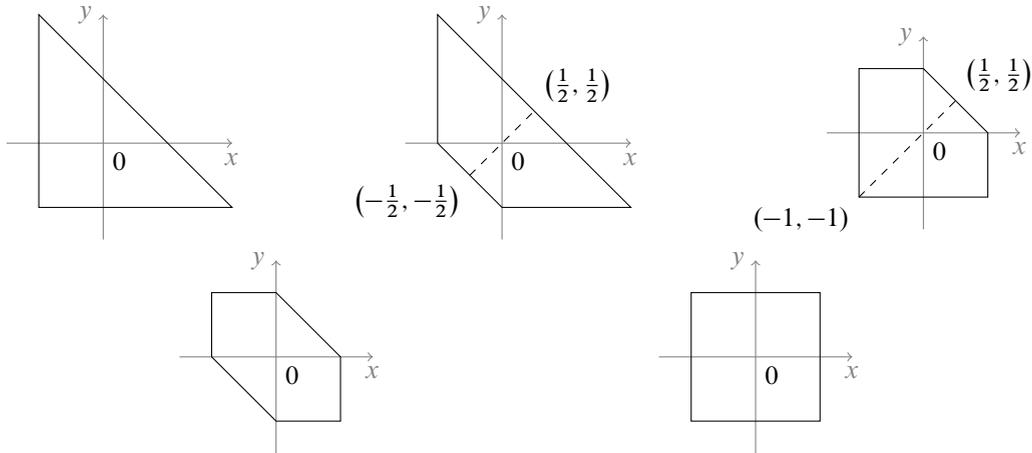


Figure 3: The polytope P for del Pezzo surfaces, namely \mathbb{P}^2 blown up at no more than three generically positioned points, and $\mathbb{P}^1 \times \mathbb{P}^1$. For the two K -unstable examples, the automorphism group $H = \text{Aut } P$ is generated by the reflection about $y = x$, and P^H is the intersection of P with this reflection axis. For the other three examples, the automorphism group $H = \text{Aut } P$ is the dihedral group associated to the polygon P , and $P^H = \{0\}$.

Example 8.2 Let P be the polytope for \mathbb{P}^2 blown up at one point, ie with v_1, v_2 , and v_3 as above and $v_4 = (1, 1)$. Then $\text{Aut } P \cong \mathbb{Z}_2$, as shown in Figure 3. There are therefore two choices for H . When $H = \text{Aut } P$, P^H is the line segment $\text{co}\{(-\frac{1}{2}, -\frac{1}{2}), (\frac{1}{2}, \frac{1}{2})\}$. By Theorem 1.4, $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{2}$, with minimum achieved at $\pm(\frac{1}{2}, \frac{1}{2})$. When H is the trivial group $P^H = P$ and Theorem 1.4 gives $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{3}$, with minimum achieved at either of the vertices $(-1, 2)$ or $(2, -1)$.

Example 8.3 Let P be the polytope for \mathbb{P}^2 blown up at two points, ie with v_1, v_2 , and v_3 as above, $v_4 = (-1, 0)$, and $v_5 = (0, -1)$. Again $\text{Aut } P \cong \mathbb{Z}_2$, as shown in Figure 3. For $H = \text{Aut } P$, P^H is the line segment $\text{co}\{(-1, -1), (\frac{1}{2}, \frac{1}{2})\}$. By Theorem 1.4, $\alpha_{k,G} = \alpha_G = \frac{1}{3}$, with minimum achieved at $(-1, -1)$. When H is the trivial group $P^H = P$ and Theorem 1.4 gives $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{3}$, with minimum achieved still at $(-1, -1)$.

Example 8.4 Let P be the polytope for \mathbb{P}^2 blown up at three noncolinear points, ie with v_1, \dots, v_5 as in the previous example and $v_6 = (1, 1)$. Now $\text{Aut } P \cong D_{12}$, the dihedral group of order 12 consisting of the cyclic permutations of the 6 vertices and the 6 reflections, as in Figure 3. For $H = \text{Aut } P$, P^H is the origin, so by Theorem 1.4, $\alpha_{k,G(H)} = \alpha_{G(H)} = 1$. When H is the trivial group $P^H = P$ and Theorem 1.4 gives $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{2}$, with minimum achieved at any vertex of P . When H is the subgroup \mathbb{Z}_2 given by reflection about $y = x$, P^H is the line segment $\text{co}\{(-\frac{1}{2}, -\frac{1}{2}), (\frac{1}{2}, \frac{1}{2})\}$ and $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{2}$, with minimum achieved at $\pm(\frac{1}{2}, \frac{1}{2})$. When H is the subgroup \mathbb{Z}_2 given by reflection about $y = -x$, $P^H = \text{co}\{(-1, 1), (1, -1)\}$, and $\alpha_{k,G(H)} = \alpha_{G(H)} = \frac{1}{2}$, with minimum achieved at $(\pm 1, \mp 1)$. The same holds for the other reflections. When H contains a cyclic permutation, or more than one reflection, P^H is the origin and $\alpha_{k,G(H)} = \alpha_{G(H)} = 1$.

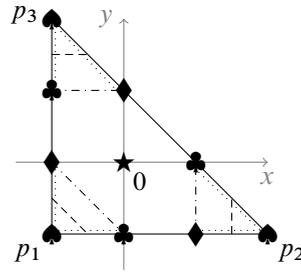


Figure 4: The four orbits $O_1^{(1)}, \dots, O_4^{(1)}$ of the action of the cyclic group of order 3 (generated by cyclic permutation of the homogeneous coordinates of \mathbb{P}^2) on the polytope corresponding to \mathbb{P}^2 with $k = 1$. The dashed-dotted lines depict the sets on which h_{Δ_1} is constant equal to 1 and that compute $\alpha_{1,2,(S^1)^2} = \frac{1}{2}$. These sets are the three components of $-\partial P \cap P$. The dotted lines depict the sets on which h_{Δ_1} is not constant but whose minimum is 1 and that also compute $\alpha_{1,2,(S^1)^2} = \frac{1}{2}$. The dashed lines depict the sets on which h_{Δ_1} is constant equal to $\frac{3}{2}$ and that compute $\alpha_{2,2,(S^1)^2} = \frac{2}{5}$. These sets are the three components of $-\frac{3}{2}\partial P \cap P$.

Example 8.5 Let P be the polytope for $\mathbb{P}^1 \times \mathbb{P}^1$, ie with v_1 and v_2 as above, $v_3 = (-1, 0)$, and $v_4 = (0, -1)$, so $P = [-1, 1]^2$ and $\text{Aut } P \cong D_8$, the dihedral group of order 8 generated by 4 reflections shown in Figure 3. Similar to previous examples, we obtain $\alpha_{k,G(H)} = \alpha_{G(H)} = 1$ whenever H contains a cyclic permutation or more than one reflection. Otherwise it equals $\frac{1}{2}$.

Example 8.6 Let P be the polytope for $X = \mathbb{P}^2$ from Example 8.1. Recall that $\text{Aut } P = S_3$. Let $H = A_3$, the alternating group of order 3, generated by a cyclic permutation of the three vertices. The group is represented by $\{I, A, A^2\}$ where

$$A = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}.$$

All orbits of H on $M \cong \mathbb{Z}^2$ have cardinality 3 except the orbit of the origin, which has cardinality 1. Any $G(H)$ -invariant subspace of $H^0(X, -kK_X)$ is spanned by a collection of monomials indexed by a union $\bigcup_{i=1}^{\ell} O_i^{(k)}$ of H -orbits on $kP \cap M$. This forces the subspace to have dimension $m = 3\ell$ or $m = 3(\ell - 1) + 1$. Thus $m \equiv_3 0, 1$ (see Figure 4 for the case $k = 1$). As a result, if $m \equiv_3 2$ then $\alpha_{k,m,G(H)} = \infty$. For this reason we only consider in Section 7 the case $H = \{\text{id}\}$.

Example 8.7 Let P be the polytope for $X = \mathbb{P}^2$ from Example 8.1. We compute the obstruction $(*P)$, ie compute $\text{argmax}_P \|\cdot\|_{-P} = \text{argmax}_P h_Q = \text{argmax}_P h_{\Delta_1}$. Recall $\Delta_1 = \{v_1, v_2, v_3\} = \{(1, 0), (0, 1), (-1, -1)\}$. Let $y = (y_1, y_2)$ be coordinates on P . Then

$$h_{v_1}(y) = y_1, \quad h_{v_2}(y) = y_2, \quad \text{and} \quad h_{v_3}(y) = -y_1 - y_2,$$

so

$$h_{\Delta_1}(y) = \max\{y_1, y_2, -y_1 - y_2\}.$$

By Lemma 2.4, $\text{Ver } P \cap \text{argmax}_P h_{\Delta_1} \neq \emptyset$. Note,

$$\text{Ver } P = \{p_1, p_2, p_3\} = \{(-1, -1), (2, -1), (-1, 2)\}.$$

So

$$\max_P h_Q = \max\{h_Q(p_1), h_Q(p_2), h_Q(p_3)\} = \max\{2, 2, 2\} = 2,$$

and $\text{Ver } P \subset \text{argmax}_P h_{\Delta_1}$. Suppose that $h_Q(y) = 2$. Then a computation shows that $y \in \text{Ver } P$. Thus $\text{Ver } P = \text{argmax}_P h_{\Delta_1}$. By Theorem 1.6, $\alpha_{k,m,(S^1)^n} > \alpha = \frac{1}{3}$ for all $k \in \mathbb{N}$ and $m \geq 2$.

Let us compute $\alpha_{1,2,(S^1)^n}$. Let $\mathcal{F} = \{f_1, f_2\} \subset P \cap M$. By Corollary 7.6 it suffices to consider “minimal” \mathcal{F} in the sense that for no other \mathcal{F}' , $\text{co } \mathcal{F}' \subset \text{co } \mathcal{F}$. So for instance, we do not need to consider $\{p_1, p_2\}$ but instead

$$\{p_1, (0, -1)\}, \{(0, -1), (1, -1)\}, \{(1, -1), p_2\}.$$

In fact, $\min_{\text{co}\{p_1, (0, -1)\}} h_Q = h_Q(0, -1) = 1$ while $\min_{\{p_1, p_2\}} h_Q = h_Q(\frac{1}{2}, -1) = \frac{1}{2}$.

Also, again by Corollary 7.6, we do not need to consider any \mathcal{F} such that $0 \in \text{co } \mathcal{F}$ since on such an interval the support function attains its absolute minimum, ie vanishes.

Next, observe that $h_{\Delta_1} = 1$ on the vertices of the hexagon

$$P_1 := \text{co}\{\pm(1, 0), \pm(0, 1), \pm(-1, 1)\}.$$

By Lemma 2.4,

$$h_{\Delta_1}|_{P_1} \leq h_{\Delta_1}|_{\text{Ver } P_1} = 1.$$

Thus (by Corollary 7.6) we do not need to consider any \mathcal{F} such that $\text{co } \mathcal{F}$ intersects P_1 , as such an \mathcal{F} will satisfy $\min_{\text{co } \mathcal{F}} h_Q \leq 1$. Since every \mathcal{F} intersects P_1 it follows that $\alpha_{1,2,(S^1)^2} = \frac{1}{2}$ by Corollary 7.6; see Figure 4.

By the same reasoning we can compute $\alpha_{k,2,(S^1)^2}$ using the hexagons

$$\begin{aligned} P_k &:= \text{co}\left\{\left(2 - \frac{1}{k}, \frac{1}{k} - 1\right), \left(2 - \frac{1}{k}, -1\right), \left(-1 + \frac{1}{k}, -1\right), \left(-1, -1 + \frac{1}{k}\right), \left(-1, 2 - \frac{1}{k}\right), \left(-1 + \frac{1}{k}, 2 - \frac{1}{k}\right)\right\} \\ &= \frac{1}{k} \text{co}\{(2k - 1, 1 - k), (2k - 1, -k), (-k + 1, -k), (-k, -k + 1), (-k, 2k - 1), (-k + 1, 2k - 1)\} \\ &\subset P \cap M/k. \end{aligned}$$

By Lemma 2.4,

$$h_{\Delta_1}|_{P_k} \leq h_{\Delta_1}|_{\text{Ver } P_k} = 2 - \frac{1}{k}.$$

Since every \mathcal{F} intersects P_k it follows that

$$\alpha_{k,2,(S^1)^2} = \left[1 + 2 - \frac{1}{k}\right]^{-1} = \frac{k}{3k - 1}$$

by Corollary 7.6. Note that $\lim_k k/(3k - 1) = \frac{1}{3} = \alpha$ (recall Example 8.1 and (69)) in accordance with Proposition 7.10 and that this is a monotone sequence in accordance with Proposition 7.8.

8.2 Central symmetry

The next result should be well known, but we could not find a reference.

Lemma 8.8 *Let X be toric Fano. Let $P \subset M_{\mathbb{R}}$ (see (18) and (27)) be the polytope associated to $(X, -K_X)$. Then $\alpha \leq \frac{1}{2}$, and P is centrally symmetric (ie $P = -P$) if and only if $\alpha = \frac{1}{2}$. If P is not centrally symmetric then $\alpha \leq \frac{1}{3}$.*

Proof If $P \neq -P$, then there is $u_0 \in P \setminus -P$. By (28), $-P = \{y \in M_{\mathbb{R}} : \max_j \langle v_j, y \rangle \leq 1\}$. Thus $\langle u_0, v_{i_0} \rangle > 1$ for some $v_{i_0} \in \Delta_1 \subset N$ (recall (21)). By Theorem 1.4 and (69), or [6, Corollary 7.16],

$$\alpha = \min_{u \in P} \min_i \frac{1}{1 + \langle u, v_i \rangle} \leq \frac{1}{1 + \langle u_0, v_{i_0} \rangle} < \frac{1}{2}.$$

In fact, by convexity one may take $u_0 \in \text{Ver } P \setminus -P$, so $u_0 \in M$ by Lemma 2.3 and hence $\mathbb{Z} \ni \langle u_0, v_{i_0} \rangle \geq 2$, ie $\alpha \leq \frac{1}{3}$.

If $P = -P$, by (68) and (69),

$$\alpha = \frac{1}{1 + \max_P \|\cdot\|_{-P}} = \frac{1}{1 + \max_P \|\cdot\|_P} = \frac{1}{1+1} = \frac{1}{2}$$

by (14) and (16). □

It turns out that any centrally symmetric toric Fano manifold can be written as the product of factors that are either \mathbb{P}^1 or the so-called toric del Pezzo variety V^{2k} , namely the unique indecomposable centrally symmetric toric Fano manifold of even dimension $2k$ [39, Theorem 6]. For example, V^2 is \mathbb{P}^2 blown up at three noncolinear points, and the only centrally symmetric toric Fano 3-folds are $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ and $\mathbb{P}^1 \times V^2$.

Example 8.9 The purpose of this paragraph is to show that when $n = 2$, condition $(*_P)$ holds for a Delzant polytope P if and only if P is not centrally symmetric, ie $P \neq -P$. Alternatively, if $n = 2$, Conjecture 1.5 holds if and only if $\alpha = \frac{1}{2}$, by Lemma 8.8. However, in dimension $n \geq 3$ this is no longer the case: for instance $\mathbb{P}^2 \times \mathbb{P}^1$ has a noncentrally symmetric polytope P but condition $(*_P)$ fails. For a counterexample that does not admit a centrally symmetric factor, consider the Fano 3-fold 4-11 [2, Table 1.3], namely the blowup of $\mathbb{F}_1 \times \mathbb{P}^1$ along $E \times \{0\}$, where E is the -1 -curve on \mathbb{F}_1 . Condition $(*_P)$ fails along three edges of the polytope.

When $n = 2$, the centrally symmetric Delzant polytopes are associated to \mathbb{P}^2 blown up at the three noncolinear points, and $\mathbb{P}^1 \times \mathbb{P}^1$ (Examples 8.4–8.5). Recall the description of $(*_P)$ given in Section 7.1. Since $P = -P$, the intersection of P and $\max_P \|\cdot\|_{-P} \partial(-P)$ is precisely all of ∂P , ie $(*_P)$ fails.

Thus it remains to check \mathbb{P}^2 blown up at up to two points. We have already showed that $(*_P)$ holds for (the noncentrally symmetric) P coming from \mathbb{P}^2 (Example 8.7). It thus remains to check the remaining two cases. The polytope for the 1-point blowup is a subset of the one for \mathbb{P}^2 by chopping a corner. Thus the intersection of P and $\max_P \|\cdot\|_{-P} \partial(-P) = -2\partial P$ is still just the two vertices $(-1, 2)$ and $(2, -1)$. For the 2-point blowup, the intersection of P and $\max_P \|\cdot\|_{-P} \partial(-P) = -2\partial P$ is just the vertex $(-1, -1)$, concluding the proof; see Figure 5.

8.3 The asymmetry constant and a remarkable Fano 3-fold

The asymmetry constant measures how far a body is from being centrally symmetric [21; 24]:

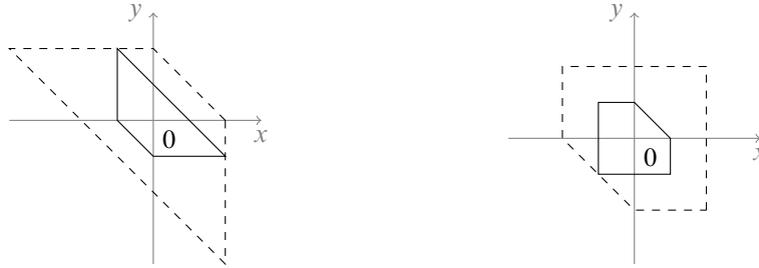


Figure 5: The polytope P for \mathbb{P}^2 blown up 1 or 2 points, and the level set $\{\|\cdot\|_{-P} = 2\}$.

Definition 8.10 For a convex polytope $P \ni 0$, define

$$\text{asym}(P) := \min_{a \in P} \max_P \|\cdot\|_{-(P-a)} = \min_{a \in P} \min\{\lambda \geq 0 : P \subseteq -\lambda(P-a)\} \geq 1.$$

In other words, $\text{asym}(P)(-P)$ is the smallest dilation of $-P$ that admits a translation containing P .

A corollary of (68)–(69) is

$$(77) \quad \alpha \leq [1 + \text{asym}(P)]^{-1}.$$

As we will see in Example 8.11, this bound is often not sharp. In order to see that, let us first describe how to compute $\text{asym}(P)$ in general.

In practice, $\text{asym}(P)$ is the smallest $\lambda \geq 0$ such that

$$-\lambda P + \tau \supseteq P$$

for some $\tau \in \mathbb{R}^n$. In other words, it is the smallest λ among all the pairs $(\tau, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}$ satisfying

$$-\lambda P + \tau \supseteq P.$$

Hence it is the solution to the following linear programming problem: minimize

$$f(\tau, \lambda) := \lambda$$

on the region

$$\{(\tau, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0} : -\lambda P + \tau \supseteq P\}.$$

More precisely, write $\text{Ver } P = \{u_i\}_{i=1}^p$ and recall (29) that the d facets of P are given by $\langle \cdot, v_j \rangle = 1$ for $j \in \{1, \dots, d\}$. Then by (28), $\text{asym}(P)$ is the solution to the following linear programming problem: minimize

$$f(\tau, \lambda) := \lambda$$

under the constraints that $(\tau, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}$ satisfies

$$(78) \quad \langle \tau, v_j \rangle - \lambda \leq \langle u_i, v_j \rangle \quad \text{for all } j \in \{1, \dots, d\} \text{ and } i \in \{1, \dots, p\}.$$

Example 8.11 For the polytopes of Figure 5, one has $\text{asym}(P) = \frac{5}{3}$ for the 1-point blowup and $\text{asym}(P) = \frac{4}{3}$ for the 2-point blowup, confirming the intuition that the latter is “more centrally symmetric” than the former. This also serves to show that in many instances

$$\max_{x \in P} \|\cdot\|_{-P} > \text{asym}(P),$$

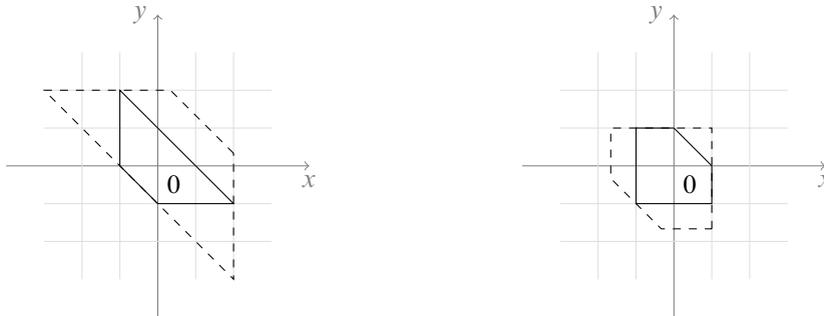


Figure 6: The polytope P for \mathbb{P}^2 blown up 1 or 2 points, and the translation of the smallest dilation of $-P$ that contains P . We can read from the figure that $\text{asym}(P) = \frac{5}{3}, \frac{4}{3}$, respectively.

ie (77) is rarely sharp for Delzant polytopes. The only examples for which we know it to be sharp are \mathbb{P}^n and centrally symmetric P .

By John's theorem [26], $\text{asym}(P) \leq n$, with equality attained for the simplex associated to \mathbb{P}^n [21]. However, the next example shows that for Delzant polytopes arising from toric Fano manifolds one may have $\max_{x \in P} \|\cdot\|_{-P} > n$, even though this seems rare (at least in small dimensions). This first happens in $n = 3$ and just for a unique Fano 3-fold.

Example 8.12 Consider the Fano 3-fold 2-36 [2, Table 1.3], namely the Calabi–Hirzebruch manifold $\mathbb{F}_{3,2} := \mathbb{P}(\mathcal{O}_{\mathbb{P}^2}(2) \oplus \mathcal{O}_{\mathbb{P}^2})$ [8]. It is a toric manifold with [40, (4), page 42]

$$\Delta_1 = \{v_1, \dots, v_5\} = \{(0, 0, 1), (0, 0, -1), (1, 0, 0), (0, 1, 0), (-1, -1, -2)\}.$$

Let P be the corresponding polytope. We have

$$\text{Ver } P = \{(-1, -1, 1), (-1, 0, 1), (0, -1, 1), (-1, -1, -1), (-1, 4, -1), (4, -1, -1)\}.$$

To see this, first notice that

$$\bigcap_{i=1}^4 \{\langle \cdot, v_i \rangle \geq -1\} = [-1, +\infty) \times [-1, +\infty) \times [-1, 1],$$

which is an infinite cuboid with vertices $(-1, -1, \pm 1)$. The polytope P is obtained by cutting it with the hyperplane $\langle \cdot, v_5 \rangle = -1$, introducing the other four vertices $(-1, 0, 1), (0, -1, 1), (-1, 4, -1)$, and $(4, -1, -1)$, namely the vertices of the isosceles trapezoid which is the facet cut out by the last equation $\langle \cdot, v_5 \rangle = -1$; see Figure 7. Since

$$(-4\partial P) \cap P = \{(-1, -1, -1), (-1, 4, -1), (4, -1, -1)\},$$

we have $\alpha = \alpha_{k,1,(S^1)^n} = \frac{1}{5} < \alpha_{k,m,(S^1)^n}$, for $m \geq 2$. Solving the linear programming problem (78) yields $\text{asym}(P) = \frac{14}{5} < 3$.

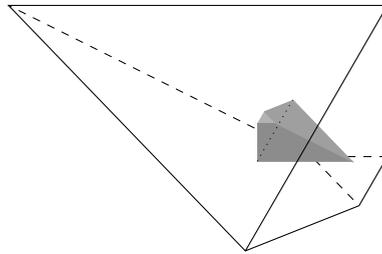


Figure 7: The frustum P (the solid inside) with upper base $\text{co}\{(-1, -1, 1), (-1, 0, 1), (0, -1, 1)\}$ and lower base $\text{co}\{(-1, -1, -1), (-1, 4, -1), (4, -1, -1)\}$, and the boundary of $-4P$ (edges in solid and dashed lines). Their intersection consists of the three vertices of the lower base of P .

References

- [1] **H Ahmadinezhad, I Cheltsov, J Schicho**, *On a conjecture of Tian*, *Math. Z.* 288 (2018) 217–241 MR
- [2] **C Araujo, A-M Castravet, I Cheltsov, K Fujita, A-S Kaloghiros, J Martinez-Garcia, C Shramov, H Süß, N Viswanathan**, *The Calabi problem for Fano threefolds*, London Mathematical Society Lecture Note Series 485, Cambridge Univ. Press (2023) MR
- [3] **V V Batyrev, E N Selivanova**, *Einstein–Kähler metrics on symmetric toric Fano manifolds*, *J. Reine Angew. Math.* 512 (1999) 225–236 MR
- [4] **B Berndtsson, V Mastrantonis, Y A Rubinstein**, *L^p -polarity, Mahler volumes, and the isotropic constant*, *Anal. PDE* 17 (2024) 2179–2245 MR
- [5] **C Birkar**, *Singularities of linear systems and boundedness of Fano varieties*, *Ann. of Math.* 193 (2021) 347–405 MR
- [6] **H Blum, M Jonsson**, *Thresholds, valuations, and K -stability*, *Adv. Math.* 365 (2020) art. id. 107062 MR
- [7] **E Calabi**, *The variation of Kähler metrics, I: The structure of the space; II: A minimum problem*, *Bull. Amer. Math. Soc.* 60 (1954) 167–168
- [8] **E Calabi**, *Métriques kählériennes et fibrés holomorphes*, *Ann. Sci. École Norm. Sup.* 12 (1979) 269–294 MR
- [9] **E Calabi**, *On geodesics in a function space of Kähler metrics*, oral communication in: International Congress on Differential Geometry in memory of Alfred Gray, September 18–23, 2000, Universidad del País Vasco, Bilbao
- [10] **D Catlin**, *The Bergman kernel and a theorem of Tian*, from “Analysis and geometry in several complex variables” (G Komatsu, M Kuranishi, editors), Birkhäuser, Boston, MA (1999) 1–23 MR
- [11] **I Cheltsov**, *Log canonical thresholds of del Pezzo surfaces*, *Geom. Funct. Anal.* 18 (2008) 1118–1144 MR
- [12] **IA Cheltsov, Y A Rubinstein**, *Asymptotically log Fano varieties*, *Adv. Math.* 285 (2015) 1241–1300 MR
- [13] **IA Cheltsov, KA Shramov**, *Log-canonical thresholds for nonsingular Fano threefolds*, *Uspekhi Mat. Nauk* 63 (2008) 73–180 MR In Russian; translated in *Russian Math. Surveys* 63 (2008) 859–958
- [14] **T Delcroix**, *Alpha-invariant of toric line bundles*, *Ann. Polon. Math.* 114 (2015) 13–27 MR
- [15] **T Delcroix**, *Log canonical thresholds on group compactifications*, *Algebr. Geom.* 4 (2017) 203–220 MR

- [16] **E Ehrhart**, *Démonstration de la loi de réciprocité pour un polyèdre entier*, C. R. Acad. Sci. Paris Sér. A-B 265 (1967) 5–7 MR
- [17] **E Ehrhart**, *Sur un problème de géométrie diophantienne linéaire, I: Polyèdres et réseaux*, J. Reine Angew. Math. 226 (1967) 1–29 MR
- [18] **G B Folland**, *Real analysis: modern techniques and their applications*, 2nd edition, Wiley, New York (1999) MR
- [19] **K Fujita, Y Odaka**, *On the K-stability of Fano varieties and anticanonical divisors*, Tohoku Math. J. 70 (2018) 511–521 MR
- [20] **W Fulton**, *Introduction to toric varieties*, Annals of Mathematics Studies 131, Princeton Univ. Press (1993) MR
- [21] **E D Gluskin, A E Litvak**, *Asymmetry of convex polytopes and vertex index of symmetric convex bodies*, Discrete Comput. Geom. 40 (2008) 528–536 MR
- [22] **P Griffiths, J Harris**, *Principles of algebraic geometry*, Wiley-Interscience, New York (1978) MR
- [23] **P M Gruber**, *Convex and discrete geometry*, Grundlehren Math. Wissen. 336, Springer (2007) MR
- [24] **B Grünbaum**, *Measures of symmetry for convex sets*, from “Proceedings of Symposia in Pure Mathematics, VII: Convexity”, Amer. Math. Soc., Providence, RI (1963) 233–270 MR
- [25] **C Jin, Y A Rubinstein**, *Asymptotics of quantized barycenters of lattice polytopes with applications to algebraic geometry*, Math. Z. 309 (2025) art. id. 49 MR
- [26] **F John**, *Extremum problems with inequalities as subsidiary conditions*, from “Studies and Essays Presented to R Courant on his 60th Birthday, January 8, 1948”, Interscience, New York (1948) 187–204 MR
- [27] **H Li, Y Shi, Y Yao**, *A criterion for the properness of the K-energy in a general Kähler class*, Math. Ann. 361 (2015) 135–156 MR
- [28] **Y Li, X Zhu**, *Tian's $\alpha_{m,k}^{\hat{K}}$ -invariants on group compactifications*, Math. Z. 298 (2021) 231–259 MR
- [29] **R T Rockafellar**, *Convex analysis*, Princeton Mathematical Series 28, Princeton Univ. Press (1970) MR
- [30] **Y A Rubinstein, G Tian, K Zhang**, *Basis divisors and balanced metrics*, J. Reine Angew. Math. 778 (2021) 171–218 MR
- [31] **Y Shi**, *On the α -invariants of cubic surfaces with Eckardt points*, Adv. Math. 225 (2010) 1285–1307 MR
- [32] **A Cannas da Silva**, *Symplectic toric manifolds*, from “Symplectic geometry of integrable Hamiltonian systems”, Birkhäuser, Basel (2003) 85–173 MR
- [33] **J Song**, *The α -invariant on toric Fano manifolds*, Amer. J. Math. 127 (2005) 1247–1259 MR
- [34] **G Tian**, *On Kähler–Einstein metrics on certain Kähler manifolds with $C_1(M) > 0$* , Invent. Math. 89 (1987) 225–246 MR
- [35] **G Tian**, *Kähler metrics on algebraic manifolds*, PhD thesis, Harvard University (1988) MR Available at <https://www.proquest.com/docview/303687845>
- [36] **G Tian**, *On a set of polarized Kähler metrics on algebraic manifolds*, J. Differential Geom. 32 (1990) 99–130 MR
- [37] **G Tian**, *On one of Calabi's problems*, from “Several complex variables and complex geometry, II” (E Bedford, J P D'Angelo, R E Greene, S G Krantz, editors), Proc. Sympos. Pure Math. 52, Part 2, Amer. Math. Soc., Providence, RI (1991) 543–556 MR

- [38] **G Tian**, *Existence of Einstein metrics on Fano manifolds*, from “Metric and differential geometry” (X Dai, X Rong, editors), Progr. Math. 297, Springer (2012) 119–159 MR
- [39] **V E Voskresenski, A A Klyachko**, *Toric Fano varieties and systems of roots*, Izv. Akad. Nauk SSSR Ser. Mat. 48 (1984) 237–263 MR In Russian; translated in Math. USSR-Izv. 24 (1985) 221–244
- [40] **K Watanabe, M Watanabe**, *The classification of Fano 3-folds with torus embeddings*, Tokyo J. Math. 5 (1982) 37–48 MR
- [41] **S-T Yau**, *Nonlinear analysis in geometry*, Enseign. Math. 33 (1987) 109–158 MR
- [42] **S Zelditch**, *Szegö kernels and a theorem of Tian*, Internat. Math. Res. Notices (1998) 317–331 MR
- [43] **K Zhang**, *Continuity of delta invariants and twisted Kähler–Einstein metrics*, Adv. Math. 388 (2021) art. id. 107888 MR

*Department of Mathematics, University of Maryland
College Park, MD, United States*

*Department of Mathematics, University of Maryland
College Park, MD, United States*

cjin123@terpmail.umd.edu, yanir@alum.mit.edu

Proposed: Gang Tian

Seconded: Arend Bayer, Tobias H Colding

Received: 30 March 2024

Revised: 10 October 2024

On the logarithmic slice filtration

FEDERICO BINDA
DOOSUNG PARK
PAUL ARNE ØSTVÆR

We consider slice filtrations in logarithmic motivic homotopy theory. Our main results establish conjectured compatibilities with the Beilinson, BMS, and HKR filtrations on (topological, log) Hochschild homology and related invariants. In the case of perfect fields admitting resolution of singularities, we show that the slice filtration realizes the BMS filtration on the p -completed topological cyclic homology. Furthermore, the motivic trace map is compatible with the slice and BMS filtrations, yielding a natural morphism from the motivic slice spectral sequence to the BMS spectral sequence. Finally, we consider the Kummer étale hypersheafification of logarithmic K -theory and show that its very effective slices compute Lichtenbaum étale motivic cohomology.

14A21, 14F30, 14F42; 19E20

1. Introduction	2653
2. Slices for logarithmic motives	2660
3. Slice filtration on K -theory	2664
4. HKR filtration on logarithmic Hochschild homology	2666
5. BMS filtration on logarithmic topological Hochschild homology	2672
6. The characteristic- p case	2680
7. Very effective slices of Kummer étale K -theory	2685
References	2691

1 Introduction

Logarithmic motives provide a framework for studying non- \mathbb{A}^1 -invariant cohomology theories in arithmetic geometry; see Binda, Lundemo, Merici and Park [11], Binda, Lundemo, Park and Østvær [12; 13], Binda and Merici [14], Binda, Park and Østvær [15; 16; 17], Merici [39] and Saito [48]. This paper introduces logarithmic analogs of slice filtrations in motivic homotopy theory defined by Voevodsky [50] and Spitzweck and Østvær [49]. The slice perspective produces explicit calculations, see Isaksen and Østvær [30] for a survey, and remains a powerful tool in homotopy theory following the solution of the Kervaire invariant one problem by Hill, Hopkins and Ravenel [27]. Examples of applications of motivic slices include a proof of the Milnor conjecture on quadratic forms by Röndigs and Østvær [44] and

calculations of universal motivic invariants by Röndigs, Spitzweck and Østvær [46; 47]. One of the main purposes of this paper is to compare the logarithmic slice filtration with the Bhatt–Morrow–Scholze [9] filtration on topological Hochschild homology and refinements.

Assuming that k is a perfect field of characteristic p and A is a smooth k -algebra, Bhatt, Morrow and Scholze [9, Section 1.4] expected that their filtrations on topological Hochschild homology $\mathrm{THH}(A; \mathbb{Z}_p)$ (and the related theories $\mathrm{TC}^-(A; \mathbb{Z}_p)$, $\mathrm{TP}(A; \mathbb{Z}_p)$, and $\mathrm{TC}(A; \mathbb{Z}_p)$) afford a precise relation with filtrations on algebraic K -theory $\mathrm{K}(A)$ (see Friedlander and Suslin [24] and Levine [33]) via the trace maps

$$\mathrm{K}(A) \rightarrow \mathrm{THH}(A; \mathbb{Z}_p), \mathrm{TC}^-(A; \mathbb{Z}_p), \mathrm{TP}(A; \mathbb{Z}_p), \mathrm{TC}(A; \mathbb{Z}_p).$$

In the literature, these filtrations are often called “motivic filtrations”. Mathew [37, page 4] verified this expectation by showing that the filtrations are compatible using Postnikov towers in (pro-)Nisnevich and (pro-)étale topologies. Furthermore, Mathew raised a deeper question regarding the existence of a unified construction that could realize both filtrations.

This paper aims to explore the motivic properties of these filtrations using logarithmic motivic homotopy theory developed in [17] and provide a first positive answer in this direction. A brief recapitulation of Voevodsky’s slice filtration in \mathbb{A}^1 -homotopy theory will help clarify our approach. In algebraic topology, a standard method to understand a spectrum E is to study its Postnikov tower, ie how it is built out of topological Eilenberg–Mac Lane spectra. Sending E to its $(n-1)$ -connected cover $E^{(n)}$ defines a functor from SH to the full subcategory $\Sigma^n \mathrm{SH}_{\geq 0}$ of $(n-1)$ -connected spectra, which is right adjoint to the obvious inclusion. Replacing the category $\Sigma^n \mathrm{SH}_{\geq 0}$ with the category $\Sigma^{2n, n} \mathrm{SH}(S)^{\mathrm{eff}}$ measuring \mathbb{G}_m -effectivity, Voevodsky defined a motivic analog of the Postnikov tower, known as the slice tower

$$f_{n+1} E \rightarrow f_n E \rightarrow f_{n-1} E \rightarrow \cdots \rightarrow E$$

for any motivic spectrum E . When $E = \mathbf{KGL}$ is the spectrum representing algebraic K -theory, the induced spectral sequence is the motivic Atiyah–Hirzebruch spectral sequence of Levine [33] and Voevodsky [51].

Since (topological) Hochschild homology and its variants are not \mathbb{A}^1 -homotopy invariant (in fact, the \mathbb{A}^1 -localization of THH is trivial, see eg Elmanto [21]), there is no motivic spectrum in $\mathrm{SH}(S)$ representing them and thus Voevodsky’s slice machinery is inapplicable. In [17], we developed an extension of $\mathrm{SH}(S)$ using the language of log schemes and proved that there are indeed \mathbb{P}^1 -spectra representing THH and variants. The HKR filtration, the Beilinson filtration, and the Bhatt–Morrow–Scholze filtration on HH , HC^- , THH , TC^- and so on are all compatible with the bounding maps defining the motivic spectra \mathbf{HH} , \mathbf{HC}^- , \mathbf{THH} , \mathbf{TC}^- and relatives, giving rise to filtered \mathbb{P}^1 -motivic spectra. See Propositions 4.9 and 5.8. In [17, Definition 6.5.6] we also constructed a \mathbb{P}^1 -motivic spectrum \mathbf{KGL} (denoted by \mathbf{logKGL} in [17]), representing algebraic K -theory, equipped with a motivic lift of the (logarithmic) cyclotomic trace map when $S = k$ is a perfect field admitting resolution of singularities. See [17, Theorem 8.6.3, Remark 8.6.4].

In this work, we propose an analog of Voevodsky’s slice filtration within the context of $\mathrm{logSH}(S)$ and its linearized variant $\mathrm{logDA}(S)$. We define the effective category $\mathrm{logSH}(S)^{\mathrm{eff}}$ as the localizing subcategory

of $\log\mathrm{SH}(S)$ generated by \mathbb{P}^1 -suspensions of $(X, \partial X) \in \mathrm{SmlSm}/S$, where $X \in \mathrm{Sm}/S$ and ∂X is a strict normal crossing divisor on X . Similarly, we define the coeffective category $\log\mathrm{SH}(S)^{\mathrm{coeff}}$ as its right orthogonal in the obvious sense. Through (de)suspension, we obtain the n -effective and the m -coeffective categories for every $n, m \in \mathbb{Z}$. Following Spitzweck and Østvær [49], we will consider the very effective version of the slice filtration in $\log\mathrm{SH}(S)$. Rather than measuring \mathbb{G}_m -effectivity, the very effective slice filtration measures \mathbb{P}^1 -effectivity and has strong convergence properties. Our main results are the following. Below, when $\mathrm{Fil}_\bullet \mathcal{F}$ is a filtration on \mathcal{F} , we set $\mathrm{Fil}^n \mathcal{F} := \mathrm{cofib}(\mathrm{Fil}_{n+1} \mathcal{F} \rightarrow \mathcal{F})$ for $n \in \mathbb{Z}$.

Theorem 1.1 (see Theorem 4.14) *Let R be any ring, and let $S \rightarrow \mathrm{Spec}(R)$ be a smooth R -scheme. For $n \in \mathbb{Z}$, the complexes*

$$\mathrm{Fil}_{\mathrm{HKR}}^n \mathbf{HH}(-/R), \quad \mathrm{Fil}_{\mathbb{B}}^n \mathbf{HC}^(-/R), \quad \mathrm{Fil}_{\mathbb{B}}^n \mathbf{HP}(-/R)$$

are n -co-very effective, that is, they belong to the subcategory $\log\mathrm{DA}(S)_{\leq n}^{\mathrm{veff}}$ of Construction 2.4. In particular, there are canonical morphisms

$$\tilde{f}_n \mathbf{HH}(-/R) \simeq \Sigma^{2n, n} \tilde{f}_0 \mathbf{HH}(-/R) \rightarrow \mathrm{Fil}_n^{\mathrm{HKR}} \mathbf{HH}(-/R),$$

and similarly for \mathbf{HC}^- and \mathbf{HP} , where \tilde{f}_n denotes the n^{th} very effective slice functor in $\log\mathrm{DA}(S)$.

Theorem 1.2 (see Theorem 5.17) *Let S be the spectrum of a p -adic quasisyntomic ring R . For $n \in \mathbb{Z}$, the filtered spectra*

$$\mathrm{Fil}_{\mathrm{BMS}}^n \mathbf{THH}(-; \mathbb{Z}_p), \quad \mathrm{Fil}_{\mathrm{BMS}}^n \mathbf{TC}^(-; \mathbb{Z}_p), \quad \mathrm{Fil}_{\mathrm{BMS}}^n \mathbf{TP}(-; \mathbb{Z}_p), \quad \mathrm{Fil}_{\mathrm{BMS}}^n \mathbf{TC}(-; \mathbb{Z}_p)$$

are in $\log\mathrm{SH}(S)_{\leq n}^{\mathrm{veff}}$. In particular, there are canonical morphisms

$$\tilde{f}_n \mathbf{THH}(-; \mathbb{Z}_p) \simeq \Sigma^{2n, n} \tilde{f}_0 \mathbf{THH}(-; \mathbb{Z}_p) \rightarrow \mathrm{Fil}_n^{\mathrm{BMS}} \mathbf{THH}(-; \mathbb{Z}_p),$$

and similarly for \mathbf{TC}^- , \mathbf{TP} and \mathbf{TC} .

Assume now that $S = \mathrm{Spec}(k)$ is a field of positive characteristic p , admitting resolution of singularities. Since topological cyclic homology and variants are, in any case, étale sheaves, we can consider an étale variant of the very effective slice filtration in the subcategory spanned by p -completed objects in $\log\mathrm{SH}(k)$. Our result is then the following.

Theorem 1.3 (see Theorem 7.5) *Let k be a perfect field admitting resolution of singularities. Then, the induced morphism*

$$\tilde{f}_i^{\mathrm{két}} \mathbf{TC}(-; \mathbb{Z}_p) \rightarrow \mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{TC}(-; \mathbb{Z}_p)$$

is an equivalence in the ∞ -category of p -complete Kummer étale motives $\log\mathrm{SH}_{\mathrm{két}}^\wedge(k)_p^\wedge$ for every integer i . Here, we denote by $\tilde{f}_i^{\mathrm{két}}$ the i^{th} very effective cover in $\log\mathrm{SH}_{\mathrm{két}}^\wedge(k)_p^\wedge$.

In other words, despite having very different origins, we can identify the slice and the BMS filtration. We see this as an answer to Mathew’s question for $\mathrm{TC}(-; \mathbb{Z}_p)$, ie the slice filtrations realize the two

“motivic filtrations” on \mathbf{K} and $\mathbf{TC}(-; \mathbb{Z}_p)$. We remark that the proof of the above theorem crucially relies on Geisser and Levine [26], together with a motivic version of the prismatic–crystalline comparison by Binda, Lundemo, Merici and Park [11].

It is a natural question to compare the slice towers in $\log\mathrm{SH}(S)$ and $\mathrm{SH}(S)$. When the base S is the spectrum of a field that admits resolution of singularities, we can combine the above results with the trace methods and obtain the following.

Theorem 1.4 *Let k be a perfect field of characteristic $p > 0$ admitting resolution of singularities.*

- (1) *The log motivic trace maps in $\log\mathrm{SH}(k)$*

$$\begin{aligned} \mathbf{KGL} &\rightarrow \mathbf{THH}(-; \mathbb{Z}_p), & \mathbf{KGL} &\rightarrow \mathbf{TC}^-(-; \mathbb{Z}_p), \\ \mathbf{KGL} &\rightarrow \mathbf{TP}(-; \mathbb{Z}_p), & \mathbf{KGL} &\rightarrow \mathbf{TC}(-; \mathbb{Z}_p) \end{aligned}$$

are compatible with the slice filtrations on the left-hand sides and the BMS filtrations on the right-hand sides.

- (2) *The natural maps*

$$\mathbf{KGL} \rightarrow \mathbf{TC}(-; \mathbb{Z}_p) \rightarrow \mathbf{TC}^-(-; \mathbb{Z}_p) \rightarrow \mathbf{TP}(-; \mathbb{Z}_p), \mathbf{THH}(-; \mathbb{Z}_p)$$

on the graded pieces give rise to the natural maps

$$\mathbf{MZ}(i)[2i] \rightarrow L_{\mathrm{sproét}} \mathbf{W}\Omega_{\log}^i[i] \rightarrow \mathrm{Fil}_i^{\mathbf{N}} \mathbf{W}\Omega[2i] \rightarrow \mathbf{W}\Omega[2i], \mathrm{Fil}_i^{\mathrm{conj}} \Omega[2i].$$

- (3) *The graded pieces of the log trace maps yield natural morphisms for every $X \in \mathrm{SmlSm}/k$*

$$\begin{aligned} R\Gamma_{\mathrm{mot}}(X - \partial X, \mathbb{Z}(i)) &\rightarrow R\Gamma_{\mathrm{set}}(X, \tau^{\leq i} \Omega), \\ R\Gamma_{\mathrm{mot}}(X - \partial X, \mathbb{Z}(i)) &\rightarrow R\Gamma_{\mathrm{set}}(X, \mathrm{Fil}_i^{\mathbf{N}} W\Omega), \\ R\Gamma_{\mathrm{mot}}(X - \partial X, \mathbb{Z}(i)) &\rightarrow R\Gamma_{\mathrm{set}}(X, W\Omega), \\ R\Gamma_{\mathrm{mot}}(X - \partial X, \mathbb{Z}(i)) &\rightarrow R\Gamma_{\mathrm{sproét}}(X, W\Omega_{\log}^i[-i]), \end{aligned}$$

where the left-hand side denotes Voevodsky–Suslin–Friedlander motivic cohomology [53] with integral coefficients.

In light of Levine’s comparison [33, Theorem 9.0.3] between the homotopy coniveau tower and the slice tower, we obtain in particular that for $X \in \mathrm{SmlSm}/k$, the log trace maps

$$\begin{aligned} \mathbf{K}(X - \partial X) &\rightarrow \mathbf{THH}(X; \mathbb{Z}_p), & \mathbf{K}(X - \partial X) &\rightarrow \mathbf{TC}^-(X; \mathbb{Z}_p), \\ \mathbf{K}(X - \partial X) &\rightarrow \mathbf{TP}(X; \mathbb{Z}_p), & \mathbf{K}(X - \partial X) &\rightarrow \mathbf{TC}(X; \mathbb{Z}_p) \end{aligned}$$

are compatible with the homotopy coniveau filtrations on the left-hand sides and the BMS filtrations on the right-hand sides. As an immediate corollary, we obtain a natural morphism from the motivic Atiyah–Hirzebruch spectral sequence [33, Section 11.3]

$$E_2^{pq} = H_{\mathrm{mot}}^{i-j}(X - \partial X, \mathbb{Z}(-j)) \Rightarrow \pi_{-i-j} \mathbf{K}(X - \partial X)$$

to the log version of the BMS spectral sequence [9, Theorem 1.12(5)] (see [13, Theorem 1.3] for the log case)

$$E_2^{ij} = H_{\text{syn}}^{i-j}(X, \mathbb{Z}_p(-j)) \Rightarrow \pi_{-i-j} \text{TC}(X; \mathbb{Z}_p)$$

for $X \in \text{SmlSm}/k$ and in particular for $X \in \text{Sm}/k$. In particular, we get directly that the morphism

$$H_{\text{mot}}^i(X - \partial X, \mathbb{Z}(j)) \rightarrow H_{\text{syn}}^i(X - \partial X, \mathbb{Z}_p(j))$$

considered in [9, Corollary 8.21] and [5, Theorem 6.15] (using Geisser–Levine to identify p -adic motivic cohomology) factors canonically through the log syntomic cohomology $H_{\text{syn}}^i(X, \mathbb{Z}_p(j))$. Note that the input of [26] is not necessary to obtain the map: it is just a consequence of the compatibility between the slice filtration and the BMS filtration, together with [33, Theorem 6.4.2]. For p -adic motivic cohomology with \mathbb{Q}_p -coefficients, a similar refinement has been considered by Ertl and Niziol in [23]; see Remark 7.11. The comparison between motivic cohomology and syntomic cohomology (the non-log case) is also considered in the recent work of Elmanto and Morrow [22].

The nontopological counterpart (for classical cyclic, periodic and Hochschild homology) also holds.

Theorem 1.5 *Let k be a perfect field of admitting resolution of singularities (eg $\text{char } k = 0$).*

- (1) *The log motivic trace maps in $\text{logSH}(k)$*

$$\mathbf{KGL} \rightarrow \mathbf{HH}(-/k), \quad \mathbf{KGL} \rightarrow \mathbf{HC}^-(-/k), \quad \mathbf{KGL} \rightarrow \mathbf{HP}(-/k)$$

are compatible with the slice filtrations on the left-hand sides, the HKR filtration on \mathbf{HH} , and the Beilinson filtrations on \mathbf{HC}^- and \mathbf{HP} .

- (2) *The graded pieces of the log motivic trace maps are identified with*

$$\mathbf{MZ}(i)[2i] \rightarrow \mathbf{\Omega}^i[i], \quad \mathbf{MZ}(i)[2i] \rightarrow \mathbf{\Omega}^{\geq i}[2i], \quad \mathbf{MZ}(i)[2i] \rightarrow \mathbf{\Omega}[2i].$$

- (3) *The graded pieces of the log trace maps yield natural morphisms*

$$\begin{aligned} R\Gamma_{\text{mot}}(X - \partial X, \mathbb{Z}(i)) &\rightarrow R\Gamma_{\text{sNis}}(X, \mathbf{\Omega}^i)[-i], \\ R\Gamma_{\text{mot}}(X - \partial X, \mathbb{Z}(i)) &\rightarrow R\Gamma_{\text{sNis}}(X, \mathbf{\Omega}^{\geq i}), \\ R\Gamma_{\text{mot}}(X - \partial X, \mathbb{Z}(i)) &\rightarrow R\Gamma_{\text{sNis}}(X, \mathbf{\Omega}). \end{aligned}$$

We remark that the compatibility between the slice filtration and the BMS filtration along the trace map has been established using completely different methods; see again [22] and the upcoming work of Bachmann, Elmanto and Morrow. Our results refine such statements (at least over a field), since the trace map further factors through the log invariants, in a way that is compatible with the motivic filtrations on the source and target.

Finally, in the last section of the paper, we consider the slice filtration in the (hypercomplete) Kummer étale version $\text{logSH}_{\text{két}}^{\wedge}(k)$ of $\text{logSH}(k)$.

Theorem 1.6 (see Theorem 7.7) *Let k be a perfect field admitting resolution of singularities, and assume that the étale cohomological dimension of k is finite.*

There are natural equivalences

$$S_i^{\text{két}} L_{\text{két}} \mathbf{KGL} \simeq \tilde{S}_i^{\text{két}} L_{\text{két}} \mathbf{KGL} \simeq L_{\text{két}} S_i \mathbf{KGL} \simeq L_{\text{két}} \tilde{S}_i \mathbf{KGL} \simeq \Sigma^{2i,i} L_{\text{két}} \mathbf{MZ}$$

in $\log\text{SH}_{\text{két}}^{\wedge}(k)$. Moreover, $L_{\text{két}} \mathbf{MZ}$ represents Lichtenbaum étale motivic cohomology (see Mazza, Voevodsky and Weibel [38, Lecture 10]) when restricted to Sm/k .

For $X \in \text{SmlSm}/k$ note moreover that there is a natural equivalence

$$\text{map}_{\log\text{SH}_{\text{két}}^{\wedge}(k)}(\Sigma^{\infty} X_+, L_{\text{két}} \mathbf{KGL}) \simeq L_{\text{két}} \mathbf{K}(X),$$

where the right-hand side is the Kummer étale hypersheafification of the functor $X \mapsto \mathbf{K}(X - \partial X)$ for $X \in \text{SmlSm}/k$. In particular, for $X \in \text{Sm}/k$ it is just étale K -theory.

Remark 1.7 Theorem 1.4 provides in particular a natural map for every $X \in \text{SmlSm}/k$

$$R\Gamma(X - \partial X, \mathbb{Z}(i)) \rightarrow R\Gamma_{\text{spét}}(X, W\Omega_{\log}^i[-i]).$$

As pointed out by a referee, one could also consider the composite morphism

$$R\Gamma(X - \partial X, \mathbb{Z}(i)) \rightarrow R\Gamma(X - \partial X, \tau_{\geq i} \mathbb{Z}(i)) \rightarrow R\Gamma(X - \partial X, \mathcal{K}_i^M[-i]) \xrightarrow{\text{dlog}} R\Gamma_{\text{spét}}(X, W\Omega_{\log}^i[-i]),$$

where \mathcal{K}_i^M is the sheaf of Milnor K -theory, and the dlog map can be constructed using for instance the argument in Jannsen, Saito and Zhao [31, Section 1]. It is reasonable to expect that the maps agree. However, this does not follow directly from our results since one would have to explicitly check the shapes of the cyclotomic trace and the resulting map after applying the slice filtration. We leave this problem to the interested reader.

Remark 1.8 (on resolution of singularities) For the K -theoretic applications we have assumed to be working over a field k admitting resolution of singularities. This is used in the explicit computation of the right adjoint ω^* from Voevodsky’s motives to log motives, that on a smooth k -scheme X satisfies $\omega^*(X) = (\bar{X}, \partial X)$, where \bar{X} is a smooth proper compactification of X , with normal crossing boundary ∂X . This simple formula allows us, for example, to identify $\omega^* \mathbf{KGL}$ as well as $\omega^* \mathbf{MZ}$, and to show that ω^* commutes with taking (very effective) slices. Using the above, we can directly invoke the known results in $\text{SH}(k)$. One may ask about compatibility between ω^* and f_n without assuming resolution of singularities. We expect that the analog of Levine’s computation holds for the log K -theory spectrum $\log \mathbf{KGL}$ defined in [17, Section 6.5]. We plan to address this question in future work.

1.9 Outline of the proofs Let us discuss the main idea of the proof of Theorems 1.1 and 1.2. In [17], we constructed motivic spectra in $\log\text{SH}(S)$ representing logarithmic Hochschild, periodic, and cyclic homology. Such spectra are Bott periodic spectra, in the sense that the strict Nisnevich sheaf $\text{HH}(-)$ of spectra on ISch/S satisfies $\Omega_{\mathbb{P}^1} \text{HH} \simeq \text{HH}$ (and similarly for $\text{HP}(-)$ and $\text{HC}^-(-)$). The compatibility of

the \mathbb{P}^1 -bundle formula with the HKR and the Beilinson filtration allows us to show (see Proposition 4.11) that we can promote the construction and obtain filtered motivic \mathbb{P}^1 -spectra, satisfying the equivalence

$$\mathrm{Fil}_i^{\mathrm{HKR}} \mathrm{HH}(-/R)(1)[2] \simeq \mathrm{Fil}_{i+1}^{\mathrm{HKR}} \mathrm{HH}(-/R),$$

and similarly for HP and HC^- with the Beilinson filtration, for any ring R . At this point, the coeffectivity of $\mathrm{Fil}_{\mathrm{HKR}}^n \mathrm{HH}$ with respect to the t -structure given by the very effective slice tower is a direct consequence of the vanishing of the sheaf cohomology $H^{2j}(X, \Omega_{X/R}^j)$ for negative j .

The topological counterpart, that is, for THH and variants, is proven by a conceptually similar method. More precisely, one shows using the \mathbb{P}^1 bundle formula that the BMS filtration on THH gives an equivalence

$$\mathrm{Fil}_i^{\mathrm{BMS}} \mathrm{THH}(-; \mathbb{Z}_p)(1)[2] \simeq \mathrm{Fil}_{i+1}^{\mathrm{BMS}} \mathrm{THH}(-; \mathbb{Z}_p),$$

and similarly for TP, TC^- and TC; see Proposition 5.11. The coeffectivity statement can then be deduced by passing to the associated graded, reducing to vanishing in (log) prismatic cohomology. This is shown by log quasisyntomic descent. Overall, the coeffectivity results immediately imply the existence of the natural morphisms between the very effective slice tower and the BMS, HKR and Beilinson filtrations.

Let us now consider the natural functor $\omega^*: \mathrm{SH}(S) \rightarrow \mathrm{logSH}(S)$, right adjoint to the \mathbb{A}^1 -localization functor from $\mathrm{logSH}(S)$ to $\mathrm{SH}(S)$ for S a scheme. When $S = \mathrm{Spec}(k)$ is a perfect field that admits resolution of singularities, it is easy to show that ω^* commutes with the (very effective) slice filtration on both sides. Using the results in [17, Section 7.7] and Levine [33, Theorems 6.4.2 and 9.0.2], we can then identify the very effective slices of $\omega^* \mathbf{KGL}$ with motivic cohomology in $\mathrm{logSH}(k)$. Combining this with Theorems 1.1 and 1.2 we obtain Theorems 1.4 and 1.5.

Finally, in Section 7, we consider the very effective slices of Kummer étale K -theory. This allows us to consider a spectrum representing Lichtenbaum étale motivic cohomology in $\mathrm{logSH}_{\mathrm{két}}(k)$: if k admits resolution of singularities, this is computed by the zeroth very effective slice of the Kummer étale sheafification of \mathbf{KGL} . See Theorem 7.7. Using Geisser–Levine as input, together with the identification $\mathbf{MZ}_p^{\mathrm{syn}}(i) \simeq L_{\mathrm{sproét}} \mathbf{W}\Omega_{\mathrm{log}}^i[-i]$ as motivic spectra from [11], we can further show that in the Kummer étale (or in the strict étale) p -complete category, we have $\mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{TC}(-; \mathbb{Z}_p) \in (\mathrm{logSH}_{\mathrm{két}}^\wedge(k)_p^\wedge)_{\geq i}^{\mathrm{veff}}$. Since we have already proved in general that $\mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{TC}(-; \mathbb{Z}_p)$ is also i -co-very effective, we conclude.

Acknowledgements We thank Marc Levine and Alberto Merici for inspiring discussions, and Matthew Morrow for helpful comments on a preliminary version of the paper. We are also grateful to the referee for their report, which significantly helped improve the exposition. Binda is partially supported by PRIN 2022 “*The arithmetic of motives and L-functions*” at MUR (Italy). Park is partially supported by the research training group GRK 2240 “*Algebra-geometric methods in algebra, arithmetic and topology*”. Østvær was partially supported by the European Commission – Horizon-MSCA-PF-2022 “*Motivic integral p-adic cohomologies*”, the RCN Project 312472 “*Equations in motivic homotopy theory*”, and a Guest Professorship awarded by the Radboud Excellence Initiative.

2 Slices for logarithmic motives

Throughout this section, let S be a quasicompact and quasiseparated scheme, and let τ be any topology on the category ISch/S generated by a family of morphisms in ISm/S . Here, ISch/S denotes the category of fs log schemes of finite type over S , and ISm/S denotes the category of log smooth fs log schemes of finite type over S . Recall that the category SmlSm/S is the full subcategory of ISm/S consisting of $X \in \mathrm{ISm}/S$ such that $\underline{X} \in \mathrm{Sm}/S$. By [17, Remark 2.2.19], [16, Lemma A.5.10], it can be equivalently described as the category of pairs $X = (\underline{X}, \partial X)$ such that $\underline{X} \in \mathrm{Sm}/S$, $X - \partial X$ is an open subscheme of \underline{X} , and the closed complement ∂X is a strict normal crossing divisor on \underline{X} . We see $(\underline{X}, \partial X)$ as a log scheme with the compactifying log structure given by the embedding $X - \partial X \subset \underline{X}$.

We work with the Nisnevich sheaf of stable presentable symmetric monoidal ∞ -categories

$$\mathrm{Sch}^{\mathrm{op}} \rightarrow \mathrm{CAlg}(\mathrm{Pr}_{\mathrm{Mod}\Lambda}^{\mathrm{L}}), \quad S \mapsto \mathrm{logSH}_{\tau}(S, \Lambda)^{\otimes},$$

given by

$$\mathrm{logSH}_{\tau}(S, \Lambda)^{\otimes} := \mathrm{Sp}_{\mathbb{P}^1}(\mathrm{Sh}_{\tau}(\mathrm{SmlSm}/S, \mathrm{Mod}\Lambda)[(\mathbb{P}^{\bullet}, \mathbb{P}^{\bullet-1})^{-1}])^{\otimes}$$

introduced in [17, Section 3.4], where Λ is an \mathbb{E}_{∞} -ring, and $\mathrm{Mod}\Lambda$ denotes the ∞ -category of Λ -module objects in Sp . Note that this ∞ -category is equivalent to the model obtained by looking at (\mathbb{P}^1, ∞) -local dividing Nisnevich sheaves on ISm/S by [17, Corollary 3.4.7]. It admits a symmetric monoidal structure and is presentable and stable by [36, Theorems 4.5.2.1(1) and 7.1.2.1, and Corollary 7.1.1.5]. To simplify the notation, we will often omit the superscript \otimes . If Λ is the sphere spectrum \mathbb{S} , then we omit Λ in $\mathrm{logSH}_{\tau}(S, \Lambda)$. If Λ is the Eilenberg–Mac Lane spectrum HR for a commutative ring R , then we set

$$\mathrm{logDA}_{\tau}(S) := \mathrm{logSH}_{\tau}(S, H\mathbb{Z}) \quad \text{and} \quad \mathrm{logDA}_{\tau}(S, R) := \mathrm{logSH}_{\tau}(S, HR).$$

If τ is the strict Nisnevich topology, we omit the subscript τ .

We will be using the following conventions for filtrations. Let \mathcal{F} be an object of an ∞ -category \mathcal{C} . The ∞ -category of filtered objects of \mathcal{C} is $\mathrm{Fun}(\mathbb{Z}^{\mathrm{op}}, \mathcal{C})$, where \mathbb{Z}^{op} is the category consisting of the integers such that $\mathrm{Hom}_{\mathbb{Z}^{\mathrm{op}}}(i, j)$ is $*$ if $i \geq j$ and \emptyset if $i < j$. A filtered object $\mathcal{F}(-)$ of \mathcal{C} is *complete* if $\lim_i \mathcal{F}(i) \simeq 0$. The *underlying object* of $\mathcal{F}(-)$ is $\mathcal{F}(-\infty) := \mathrm{colim}_i \mathcal{F}(i)$. The filtration $\mathcal{F}(-)$ on $\mathcal{F}(-\infty)$ is *exhaustive*. We write

$$\mathrm{Fil}_i \mathcal{F} := \mathcal{F}(i), \quad \mathrm{Fil}^i \mathcal{F} := \mathrm{cofib}(\mathrm{Fil}_{i+1} \mathcal{F} \rightarrow \mathcal{F}), \quad \mathrm{gr}_i \mathcal{F} := \mathrm{cofib}(\mathrm{Fil}_{i+1} \mathcal{F} \rightarrow \mathrm{Fil}_i \mathcal{F}).$$

Observe that we have $\mathrm{gr}_i \mathcal{F} \simeq \mathrm{fib}(\mathrm{Fil}^i \mathcal{F} \rightarrow \mathrm{Fil}^{i-1} \mathcal{F})$.

The slice filtration on $\mathrm{SH}(S)$ is due to Voevodsky [50, Section 2]. We have its direct analog on $\mathrm{logSH}(S)$ as follows.

Construction 2.1 Let $\mathrm{logSH}_{\tau}(S, \Lambda)^{\mathrm{eff}} \subset \mathrm{logSH}_{\tau}(S, \Lambda)$ be the category of *effective spectra*, that is, the full stable ∞ -subcategory of $\mathrm{logSH}_{\tau}(S, \Lambda)$ generated under colimits by $\Sigma^{n,0} \Sigma_{\mathbb{P}^1}^{\infty} X_+$; here $\Sigma_{\mathbb{P}^1}^{\infty} X_+$ (denoted by $\Sigma_{\mathbb{T}}^{\infty} X_+$ in [17]) is the \mathbb{P}^1 -suspension spectrum of $X \in \mathrm{SmlSm}/S$ equipped with an extra basepoint, n is an integer, and $\Sigma^{p,q}$ tensoring with the log motivic sphere $S^{p-q} \otimes (\mathbb{A}^1/(\mathbb{A}^1, 0))^{\otimes q}$

(equivalently, it is the localizing subcategory as a triangulated generated by $\Sigma_{\mathbb{P}^1}^\infty X_+$ for varying X , which means the smallest full triangulated subcategory containing those objects and closed under direct sums). See also [15, Section 4] for a quick recollection.

For every integer $i \in \mathbb{Z}$, let $\Sigma_{\mathbb{P}^1}^i \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}} \subset \log\mathrm{SH}_\tau(S, \Lambda)$ be the full subcategory of $\log\mathrm{SH}_\tau(S, \Lambda)$ generated under colimits by $\Sigma_{\mathbb{P}^1}^i E$ for $E \in \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}}$. These categories form an exhaustive filtration of $\log\mathrm{SH}_\tau(S, \Lambda)$ by subcategories

$$\cdots \subset \Sigma_{\mathbb{P}^1}^{i+1} \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}} \subset \Sigma_{\mathbb{P}^1}^i \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}} \subset \cdots \subset \log\mathrm{SH}_\tau(S, \Lambda).$$

Definition 2.2 We will write $\log\mathrm{SH}_\tau(S, \Lambda)_{\geq i}^{\mathrm{eff}}$ for $\Sigma_{\mathbb{P}^1}^i \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}}$ and call it the ∞ -category of i -effective spectra.

Similarly, let $\log\mathrm{SH}_\tau(S, \Lambda)_{\leq i}^{\mathrm{eff}}$ be the ∞ -category of i -coeffective spectra, that is, the full subcategory of $\log\mathrm{SH}_\tau(S, \Lambda)$ spanned by those objects \mathcal{G} such that $\mathrm{Hom}_{\log\mathrm{SH}_\tau(S, \Lambda)}(\mathcal{F}, \mathcal{G}) \simeq 0$ for every $\mathcal{F} \in \log\mathrm{SH}_\tau(S, \Lambda)_{\geq i+1}^{\mathrm{eff}}$.

The inclusion $\iota_i : \Sigma_{\mathbb{P}^1}^i \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}} \subset \log\mathrm{SH}_\tau(S, \Lambda)$ preserves colimits and so admits a right adjoint r_i . Let $f_i^\tau \rightarrow \mathrm{id}$ be the counit of this adjunction pair, and let $f_i^{\tau-1}$ be the cofiber of this natural transformation. Observe that the essential image of f_i^τ is in $\Sigma_{\mathbb{P}^1}^i \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}}$. We set $s_i^\tau := \mathrm{cofib}(f_{i+1}^\tau \rightarrow f_i^\tau)$. We omit τ in f_i^τ , $f_i^{\tau-1}$ and s_i^τ when $\tau = \mathrm{sNis}$.

We also remark that the inclusion ι_i preserves compact objects. Thus, its right adjoint r_i preserves filtered colimits [35, Proposition 5.5.7.2], and hence the composite $f_i = \iota_i r_i$ preserves filtered colimits too.

Remark 2.3 The category $\log\mathrm{SH}(S)^{\mathrm{eff}}$ is a stable symmetric monoidal ∞ -subcategory of $\log\mathrm{SH}(S)$.

We also consider the log version of the very effective slice filtration following Section 5 of Spitzweck and Østvær [49].

Construction 2.4 Let $\log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{veff}}$ be the smallest full subcategory of $\log\mathrm{SH}_\tau(S, \Lambda)$ containing $\Sigma_{\mathbb{P}^1}^\infty X_+$, closed under colimits and closed under extensions in the sense that if $X \rightarrow Y \rightarrow Z$ is a cofiber sequence in $\log\mathrm{SH}_\tau(S, \Lambda)$ with X and Z in $\log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{veff}}$, then also $Y \in \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{veff}}$. We note that $\log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{veff}}$ is a symmetric monoidal ∞ -subcategory of $\log\mathrm{SH}_\tau(S, \Lambda)$, but unlike $\log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{eff}}$ it is not stable.

We can consider the colimit-preserving inclusions (we omit τ and Λ for simplicity)

$$\Sigma_{\mathbb{P}^1}^{i+1} \log\mathrm{SH}(S)^{\mathrm{veff}} \subset \Sigma_{\mathbb{P}^1}^i \log\mathrm{SH}(S)^{\mathrm{veff}} \subset \Sigma_{\mathbb{P}^1}^{i-1} \log\mathrm{SH}(S)^{\mathrm{veff}} \subset \cdots \subset \log\mathrm{SH}(S),$$

giving rise via their right adjoints to functorial filtrations

$$\tilde{f}_{i+1}^\tau E \rightarrow \tilde{f}_i^\tau E \rightarrow \tilde{f}_{i-1}^\tau E \rightarrow \cdots \rightarrow E,$$

where each \tilde{f}_i^τ is the right adjoint to the inclusion $\Sigma_{\mathbb{P}^1}^i \log\mathrm{SH}_\tau(S, \Lambda)^{\mathrm{veff}} \subset \log\mathrm{SH}_\tau(S, \Lambda)$. The very effective slice functors $\tilde{s}_i^\tau : \log\mathrm{SH}_\tau(S, \Lambda) \rightarrow \log\mathrm{SH}_\tau(S, \Lambda)$ are then defined by the cofiber sequence

$$\tilde{f}_{i+1}^\tau \rightarrow \tilde{f}_i^\tau \rightarrow \tilde{s}_i^\tau.$$

Note that one can effectively “compute” the i^{th} slice as $\tilde{s}_i^\tau E \simeq \Sigma_{\mathbb{P}^1}^i \tilde{s}_0^\tau E$ for every motivic spectrum E that satisfies Bott periodicity $\Sigma_{\mathbb{P}^1} E \simeq E$. As before, we omit τ in \tilde{f}_i^τ and f_i^τ when $\tau = \text{sNis}$.

Definition 2.5 We will write $\log\text{SH}_\tau(S, \Lambda)_{\geq i}^{\text{veff}}$ for $\Sigma_{\mathbb{P}^1}^i \log\text{SH}_\tau(S, \Lambda)^{\text{veff}}$. We define for every integer i the category $\log\text{SH}_\tau(S, \Lambda)_{\leq 0}^{\text{veff}}$ to be the category of *co-very effective* spectra, that is the full subcategory of $\log\text{SH}_\tau(S, \Lambda)$ that is spanned by those objects \mathcal{G} such that $\text{Hom}_{\log\text{SH}_\tau(S, \Lambda)}(\mathcal{F}, \mathcal{G}) \simeq 0$ for every $\mathcal{F} \in \log\text{SH}_\tau(S, \Lambda)_{\geq 1}^{\text{veff}}$. We similarly write $\log\text{SH}_\tau(S, \Lambda)_{\leq i}^{\text{veff}}$ for the right orthogonal subcategory to $\log\text{SH}_\tau(S, \Lambda)_{\geq i+1}^{\text{veff}}$.

Remark 2.6 For every integer i , observe that we have the obvious inclusions

$$\log\text{SH}(S)_{\geq i}^{\text{veff}} \subset \log\text{SH}(S)_{\geq i}^{\text{eff}} \quad \text{and} \quad \log\text{SH}(S)_{\leq i}^{\text{eff}} \subset \log\text{SH}(S)_{\leq i}^{\text{veff}}.$$

Remark 2.7 Assume that $S = \text{Spec}(k)$ is a field and that $\tau = \text{dNis}$. As in \mathbb{A}^1 -motivic homotopy theory, $\log\text{SH}_\tau(k, \Lambda)^{\text{veff}}$ is the connective part of t -structures on $\log\text{SH}_\tau(S, \Lambda)$ and on $\log\text{SH}_\tau(S, \Lambda)^{\text{eff}}$ by [36, Proposition 1.4.4.11]. Note that the analog of Morel’s connectivity theorem in the log setting [14, Theorem 3.2], together with semipurity [14, Theorem 4.4], gives that this is exactly the nonnegative part of the standard homotopy t -structure; see [40, Theorem 5.2.3].

Remark 2.8 A similar construction can be performed in other motivic settings. For example, one could consider, in the category of \mathbb{P}^1 -spectra MS_S introduced by Annala and Iwasa [2], the subcategory $\text{MS}_S^{\text{veff}}$ generated by $\Sigma_{\mathbb{P}^1}^\infty X_+$ for $X \in \text{Sm}/S$ and closed under colimits and extensions. Again by [36, Proposition 1.4.4.11], this is the connective part of a t -structure on MS_S (that we can call the homotopy t -structure, if the expected connection to homotopy sheaves holds). We can consider the colimit-preserving inclusions

$$\Sigma_{\mathbb{P}^1}^{i+1} \text{MS}_S^{\text{veff}} \subset \Sigma_{\mathbb{P}^1}^i \text{MS}_S^{\text{veff}} \subset \dots \subset \text{MS}_S,$$

giving rise via their right adjoints to functorial filtrations

$$\tilde{f}_{i+1} E \rightarrow \tilde{f}_i E \rightarrow \dots \rightarrow E$$

for any spectrum E .

Recall from [17, Construction 4.0.19] that there is a symmetric monoidal functor

$$\lambda_\# : \text{MS}_S \rightarrow \log\text{SH}(S)$$

induced by the functor $\lambda : \text{Sm}/S \rightarrow \text{SmlSm}/S$ given by $\lambda(X) := X$ (seen as log scheme with trivial log structure). The following is immediate from the definitions.

Proposition 2.9 *Let S be any quasicompact quasiseparated scheme. Then for every integer n we have the inclusion $\lambda_\# \Sigma_{\mathbb{P}^1}^n \text{MS}_S^{\text{veff}} \subset \Sigma_{\mathbb{P}^1}^n \log\text{SH}(S)^{\text{veff}}$.*

We record some elementary properties of the slice filtration.

Lemma 2.10 *Let τ' be a topology on SmlSm/S finer than τ . Then for every integer i , the τ' -localization functor*

$$L_{\tau'} : \log\text{SH}_{\tau}(S, \Lambda) \rightarrow \log\text{SH}_{\tau'}(S, \Lambda)$$

sends $\log\text{SH}_{\tau}(S, \Lambda)_{\geq i}^{\text{eff}}$ into $\log\text{SH}_{\tau'}(S, \Lambda)_{\geq i}^{\text{eff}}$, and its right adjoint

$$\iota : \log\text{SH}_{\tau'}(S, \Lambda) \hookrightarrow \log\text{SH}_{\tau}(S, \Lambda)$$

sends $\log\text{SH}_{\tau'}(S, \Lambda)_{\leq i}^{\text{eff}}$ into $\log\text{SH}_{\tau}(S, \Lambda)_{\leq i}^{\text{eff}}$. A similar result holds for the very effective version, too.

Proof We only give proof for the effective version. The statement about the left adjoint is a consequence of the fact that $L_{\tau'}$ preserves colimits and sends $\Sigma^{2n,n}\Sigma^{\infty}X_+$ to $\Sigma^{2n,n}\Sigma^{\infty}X_+$ for every integer n . Use the natural isomorphism

$$\text{Hom}_{\log\text{SH}_{\tau}(S, \Lambda)}(\mathcal{F}, \iota\mathcal{G}) \cong \text{Hom}_{\log\text{SH}_{\tau'}(S, \Lambda)}(L_{\tau'}\mathcal{F}, \mathcal{G})$$

for $\mathcal{F} \in \log\text{SH}_{\tau}(S, \Lambda)_{\geq i+1}^{\text{eff}}$ and $\mathcal{G} \in \log\text{SH}_{\tau'}(S, \Lambda)_{\leq i}^{\text{eff}}$ to deduce the statement about the right adjoint. \square

Lemma 2.11 *Let $\Lambda \rightarrow \Lambda'$ be a map of \mathbb{E}_{∞} -rings. Then the base-change functor*

$$-\otimes_{\Lambda} \Lambda' : \log\text{SH}_{\tau}(S, \Lambda) \rightarrow \log\text{SH}_{\tau}(S, \Lambda')$$

sends $\log\text{SH}_{\tau}(S, \Lambda)_{\geq i}^{\text{eff}}$ into $\log\text{SH}_{\tau}(S, \Lambda')_{\geq i}^{\text{eff}}$, and the restriction functor

$$\log\text{SH}_{\tau}(S, \Lambda') \rightarrow \log\text{SH}_{\tau}(S, \Lambda)$$

sends $\log\text{SH}_{\tau}(S, \Lambda')_{\leq i}^{\text{eff}}$ to $\log\text{SH}_{\tau}(S, \Lambda)_{\leq i}^{\text{eff}}$. A similar result holds for the very effective version, too.

Proof This is a consequence of the fact that $-\otimes_{\Lambda} \Lambda'$ preserves colimits and sends $\Sigma^{2n,n}\Sigma^{\infty}X_+$ to $\Sigma^{2n,n}\Sigma^{\infty}X_+$ for every $X \in \text{SmlSm}/S$ and integer n . \square

In particular, the restriction functor for the morphism $\mathbb{S} \rightarrow H\mathbb{Z}$

$$\log\text{DA}(S) = \log\text{SH}(S, H\mathbb{Z}) \rightarrow \log\text{SH}(S)$$

sends $\log\text{DA}(S)_{\leq i}^{\text{eff}}$ to $\log\text{SH}(S)_{\leq i}^{\text{eff}}$.

The following easy lemma will be useful for the main theorem later in the text.

Lemma 2.12 *If $\mathcal{F} \in \log\text{SH}_{\tau}(S, \Lambda)$ admits a filtration such that $\text{Fil}^i \mathcal{F} \in \log\text{SH}_{\tau}(S, \Lambda)_{\leq i}^{\text{eff}}$ for every integer i , then there exists a unique natural morphism of filtered objects*

$$f_{\bullet}^{\tau} \mathcal{F} \rightarrow \text{Fil}_{\bullet} \mathcal{F}$$

whose morphism of underlying objects is $\text{id}: \mathcal{F} \rightarrow \mathcal{F}$. Similarly, if $\mathcal{F} \in \log\text{SH}_{\tau}(S, \Lambda)$ admits a filtration such that $\text{Fil}^i \mathcal{F} \in \log\text{SH}_{\tau}(S, \Lambda)_{\leq i}^{\text{veff}}$ for every i , then there exists a unique natural morphism of filtered objects

$$\tilde{f}_{\bullet}^{\tau} \mathcal{F} \rightarrow \text{Fil}_{\bullet} \mathcal{F}$$

whose morphism of underlying objects is $\text{id}: \mathcal{F} \rightarrow \mathcal{F}$.

Proof We only give proof for the effective version. Consider the naturally induced maps

$$f_{\tau}^i \mathcal{F} \rightarrow f_{\tau}^i(\mathrm{Fil}^i \mathcal{F}) \xrightarrow{\simeq} \mathrm{Fil}^i \mathcal{F},$$

where the first map is obtained by functoriality and the second map is the inverse of the counit of the adjunction, which is an equivalence since $\mathrm{Fil}^i \mathcal{F} \in \mathrm{logSH}_{\tau}(S, \Lambda)_{\leq i}^{\mathrm{eff}}$ for every integer i . By taking $\mathrm{fib}(\mathcal{F} \rightarrow (-))$, we get the desired morphism $f_{\bullet}^{\tau} \mathcal{F} \rightarrow \mathrm{Fil}_{\bullet} \mathcal{F}$. For its uniqueness, consider the induced exact sequence

$$\begin{aligned} \mathrm{Hom}_{\mathrm{logSH}_{\tau}(S, \Lambda)}(\Sigma^{1,0} f_{i+1}^{\tau} \mathcal{F}, \mathrm{Fil}^i \mathcal{F}) &\rightarrow \mathrm{Hom}_{\mathrm{logSH}_{\tau}(S, \Lambda)}(f_{\tau}^i \mathcal{F}, \mathrm{Fil}^i \mathcal{F}) \\ &\rightarrow \mathrm{Hom}_{\mathrm{logSH}_{\tau}(S, \Lambda)}(\mathcal{F}, \mathrm{Fil}^i \mathcal{F}) \rightarrow \mathrm{Hom}_{\mathrm{logSH}_{\tau}(S, \Lambda)}(f_{i+1}^{\tau} \mathcal{F}, \mathrm{Fil}^i \mathcal{F}). \end{aligned}$$

The first and fourth terms are 0 since $\Sigma^{1,0}$ sends $\mathrm{logSH}_{\tau}(S, \Lambda)_{\geq i+1}^{\mathrm{eff}}$ into itself (note that this works verbatim for $\mathrm{logSH}_{\tau}(S, \Lambda)_{\geq i+1}^{\mathrm{veff}}$ as well, since it is closed under colimits, hence in particular under positive suspensions). This establishes the uniqueness of $f^{\bullet} \mathcal{F} \rightarrow \mathrm{Fil}^{\bullet} \mathcal{F}$ and hence the uniqueness of $f_{\bullet}^{\tau} \mathcal{F} \rightarrow \mathrm{Fil}_{\bullet} \mathcal{F}$. □

3 Slice filtration on K-theory

One of the fundamental results in \mathbb{A}^1 -homotopy theory is the equivalence $s_i \mathbf{KGL} \simeq \Sigma^{2i,i} \mathbf{MZ}$ in $\mathrm{SH}(k)$, where k is a perfect field, due to Voevodsky [52] in characteristic zero and Levine [33] in general (this is now known more generally, eg over Dedekind schemes by [7] and [46, Theorem 2.19]). The effective and the very effective slices of \mathbf{KGL} coincide, see [1, Lemma 2.4], so that, in particular, the motivic Eilenberg–Mac Lane spectrum \mathbf{MZ} is very effective.

In this section, we promote this to $\mathrm{logSH}(k)$, assuming resolution of singularities.

Proposition 3.1 *Let S be a quasicompact quasiseparated scheme. Then, for every integer i , we have the inclusion*

$$\omega^* \mathrm{SH}(S)_{\leq i}^{\mathrm{eff}} \subset \mathrm{logSH}(S)_{\leq i}^{\mathrm{eff}}.$$

A similar result holds for the very effective version, too.

Proof By adjunction, it suffices to show $\omega_{\#} \mathrm{logSH}(S)_{\geq i}^{\mathrm{eff}} \subset \mathrm{SH}(S)_{\geq i}^{\mathrm{eff}}$ and $\omega_{\#} \mathrm{logSH}(S)_{\geq i}^{\mathrm{veff}} \subset \mathrm{SH}(S)_{\geq i}^{\mathrm{veff}}$. This holds since $\omega_{\#}$ preserves colimits and $\omega_{\#} \Sigma^{n,0} \Sigma^{\infty} Y_+ \simeq \Sigma^{n,0} \Sigma^{\infty} (Y - \partial Y)_+$ for $Y \in \mathrm{SmlSm}/S$ and integer n . □

Proposition 3.2 *Let k be a perfect field admitting resolution of singularities. Then, for every integer i , we have the inclusion*

$$\omega^* \mathrm{SH}(k)_{\geq i}^{\mathrm{eff}} \subset \mathrm{logSH}(k)_{\geq i}^{\mathrm{eff}}.$$

A similar result holds for the very effective version, too.

Proof For $X \in \text{Sm}/k$, there exists proper $Y \in \text{SmlSm}/k$ such that $Y - \partial Y \cong X$ by resolution of singularities. Due to [43, Theorem 1.1(2)], we have an equivalence $\omega^* \Sigma^{n,0} \Sigma^\infty X_+ \simeq \Sigma^{n,0} \Sigma^\infty Y_+$ for every integer n . To conclude, observe that ω^* preserves colimits by [43, Theorem 1.1(4)]. \square

Proposition 3.3 *Let k be a perfect field admitting resolution of singularities. For $\mathcal{F} \in \text{SH}(k)$ and integer i , there is a natural equivalence*

$$\omega^* f_i \overline{\mathcal{F}} \simeq f_i \omega^* \overline{\mathcal{F}}$$

in $\text{logSH}(k)$. A similar result holds for the very effective version, too.

Proof We have the natural fiber sequence

$$f_i \omega^* f_i \overline{\mathcal{F}} \rightarrow f_i \omega^* \overline{\mathcal{F}} \rightarrow f_i \omega^* \Gamma^{i-1} \overline{\mathcal{F}}$$

in $\text{logSH}(k)$. Proposition 3.2 implies $f_i \omega^* f_i \overline{\mathcal{F}} \simeq \omega^* f_i \overline{\mathcal{F}}$, and Proposition 3.1 implies $f_i \omega^* \Gamma^{i-1} \overline{\mathcal{F}} \simeq 0$. From these, we obtain the desired equivalence. \square

Let now \mathbf{KGL} denote the motivic \mathbb{P}^1 -spectrum representing algebraic K -theory in $\text{logSH}(k)$ constructed in [17, Definition 6.5.6] (denoted by \mathbf{logKGL} in *loc. cit.*), and let $\mathbf{MZ} \in \text{logSH}(k)$ denote the motivic \mathbb{P}^1 -spectrum representing integral motivic cohomology constructed in [17, Definition 7.7.5], denoted by \mathbf{logMZ} in [*loc. cit.*]). Since \mathbf{KGL} and \mathbf{MZ} satisfy the property that $\mathbf{MZ} \simeq \omega^* \mathbf{MZ}$ when k admits resolution of singularities, and $\mathbf{KGL} \simeq \omega^* \mathbf{KGL}$ in general, we may unambiguously drop the prefix \mathbf{log} from the notation.

Theorem 3.4 *Let k be a perfect field admitting resolution of singularities. Then the slice filtration and the very effective slice filtration on $\mathbf{KGL} \in \text{logSH}(k)$ agree and are complete and exhaustive. Furthermore, there are natural equivalences*

$$s_i \mathbf{KGL} \simeq \widetilde{s}_i \mathbf{KGL} \simeq \mathbf{MZ}(i)[2i]$$

in $\text{logSH}(k)$ for every integer i , where $\mathbf{MZ}(i)[2i]$ denotes the suspension $\Sigma^{2i,i} \mathbf{MZ}$.

Proof We have equivalences

$$s_i \mathbf{KGL} \stackrel{(1)}{\simeq} s_i \omega^* \mathbf{KGL} \stackrel{(2)}{\simeq} \omega^* s_i \mathbf{KGL} \stackrel{(3)}{\simeq} \omega^* \Sigma^{2i,i} \mathbf{MZ} \stackrel{(4)}{\simeq} \Sigma^{2i,i} \mathbf{MZ}$$

in $\text{logSH}(k)$, where (1) is due to [17, Definition 6.5.6], (2) is due to Proposition 3.3, (3) is due to [33, Theorems 6.4.2 and 9.0.3], and (4) is due to [17, Proposition 7.7.6].

We also have $f_i \mathbf{KGL} \simeq \omega^* f_i \mathbf{KGL}$. To show that the slice filtration on $\mathbf{KGL} \in \text{logSH}(k)$ is complete and exhaustive, it suffices to show that the slice filtration on $\mathbf{KGL} \in \text{SH}(k)$ is complete and exhaustive since ω^* preserves colimits and limits. The fact that \mathbf{KGL} is slice complete can be found, for instance, in [45, Lemma 3.11].

For the very effective slice of $\mathbf{KGL} \in \text{logSH}(k)$, use [1, Lemma 2.4] and argue similarly as above. \square

4 HKR filtration on logarithmic Hochschild homology

Throughout this section, we fix (R, P) a pre-log ring, and S a quasicompact quasiseparated scheme over $\text{Spec}(R)$. In practice, the reader can ignore P in most statements and assume that R is always considered with a trivial log structure.

Our goal is to construct the HKR filtration on $\mathbf{HH}(-/R)$ and Beilinson filtrations on $\mathbf{HC}^-(-/R)$ and $\mathbf{HP}(-/R)$ in $\text{logDA}(S)$, which are the \mathbb{P}^1 -stabilized versions of the corresponding filtrations on $\text{HH}(-/R)$, $\text{HC}^-(-/R)$ and $\text{HP}(-/R)$, built, in the log setting, in our previous works [12] and [13]. We also explore their fundamental properties.

Definition 4.1 For a map of pre-log rings $(R, P) \rightarrow (A, M)$ and integer i , we set

$$\mathbb{L}\Omega_{(A,M)/(R,P)}^i := \wedge_A^i \mathbb{L}_{(A,M)/(R,P)},$$

where $\mathbb{L}_{(A,M)/(R,P)}$ is Gabber’s logarithmic cotangent complex [42, Section 8], and $\wedge_A^i(-)$ is the i^{th} derived exterior power. The *logarithmic derived de Rham cohomology* of $(R, P) \rightarrow (A, M)$ is defined as the total complex

$$\mathbb{L}\Omega_{(A,M)/(R,P)} := \text{Tot}(\mathbb{L}\Omega_{(A,M)/(R,P)}^0 \rightarrow \mathbb{L}\Omega_{(A,M)/(R,P)}^1 \rightarrow \dots).$$

This admits the Hodge filtration given by

$$\mathbb{L}\Omega_{(A,M)/(R,P)}^{\geq i} := \text{Tot}(0 \rightarrow \dots \rightarrow 0 \rightarrow \mathbb{L}\Omega_{(A,M)/(R,P)}^i \rightarrow \mathbb{L}\Omega_{(A,M)/(R,P)}^{i+1} \rightarrow \dots),$$

where $\mathbb{L}\Omega_{(A,M)/(R,P)}^i$ is in cohomological degree i for $i \geq 0$ and $\mathbb{L}\Omega_{(A,M)/(R,P)}^{\geq i} := \mathbb{L}\Omega_{(A,M)/(R,P)}$ for $i < 0$. Let $\widehat{\mathbb{L}\Omega}_{(A,M)/(R,P)}$ be the completion of $\mathbb{L}\Omega_{(A,M)/(R,P)}$ with respect to the Hodge filtration, and let $\widehat{\mathbb{L}\Omega}_{X/(R,P)}^{\geq n}$ be the n^{th} term of the induced Hodge filtration of $\widehat{\mathbb{L}\Omega}_{(A,M)/(R,P)}$.

Remark 4.2 By Zariski descent, we can also define the sheaf of complexes $\mathbb{L}\Omega_{X/(R,P)}^i$, $\mathbb{L}\Omega_{X/(R,P)}$, $\mathbb{L}\Omega_{X/(R,P)}^{\geq i}$, $\widehat{\mathbb{L}\Omega}_{X/(R,P)}$ and $\widehat{\mathbb{L}\Omega}_{X/(R,P)}^{\geq i}$ on X_{Zar} for every log scheme X over (R, P) .

We now list some facts that we will use in the paper.

- (1) For any map of pre-log rings $(R, P) \rightarrow (A, M)$, the logarithmic Hochschild homology

$$\mathbf{HH}((A, M)/(R, P))$$

of [12, Definition 5.3] admits a separated, descending filtration with graded pieces

$$\wedge_A^i \mathbb{L}_{(A,M)/(R,P)}[i].$$

This is obtained by Kan extension of the Postnikov filtration in the polynomial case. We shall refer to this as the log HKR filtration; see [12, Theorem 1.1].

(2) The negative cyclic homology and the periodic homology

$$\mathrm{HC}^-((A, M)/(R, P)) \quad \text{and} \quad \mathrm{HP}((A, M)/(R, P))$$

admit the complete exhaustive *Beilinson filtrations*

$$\mathrm{Fil}_{\geq \bullet}^{\mathrm{B}} \mathrm{HC}^-((A, M)/(R, P)) \quad \text{and} \quad \mathrm{Fil}_{\geq \bullet}^{\mathrm{B}} \mathrm{HP}((A, M)/(R, P))$$

with graded pieces $\widehat{\mathbb{L}\Omega}_{(A, M)/(R, P)}^{\geq n}[2n]$ and $\widehat{\mathbb{L}\Omega}_{(A, M)/(R, P)}[2n]$. The non-log case is due to Antieau [4, Theorem 1.1], while the log case is [13, Theorem 1.1]. If $(A, M) \in \mathrm{Poly}_{(R, P)}$, then

$$\begin{aligned} \mathrm{Fil}_i^{\mathrm{B}} \mathrm{HC}^-((A, M)/(R, P)) &:= \tau_{\geq 2i}^{\mathrm{B}} \tau_{\geq \bullet} \mathrm{HC}^-((A, M)/(R, P)), \\ \mathrm{Fil}_i^{\mathrm{B}} \mathrm{HP}((A, M)/(R, P)) &:= \tau_{\geq 2i}^{\mathrm{B}} \tau_{\geq \bullet} \mathrm{HP}((A, M)/(R, P)), \end{aligned}$$

where $\tau_{\geq i}^{\mathrm{B}}$ denotes the truncation functor for the Beilinson t -structure on the filtered derived category $\mathrm{DF}(A)$; see [4, Definition 2.2]. We obtain the filtrations for general (A, M) by left Kan extension and regarding HC^- and HP as bicomplete bifiltered complexes; see [4, Remark 4.4] or [13, Section 2.2].

Construction 4.3 Let X be a log scheme. The inclusion functor from the category of line bundles over \underline{X} to the category of vector bundles over \underline{X} yields the canonical morphism

$$c_1^{\mathrm{K}}: R\Gamma_{\mathrm{Zar}}(\underline{X}, \mathbb{G}_m)[1] \rightarrow \mathrm{K}(\underline{X}).$$

The *first Chern class* for TC is the composite morphism

$$c_1^{\mathrm{TC}}: R\Gamma_{\mathrm{Zar}}(\underline{X}, \mathbb{G}_m)[1] \xrightarrow{c_1^{\mathrm{K}}} \mathrm{K}(\underline{X}) \xrightarrow{\mathrm{Tr}} \mathrm{TC}(\underline{X}) \xrightarrow{p^*} \mathrm{TC}(X),$$

where $p: X \rightarrow \underline{X}$ is the morphism removing the log structure. We similarly obtain the morphisms c_1^{THH} , $c_1^{\mathrm{TC}^-}$, c_1^{TP} , c_1^{HH} , $c_1^{\mathrm{HC}^-}$ and c_1^{HP} to $\mathrm{THH}(X)$, $\mathrm{TC}^-(X)$, $\mathrm{TP}(X)$, $\mathrm{HH}(X)$, $\mathrm{HC}^-(X)$ and $\mathrm{HP}(X)$.

Remark 4.4 We will set the following notation for convenience when dealing with projective bundle formulas below: For a morphism of commutative ring spectra $f: \mathcal{F} \rightarrow \mathcal{G}$ with $c \in \pi_0(\mathcal{G})$, we will often write fc (or simply c if f is clear from the context) for the composite morphism of spectra

$$(4-1) \quad fc: \mathcal{F} \xrightarrow{\simeq} \mathcal{F} \otimes_{\mathbb{S}} \mathbb{S} \xrightarrow{f \otimes c} \mathcal{G} \otimes_{\mathbb{S}} \mathcal{G} \xrightarrow{\mu} \mathcal{G},$$

where μ is the multiplication. We also consider the unit $1 \in \pi_0(\mathcal{G})$.

Recall the projective bundle formula for $\mathrm{K}(\underline{X} \times \mathbb{P}^1)$: The morphism of spectra

$$(1, c_1^{\mathrm{K}}(\mathcal{O}(1))) : \mathrm{K}(\underline{X}) \oplus \mathrm{K}(\underline{X}) \rightarrow \mathrm{K}(\underline{X} \times \mathbb{P}^1)$$

is an equivalence, where $\mathcal{O}(1) \in \mathrm{Pic}(\underline{X} \times \mathbb{P}^1) \cong \pi_0(R\Gamma_{\mathrm{Zar}}(\underline{X}, \mathbb{G}_m)[1])$ and $c_1^{\mathrm{K}}(\mathcal{O}(1)) \in \pi_0(\mathrm{K}(\underline{X} \times \mathbb{P}^1))$ (implicitly we are writing 1 for the pullback map along $\pi: \underline{X} \times \mathbb{P}^1 \rightarrow \underline{X}$). Due to [19, Theorem 1.5], we similarly have the projective bundle formula for $\mathrm{TC}(\underline{X} \times \mathbb{P}^1)$: The morphism of spectra

$$(4-2) \quad (1, c_1^{\mathrm{TC}}(\mathcal{O}(1))) : \mathrm{TC}(\underline{X}) \oplus \mathrm{TC}(\underline{X}) \rightarrow \mathrm{TC}(\underline{X} \times \mathbb{P}^1)$$

is an equivalence. Analogous results hold for THH, TC^- , TP, HH, HC^- and HP.

Proposition 4.5 *There is an S^1 -equivariant equivalence of filtered complexes*

$$(4-3) \quad \mathbb{Z} \oplus (\mathbb{Z}[1])\{-1\} \simeq \mathrm{HH}(\mathbb{P}^1/\mathbb{Z}),$$

where the filtrations on \mathbb{Z} and $\mathbb{Z}[1]$ are the Postnikov filtrations. Here, the notation $(\mathbb{Z}[1])\{-1\}$ means

$$\mathrm{Fil}_i((\mathbb{Z}[1])\{-1\}) = \begin{cases} \mathbb{Z} & \text{if } i \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof Every S^1 -action on $\mathbb{Z} \oplus \mathbb{Z}$ is trivial, so it suffices to show the claim after forgetting the S^1 -actions. We have the equivalences of filtered complexes

$$\begin{aligned} \mathbb{Z}[x] \oplus \mathbb{Z}[x]dx[1] &\simeq \mathrm{HH}(\mathbb{Z}[x]), \\ \mathbb{Z}[x^{-1}] \oplus \mathbb{Z}[x^{-1}]d(x^{-1})[1] &\simeq \mathrm{HH}(\mathbb{Z}[x^{-1}]), \\ \mathbb{Z}[x, x^{-1}] \oplus \mathbb{Z}[x, x^{-1}]dx[1] &\simeq \mathrm{HH}(\mathbb{Z}[x, x^{-1}]), \end{aligned}$$

where the filtrations are the Postnikov filtrations on both sides. Consider the standard cover on \mathbb{P}^1 , and use the equivalences of complexes

$$\mathbb{Z} \simeq \mathrm{fib}(\mathbb{Z}[x] \oplus \mathbb{Z}[x^{-1}] \rightarrow \mathbb{Z}[x, x^{-1}]) \quad \text{and} \quad \mathbb{Z}[-1] \simeq \mathrm{fib}(\mathbb{Z}[x]dx \oplus \mathbb{Z}[x^{-1}]d(x^{-1}) \rightarrow \mathbb{Z}[x, x^{-1}]dx)$$

to obtain the desired equivalence. □

Proposition 4.6 *Let X be a log scheme. Then the morphism of spectra*

$$(4-4) \quad (1, c_1^{\mathrm{TC}}(\mathcal{O}(1))) : \mathrm{TC}(X) \oplus \mathrm{TC}(X) \rightarrow \mathrm{TC}(X \times \mathbb{P}^1)$$

is an equivalence, and this is obtained by applying $\mathrm{TC}(X) \otimes_{\mathbb{Z}} -$ to (4-3). We have similar results for THH, TC^- , TP, HH, HC^- and HP.

Proof We focus on TC since the proofs are similar. By base-change along $\mathrm{TC}(\underline{X}) \rightarrow \mathrm{TC}(X)$ from (4-2) we get (4-4). Hence we reduce to the case when X has a trivial log structure.

By [19, Theorem 9.1 and page 1102], we have the commutative diagram of spectra

$$\begin{array}{ccccc} \mathrm{K}(X)[1] & \xrightarrow{\tau_K} & \mathrm{K}(X \times \mathbb{G}_m) & \xrightarrow{\partial} & \mathrm{K}(X \times \mathbb{P}^1)[1] \\ \mathrm{Tr} \downarrow & & \downarrow \mathrm{Tr} & & \downarrow \mathrm{Tr} \\ \mathrm{TC}(X)[1] & \xrightarrow{\tau_{\mathrm{TC}}} & \mathrm{TC}(X \times \mathbb{G}_m) & \xrightarrow{\partial} & \mathrm{TC}(X \times \mathbb{P}^1)[1] \end{array}$$

where τ_K and τ_{TC} are obtained by the Bass functor structures on K and TC, and the boundary morphisms ∂ are obtained by the standard cover of \mathbb{P}^1 . The composite $\partial \circ \tau_K$ sends the element $1 \in \pi_0\mathrm{K}(X)$ to

$c_1^K(\mathbb{O}(1)) \in \pi_0 \mathbf{K}(X \times \mathbb{P}^1)$, so the composite $\partial \circ \tau_{\text{TC}}$ sends $1 \in \pi_0 \text{TC}(X)$ to $c_1^{\text{TC}}(\mathbb{O}(1)) \in \pi_0 \text{TC}(X \times \mathbb{P}^1)$. It follows that the morphism

$$(1, \partial \circ \tau_{\text{TC}}): \text{TC}(X) \oplus \text{TC}(X) \rightarrow \text{TC}(X \times \mathbb{P}^1)$$

agrees with $(1, c_1^{\text{TC}}(\mathbb{O}(1)))$.

By [19, pages 1102–1103], we see that $\tau_{\text{THH}}: \text{THH}(X)[1] \rightarrow \text{THH}(X \times \mathbb{G}_m)$ is obtained by applying $\text{THH}(X) \otimes_{\mathbb{Z}} -$ to $\tau_{\text{HH}}: \text{HH}(\mathbb{Z})[1] \rightarrow \text{HH}(\mathbb{Z}[x, x^{-1}])$ that sends $1 \in \pi_0 \text{HH}(\mathbb{Z})$ to $dx/x \in \Omega_{\mathbb{Z}[x, x^{-1}]/\mathbb{Z}} \cong \pi_1 \text{HH}(\mathbb{Z}[x, x^{-1}])$. Furthermore, τ_{TC} is obtained by applying TC to τ_{THH} . Compare this with (4-3) to conclude. \square

Proposition 4.7 *Let (R, P) be a pre-log ring. For a quasicompact quasiseparated log scheme X over (R, P) and $i \in \mathbb{Z} \cup \{-\infty\}$, the projective bundle formula for $\text{HH}(X \times \mathbb{P}^1/(R, P))$ restricts to natural equivalences of complexes*

$$\begin{aligned} \text{Fil}_i^{\text{HKR}} \text{HH}(X \times \mathbb{P}^1/(R, P)) &\simeq \text{Fil}_i^{\text{HKR}} \text{HH}(X/(R, P)) \oplus \text{Fil}_{i-1}^{\text{HKR}} \text{HH}(X/(R, P)), \\ \text{Fil}_i^{\text{B}} \text{HC}^-(X \times \mathbb{P}^1/(R, P)) &\simeq \text{Fil}_i^{\text{B}} \text{HC}^-(X/(R, P)) \oplus \text{Fil}_{i-1}^{\text{B}} \text{HC}^-(X/(R, P)), \\ \text{Fil}_i^{\text{B}} \text{HP}(X \times \mathbb{P}^1/(R, P)) &\simeq \text{Fil}_i^{\text{B}} \text{HP}(X/(R, P)) \oplus \text{Fil}_{i-1}^{\text{B}} \text{HP}(X/(R, P)). \end{aligned}$$

Proof We refer to [13, Section 2.2] for the category $\text{Poly}_{(R, P)}$ of polynomial (R, P) -algebras. By Zariski descent and left Kan extension, we reduce to the case when $X := \text{Spec}(A, M)$ with $(A, M) \in \text{Poly}_{(R, P)}$. In this case, the HKR filtration on $\text{HH}(X/(R, P))$ is the Postnikov filtration. Hence we have the S^1 -equivariant equivalence of filtered complexes

$$\text{HH}(X \times \mathbb{P}^1/(R, P)) \simeq \text{HH}(X/(R, P)) \otimes \text{HH}(\mathbb{P}^1/\mathbb{Z}).$$

Together with Proposition 4.5, we obtain the S^1 -equivariant decomposition of filtered complexes

$$(4-5) \quad \text{HH}(X \times \mathbb{P}^1/(R, P)) \simeq \text{HH}(X/(R, P)) \oplus (\text{HH}(X/(R, P))[1])\{-1\},$$

where the filtrations on $\text{HH}(X/(R, P))$ and $\text{HH}(X/(R, P))[1]$ are the Postnikov filtrations. This implies the claim for HH.

By applying $(-)^{hS^1}$ on both sides of (4-5), we obtain an equivalence of complexes induced by the homotopy fixed point spectral sequence:

$$\text{Fil}_{\bullet}^{\text{HKR}} \text{HC}^-(X \times \mathbb{P}^1/(R, P)) \simeq \text{Fil}_{\bullet}^{\text{HKR}} \text{HC}^-(X/(R, P)) \oplus \text{Fil}_{\bullet-1}^{\text{HKR}} \text{HC}^-(X/(R, P)).$$

Take $\tau_{\geq 2i}^{\text{B}}$ on both sides to show the claim for HC^- . The claim for HP can be proven similarly. \square

Construction 4.8 The local-to-global spectral sequence $E_2^{st} = H^s(X, \Omega_X^t) \Rightarrow \pi_{t-s} \text{HH}(X)$ for $X = \mathbb{P}^1$, together with the decomposition of $\text{HH}(\mathbb{P}^1/\mathbb{Z})$ of Proposition 4.5, induces by restriction of $c_1^{\text{HH}}(\mathbb{O}(1)) \in \pi_0 \text{HH}(\mathbb{P}^1)$ a class

$$c_1^{\text{Hod}}(\mathbb{O}(1)) \in H_{\text{Zar}}^1(\mathbb{P}^1, \Omega_{\mathbb{P}^1/\mathbb{Z}}^1),$$

which agrees with the classical first Chern class in Hodge cohomology. After pulling back, we obtain

$$c_1^{\text{Hod}}(\mathbb{O}(1)) \in H_{\text{Zar}}^1(X \times \mathbb{P}^1, \mathbb{L}\Omega_{X/(R,P)}^1)$$

for a log scheme X over a pre-log ring (R, P) . By descent and reduction to the polynomial case, we have the following \mathbb{P}^1 -bundle formulas: the natural morphisms of complexes

$$\begin{aligned} R\Gamma_{\text{Zar}}(X, \mathbb{L}\Omega_{X/(R,P)}^i) \oplus R\Gamma_{\text{Zar}}(X, \mathbb{L}\Omega_{X/(R,P)}^{i-1})[-1] &\rightarrow R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \mathbb{L}\Omega_{X \times \mathbb{P}^1/(R,P)}^i), \\ R\Gamma_{\text{Zar}}(X, \widehat{\mathbb{L}\Omega}_{X/(R,P)}^{\geq i}) \oplus R\Gamma_{\text{Zar}}(X, \widehat{\mathbb{L}\Omega}_{X/(R,P)}^{\geq i-1})[-1] &\rightarrow R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \widehat{\mathbb{L}\Omega}_{X \times \mathbb{P}^1/(R,P)}^{\geq i}), \\ R\Gamma_{\text{Zar}}(X, \widehat{\mathbb{L}\Omega}_{X/(R,P)}) \oplus R\Gamma_{\text{Zar}}(X, \widehat{\mathbb{L}\Omega}_{X/(R,P)})[-1] &\rightarrow R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \widehat{\mathbb{L}\Omega}_{X \times \mathbb{P}^1/(R,P)}), \end{aligned}$$

induced by 1 and $c_1^{\text{Hod}}(\mathbb{O}(1))$ are equivalences; see proof of [11, Proposition 2.20].

If X is quasicompact and quasiseparated, these morphisms are identified with the natural morphisms

$$\begin{aligned} \text{gr}_i^{\text{HKR}}\text{HH}(X/(R, P)) \oplus \text{gr}_{i-1}^{\text{HKR}}\text{HH}(X/(R, P)) &\simeq \text{gr}_i^{\text{HKR}}\text{HH}(X \times \mathbb{P}^1/(R, P)), \\ \text{gr}_i^{\text{B}}\text{HC}^-(X/(R, P)) \oplus \text{gr}_{i-1}^{\text{B}}\text{HC}^-(X/(R, P)) &\simeq \text{gr}_i^{\text{B}}\text{HC}^-(X \times \mathbb{P}^1/(R, P)), \\ \text{gr}_i^{\text{B}}\text{HP}(X/(R, P)) \oplus \text{gr}_{i-1}^{\text{B}}\text{HP}(X/(R, P)) &\simeq \text{gr}_i^{\text{B}}\text{HP}(X \times \mathbb{P}^1/(R, P)), \end{aligned}$$

obtained by Proposition 4.7. Indeed, this can be shown for HH by comparing (4-1) with the composite map

$$\begin{aligned} R\Gamma_{\text{Zar}}(X, \mathbb{L}\Omega_{X/(R,P)}^{i-1}) \otimes_{\mathbb{Z}} \mathbb{Z}[-1] &\rightarrow R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \mathbb{L}\Omega_{X \times \mathbb{P}^1/(R,P)}^{i-1}) \otimes_{\mathbb{Z}} R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \mathbb{L}\Omega_{X \times \mathbb{P}^1/(R,P)}^1) \\ &\rightarrow R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \mathbb{L}\Omega_{X \times \mathbb{P}^1/(R,P)}^i). \end{aligned}$$

A similar argument proves the claim for HC^- and HP .

Proposition 4.9 *Let (R, P) be a pre-log ring. Then the presheaves of complexes*

$$\text{Fil}_i^{\text{HKR}}\text{HH}(-/(R, P)), \quad \text{Fil}_i^{\text{B}}\text{HC}^-(-/(R, P)) \quad \text{and} \quad \text{Fil}_i^{\text{B}}\text{HP}(-/(R, P))$$

on log schemes over (R, P) are $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant for $n > 0$ and $i \in \mathbb{Z} \cup \{-\infty\}$. Moreover, the presheaves of complexes $\mathbb{L}\Omega_{X/(R,P)}^i$, $\widehat{\mathbb{L}\Omega}_{X/(R,P)}^{\geq i}$ and $\widehat{\mathbb{L}\Omega}_{X/(R,P)}$ on the category of log schemes over (R, P) given by

$$X \mapsto R\Gamma_{\text{Zar}}(X, \mathbb{L}\Omega_{X/(R,P)}^i), \quad R\Gamma_{\text{Zar}}(X, \widehat{\mathbb{L}\Omega}_{X/(R,P)}^{\geq i}) \quad \text{and} \quad R\Gamma_{\text{Zar}}(X, \widehat{\mathbb{L}\Omega}_{X/(R,P)})$$

are $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant for $n > 0$ and $i \in \mathbb{Z}$.

Proof This is basically discussed in [12, Section 7, 8], forgetting the filtrations. Let X be a log scheme over (R, P) . Using the completeness and exhaustiveness of the filtrations, it suffices to show that the natural map of complexes

$$\text{gr}_i^{\text{HKR}}\text{HH}(X/(R, P)) \rightarrow \text{gr}_i^{\text{HKR}}\text{HH}(X \times (\mathbb{P}^n, \mathbb{P}^{n-1})/(R, P))$$

and the similar maps for $\text{gr}_i^{\text{B}}\text{HC}^-$ and $\text{gr}_i^{\text{B}}\text{HP}$ are equivalences. Furthermore, we reduce to the case when $X = \text{Spec}(A, M)$ with $(A, M) \in \text{Poly}_{(R, P)}$ by left Kan extension. Then the claim is a consequence of the $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariance of Hodge cohomology; see eg the proof of [16, Proposition 9.2.1]. \square

Definition 4.10 For $X \in \text{SmlSm}/S$ and $i \in \mathbb{Z} \cup \{-\infty\}$, Proposition 4.7 yields a natural equivalence of complexes

$$(4-6) \quad \text{Fil}_i^{\text{HKR}} \mathbf{HH}(X/R) \simeq \Omega_{\mathbb{P}^1} \text{Fil}_{i+1}^{\text{HKR}} \mathbf{HH}(X/R).$$

Using this as bonding maps, Proposition 4.9 enables us to define the complete exhaustive filtration

$$\text{Fil}_i^{\text{HKR}} \mathbf{HH}(-/R) := (\text{Fil}_i^{\text{HKR}} \mathbf{HH}(-/R), \text{Fil}_{i+1}^{\text{HKR}} \mathbf{HH}(-/R), \dots)$$

on $\mathbf{HH}(-/R) \simeq \text{Fil}_{-\infty}^{\text{HKR}} \mathbf{HH}(-/R) \in \text{logDA}(S)$, where $\mathbf{HH}(-/R)$ is the \mathbb{P}^1 -spectrum representing (log) Hochschild homology constructed in [12, Section 8]. We similarly define the complete exhaustive filtrations

$$\text{Fil}_i^{\text{B}} \mathbf{HC}^-(-/R) \quad \text{and} \quad \text{Fil}_i^{\text{B}} \mathbf{HP}(-/R).$$

Proposition 4.11 For $i \in \mathbb{Z} \cup \{-\infty\}$, we have a natural equivalence of \mathbb{P}^1 -spectra

$$\text{Fil}_i^{\text{HKR}} \mathbf{HH}(-/R)(1)[2] \simeq \text{Fil}_{i+1}^{\text{HKR}} \mathbf{HH}(-/R).$$

We have similar equivalences for $\text{Fil}_i^{\text{B}} \mathbf{HC}^-$ and $\text{Fil}_i^{\text{B}} \mathbf{HP}$.

Proof For $\text{Fil}_i^{\text{HKR}} \mathbf{HH}$, this is a direct consequence of (4-6). The proofs for the other cases are similar. \square

Definition 4.12 For $i \in \mathbb{Z}$, Proposition 4.9 enables us to define the \mathbb{P}^1 -spectra

$$\begin{aligned} \mathbb{L}\Omega_{/R}^i &:= (\mathbb{L}\Omega_{/R}^i, \mathbb{L}\Omega_{/R}^{i+1}[1], \dots), \\ \widehat{\mathbb{L}\Omega}_{/R}^{\geq i} &:= (\widehat{\mathbb{L}\Omega}_{/R}^{\geq i}, \widehat{\mathbb{L}\Omega}_{/R}^{\geq i+1}[2], \dots), \\ \widehat{\mathbb{L}\Omega}_{/R} &:= (\widehat{\mathbb{L}\Omega}_{/R}, \widehat{\mathbb{L}\Omega}_{/R}[2], \dots) \end{aligned}$$

in $\text{logDA}(S)$, whose bonding maps are given by the projective bundle formulas in Construction 4.8. If $S \rightarrow \text{Spec}(R)$ is smooth, we omit \mathbb{L} in $\mathbb{L}\Omega$.

Proposition 4.13 For $i \in \mathbb{Z}$, we have natural equivalences

$$\text{gr}_i^{\text{HKR}} \mathbf{HH}(-/R) \simeq \mathbb{L}\Omega_{/R}^i[i], \quad \text{gr}_i^{\text{B}} \mathbf{HC}^-(-/R) \simeq \widehat{\mathbb{L}\Omega}_{/R}^{\geq i}[2i] \quad \text{and} \quad \text{gr}_i^{\text{B}} \mathbf{HP}(-/R) \simeq \widehat{\mathbb{L}\Omega}_{/R}[2i]$$

in $\text{logDA}(S)$.

Proof As observed in Construction 4.8, the bonding maps defining the \mathbb{P}^1 -spectra on both sides are compatible. \square

We are ready to prove our main result on HH and variants.

Theorem 4.14 *Assume that $S \rightarrow \text{Spec}(R)$ is smooth. For $i \in \mathbb{Z}$, we have*

$$(4-7) \quad \text{Fil}_{\text{HKR}}^i \mathbf{HH}(-/R), \quad \text{Fil}_{\mathbf{B}}^i \mathbf{HC}^-(-/R), \quad \text{Fil}_{\mathbf{B}}^i \mathbf{HP}(-/R) \in \log\text{DA}(S)_{\leq i}^{\text{veff}}.$$

In particular, there are canonical morphisms

$$\begin{aligned} \tilde{f}_n \mathbf{HH}(-/R) &\simeq \Sigma^{2n, n} \tilde{f}_0 \mathbf{HH}(-/R) \rightarrow \text{Fil}_n^{\text{HKR}} \mathbf{HH}(-/R), \\ \tilde{f}_n \mathbf{HC}^-(-/R) &\simeq \Sigma^{2n, n} \tilde{f}_0 \mathbf{HC}^-(-/R) \rightarrow \text{Fil}_n^{\mathbf{B}} \mathbf{HC}^-(-/R), \\ \tilde{f}_n \mathbf{HP}(-/R) &\simeq \Sigma^{2n, n} \tilde{f}_0 \mathbf{HP}(-/R) \rightarrow \text{Fil}_n^{\mathbf{B}} \mathbf{HP}(-/R). \end{aligned}$$

Proof We only need to show the first claim, namely, the statement of (4-7), since the remaining claims are a direct consequence of Lemma 2.12. By Proposition 4.11, we reduce to the case when $i = -1$. Since $\Sigma^\infty X_+$ is compact in $\log\text{DA}(S)$, it suffices to show the vanishing

$$\text{Hom}_{\log\text{DA}(S)}(\Sigma^\infty X_+, \text{gr}_{-j}^{\text{HKR}} \mathbf{HH}(-/R)) \simeq 0$$

for every integer $j > 0$, and similar vanishings for $\text{gr}_{-j}^{\mathbf{B}} \mathbf{HC}^-$ and $\text{gr}_{-j}^{\mathbf{B}} \mathbf{HP}$. By the smoothness assumption and Proposition 4.13, it suffices to show the vanishings

$$H_{\text{Zar}}^{-j}(X, \Omega_{X/R}^{-j}) = 0 \quad \text{and} \quad H_{\text{Zar}}^{-2j}(X, \Omega_{X/R}^{\geq 0}) = 0$$

for $j > 0$. This is clear. □

Remark 4.15 Since $H_{\text{Zar}}^{2j}(X, \Omega_{X/R}^{\geq 0})$ does not vanish for $j \geq 0$ and nontrivial X , the proof of Theorem 4.14 shows that $\text{Fil}_{\mathbf{B}}^i \mathbf{HC}^-(-/R)$ and $\text{Fil}_{\mathbf{B}}^i \mathbf{HP}(-/R)$ are not in $\log\text{DA}(S)_{\leq i}^{\text{eff}}$. On the other hand, we have $\text{Fil}_{\text{HKR}}^i \mathbf{HH}(-/R) \in \log\text{DA}_{\leq i}^{\text{eff}}(S)$.

Proof of Theorem 1.5 (1) is a consequence of Lemma 2.12 and Theorems 3.4 and 4.14. (2) is a consequence of (1) and Proposition 4.13, and (3) is a consequence of (2). □

Question 4.16 Despite having a very different origin, Theorem 4.14 exhibits a connection between the (very effective) slice filtration and the HKR filtration (and the Beilinson filtrations). We leave it as an open question whether the (very effective) slice and HKR filtrations agree on $\mathbf{HH}(-/(R, P))$, and whether the slice and Beilinson filtrations agree on $\mathbf{HC}^-(/(R, P))$ and $\mathbf{HP}(-/(R, P))$.

5 BMS filtration on logarithmic topological Hochschild homology

Let us fix a prime p . We refer the reader to [13] for the definitions and the main properties of the category IQSyn of log quasisyntomic rings, the category IQRSPerfd of log quasiregular semiperfectoid rings, and the log quasisyntomic topology on IQSyn .

Definition 5.1 Let \mathfrak{X} be a quasicohherent bounded p -adic formal log scheme. We say that \mathfrak{X} is *log quasisyntomic* if strict étale locally it is isomorphic to $\mathrm{Spf}(A, M)^a$ for $(A, M) \in \mathrm{IQSyn}$. Let FIQSyn denote the category of quasicompact quasiseparated log quasisyntomic formal log schemes, as defined in [11, Definition 3.1]: this is a log variant of the category of qcqs quasisyntomic p -adic formal schemes of [8] (ie qcqs p -adic formal schemes which are locally of the form $\mathrm{Spf}(R)$ for a p -adic quasisyntomic ring R).

Our goal is to define the BMS filtrations on the motivic \mathbb{P}^1 -spectra representing (log) THH, TC^- , TP and TC, defined in [17, Section 8], and explore their fundamental properties.

Proposition 5.2 *The presheaves*

$$\mathrm{THH}(-; \mathbb{Z}_p), \quad \mathrm{THH}(-; \mathbb{Z}_p)_{hS^1}, \quad \mathrm{TC}(-; \mathbb{Z}_p)^- \quad \text{and} \quad \mathrm{TP}(-; \mathbb{Z}_p)$$

on $\mathrm{IQSyn}^{\mathrm{op}}$ are log quasisyntomic sheaves.

Proof Argue as in [37, Theorem 5.5], but use the p -completed version of [13, Theorem 2.3] as in [9, Remark 4.9]. See [13, Proposition 3.10] for the non- p -completed version. \square

In particular, for every $\mathfrak{X} \in \mathrm{FIQSyn}$, we can consider $\mathrm{THH}(\mathfrak{X}; \mathbb{Z}_p)$ and its cousins.

Remark 5.3 Given any p -complete ring R , we can consider the canonical p -completion functor $\mathrm{ISm}/R \rightarrow \mathrm{FISm}/R$ from log smooth log schemes over R to formal log smooth fs log schemes over $\mathrm{Spf}(R)$ as defined in [11, Definition 2.2]. Since for any commutative ring A we have an equivalence $\mathrm{THH}(A_p^\wedge; \mathbb{Z}_p) \simeq \mathrm{THH}(A; \mathbb{Z}_p)$, the functor $\mathrm{THH}(-; \mathbb{Z}_p)$ commutes with the p -completion functor. For this reason, we will implicitly extend the \mathbb{P}^1 -spectrum **THH** in $\mathrm{logSH}(R)$ of [17, Section 8] (denoted by **logTHH** in [loc. cit.]) to a \mathbb{P}^1 -spectrum, indicated with the same letters, in $\mathrm{logFSH}(R)$, where $\mathrm{logFSH}(R)$ is the category of log formal log motives, constructed as in [11, Definition 2.6], using the strict Nisnevich topology and sheaves of spectra. For the same reason, we will implicitly consider the sheaf $\mathrm{THH}(-; \mathbb{Z}_p)$ and variants as defined on (log) schemes or on p -adic formal (log) schemes, without explicitly mentioning it.

Let us quickly review how we can extend the BMS filtrations in [9] to pre-log rings; see [13]. For $(A, M) \in \mathrm{IQRSPerf}$, the *BMS filtrations* on

$$\mathrm{THH}((A, M); \mathbb{Z}_p), \quad \mathrm{TC}^-((A, M); \mathbb{Z}_p), \quad \mathrm{TP}((A, M); \mathbb{Z}_p) \quad \text{and} \quad \mathrm{TC}((A, M); \mathbb{Z}_p)$$

are the double-speed Postnikov filtrations, that is, $\mathrm{Fil}_{\mathrm{BMS}}^i := \tau_{\leq 2i}$. Equivalently, we have $\mathrm{Fil}_i^{\mathrm{BMS}} := \tau_{\geq 2i-1}$. By log quasisyntomic descent, we obtain the *BMS filtrations* for $(A, M) \in \mathrm{IQSyn}$ too. Observe that we have a natural equivalence of spectra

$$(5-1) \quad \mathrm{gr}_i^{\mathrm{BMS}} \mathrm{TC}((A, M); \mathbb{Z}_p) \simeq \mathrm{fib}(\mathrm{gr}_i^{\mathrm{BMS}} \mathrm{TC}^-((A, M); \mathbb{Z}_p) \xrightarrow{\varphi^- \text{-can}} \mathrm{gr}_i^{\mathrm{BMS}} \mathrm{TP}((A, M); \mathbb{Z}_p)).$$

For $\mathfrak{X} \in \text{FIQSyn}$, we also obtain the BMS filtrations on

$$\text{THH}(\mathfrak{X}; \mathbb{Z}_p), \quad \text{TC}^-(\mathfrak{X}; \mathbb{Z}_p), \quad \text{TP}(\mathfrak{X}; \mathbb{Z}_p) \quad \text{and} \quad \text{TC}(\mathfrak{X}; \mathbb{Z}_p)$$

by Zariski descent. The graded pieces satisfy quasisyntomic descent as discussed in [11, Section 3].

For a p -adic formal scheme \mathfrak{X} over $\text{Spf}(R)$ for a quasisyntomic ring R , let $\mathfrak{X} \times \mathbb{P}^n$ denote the p -adic formal scheme $\mathfrak{X} \times (\mathbb{P}_{\text{Spf}(R)}^n)_p^\wedge$, to simplify notation.

Proposition 5.4 *The presheaves*

$$\text{Fil}_i^{\text{BMS}} \text{THH}(-; \mathbb{Z}_p), \quad \text{Fil}_i^{\text{BMS}} \text{TC}^-(-; \mathbb{Z}_p) \quad \text{and} \quad \text{Fil}_i^{\text{BMS}} \text{TP}(-; \mathbb{Z}_p)$$

on FIQSyn are $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant for $n \in \mathbb{Z}_{>0}$ and $i \in \mathbb{Z} \cup \{-\infty\}$.

Proof Using the completeness and exhaustiveness of the filtrations, it suffices to show that the natural map of spectra

$$\text{gr}_i^{\text{BMS}} \text{THH}(\mathfrak{X}; \mathbb{Z}_p) \rightarrow \text{gr}_i^{\text{BMS}} \text{THH}(\mathfrak{X} \times (\mathbb{P}^n, \mathbb{P}^{n-1}); \mathbb{Z}_p)$$

and the similar maps for TC^- , TP and TC are equivalences. This follows from [13, Theorem 1.3(3),(4)] and [11, Corollary 3.11]. □

Proposition 5.5 *For $(A, M) \in \text{IQSyn}$ and integer $i < 0$, we have the vanishing*

$$\text{Fil}_{\text{BMS}}^i \text{TC}((A, M); \mathbb{Z}_p) \simeq 0.$$

Proof By (log) quasisyntomic descent, we reduce to the case $(A, M) \in \text{IQRSPerfd}$. To conclude, observe that we have $\pi_i \text{TC}((A, M); \mathbb{Z}_p) \simeq 0$ for $i < -1$, using [41, Theorem II.4.10] as in the proof of [9, Proposition 7.16]. □

Construction 5.6 Let A be a quasisyntomic ring. Apply the étale sheafification and p -completion functors to the first Chern class $c_1^{\text{TC}}: R\Gamma_{\text{Zar}}(A, \mathbb{G}_m)[1] \rightarrow \text{TC}(A)$ to obtain the natural morphism

$$\widehat{c}_1^{\text{TC}}: R\Gamma_{\text{ét}}(A, \mathbb{G}_m)_p^\wedge[1] \rightarrow \text{TC}(A; \mathbb{Z}_p).$$

Apply $\pi_2(-)[2]$ to this to obtain the morphism

$$T_p(A^\times)[2] \rightarrow (\pi_2 \text{TC}(A; \mathbb{Z}_p))[2],$$

where $T_p(A^\times)$ is the p -adic Tate module of the abelian group A^\times . The étale sheafification of the left-hand side is $R\Gamma_{\text{ét}}(A, \mathbb{G}_m)_p^\wedge[1]$, and the quasisyntomic sheafification of the right-hand side is $R\Gamma_{\text{syn}}(A, \mathbb{Z}_p(1))[2]$ by [9, Proposition 7.17]. Hence we obtain the induced natural morphism

$$\widehat{c}_1^{\text{syn}}: R\Gamma_{\text{ét}}(A, \mathbb{G}_m)_p^\wedge[1] \rightarrow R\Gamma_{\text{syn}}(A, \mathbb{Z}_p(1))[2]$$

for every quasisyntomic ring A . This agrees with $\widehat{c}_1^{\text{syn}}$ in [8, Theorem 7.5.6] up to shift for quasiregular semiperfectoid $\mathbb{Z}_p^{\text{cycl}} = \mathbb{Z}_p[\mu_{p^\infty}]_p^\wedge$ -algebra by comparing [3, Theorem 6.7] and [8, Notation 7.5.1], for

quasiregular semiperfectoid ring by quasisyntomic descent for the cover $A \rightarrow A \widehat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p^{\text{cycl}}$, and for quasisyntomic ring by quasisyntomic descent again.

For a quasiregular semiperfectoid ring A such that A is w -local in the sense of [10, Definition 1.6] (in particular any Zariski cover of $\text{Spec } A$ is split) and A^\times is p -divisible (such rings form a base for the quasisyntomic topology in light of the proof of [9, Proposition 7.17]), we see that $T_p(A^\times)[2] \simeq R\Gamma_{\acute{e}t}(A, \mathbb{G}_m)_p^\wedge[1]$ is concentrated in degree 2. Hence for such a ring A , we have a natural equivalence $\tau_{[1,2]}\widehat{c}_1^{\text{TC}} \simeq \widehat{c}_1^{\text{syn}}$. In particular, we obtain the natural commutative diagram

$$\begin{array}{ccccc} & & T_p(A^\times)[2] & & \\ & \swarrow \widehat{c}_1^{\text{syn}} & \downarrow & \searrow \widehat{c}_1^{\text{TC}} & \\ \tau_{[1,2]}\text{TC}(A; \mathbb{Z}_p) & \longleftarrow & \tau_{\geq 1}\text{TC}(A; \mathbb{Z}_p) & \longrightarrow & \text{TC}(A; \mathbb{Z}_p) \end{array}$$

For $\mathfrak{X} \in \text{FIQSyn}$, we obtain the natural commutative diagram

$$(5-2) \quad \begin{array}{ccccc} & & R\Gamma_{\acute{e}t}(\mathfrak{X}, \mathbb{G}_m)_p^\wedge[1] & & \\ & \swarrow \widehat{c}_1^{\text{syn}} & \downarrow & \searrow \widehat{c}_1^{\text{TC}} & \\ R\Gamma_{\text{syn}}(\mathfrak{X}, \mathbb{Z}_p(1))[2] & \longleftarrow & \text{Fil}_1^{\text{BMS}} \text{TC}(\mathfrak{X}; \mathbb{Z}_p) & \longrightarrow & \text{TC}(\mathfrak{X}; \mathbb{Z}_p) \end{array}$$

by quasisyntomic descent.

For any p -adically complete ring A , write $A\langle x \rangle$ for the p -adic completion of the polynomial ring $A[x]$.

Lemma 5.7 For $(A, M) \in \text{IQRSPerfd}$, the BMS filtrations on

$$\text{THH}((A\langle x \rangle, M); \mathbb{Z}_p) \quad \text{and} \quad \text{THH}((A\langle x, x^{-1} \rangle, M); \mathbb{Z}_p)$$

are the double-speed Postnikov filtrations. Similar results hold for TC^- and TP too.

Proof We focus on the case of $(A\langle x \rangle, M)$, since the proofs are similar. By definition of quasiregular semiperfectoid ring, there exists a perfectoid ring R with a map $R \rightarrow A$. By base-change, we have equivalences of complexes

$$\mathbb{L}_{(A[x], M)/R} \simeq \mathbb{L}_{R[x]/R} \otimes_R^{\mathbb{L}} A \oplus \mathbb{L}_{(A, M)/R} \otimes_R^{\mathbb{L}} R[x] \simeq A[x] \oplus \mathbb{L}_{(A, M)/R} \otimes_R^{\mathbb{L}} R[x].$$

After taking p -completions, we get an equivalence of complexes

$$\mathbb{L}_{(A\langle x \rangle, M)/R} \simeq A\langle x \rangle \oplus \mathbb{L}_{(A, M)/R} \widehat{\otimes}_R^{\mathbb{L}} R\langle x \rangle.$$

Since $\mathbb{L}_{(A, M)/R}$ has p -complete Tor amplitude in degree -1 by for instance [13, Lemma 4.15], and $\wedge_{A\langle x \rangle}^j A\langle x \rangle \simeq 0$ for $j \geq 2$, $\wedge_{A\langle x \rangle}^i \mathbb{L}_{(A\langle x \rangle, M)/R}$ has p -complete Tor amplitude in $[-i, -i + 1]$ for $i \geq 0$. By [13, Proposition 7.3] for THH and [9, Theorem 7.2(3),(4), Proposition 7.8] for TC^- and TP , we see that the i^{th} graded pieces of $\text{THH}((A\langle x \rangle, M); \mathbb{Z}_p)$, $\text{TC}^-((A\langle x \rangle, M); \mathbb{Z}_p)$ and $\text{TP}((A\langle x \rangle, M); \mathbb{Z}_p)$ are in $D^{[-2i, -2i+1]}(A\langle x \rangle)$. Using this, we deduce the claim. \square

The following proposition is crucial for us.

Proposition 5.8 *For $\mathfrak{X} \in \text{FIQSyn}_{\mathbb{R}}$ and $i \in \mathbb{Z} \cup \{-\infty\}$, the projective bundle formula for $\text{THH}(\mathfrak{X} \times \mathbb{P}^1; \mathbb{Z}_p)$ restricts to an equivalence*

$$\text{Fil}_i^{\text{BMS}} \text{THH}(\mathfrak{X} \times \mathbb{P}^1; \mathbb{Z}_p) \simeq \text{Fil}_i^{\text{BMS}} \text{THH}(\mathfrak{X}; \mathbb{Z}_p) \oplus \text{Fil}_{i-1}^{\text{BMS}} \text{THH}(\mathfrak{X}; \mathbb{Z}_p).$$

We have similar equivalences for TC^- , TP and TC too.

Proof By log quasisyntomic descent, we reduce to the case when $\mathfrak{X} = \text{Spf}(A, M)$ and $(A, M) \in \text{IQRSPerfd}$. By Remark 5.3 and Lemma 5.7, the standard cover on \mathbb{P}^1 yields the cartesian square of S^1 -equivariant filtered spectra

$$\begin{array}{ccc} \text{THH}((\mathbb{P}^1_{\text{Spf}(A, M)} \wedge)_p^\wedge; \mathbb{Z}_p) & \longrightarrow & \text{THH}((A\langle x \rangle, M); \mathbb{Z}_p) \\ \downarrow & & \downarrow \\ \text{THH}((A\langle x^{-1} \rangle, M); \mathbb{Z}_p) & \longrightarrow & \text{THH}((A\langle x, x^{-1} \rangle, M); \mathbb{Z}_p) \end{array}$$

such that the filtrations on the three corners other than $\text{THH}((\mathbb{P}^1_{\text{Spf}(A, M)} \wedge)_p^\wedge; \mathbb{Z}_p)$ are the double-speed Postnikov filtrations. Next, we note that we have an S^1 -equivariant equivalence

$$(5-3) \quad \text{THH}((A\langle x \rangle, M); \mathbb{Z}_p) \simeq \text{THH}((A, M); \mathbb{Z}_p) \otimes_{\mathbb{Z}} \text{HH}(\mathbb{Z}[x]),$$

and similarly for $A\langle x^{-1} \rangle$ and $A\langle x, x^{-1} \rangle$ (note that the right-hand side is already p -complete). Hence, the descriptions of $\text{HH}(\mathbb{Z}[x])$, $\text{HH}(\mathbb{Z}[x^{-1}])$ and $\text{HH}(\mathbb{Z}[x, x^{-1}])$ in Proposition 4.5 tell us that the double-speed Postnikov filtrations on these agree with the usual Postnikov filtrations: in other words, the equivalence (5-3) is an equivalence of S^1 -equivariant filtered spectra. Hence we obtain the S^1 -equivariant equivalence of filtered spectra

$$\text{THH}((\mathbb{P}^1_{\text{Spf}(A, M)} \wedge)_p^\wedge; \mathbb{Z}_p) \simeq \text{THH}(X; \mathbb{Z}_p) \otimes_{\mathbb{Z}} \text{HH}(\mathbb{P}^1),$$

where the filtration on $\text{HH}(\mathbb{P}^1)$ is given by Proposition 4.5. We obtain the S^1 -equivariant decomposition

$$(5-4) \quad \text{THH}(\mathfrak{X} \times \mathbb{P}^1; \mathbb{Z}_p) \simeq \text{THH}(\mathfrak{X}; \mathbb{Z}_p) \oplus (\text{THH}(\mathfrak{X}; \mathbb{Z}_p)[1])\{-1\},$$

where the filtration on $\text{THH}(X; \mathbb{Z}_p)$ is the double-speed Postnikov filtration, and

$$\text{Fil}_i(\text{THH}(\mathfrak{X}; \mathbb{Z}_p)[1])\{-1\} := \text{Fil}_{i-1} \text{THH}(\mathfrak{X}; \mathbb{Z}_p).$$

From this, we obtain the desired equivalence for THH . To obtain the desired equivalences for TC^- and TP , apply $(-)^{hS^1}$ and $(-)^{tS^1}$ to (5-4), and use Lemma 5.7. For TC , use (5-1). \square

Construction 5.9 For $\mathfrak{X} \in \text{FIQSyn}_{\mathbb{R}}$, Proposition 5.8 yields the natural equivalences of spectra

$$\begin{aligned} \text{gr}_i^{\text{BMS}} \text{THH}(\mathfrak{X}; \mathbb{Z}_p) \oplus \text{gr}_{i-1}^{\text{BMS}} \text{THH}(\mathfrak{X}; \mathbb{Z}_p) &\simeq \text{gr}_i^{\text{BMS}} \text{THH}(\mathfrak{X} \times \mathbb{P}^1; \mathbb{Z}_p), \\ \text{gr}_i^{\text{BMS}} \text{TC}^-(\mathfrak{X}; \mathbb{Z}_p) \oplus \text{gr}_{i-1}^{\text{BMS}} \text{TC}^-(\mathfrak{X}; \mathbb{Z}_p) &\simeq \text{gr}_i^{\text{BMS}} \text{TC}^-(\mathfrak{X} \times \mathbb{P}^1; \mathbb{Z}_p), \\ \text{gr}_i^{\text{BMS}} \text{TP}(\mathfrak{X}; \mathbb{Z}_p) \oplus \text{gr}_{i-1}^{\text{BMS}} \text{TP}(\mathfrak{X}; \mathbb{Z}_p) &\simeq \text{gr}_i^{\text{BMS}} \text{TP}(\mathfrak{X} \times \mathbb{P}^1; \mathbb{Z}_p), \\ \text{gr}_i^{\text{BMS}} \text{TC}(\mathfrak{X}; \mathbb{Z}_p) \oplus \text{gr}_{i-1}^{\text{BMS}} \text{TC}(\mathfrak{X}; \mathbb{Z}_p) &\simeq \text{gr}_i^{\text{BMS}} \text{TC}(\mathfrak{X} \times \mathbb{P}^1; \mathbb{Z}_p). \end{aligned}$$

As in Construction 4.8, these are identified with the projective bundle formulas for $\mathfrak{X} \times \mathbb{P}^1$

$$\begin{aligned} \mathrm{gr}_i^N R\Gamma_{\widehat{\Delta}}(\mathfrak{X})\{i\} \oplus \mathrm{gr}_{i-1}^N R\Gamma_{\widehat{\Delta}}(\mathfrak{X})\{i-1\}[-2] &\xrightarrow{\simeq} \mathrm{gr}_i^N R\Gamma_{\widehat{\Delta}}(\mathfrak{X} \times \mathbb{P}^1)\{i\}, \\ \mathrm{Fil}_i^N R\Gamma_{\widehat{\Delta}}(\mathfrak{X})\{i\} \oplus \mathrm{Fil}_{i-1}^N R\Gamma_{\widehat{\Delta}}(\mathfrak{X})\{i-1\}[-2] &\xrightarrow{\simeq} \mathrm{Fil}_i^N R\Gamma_{\widehat{\Delta}}(\mathfrak{X} \times \mathbb{P}^1)\{i\}, \\ R\Gamma_{\widehat{\Delta}}(\mathfrak{X})\{i\} \oplus R\Gamma_{\widehat{\Delta}}(\mathfrak{X})\{i-1\}[-2] &\xrightarrow{\simeq} R\Gamma_{\widehat{\Delta}}(\mathfrak{X} \times \mathbb{P}^1)\{i\}, \\ R\Gamma_{\mathrm{syn}}(\mathfrak{X}, \mathbb{Z}_p(i)) \oplus R\Gamma_{\mathrm{syn}}(\mathfrak{X}, \mathbb{Z}_p(i-1))[-2] &\xrightarrow{\simeq} R\Gamma_{\mathrm{syn}}(\mathfrak{X} \times \mathbb{P}^1, \mathbb{Z}_p(i)), \end{aligned}$$

induced by $(1, \widehat{c}_1^{\mathrm{syn}}(\mathcal{O}(1)))$. These are special cases of [8, Lemma 9.1.4(4)–(6)] when \mathfrak{X} has trivial log structure, since the two constructions of $\widehat{c}_1^{\mathrm{syn}}$ discussed in Construction 5.6 agree.

We now fix the spectrum S of a quasisyntomic ring R .

Definition 5.10 For $X \in \mathrm{SmlSm}/S$ and $i \in \mathbb{Z} \cup \{-\infty\}$, Proposition 5.8 yields a natural equivalence of spectra

$$(5-5) \quad \mathrm{Fil}_i^{\mathrm{BMS}} \mathrm{THH}(X; \mathbb{Z}_p) \simeq \Omega_{\mathbb{P}^1} \mathrm{Fil}_{i+1}^{\mathrm{BMS}} \mathrm{THH}(X; \mathbb{Z}_p).$$

Here, we are implicitly composing with the p -completion functor. Note that every $X \in \mathrm{SmlSm}/S$ automatically satisfies the property that $X_p^\wedge \in \mathrm{FIQSyn}$. Using (5-5) as bonding maps, Proposition 5.4 enables us to define the complete exhaustive filtration

$$\mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{THH}(-; \mathbb{Z}_p) := (\mathrm{Fil}_i^{\mathrm{BMS}} \mathrm{THH}(-; \mathbb{Z}_p), \mathrm{Fil}_{i+1}^{\mathrm{BMS}} \mathrm{THH}(-; \mathbb{Z}_p), \dots)$$

on the \mathbb{P}^1 -spectrum $\mathbf{THH}(-; \mathbb{Z}_p) \simeq \mathrm{Fil}_{-\infty}^{\mathrm{BMS}} \mathbf{THH}(-; \mathbb{Z}_p) \in \mathrm{logSH}(S)$. We similarly define the complete exhaustive filtrations

$$\mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{TC}^-(-; \mathbb{Z}_p), \mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{TP}(-; \mathbb{Z}_p), \mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{TC}(-; \mathbb{Z}_p).$$

Proposition 5.11 For $i \in \mathbb{Z} \cup \{-\infty\}$, we have a natural equivalence in $\mathrm{logSH}(S)$

$$\mathrm{Fil}_i^{\mathrm{BMS}} \mathbf{THH}(-; \mathbb{Z}_p)(1)[2] \simeq \mathrm{Fil}_{i+1}^{\mathrm{BMS}} \mathbf{THH}(-; \mathbb{Z}_p).$$

We have similar equivalences for \mathbf{TC}^- , \mathbf{TP} and \mathbf{TC} too.

Proof For \mathbf{THH} , this is a direct consequence of (5-5). The proofs of the other cases are similar. \square

We recall now from [13, Theorem 1.3 and Section 7.4] (see also [11, Section 3]) that the Nygaard-completed absolute log prismatic cohomology $\widehat{\Delta} := R\Gamma_{\mathrm{lqsyn}}(-, \pi_0 \mathrm{TP}(-)_p^\wedge)$ equipped with the Nygaard filtration define functors

$$\widehat{\Delta} \in \mathrm{Fun}(\mathrm{IQSyn}, \mathcal{D}(\mathbb{Z}_p)) \quad \text{and} \quad \mathrm{Fil}_i^N \widehat{\Delta} \in \mathrm{Fun}(\mathrm{IQSyn}, \widehat{\mathcal{DF}}(\mathbb{Z}_p))$$

analogously to [9, Section 7], where $\mathrm{Fil}_i^N \widehat{\Delta}$ denotes the filtered piece $\geq i$ with respect to the Nygaard filtration. By [13, Theorem 2.9], these functors are in fact quasisyntomic sheaves. Define the Breuil–Kisin twists

$$\widehat{\Delta}\{1\} := R\Gamma_{\mathrm{lqsyn}}(-, \pi_2 \mathrm{TP}(-)_p^\wedge)[-2] \quad \text{and} \quad \widehat{\Delta}\{-1\} := R\Gamma_{\mathrm{lqsyn}}(-, \pi_{-2} \mathrm{TP}(-)_p^\wedge)[2],$$

with the Nygaard filtrations given by unfolding the double-speed Postnikov filtration. We get the twisted prismatic cohomology $\widehat{\Delta}\{i\}$ as in [9, Theorem 1.12(3)] by taking tensor powers in $\text{Fun}(\text{IQSyn}, \widehat{\mathcal{DF}}(\mathbb{Z}_p))$, with the induced filtration. In light of [11, Theorems 3.15, 3.16], we can make the following definition:

Definition 5.12 For integers d and i , we define \mathbb{P}^1 -spectra

$$\begin{aligned} \text{Fil}_i^{\mathbb{N}} \widehat{\Delta}\{d\} &:= (\text{Fil}_i^{\mathbb{N}} \widehat{\Delta}\{d\}, \text{Fil}_{i+1}^{\mathbb{N}} \widehat{\Delta}\{d+1\}[2], \dots), \\ \widehat{\Delta}\{d\} &:= (\widehat{\Delta}\{d\}, \widehat{\Delta}\{d+1\}[2], \dots), \\ \mathbf{MZ}_p^{\text{syn}}(d) &:= \text{fib}(\varphi_p - \text{can}: \text{Fil}_d^{\mathbb{N}} \widehat{\Delta}\{d\} \rightarrow \widehat{\Delta}\{d\}), \end{aligned}$$

where the bonding maps for $\text{Fil}_i^{\mathbb{N}} \widehat{\Delta}\{d\}$, $\widehat{\Delta}\{d\}$ and $\mathbf{MZ}_p^{\text{syn}}(d)$ are obtained by the projective bundle formulas in Construction 5.9, and the morphisms φ_p and can are the pointwise cyclotomic trace and canonical morphisms. We will omit $\{d\}$ and (d) if $d = 0$.

Note that in [11, Theorems 3.15, 3.16], (formal variants of) $\widehat{\Delta}$, $\text{Fil}_0^{\mathbb{N}} \widehat{\Delta}$ and $\mathbf{MZ}_p^{\text{syn}}$ were denoted by $\mathbf{E}^{\widehat{\Delta}}$, $\mathbf{E}^{\text{Fil } \widehat{\Delta}}$ and \mathbf{E}^{Fsyn} .

Proposition 5.13 For $i \in \mathbb{Z}$, we have natural equivalences

$$\begin{aligned} \text{gr}_i^{\text{BMS}} \mathbf{THH}(-; \mathbb{Z}_p) &\simeq \text{gr}_i^{\mathbb{N}} \widehat{\Delta}\{i\}[2i], & \text{gr}_i^{\text{BMS}} \mathbf{TC}^-(-; \mathbb{Z}_p) &\simeq \text{Fil}_i^{\mathbb{N}} \widehat{\Delta}\{i\}[2i], \\ \text{gr}_i^{\text{BMS}} \mathbf{TP}(-; \mathbb{Z}_p) &\simeq \widehat{\Delta}\{i\}[2i], & \text{gr}_i^{\text{BMS}} \mathbf{TC}(-; \mathbb{Z}_p) &\simeq \mathbf{MZ}_p^{\text{syn}}(i)[2i], \end{aligned}$$

in $\text{logSH}(S)$.

Proof By [13, Theorem 1.3(3),(4)], it remains to show that the bonding maps for the left-hand sides can be identified with the bonding maps for the right-hand sides. This is a consequence of Construction 5.9. \square

Remark 5.14 The right-hand side of the displayed equivalences in Proposition 5.13 are naturally constructed as objects of $\text{logDA}(S, \mathbb{Z}_p)$, while the filtered spectra $\text{Fil}_{i+1}^{\text{BMS}} \mathbf{THH}(-; \mathbb{Z}_p)$ and variants are in $\text{logSH}(S, \mathbb{S}_p)$, where \mathbb{S}_p denotes the p -completed sphere spectrum. In particular, we obtain that each graded piece gr_i^{BMS} is in $\text{logDA}(S, \mathbb{Z}_p)$.

The Tate twist is related to the Breuil–Kisin twist in the following sense:

Proposition 5.15 For $d \in \mathbb{Z}$ and $i \in \mathbb{Z} \cup \{-\infty\}$, we have a natural equivalence

$$\text{Fil}_i^{\mathbb{N}} \widehat{\Delta}(d) \simeq \text{Fil}_{i+d}^{\mathbb{N}} \widehat{\Delta}\{d\}$$

in $\text{logSH}(S)$.

Proof We have

$$\text{Fil}_i^{\mathbb{N}} \widehat{\Delta} := (\text{Fil}_i^{\mathbb{N}} \widehat{\Delta}, \text{Fil}_{i+1}^{\mathbb{N}} \widehat{\Delta}\{1\}[2], \dots).$$

Tensoring with $(d)[2d]$ in $\text{logSH}(S)$ means shifting d terms, so we have

$$\text{Fil}_i^{\mathbb{N}} \widehat{\Delta}(d)[2d] \simeq (\text{Fil}_{i+d}^{\mathbb{N}} \widehat{\Delta}\{d\}[2d], \text{Fil}_{i+d+1}^{\mathbb{N}} \widehat{\Delta}\{d+1\}[2d+1], \dots).$$

We obtain the desired equivalence after applying $[-2d]$. \square

If $i = -\infty$, observe that Proposition 5.15 yields an equivalence

$$\widehat{\Delta}(d) \simeq \widehat{\Delta}\{d\}$$

in $\log\mathrm{SH}(S)$ for $d \in \mathbb{Z}$.

The notation $\mathbf{MZ}_p^{\mathrm{syn}}(i)$ is compatible with the Tate twist in the following sense:

Proposition 5.16 For $d, i \in \mathbb{Z}$, we have a natural equivalence

$$(\mathbf{MZ}_p^{\mathrm{syn}}(i))(d) \simeq \mathbf{MZ}_p^{\mathrm{syn}}(i + d)$$

in $\log\mathrm{SH}(S)$.

Proof This is a direct consequence of Proposition 5.15. □

Theorem 5.17 For $i \in \mathbb{Z}$,

$$\mathrm{Fil}_{\mathrm{BMS}}^i \mathbf{THH}(-; \mathbb{Z}_p), \quad \mathrm{Fil}_{\mathrm{BMS}}^i \mathbf{TC}^-(; \mathbb{Z}_p), \quad \mathrm{Fil}_{\mathrm{BMS}}^i \mathbf{TP}(-; \mathbb{Z}_p), \quad \mathrm{Fil}_{\mathrm{BMS}}^i \mathbf{TC}(-; \mathbb{Z}_p)$$

are in $\log\mathrm{SH}(S)_{\leq i}^{\mathrm{veff}}$. In particular, there are canonical morphisms

$$\tilde{\mathfrak{f}}_n \mathbf{THH}(-; \mathbb{Z}_p) \simeq \Sigma^{2n, n} \tilde{\mathfrak{f}}_0 \mathbf{THH}(-; \mathbb{Z}_p) \rightarrow \mathrm{Fil}_n^{\mathrm{BMS}} \mathbf{THH}(-; \mathbb{Z}_p),$$

and similarly for \mathbf{TC}^- , \mathbf{TP} and \mathbf{TC} .

Proof By Proposition 5.11, we reduce to the case when $i = -1$. It suffices to show the vanishing

$$\mathrm{Hom}_{\log\mathrm{SH}(S)}(\Sigma^\infty X_+, \mathrm{Fil}_{\mathrm{BMS}}^{-1} \mathbf{THH}(-; \mathbb{Z}_p)) = 0,$$

and similar vanishings for \mathbf{TC}^- , \mathbf{TP} and \mathbf{TC} . Since $\mathrm{Fil}_{\mathrm{BMS}}^{-1} \mathbf{THH}(-; \mathbb{Z}_p) = 0$ by construction and $\mathrm{Fil}_{\mathrm{BMS}}^{-1} \mathbf{TC}(-; \mathbb{Z}_p) = 0$ by Proposition 5.5, the claim holds for \mathbf{THH} and \mathbf{TC} . Using the compactness of $\Sigma^\infty X_+$ in $\log\mathrm{SH}(S)$, it suffices to show the vanishing

$$\mathrm{Hom}_{\log\mathrm{SH}(S)}(\Sigma^\infty X_+, \mathrm{gr}_{-j}^{\mathrm{BMS}} \mathbf{TC}^-(; \mathbb{Z}_p)) = 0$$

for every integer $j > 0$, and similar vanishing for \mathbf{TP} . By Proposition 5.13, it suffices to show the vanishings

$$H_{\mathrm{sNis}}^{-2j}(X, \widehat{\Delta}\{-j\}) = 0.$$

This follows from $\widehat{\Delta}_{(A, M)}\{-j\} \in \mathbf{D}^{\geq 0}(A)$ for $(A, M) \in \mathrm{IQSyn}$, which can be deduced from the case of $(A, M) \in \mathrm{IQRSPerfd}$ by log quasisyntomic descent. □

Question 5.18 As for the HKR filtration, we leave it as an open question whether the slice and BMS filtrations agree on $\mathbf{THH}(-; \mathbb{Z}_p)$, $\mathbf{TC}^-(; \mathbb{Z}_p)$, $\mathbf{TP}(-; \mathbb{Z}_p)$ and $\mathbf{TC}(-; \mathbb{Z}_p)$. Since T is the analog of S^2 , the double-speed convergence of the very effective slice filtration on \mathbf{KGL} gives at least an analogy with the double-speed Postnikov filtration defining the BMS filtration.

6 The characteristic- p case

Throughout this section, k is a perfect field of characteristic p . Our goal is to identify the graded pieces of the filtered spectra $\mathbf{THH}(-; \mathbb{Z}_p)$, $\mathbf{TC}^-(-; \mathbb{Z}_p)$, $\mathbf{TP}(-; \mathbb{Z}_p)$ and $\mathbf{TC}(-; \mathbb{Z}_p)$ in this case. We will use [9] for the identifications at the level of S^1 -spectra, and then we will use the projective bundle formulas in [8] to identify the bonding maps.

Construction 6.1 For a log smooth saturated pre-log k -algebra (A, M) , recall from [11, Section 4.3] that the chain complex $W\Omega_{(A,M)/k}$ comes equipped with the Nygaard filtration $\text{Fil}_N^i W\Omega_{(A,M)/k}$ given by the subcomplex

$$p^{i-1} V W(A) \rightarrow p^{i-2} V W\Omega_{(A,M)/k}^1 \rightarrow \dots \rightarrow p V W\Omega_{(A,M)/k}^{i-2} \rightarrow V W\Omega_{(A,M)/k}^{i-1} \rightarrow W\Omega_{(A,M)/k}^i \rightarrow W\Omega_{(A,M)/k}^{i+1} \rightarrow \dots$$

For $X \in \text{ISm}/k$, the log crystalline cohomology $R\Gamma_{\text{crys}}(X) := R\Gamma_{\text{Zar}}(X, W\Omega_{X/k})$ is then equipped with the Nygaard filtration $\text{Fil}_i^N R\Gamma_{\text{crys}}$ by Zariski descent. Furthermore, we obtain a natural equivalence of complexes

$$(6-1) \quad \text{gr}_i^N R\Gamma_{\text{crys}}(X) \simeq R\Gamma_{\text{Zar}}(X, \tau^{\leq i} \Omega_{X/k})$$

from [11, (4.22.1) and (4.22.2)].

By [11, Theorem 4.30], we have a natural equivalence of filtered \mathbb{E}_∞ -rings

$$R\Gamma_{\widehat{\Delta}}(X) \simeq R\Gamma_{\text{crys}}(X).$$

Construction 6.2 For $X \in \text{ISm}/k$, recall from [8, Construction 7.3.1] the morphism of complexes

$$c_1^{\text{crys}}: R\Gamma_{\text{ét}}(\underline{X}, \mathbb{G}_m)[1] \rightarrow \text{Fil}_1^N R\Gamma_{\text{crys}}(\underline{X})[2].$$

Compose this with the canonical morphism

$$\text{Fil}_1^N R\Gamma_{\text{crys}}(\underline{X})[2] \rightarrow \text{Fil}_1^N R\Gamma_{\text{crys}}(X)[2]$$

to obtain

$$c_1^{\text{crys}}: R\Gamma_{\text{ét}}(\underline{X}, \mathbb{G}_m)[1] \rightarrow \text{Fil}_1^N R\Gamma_{\text{crys}}(X)[2].$$

As a consequence of [8, Proposition 7.5.5], we see that the triangle

$$(6-2) \quad \begin{array}{ccc} & R\Gamma_{\text{ét}}(\underline{X}, \mathbb{G}_m)[1] & \\ c_1^{\widehat{\Delta}} \swarrow & & \searrow c_1^{\text{crys}} \\ \text{Fil}_1^N R\Gamma_{\widehat{\Delta}}(X)[2] & \xrightarrow{\simeq} & \text{Fil}_1^N R\Gamma_{\text{crys}}(X)[2] \end{array}$$

commutes.

Construction 6.3 For $X \in \text{ISm}/k$, recall from [32, Theorem 4.12] that we have the inverse Cartier operator, which is a natural isomorphism of graded \mathbb{O}_X -algebras

$$C_{X/k}^{-1} : \Omega_{X/k}^* \xrightarrow{\cong} \mathcal{H}^*(\Omega_{X/k})$$

sending $d \log x$ to $d \log x$ for every local section x of \mathcal{M}_X . Using this, we have the composite morphism

$$c_1^{\text{Hod}} : R\Gamma_{\text{Zar}}(X, \mathbb{G}_m)[1] \xrightarrow{c_1^{\text{Hod}}} R\Gamma_{\text{Zar}}(X, \Omega_{X/k}^1)[1] \xrightarrow{C_{X/k}^{-1}} R\Gamma_{\text{Zar}}(X, \mathcal{H}^1(\Omega_{X/k}))[1],$$

and we have the projective bundle formulas for \mathbb{P}^1

$$\begin{aligned} R\Gamma_{\text{Zar}}(X, \mathcal{H}^i(\Omega_{X/k})) \oplus R\Gamma_{\text{Zar}}(X, \mathcal{H}^{i-1}(\Omega_{X/k}))[-1] &\xrightarrow{\cong} R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \mathcal{H}^i(\Omega_{X/k})), \\ R\Gamma_{\text{Zar}}(X, \tau^{\leq i} \Omega_{X/k}) \oplus R\Gamma_{\text{Zar}}(X, \tau^{\leq i-1} \Omega_{X/k})[-1] &\xrightarrow{\cong} R\Gamma_{\text{Zar}}(X \times \mathbb{P}^1, \tau^{\leq i} \Omega_{X/k}), \end{aligned}$$

induced by $(1, c_1^{\text{Hod}}(\mathbb{O}(1)))$. Using (6-1), the last equivalence can be identified with

$$\text{gr}_i^{\text{N}} R\Gamma_{\text{crys}}(X) \oplus \text{gr}_{i-1}^{\text{N}} R\Gamma_{\text{crys}}(X)[-1] \xrightarrow{\cong} \text{gr}_i^{\text{N}} R\Gamma_{\text{crys}}(X \times \mathbb{P}^1).$$

Proposition 6.4 For every $X \in \text{ISm}/k$ and integer i , the natural square

$$\begin{array}{ccc} \text{Fil}_i^{\text{N}} R\Gamma_{\text{crys}}(X) \oplus \text{Fil}_{i-1}^{\text{N}} R\Gamma_{\text{crys}}(X)[-1] & \xrightarrow{(1, c_1^{\text{crys}}(\mathbb{O}(1)))} & \text{Fil}_i^{\text{N}} R\Gamma_{\text{crys}}(X \times \mathbb{P}^1) \\ \downarrow & & \downarrow \\ \text{gr}_i^{\text{N}} R\Gamma_{\text{crys}}(X) \oplus \text{gr}_{i-1}^{\text{N}} R\Gamma_{\text{crys}}(X)[-1] & \xrightarrow{(1, c_1^{\text{Hod}}(\mathbb{O}(1)))} & \text{gr}_i^{\text{N}} R\Gamma_{\text{crys}}(X \times \mathbb{P}^1) \end{array}$$

commutes.

Proof It suffices to show that the image of

$$c_1^{\text{crys}}(\mathbb{O}(1)) \in H^2(\text{Fil}_1^{\text{N}} R\Gamma_{\text{crys}}(X \times \mathbb{P}^1))$$

in $H^2(\text{gr}_1^{\text{N}} R\Gamma_{\text{crys}}(X \times \mathbb{P}^1))$ agrees with $c_1^{\text{Hod}}(\mathbb{O}(1))$ since these induce the projective bundle formulas for \mathbb{P}^1 . For this, we reduce to the case when $X = \text{Spec}(k)$, since the general case can be shown by considering the pullbacks of the first Chern classes.

We set $A := k[x, x^{-1}]$ for simplicity of notation. Since $\mathbb{O}(1) \in R\Gamma_{\text{Zar}}(\mathbb{P}_k^1, \mathbb{G}_m)$ is the image of $x^{-1} \in A^*$ under the boundary map $A^* \cong R\Gamma_{\text{Zar}}(\mathbb{G}_{m,k}, \mathbb{G}_m) \rightarrow R\Gamma_{\text{Zar}}(\mathbb{P}_k^1, \mathbb{G}_m)$, it suffices to show that the image of $c_1^{\text{crys}}(x^{-1}) \in H^1(\text{Fil}_1^{\text{N}} W\Omega_{A/k})$ in $H^1(\text{gr}_1^{\text{N}} W\Omega_{A/k})$ agrees with $c_1^{\text{Hod}}(x^{-1})$.

Consider the canonical morphism $W\Omega_{A/k} \rightarrow \Omega_{A/k}$, which induces $\text{Fil}_1^{\text{N}} W\Omega_{A/k} \rightarrow \text{Fil}_1^{\text{Hod}} \Omega_{A/k}$ and hence $\gamma^{\text{dR}} : \text{gr}_1^{\text{N}} W\Omega_{A/k} \rightarrow \text{gr}_1^{\text{Hod}} \Omega_{A/k}$. The equivalence $\text{gr}_1^{\text{N}} W\Omega_{A/k} \simeq \tau^{\leq 1} \Omega_{A/k}$ is explicitly given by the isomorphism of chain complexes

$$\begin{array}{ccc} VW(A)/pVW(A) & \xrightarrow{F/p} & A \\ d \downarrow & & \downarrow d \\ W\Omega_{A/k}^1/VW\Omega_{A/k}^1 & \xrightarrow{F} & Z\Omega_{A/k}^1 \end{array}$$

where the lower horizontal map is due to [29, (I.3.11.3)]. Together with [29, Proposition I.3.3], we see that γ^{dR} is given by the composite morphism

$$\tau^{\leq 1} \Omega_{A/k} \rightarrow H^1(\Omega_{A/k})[-1] \xrightarrow{C_{X/k}} \Omega_{A/k}^1.$$

In particular, $H^1(\gamma^{\text{dR}})$ is an isomorphism. So to compare $c_1^{\text{crys}}(x^{-1})$ and $c_1^{\text{Hod}}(x^{-1})$ in $H^1(\text{gr}_1^{\text{N}} W \Omega_{A/k})$, it suffices to compare their images in $H^1(\text{gr}_1^{\text{Hod}} \Omega_{A/k})$.

By [8, Example 5.2.4, Theorem 7.6.2], the image of $c_1^{\text{crys}}(x^{-1})$ in $H^1(\text{gr}_1^{\text{Hod}} \Omega_{A/k})$ is identified with $c_1^{\text{Hod}}(x^{-1}) = d \log x \in \Omega_{A/k}^1$. To conclude, observe that $H^1(\gamma^{\text{dR}}) = C_{X/k}: H^1(\Omega_{A/k}) \rightarrow \Omega_{A/k}$ sends $c_1^{\text{Hod}}(x^{-1}) = d \log x$ to $d \log x$. □

Definition 6.5 The *strict pro-étale topology* on the category of log schemes is the topology generated by the families $\{X_i \rightarrow X\}_{i \in I}$ such that $\{\underline{X}_i \rightarrow \underline{X}\}_{i \in I}$ is a pro-étale covering. Let sproét be the shorthand for this topology. We denote by $L_{\text{sproét}}(-)$ the strict pro-étale sheafification functor.

Recall the following construction from [34].

Definition 6.6 Let X be a fine log scheme over k and let $d \log: \mathcal{M}_X^{\text{gp}} \rightarrow \Omega_{X/k}^1$ be the natural map. For every integer $r \geq 1$, write $d \log[\cdot]: (\mathcal{M}_X^{\text{gp}})^{\otimes i} \rightarrow W_r \Omega_{X/k}^i$ for the map of strict étale sheaves locally given by $m_1 \otimes \cdots \otimes m_i \rightarrow d \log[m_1] \wedge \cdots \wedge d \log[m_i]$, where $[m]$ denotes the lift of m as in [34, page 256]. We denote by $W_r \Omega_{X/k, \log}^i \subseteq W_r \Omega_{X/k}^i$ the strict étale subsheaf generated by the image of $d \log[\cdot]$, and by $W \Omega_{X/k, \log}^i$ the limit $\lim_r W_r \Omega_{X/k, \log}^i$ as a strict pro-étale sheaf.

Proposition 6.7 For $X \in \text{SmlSm}/k$ and integer i , the sequence of chain complexes of strict pro-étale sheaves

$$0 \rightarrow W \Omega_{X/k, \log}^i[-i] \rightarrow \text{Fil}_i^{\text{N}} W \Omega_{X/k}^\bullet \xrightarrow{\varphi/p^{i-1}} W \Omega_{X/k}^\bullet \rightarrow 0$$

is exact in each degree. Furthermore, we have a natural equivalence of complexes of strict pro-étale sheaves $W \Omega_{X/k, \log}^i \simeq R \lim_m W_m \Omega_{X/k, \log}^i$ (that is, $W \Omega_{X/k, \log}^i$ is also the derived inverse limit).

Proof This is formally identical to [9, Proposition 8.4], using [34, Corollary 2.14] in the place of [29, Theorem I.5.7.2]. □

Remark 6.8 Recall the following fact. For any fs log scheme X and any quasicoherent sheaf \mathcal{F} on \underline{X} , we have $R\Gamma_{\text{sét}}(X, \mathcal{F}) \simeq R\Gamma_{\text{két}}(X, \mathcal{F})$ (this is due to Kato, see eg [16, Proposition 9.2.3]), and this agrees with $R\Gamma_{\text{Zar}}(X, \mathcal{F})$ and $R\Gamma_{\text{sNis}}(X, \mathcal{F})$ by [16, (9.1.1)]. Moreover, we have an equivalence

$$R\Gamma(X, L_{\text{ét}} \mathcal{F}) \simeq \text{colim}_{Y \in X_{\text{div}}} R\Gamma_{\text{két}}(Y, \mathcal{F})$$

by [16, Theorem 5.1.2]. Hence if \mathcal{G} is a complex of bounded below $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant coherent sheaves on SmlSm/k , then \mathcal{G} is a complex of log étale hypersheaves.

For $X \in \text{ISm}/k$ and integer $i \geq 0$, we set

$$B\Omega_{X/k}^i := \text{im}(\Omega_{X/k}^{i-1} \xrightarrow{d} \Omega_{X/k}^i) \quad \text{and} \quad Z\Omega_{X/k}^i := \ker(\Omega_{X/k}^i \xrightarrow{d} \Omega_{X/k}^{i+1}).$$

The following result is a consequence of [39, Theorem 4.2] and Remark 6.8 above, using, for example, the fact that the above presheaves restricted to smooth k -schemes are reciprocity sheaves; see [18, Section 11]. We also provide another direct proof.

Proposition 6.9 *The presheaves of complexes*

$$\Omega^i, \quad B\Omega^i, \quad Z\Omega^i, \quad \tau^{\leq i}\Omega, \quad W_m\Omega, \quad W\Omega$$

on SmlSm/k given by

$$X \mapsto R\Gamma_{\text{Zar}}(X, \Omega_{X/k}^i), R\Gamma_{\text{Zar}}(X, B\Omega_{X/k}^i), R\Gamma_{\text{Zar}}(X, Z\Omega_{X/k}^i), R\Gamma_{\text{Zar}}(X, \tau^{\leq i}\Omega_{X/k}), \\ R\Gamma_{\text{Zar}}(X, W_m\Omega_{X/k}), R\Gamma_{\text{Zar}}(X, W\Omega_{X/k}),$$

are $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and log étale hypersheaves for all integers $m, n > 0$ and i .

Proof By [16, Corollary 9.2.2, Proposition 9.2.6], Ω^i is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and a log étale hypersheaf. We proceed by induction to show that $B\Omega^i$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and a log étale hypersheaf. The claim is clear if $i = 0$ since $B\Omega_{X/k}^0 = 0$ for $X \in \text{SmlSm}/k$. If the claim holds for i , then use the obvious exact sequence

$$0 \rightarrow B\Omega_{X/k}^i \rightarrow \Omega_{X/k}^i \rightarrow Z\Omega_{X/k}^{i+1} \rightarrow 0$$

and the exact sequence

$$0 \rightarrow B\Omega_{X/k}^{i+1} \rightarrow Z\Omega_{X/k}^{i+1} \rightarrow \Omega_{X/k}^{i+1} \rightarrow 0$$

obtained by [28, (2.12.2)] to show the claim for $i + 1$. We also deduce that $Z\Omega^i$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and a log étale hypersheaf. It follows that $\tau^{\leq i}\Omega$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and a log étale hypersheaf.

The claim for Ω^i for all i implies the claim for Ω . By induction on m , we deduce the claim for $W_m\Omega$. After taking limits over m , we deduce the claim for $W\Omega$. \square

We note that $R\Gamma_{\text{Zar}}(X, W\Omega_{X/k})$ represents Hyodo–Kato cohomology with the trivial log structure on k ; see Lorenzon [34, Corollary 1.23].

Proposition 6.10 *The presheaves of complexes*

$$L_{\text{sét}}W_m\Omega_{\log}^i, \quad L_{\text{sproét}}W\Omega_{\log}^i$$

on SmlSm/k given by

$$X \mapsto R\Gamma_{\text{sét}}(X, W_m\Omega_{X/k, \log}^i), R\Gamma_{\text{sproét}}(X, W\Omega_{X/k, \log}^i)$$

are $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and log étale hypersheaves for all integers $m, n > 0$ and i .

Proof Consider the exact sequence of strict étale sheaves

$$0 \rightarrow \Omega_{X/k,\log}^i \rightarrow \Omega_{X/k}^i \rightarrow \Omega_{X/k}^i / B\Omega_{X/k}^i \rightarrow 0$$

obtained by [34, Proposition 2.13]. By [34, (2.12)], Remark 6.8 and induction on m , we see that $L_{\text{sét}} W_m \Omega_{\log}^i$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and a log étale hypersheaf. Together with Proposition 6.7, we deduce that $L_{\text{sproét}} W \Omega_{\log}^i$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant and a log étale hypersheaf. \square

Construction 6.11 For every $X \in \text{ISm}/k$ and integer i , we have a natural equivalence of complexes

$$(6-3) \quad R\Gamma_{\text{syn}}(X, \mathbb{Z}_p(i)) \simeq R\Gamma_{\text{sproét}}(X, W\Omega_{X/k,\log}^i)$$

by [11, Theorem 4.30] and Proposition 6.7, and see [11, Theorem 4.27] for the φ -compatibility. Together with [8, Theorem 7.3.5], we see that $c_1^{\text{crys}} : R\Gamma_{\text{ét}}(X, \mathbb{G}_m)[1] \rightarrow R\Gamma_{\text{crys}}(X)[2]$ naturally factors through a morphism of complexes

$$c_1^{W\Omega_{\log}} : R\Gamma_{\text{ét}}(X, \mathbb{G}_m)[1] \rightarrow R\Gamma_{\text{sproét}}(X, W\Omega_{X/k,\log}^1)[1].$$

As a consequence of [8, Proposition 7.5.5], we see that the triangle

$$(6-4) \quad \begin{array}{ccc} & R\Gamma_{\text{ét}}(\underline{X}, \mathbb{G}_m)[1] & \\ c_1^{\text{syn}} \swarrow & & \searrow c_1^{W\Omega_{\log}} \\ R\Gamma_{\text{syn}}(X)[2] & \xrightarrow{\simeq} & R\Gamma_{\text{sproét}}(X, W\Omega_{X/k,\log}^1)[2] \end{array}$$

commutes.

Construction 6.12 For $X \in \text{ISm}/k$, recall from [8, Construction 7.3.1] the morphism of complexes

$$c_1^{\text{syn}} : R\Gamma_{\text{ét}}(\underline{X}, \mathbb{G}_m)[1] \rightarrow \text{Fil}_1^{\mathbb{N}} R\Gamma_{\text{crys}}(\underline{X})[2].$$

Compose this with the canonical morphism $\text{Fil}_1^{\mathbb{N}} R\Gamma_{\text{crys}}(\underline{X})[2] \rightarrow \text{Fil}_1^{\mathbb{N}} R\Gamma_{\text{crys}}(X)[2]$ to obtain

$$c_1^{\text{crys}} : R\Gamma_{\text{ét}}(\underline{X}, \mathbb{G}_m)[1] \rightarrow \text{Fil}_1^{\mathbb{N}} R\Gamma_{\text{crys}}(X)[2].$$

Definition 6.13 By Proposition 6.9, we can define the objects

$$\begin{aligned} \text{Fil}_i^{\text{conj}} \Omega &:= (\tau_{\geq i} \Omega, (\tau_{\geq i+1} \Omega)[1], \dots), \\ \text{Fil}_i^{\mathbb{N}} W\Omega &:= (\text{Fil}_i^{\mathbb{N}} W\Omega, (\text{Fil}_{i+1}^{\mathbb{N}} W\Omega)[1], \dots), \\ L_{\text{sproét}} W\Omega_{\log}^i &:= (L_{\text{sproét}} W\Omega_{\log}^i, L_{\text{sproét}} W\Omega_{\log}^{i+1}[1], \dots) \end{aligned}$$

in $\text{logSH}(k)$, where the bonding maps are given by the composite morphisms

$$\begin{aligned} \tau_{\geq i} \Omega &\xrightarrow{c_1^{\text{Hod}}} \tau_{\geq i}(\Omega_{\mathbb{P}^1} \Omega[1]) \xrightarrow{\simeq} \Omega_{\mathbb{P}^1}(\tau_{\geq i+1} \Omega)[1], \\ \text{Fil}_i^{\mathbb{N}} W\Omega &\xrightarrow{c_1^{\text{crys}}} \text{Fil}_i^{\mathbb{N}}(\Omega_{\mathbb{P}^1} W\Omega[1]) \xrightarrow{\simeq} \Omega_{\mathbb{P}^1}(\text{Fil}_{i+1}^{\mathbb{N}} W\Omega)[1], \\ L_{\text{sproét}} W\Omega_{\log}^i &\xrightarrow{c_1^{W\Omega_{\log}}} L_{\text{sproét}} \Omega_{\mathbb{P}^1} W\Omega_{\log}^{i+1}[1] \xrightarrow{\simeq} \Omega_{\mathbb{P}^1} L_{\text{sproét}} W\Omega_{\log}^{i+1}[1]. \end{aligned}$$

Proposition 6.14 For $i \in \mathbb{Z}$, we have natural equivalences

$$\mathrm{gr}_i^{\mathrm{N}} \widehat{\Delta} \simeq \mathrm{Fil}_i^{\mathrm{conj}} \Omega, \quad \mathrm{Fil}_i^{\mathrm{N}} \widehat{\Delta} \simeq \mathrm{Fil}_i^{\mathrm{N}} \mathbf{W}\Omega, \quad \widehat{\Delta} \simeq \mathbf{W}\Omega, \quad \mathbf{M}\mathbb{Z}_p^{\mathrm{syn}}(i) \simeq L_{\mathrm{sproét}} \mathbf{W}\Omega_{\log}^i[-i]$$

in $\mathrm{logSH}(k)$.

Proof By [11, Theorem 4.30, Proposition 5.1] and (6-3), it suffices to show that the bonding maps on both sides are compatible. This follows from Proposition 6.4 and the commutativity of (6-2) and (6-4). \square

Proof of Theorem 1.4 (1) is a consequence of Lemma 2.12 and Theorems 3.4 and 5.17. (2) is a consequence of (1) and Propositions 5.13 and 6.14, and (3) is a consequence of (2). \square

7 Very effective slices of Kummer étale K-theory

Let k be a perfect field. In this section, we construct a morphism in $\mathrm{logSH}_{\mathrm{két}}^{\wedge}(k)$ (the hypercomplete version of $\mathrm{logSH}_{\mathrm{két}}(k)$) representing a morphism from the Lichtenbaum étale cohomology to the syntomic cohomology induced by the cyclotomic trace. We also discuss the very effective slices of Kummer étale K-theory.

By Proposition 6.10, the functors $L_{\mathrm{sproét}} \mathbf{W}\Omega_{\log}^i$ are log étale hypersheaves, and represent p -adic log syntomic cohomology. Hence we obtain the morphism $L_{\mathrm{két}} \mathbf{M}\mathbb{Z} \rightarrow \mathbf{M}\mathbb{Z}_p^{\mathrm{syn}}$ in $\mathrm{logSH}_{\mathrm{két}}^{\wedge}(k)$ by adjunction from the morphism $\mathbf{M}\mathbb{Z} \rightarrow \mathbf{M}\mathbb{Z}_p^{\mathrm{syn}}$ in $\mathrm{logSH}(k)$ induced by the motivic cyclotomic trace map. We will show that $L_{\mathrm{két}} \mathbf{M}\mathbb{Z}$ represents the Lichtenbaum étale motivic cohomology.

For every integer i and a commutative ring R , let $R(i)$ be the Nisnevich sheaf of complexes on Sm/k given by the motivic cohomology

$$X \mapsto R\Gamma_{\mathrm{mot}}(X, R(i)).$$

Consider again the functor

$$\omega^* : \mathrm{Sh}_{\mathrm{Nis}}(\mathrm{Sm}/S, \mathrm{Sp}) \rightarrow \mathrm{Sh}_{\mathrm{sNis}}(\mathrm{SmlSm}/S, \mathrm{Sp})$$

such that $\omega^* \mathcal{F}(X) := \mathcal{F}(X - \partial X)$ for $\mathcal{F} \in \mathrm{Sh}_{\mathrm{Nis}}(\mathrm{Sm}/k, \mathrm{Sp})$ and $X \in \mathrm{SmlSm}/k$.

Proposition 7.1 The hypersheaf of complexes $L_{\mathrm{két}} \omega^* \mathbb{Z}(i)$ on SmlSm/k is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant for every integer $n > 0$ and i .

Proof For every complex \mathcal{F} , we have the arithmetic fracture square

$$\begin{array}{ccc} \mathcal{F} & \longrightarrow & \prod_{\ell} \mathcal{F}_{\ell}^{\wedge} \\ \downarrow & & \downarrow \\ \mathcal{F}_{\mathbb{Q}} & \longrightarrow & \prod_{\ell} (\mathcal{F}_{\ell}^{\wedge})_{\mathbb{Q}} \end{array}$$

that is cartesian, where ℓ runs over primes, $(-)^{\wedge}_{\ell}$ is the ℓ -adic completion, and $(-)_{\mathbb{Q}}$ is the rationalization.

Use this square for

$$\mathcal{F} := \text{fib}(L_{\text{két}}\omega^*\mathbb{Z}(i)(X) \rightarrow L_{\text{két}}\omega^*\mathbb{Z}(i)(X \times (\mathbb{P}^n, \mathbb{P}^{n-1})))$$

to reduce to showing that the presheaves of complexes $L_{\text{két}}\omega^*\mathbb{Q}(i)$ and $(L_{\text{két}}\omega^*\mathbb{Z}(i))_\ell^\wedge$ are $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant.

If $X_\bullet \rightarrow X$ is a Kummer étale hypercover, then $X_\bullet - \partial X_\bullet \rightarrow X - \partial X$ is an étale hypercover. By [38, Theorem 14.30], we have an equivalence $\text{colim}_{i \in \Delta} M(X_i - \partial X_i) \simeq M(X - \partial X)$ in the ∞ -category of Voevodsky’s motives $\text{DM}^{\text{eff}}(k, \mathbb{Q})$. Apply $\text{map}_{\text{DM}^{\text{eff}}(k, \mathbb{Q})}(-, \mathbb{Q}(i)[2i])$ to this equivalence to deduce that $X \in \text{SmlSm}/k \mapsto R\Gamma_{\text{mot}}(X - \partial X, \mathbb{Q}(i))$ is a Kummer étale hypersheaf. It follows that we have an equivalence of presheaves of complexes with rational coefficients $L_{\text{két}}\omega^*\mathbb{Q}(i) \simeq L_{\text{Snis}}\omega^*\mathbb{Q}(i)$. The latter is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant by [16, Proposition 8.1.12]. Hence it remains to show that $(L_{\text{két}}\omega^*\mathbb{Z}(i))_\ell^\wedge$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant. Since $(L_{\text{két}}\omega^*\mathbb{Z}(i))_\ell^\wedge \simeq \lim_d L_{\text{két}}\omega^*\mathbb{Z}/\ell^d(i)$, it suffices to show that $L_{\text{két}}\omega^*\mathbb{Z}/\ell^d(i)$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant. By induction and using the exact sequence of abelian groups

$$0 \rightarrow \mathbb{Z}/\ell \rightarrow \mathbb{Z}/\ell^n \rightarrow \mathbb{Z}/\ell^{n-1} \rightarrow 0,$$

we reduce to showing that $L_{\text{két}}\omega^*\mathbb{Z}/\ell(i)$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant.

Assume $\ell \neq p$. We have an equivalence

$$L_{\text{két}}\omega^*\mathbb{Z}/\ell(i) \simeq L_{\text{két}}\omega^*\mu_\ell^{\otimes i}$$

by [38, Theorem 10.2], since the Kummer étale topology is finer than the strict étale topology. Hence $L_{\text{két}}\omega^*\mathbb{Z}/\ell(i)(X)$ is equivalent to the Kummer étale cohomology $R\Gamma_{\text{két}}(X, \mathbb{Z}/\ell(i))$. This is \square -invariant and a log étale hypersheaf by [17, Theorem 9.1.5 and Proposition 9.1.6], so this is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant by [17, Section 3.4].

Assume $\ell = p$. Then we have an equivalence of Nisnevich sheaves $\mathbb{Z}/p^n(i) \simeq W_n\Omega_{\log}^i[-i]$ on Sm/k by Geisser and Levine [26, Theorem 8.5]. Hence we have an equivalence of Kummer étale sheaves

$$(7-1) \quad L_{\text{két}}\omega^*\mathbb{Z}/p^n(i) \simeq L_{\text{két}}W_n\Omega_{\log}^i[-i]$$

on SmlSm/k , since $\omega^*W_n\Omega_{\log}^i \simeq W_n\Omega_{\log}^i$ by the definition of logarithmic forms. This is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant by Proposition 6.10. \square

Definition 7.2 For $X \in \text{SmlSm}/k$, the *Kummer étale motivic cohomology* $R\Gamma_L(X, \mathbb{Z}(i))$ is the Kummer étale hypersheafification of $X \mapsto R\Gamma_{\text{mot}}(X, \mathbb{Z}(i)) := R\Gamma_{\text{mot}}(X - \partial X, \mathbb{Z}(i))$, ie

$$R\Gamma_L(X, \mathbb{Z}(i)) := (L_{\text{két}}\omega^*\mathbb{Z}(i))(X).$$

For every integer i , Proposition 7.1 yields a natural equivalence

$$(7-2) \quad \text{map}_{\log\text{SH}_{\text{két}}^\wedge(k)}(\Sigma^\infty X_+, \Sigma^{0,i}L_{\text{két}}\mathbf{MZ}) \simeq R\Gamma_L(X, \mathbb{Z}(i)),$$

where $\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)$ is the Kummer étale hyperlocalization of $\log\mathrm{SH}(k)$, and $L_{\mathrm{k\acute{e}t}}: \log\mathrm{SH}(k) \rightarrow \log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)$ is the localization functor. The *Kummer étale K-theory* $L_{\mathrm{k\acute{e}t}}\mathbf{K}(X)$ is the Kummer étale hypersheafification of $X \mapsto \mathbf{K}(X) := \mathbf{K}(X - \partial X)$.

Assume that X has a trivial log structure. Then the small Kummer étale site $X_{\mathrm{k\acute{e}t}}$ and small étale site $X_{\acute{e}t}$ agree. Hence $R\Gamma_{\mathrm{L}}(X, \mathbb{Z}(i))$ agrees with the Lichtenbaum étale motivic cohomology, and $L_{\mathrm{k\acute{e}t}}\mathbf{K}(X)$ agrees with the étale K -theory of X .

We also consider the localization functor $L_{\mathrm{s\acute{e}t}}: \log\mathrm{SH}(k) \rightarrow \log\mathrm{SH}_{\mathrm{s\acute{e}t}}^{\wedge}(k)$.

Definition 7.3 For a topology τ on SmlSm/k , let $\log\mathrm{SH}_{\tau}(k)_p^{\wedge}$ be the full subcategory of $\log\mathrm{SH}_{\tau}(k)$ spanned by p -complete objects, and let $(\log\mathrm{SH}_{\tau}(k)_p^{\wedge})^{\mathrm{eff}}$ be the full stable ∞ -subcategory of $\log\mathrm{SH}_{\tau}(k)_p^{\wedge}$ generated under colimits by $(\Sigma^{n,0}\Sigma_{\mathbb{P}^1}^{\infty} X_+)_p^{\wedge}$ for $X \in \mathrm{SmlSm}/k$ and $n \in \mathbb{Z}$. We define the slice filtration on $\log\mathrm{SH}_{\tau}(k)_p^{\wedge}$ as we did on $\log\mathrm{SH}_{\tau}(k)$.

Proposition 7.4 For every integer i , there are equivalences in $\log\mathrm{SH}(k)_p^{\wedge}$

$$(L_{\mathrm{s\acute{e}t}}\mathbf{MZ}(i))_p^{\wedge} \simeq (L_{\mathrm{k\acute{e}t}}\mathbf{MZ}(i))_p^{\wedge} \simeq \mathbf{MZ}_p^{\mathrm{syn}}(i).$$

Proof We have an equivalence in $\log\mathrm{SH}(k)$

$$\mathbf{MZ}_p^{\mathrm{syn}}(i) \simeq \lim_m (L_{\mathrm{k\acute{e}t}}\mathbf{W}_m\mathbf{\Omega}_{\log}^i[-i])$$

by Proposition 6.7. The right-hand side is equivalent to $(L_{\mathrm{k\acute{e}t}}\mathbf{MZ}(i))_p^{\wedge}$ by (7-1). To show $(L_{\mathrm{s\acute{e}t}}\mathbf{MZ}(i))_p^{\wedge} \simeq (L_{\mathrm{k\acute{e}t}}\mathbf{MZ}(i))_p^{\wedge}$, use Proposition 6.10 and (7-1). \square

Theorem 7.5 Let k be a perfect field admitting resolution of singularities. Then the induced morphism

$$\tilde{f}_i^{\mathrm{k\acute{e}t}}\mathbf{TC}(-; \mathbb{Z}_p) \rightarrow \mathrm{Fil}_i^{\mathrm{BMS}}\mathbf{TC}(-; \mathbb{Z}_p)$$

is an equivalence in $\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_p^{\wedge}$ for every integer i . Here, we denote by $\tilde{f}_i^{\mathrm{k\acute{e}t}}$ the i^{th} very effective cover in $\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_p^{\wedge}$. We have a similar result for $\tilde{f}_i^{\mathrm{s\acute{e}t}}$ and $\log\mathrm{SH}_{\mathrm{s\acute{e}t}}^{\wedge}(k)_p^{\wedge}$ too.

Proof We focus on the proof for the Kummer étale case since the proof for the strict étale case proceeds verbatim. By Lemma 2.10 and Theorem 3.4, we have $L_{\mathrm{k\acute{e}t}}\mathbf{MZ}(i)[2i] \in \log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_{\geq i}^{\mathrm{veff}}$. Since the p -completion functor $\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k) \rightarrow \log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_p^{\wedge}$ is a left adjoint, we can show that this preserves (very) effectiveness arguing as in Lemma 2.10. Hence we have $\mathrm{gr}_i^{\mathrm{BMS}}\mathbf{TC}(-; \mathbb{Z}_p) \in (\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_p^{\wedge})_{\geq i}^{\mathrm{veff}}$ using Propositions 5.13 and 7.4. For $j \geq i$, we have

$$\mathrm{fib}(\mathrm{Fil}_j^{\mathrm{BMS}}\mathbf{TC}(-; \mathbb{Z}_p) \rightarrow \mathrm{Fil}_{j-1}^{\mathrm{BMS}}\mathbf{TC}(-; \mathbb{Z}_p)) \in (\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_p^{\wedge})_{\geq i}^{\mathrm{veff}}$$

by induction on j . Take colimits on j to have $\mathrm{Fil}_i^{\mathrm{BMS}}\mathbf{TC}(-; \mathbb{Z}_p) \in (\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_p^{\wedge})_{\geq i}^{\mathrm{veff}}$. On the other hand, Theorem 5.17 implies $\mathrm{Fil}_{\mathrm{BMS}}^{i-1}\mathbf{TC}(-; \mathbb{Z}_p) \in (\log\mathrm{SH}_{\mathrm{k\acute{e}t}}^{\wedge}(k)_p^{\wedge})_{\leq i-1}^{\mathrm{veff}}$. It follows that

$$\tilde{f}^{i-1}\mathrm{Fil}_{\mathrm{BMS}}^{i-1}\mathbf{TC}(-; \mathbb{Z}_p) \simeq \mathrm{Fil}_{\mathrm{BMS}}^{i-1}\mathbf{TC}(-; \mathbb{Z}_p),$$

and hence $\tilde{f}_i \text{Fil}_{\text{BMS}}^{i-1} \mathbf{TC}(-; \mathbb{Z}_p) \simeq 0$. Now after applying \tilde{f}_i to the fiber sequence

$$\text{Fil}_i^{\text{BMS}} \mathbf{TC}(-; \mathbb{Z}_p) \rightarrow \mathbf{TC}(-; \mathbb{Z}_p) \rightarrow \text{Fil}_{\text{BMS}}^{i-1} \mathbf{TC}(-; \mathbb{Z}_p),$$

we obtain an equivalence $\text{Fil}_i^{\text{BMS}} \mathbf{TC}(-; \mathbb{Z}_p) \simeq \tilde{f}_i \mathbf{TC}(-; \mathbb{Z}_p)$. □

Remark 7.6 (1) In the proof of Theorem 7.5, we have used in particular the fact that $(L_{\text{két}} \mathbf{MZ}(i))_p^\wedge$ is very effective in the category $\log\text{SH}_{\text{két}}^\wedge(k)_p^\wedge$. Since the inclusion $\log\text{SH}_{\text{két}}^\wedge(k)_p^\wedge \subset \log\text{SH}_{\text{két}}^\wedge(k)$ does not obviously preserve the very effective subcategory, we are not claiming that the same holds when we see $(L_{\text{két}} \mathbf{MZ}(i))_p^\wedge$ as object in $\log\text{SH}_{\text{két}}^\wedge(k)$.

(2) One could reasonably ask if the canonical morphism constructed in Theorem 5.17 becomes an equivalence for k a field (say, assuming resolution of singularities) after applying the canonical functor from $\log\text{SH}(k, \mathbb{Z}_p)$ to $\log\text{SH}_{\text{két}}^\wedge(k)_p^\wedge$, and if this equivalence coincides with the one provided by Theorem 7.5. Equivalently, one could ask if the p -completion of the Kummer étale very effective slice filtration for $\mathbf{TC}(-; \mathbb{Z}_p)$ coincides with the Kummer étale very effective slice filtration computed in $\log\text{SH}_{\text{két}}^\wedge(k)_p^\wedge$. We expect this to be the case.

Theorem 7.7 *Let k be a perfect field admitting resolution of singularities, and assume that the étale cohomological dimension of k is finite.*

(1) *The filtrations*

$$f_{\bullet}^{\text{két}} L_{\text{két}} \mathbf{KGL}, \quad \tilde{f}_{\bullet}^{\text{két}} L_{\text{két}} \mathbf{KGL} \quad \text{and} \quad L_{\text{két}} \tilde{f}_{\bullet} \mathbf{KGL}$$

on $L_{\text{két}} \mathbf{KGL} \in \log\text{SH}_{\text{két}}^\wedge(k)$ agree, and are complete and exhaustive.

(2) *There are natural equivalences*

$$s_i^{\text{két}} L_{\text{két}} \mathbf{KGL} \simeq \tilde{s}_i^{\text{két}} L_{\text{két}} \mathbf{KGL} \simeq L_{\text{két}} s_i \mathbf{KGL} \simeq L_{\text{két}} \tilde{s}_i \mathbf{KGL} \simeq \Sigma^{2i,i} L_{\text{két}} \mathbf{MZ}$$

in $\log\text{SH}_{\text{két}}^\wedge(k)$.

(3) *For $X \in \text{SmlSm}/k$, there is a natural equivalence*

$$\text{map}_{\log\text{SH}_{\text{két}}^\wedge(k)}(\Sigma^\infty X_+, L_{\text{két}} \mathbf{KGL}) \simeq L_{\text{két}} \mathbf{K}(X).$$

(4) *For $X \in \text{SmlSm}/k$, there are natural equivalences*

$$\text{map}_{\log\text{SH}_{\text{két}}^\wedge(k, \mathbb{S}/p^n)}(\Sigma^\infty X_+, L_{\text{két}} \mathbf{KGL} \otimes_{\mathbb{S}} \mathbb{S}/p^n) \simeq L_{\text{két}} \mathbf{K}(X) \otimes_{\mathbb{S}} \mathbb{S}/p^n,$$

$$\text{map}_{\log\text{SH}_{\text{két}}^\wedge(k)_p^\wedge}(\Sigma^\infty X_+, (L_{\text{két}} \mathbf{KGL})_p^\wedge) \simeq (L_{\text{két}} \mathbf{K}(X))_p^\wedge,$$

without the assumption on the étale cohomological dimension.

Proof As pointed out by Spitzweck and Østvær [49, Proposition 5.11], the connectivity of $\tilde{f}_n E$ of any motivic spectrum increases with n . Hence the very effective slice filtration is automatically complete.

By Proposition 7.1, we have

$$\text{map}_{\log\text{SH}_{\text{két}}^{\wedge}(k)}(\Sigma^{\infty} X_+, \Sigma^{0,i} L_{\text{két}} \mathbf{MZ}) \simeq 0$$

for $X \in \text{SmlSm}/k$ and $i < 0$. By [16, Corollary 5.1.7] and [17, Proposition 2.6.8], $\Sigma^{\infty} X_+$ is compact. Since $L_{\text{két}}$ preserve colimits, Theorem 3.4 implies that the filtration $L_{\text{két}} \tilde{f}_{\bullet} \mathbf{KGL}$ is exhaustive. It follows that we have

$$\text{map}_{\log\text{SH}_{\text{két}}^{\wedge}(k)}(\Sigma^{\infty} X_+, L_{\text{két}} \tilde{f}^{-1} \mathbf{KGL}) \simeq 0.$$

This implies $L_{\text{két}} \tilde{f}^{-1} \mathbf{KGL} \in \log\text{SH}_{\text{két}}^{\wedge}(k)_{\leq -1}^{\text{eff}}$. On the other hand, $L_{\text{két}} \tilde{f}_0 \mathbf{KGL} \in \log\text{SH}_{\text{két}}^{\wedge}(k)_{\geq 0}^{\text{veff}}$ follows from Lemma 2.10. From these, we have natural equivalences

$$f_0 L_{\text{két}} \mathbf{KGL} \simeq \tilde{f}_0 L_{\text{két}} \mathbf{KGL} \simeq L_{\text{két}} \tilde{f}_0 \mathbf{KGL}.$$

Apply $\Sigma^{2i,i}$ to this and use Theorem 3.4 to finish the proof of (1).

(2) is an immediate consequence of (1).

Consider the localization functor

$$L_{\text{két}} : \text{Sp}_{\mathbb{P}^1}(\text{Sh}_{\text{SmlSm}}(\text{SmlSm}/k, \text{Sp})) \rightarrow \text{Sp}_{\mathbb{P}^1}(\text{Sh}_{\text{két}}(\text{SmlSm}/k, \text{Sp}))$$

without the further $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -localizations. To show (3), we only need to show that $L_{\text{két}} \mathbf{KGL}$ is $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant for every integer n . The above argument shows that the filtration $L_{\text{két}} \tilde{f}_{\geq \bullet} \mathbf{KGL}$ on $L_{\text{két}} \mathbf{KGL}$ is complete and exhaustive since this argument does not require $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -localizations. Hence it suffices to show that the graded pieces $L_{\text{két}} \tilde{f}_i \mathbf{KGL}$ are $(\mathbb{P}^n, \mathbb{P}^{n-1})$ -invariant. This is a consequence of Theorem 3.4 and Proposition 7.1.

Argue similarly as above and use \lim_n for (4), but note that the étale \mathbb{Z}/p^n -cohomological dimension (=étale \mathbb{S}/p^n -cohomological dimension) of k is always finite by [6, Théorème X.5.1]. \square

Remark 7.8 Even though $\mathbf{KGL} \in \log\text{SH}(k)$ is \mathbb{A}^1 -invariant, $L_{\text{két}} \mathbf{KGL} \in \log\text{SH}_{\text{két}}^{\wedge}(k)$ is not \mathbb{A}^1 -invariant in the above case of k with $\text{char } k > 0$ since $R\Gamma_L$ is not \mathbb{A}^1 -invariant, as observed in [38, Lecture 10].

Construction 7.9 Let k be a perfect field admitting resolution of singularities. By Proposition 6.10 and (6-3), log syntomic cohomology on SmlSm/k is a Kummer étale hypersheaf. It follows that the log motivic cyclotomic trace map

$$\text{Tr} : \mathbf{KGL} \rightarrow \mathbf{TC}$$

in $\log\text{SH}(k)$ yields the map

$$L_{\text{két}} \text{Tr} : L_{\text{két}} \mathbf{KGL} \rightarrow \mathbf{TC}$$

in $\log\text{SH}_{\text{két}}^{\wedge}(k)$.

If we apply $\text{map}_{\log\text{SH}_{\text{két}}^{\wedge}(k)}(\Sigma^{\infty} X_+, -)$ to $L_{\text{két}} \text{Tr}$, then we get

$$L_{\text{két}} \mathbf{K}(X) \rightarrow \mathbf{TC}(X)$$

by Theorem 7.7(3) when the étale cohomological dimension of k is finite, which is the Kummer étale hypersheafification of the log cyclotomic trace map $\mathbf{K}(X) \rightarrow \mathbf{TC}(X)$. Similarly, if we apply $\text{map}_{\log\text{SH}_{\text{két}}^\wedge(k)_p}(\Sigma^\infty X_+, -)$ to $(L_{\text{két}}\text{Tr})_p^\wedge$, then we get

$$(L_{\text{két}}\mathbf{K}(X))_p^\wedge \rightarrow \mathbf{TC}(X; \mathbb{Z}_p)$$

by Theorem 7.7(4).

As an application of Kummer étale K -theory, we show the following:

Proposition 7.10 *Let k be a perfect field admitting resolution of singularities. For $X \in \text{SmlSm}/k$, there is a natural equivalence*

$$(L_{\text{két}}\mathbf{K}(X))_p^\wedge \simeq \mathbf{TC}(X; \mathbb{Z}_p).$$

Proof Assume first that $X \in \text{Sm}/k$. Then the equivalence $(L_{\text{ét}}\mathbf{K}(X))_p^\wedge \simeq \mathbf{TC}(X; \mathbb{Z}_p)$ is due to Geisser and Hesselholt [25, Theorem 4.2.6]. In general, the desired equivalence for $X \in \text{SmlSm}/k$ follows from the Gysin sequence [17, Theorem 3.2.19] by induction on the number of smooth irreducible components r of ∂X . More precisely, if Z is a smooth divisor on proper $X \in \text{Sm}/k$, then we have the fiber sequence in $\log\text{SH}(k)$

$$(7-3) \quad \Sigma^\infty(X, Z)_+ \rightarrow \Sigma^\infty X_+ \rightarrow \text{Th}(\mathbf{N}_Z X),$$

where $\text{Th}(\mathbf{N}_Z X)$ is the motivic Thom space defined in [17, Definition 3.2.4]. Note that by applying the natural functor $\omega_\sharp: \log\text{SH}(k) \rightarrow \text{SH}(k)$, left adjoint to ω^* , the above sequence reduces to the standard localization sequence in $\text{SH}(k)$. In particular, applying the log K -theory spectrum \mathbf{KGL} , we obtain the localization sequence in algebraic K -theory in light of [17, Theorem 6.5.7]. On the other hand, if we apply the log \mathbf{TC} spectrum \mathbf{TC} to (7-3) we obtain by definition the Gysin or residue sequence in logarithmic topological cyclic homology. By Construction 7.9, we obtain a commutative diagram of spectra

$$\begin{array}{ccccc} (L_{\text{két}}\mathbf{K}(Z))_p^\wedge & \longrightarrow & (L_{\text{két}}\mathbf{K}(X))_p^\wedge & \longrightarrow & (L_{\text{két}}\mathbf{K}(X, Z))_p^\wedge \\ \downarrow & & \downarrow & & \downarrow \\ \mathbf{TC}(Z; \mathbb{Z}_p) & \longrightarrow & \mathbf{TC}(X; \mathbb{Z}_p) & \longrightarrow & \mathbf{TC}((X, Z); \mathbb{Z}_p) \end{array}$$

whose horizontal sequences are fiber sequences. The general case follows similarly by induction on r . \square

Remark 7.11 Assume that the boundary ∂X of $X \in \text{SmlSm}/k$ has r irreducible components. For integers $m \geq 1$ and $n \leq d - r$, the induced natural map

$$\pi_n(R\Gamma_{\text{mot}}(X, \mathbb{Z}/p^m(d))) \rightarrow \pi_n(R\Gamma_{\text{L}}(X, \mathbb{Z}/p^m(d)))$$

is an isomorphism. Indeed, if $r = 0$, then this is a consequence of [26, Theorem 8.5]. If $r > 0$, then proceed by induction on r , and use the Gysin sequence [17, Theorem 3.2.19] and the five lemma. See [23, Corollary 1.1] for a related result with \mathbb{Q}_p -coefficients and the inequality $n \leq d$.

Remark 7.12 The fact that p -adic étale K -theory is identified with topological cyclic homology holds in larger generality (namely, without smoothness assumption) in light of [20, Theorem C]. However, this does not immediately imply a generalization of Proposition 7.10.

References

- [1] **A Ananyevskiy, O Röndigs, P A Østvær**, *On very effective hermitian K -theory*, Math. Z. 294 (2020) 1021–1034 MR
- [2] **T Annala, R Iwasa**, *Motivic spectra and universality of K -theory*, preprint (2022) arXiv 2204.03434
- [3] **J Anschütz, A-C Le Bras**, *The p -completed cyclotomic trace in degree 2*, Ann. K-Theory 5 (2020) 539–580 MR
- [4] **B Antieau**, *Periodic cyclic homology and derived de Rham cohomology*, Ann. K-Theory 4 (2019) 505–519 MR
- [5] **B Antieau, A Mathew, M Morrow, T Nikolaus**, *On the Beilinson fiber square*, Duke Math. J. 171 (2022) 3707–3806 MR
- [6] **M Artin, A Grothendieck, J L Verdier**, *Théorie des topos et cohomologie étale des schémas, Tome 3: Exposés IX–XIX (SGA 4₃)*, Lecture Notes in Math. 305, Springer (1973) MR
- [7] **T Bachmann**, *The very effective covers of KO and KGL over Dedekind schemes*, J. Eur. Math. Soc. (online publication April 2024)
- [8] **B Bhatt, J Lurie**, *Absolute prismatic cohomology*, preprint (2022) arXiv 2201.06120
- [9] **B Bhatt, M Morrow, P Scholze**, *Topological Hochschild homology and integral p -adic Hodge theory*, Publ. Math. Inst. Hautes Études Sci. 129 (2019) 199–310 MR
- [10] **B Bhatt, P Scholze**, *The pro-étale topology for schemes*, from “De la géométrie algébrique aux formes automorphes, I” (J-B Bost, P Boyer, A Genestier, L Lafforgue, S Lysenko, S Morel, B C Ngo, editors), Astérisque 369, Soc. Math. France, Paris (2015) 99–201 MR
- [11] **F Binda, T Lundemo, A Merici, D Park**, *Logarithmic prismatic cohomology, motivic sheaves, and comparison theorems*, preprint (2023) arXiv 2312.13129
- [12] **F Binda, T Lundemo, D Park, P A Østvær**, *A Hochschild–Kostant–Rosenberg theorem and residue sequences for logarithmic Hochschild homology*, Adv. Math. 435 (2023) art. id. 109354 MR
- [13] **F Binda, T Lundemo, D Park, P A Østvær**, *Logarithmic prismatic cohomology via logarithmic THH*, Int. Math. Res. Not. 2023 (2023) 19641–19696 MR
- [14] **F Binda, A Merici**, *Connectivity and purity for logarithmic motives*, J. Inst. Math. Jussieu 22 (2023) 335–381 MR
- [15] **F Binda, D Park, P A Østvær**, *Motives and homotopy theory in logarithmic geometry*, C. R. Math. Acad. Sci. Paris 360 (2022) 717–727 MR
- [16] **F Binda, D Park, P A Østvær**, *Triangulated categories of logarithmic motives over a field*, Astérisque 433, Soc. Math. France, Paris (2022) MR
- [17] **F Binda, D Park, P A Østvær**, *Logarithmic motivic homotopy theory*, preprint (2023) arXiv 2303.02729 To appear under Mem. Amer. Math. Soc.

- [18] **F Binda, K Rülling, S Saito**, *On the cohomology of reciprocity sheaves*, Forum Math. Sigma 10 (2022) art. id. e72 MR
- [19] **A J Blumberg, M A Mandell**, *Localization theorems in topological Hochschild homology and topological cyclic homology*, Geom. Topol. 16 (2012) 1053–1120 MR
- [20] **D Clausen, A Mathew, M Morrow**, *K-theory and topological cyclic homology of henselian pairs*, J. Amer. Math. Soc. 34 (2021) 411–473 MR
- [21] **E Elmanto**, *THH and TC are (very) far from being homotopy functors*, J. Pure Appl. Algebra 225 (2021) art., id. 106640 MR
- [22] **E Elmanto, M Morrow**, *Motivic cohomology of equicharacteristic schemes*, preprint (2023) arXiv 2309.08463v1
- [23] **V Ertl, W a Nizioł**, *Syntomic cohomology and p -adic motivic cohomology*, Algebr. Geom. 6 (2019) 100–131 MR
- [24] **E M Friedlander, A Suslin**, *The spectral sequence relating algebraic K-theory to motivic cohomology*, Ann. Sci. École Norm. Sup. 35 (2002) 773–875 MR
- [25] **T Geisser, L Hesselholt**, *Topological cyclic homology of schemes*, from “Algebraic K-theory” (W Raskind, C Weibel, editors), Proc. Sympos. Pure Math. 67, Amer. Math. Soc., Providence, RI (1999) 41–87 MR
- [26] **T Geisser, M Levine**, *The K-theory of fields in characteristic p* , Invent. Math. 139 (2000) 459–493 MR
- [27] **M A Hill, M J Hopkins, D C Ravenel**, *On the nonexistence of elements of Kervaire invariant one*, Ann. of Math. 184 (2016) 1–262 MR
- [28] **O Hyodo, K Kato**, *Semi-stable reduction and crystalline cohomology with logarithmic poles*, from “Périodes p -adiques”, Astérisque 223 (1994) 221–268 MR
- [29] **L Illusie**, *Complexe de de Rham–Witt et cohomologie cristalline*, Ann. Sci. École Norm. Sup. 12 (1979) 501–661 MR
- [30] **D C Isaksen, P A Østvær**, *Motivic stable homotopy groups*, from “Handbook of homotopy theory” (H Miller, editor), CRC Press, Boca Raton, FL (2020) 757–791 MR
- [31] **U Jannsen, S Saito, Y Zhao**, *Duality for relative logarithmic de Rham–Witt sheaves and wildly ramified class field theory over finite fields*, Compos. Math. 154 (2018) 1306–1331 MR
- [32] **K Kato**, *Logarithmic structures of Fontaine–Illusie*, from “Algebraic analysis, geometry, and number theory” (J-I Igusa, editor), Johns Hopkins Univ. Press, Baltimore, MD (1989) 191–224 MR
- [33] **M Levine**, *The homotopy coniveau tower*, J. Topol. 1 (2008) 217–267 MR
- [34] **P Lorenzon**, *Logarithmic Hodge–Witt forms and Hyodo–Kato cohomology*, J. Algebra 249 (2002) 247–265 MR
- [35] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR
- [36] **J Lurie**, *Higher algebra*, preprint (2017) Available at <https://url.msp.org/Lurie-HA>
- [37] **A Mathew**, *Some recent advances in topological Hochschild homology*, Bull. Lond. Math. Soc. 54 (2022) 1–44 MR
- [38] **C Mazza, V Voevodsky, C Weibel**, *Lecture notes on motivic cohomology*, Clay Mathematics Monographs 2, Amer. Math. Soc., Providence, RI (2006) MR
- [39] **A Merici**, *A motivic integral p -adic cohomology*, preprint (2022) arXiv 2211.14303

- [40] **F Morel**, *An introduction to \mathbb{A}^1 -homotopy theory*, from “Contemporary developments in algebraic K -theory” (M Karoubi, A O Kuku, C Pedrini, editors), ICTP Lect. Notes 15, Abdus Salam Int. Cent. Theoret. Phys., Trieste (2004) 357–441 MR
- [41] **T Nikolaus, P Scholze**, *On topological cyclic homology*, Acta Math. 221 (2018) 203–409 MR
- [42] **M C Olsson**, *The logarithmic cotangent complex*, Math. Ann. 333 (2005) 859–931 MR
- [43] **D Park**, *On the log motivic stable homotopy groups*, Homology Homotopy Appl. 27 (2025) 1–15 MR
- [44] **O Röndigs, P A Østvær**, *Slices of hermitian K -theory and Milnor’s conjecture on quadratic forms*, Geom. Topol. 20 (2016) 1157–1212 MR
- [45] **O Röndigs, M Spitzweck, P A Østvær**, *The motivic Hopf map solves the homotopy limit problem for K -theory*, Doc. Math. 23 (2018) 1405–1424 MR
- [46] **O Röndigs, M Spitzweck, P A Østvær**, *The first stable homotopy groups of motivic spheres*, Ann. of Math. 189 (2019) 1–74 MR
- [47] **O Röndigs, M Spitzweck, P A Østvær**, *The second stable homotopy groups of motivic spheres*, Duke Math. J. 173 (2024) 1017–1084 MR
- [48] **S Saito**, *Reciprocity sheaves and logarithmic motives*, Compos. Math. 159 (2023) 355–379 MR
- [49] **M Spitzweck, P A Østvær**, *Motivic twisted K -theory*, Algebr. Geom. Topol. 12 (2012) 565–599 MR
- [50] **V Voevodsky**, *Open problems in the motivic stable homotopy theory, I*, from “Motives, polylogarithms and Hodge theory, I” (F Bogomolov, L Katzarkov, editors), Int. Press Lect. Ser. 3, I, International, Somerville, MA (2002) 3–34 MR
- [51] **V Voevodsky**, *A possible new approach to the motivic spectral sequence for algebraic K -theory*, from “Recent progress in homotopy theory” (D M Davis, J Morava, G Nishida, W S Wilson, N Yagita, editors), Contemp. Math. 293, Amer. Math. Soc., Providence, RI (2002) 371–379 MR
- [52] **V Voevodsky**, *On the zero slice of the sphere spectrum*, Tr. Mat. Inst. Steklova 246 (2004) 106–115 MR
- [53] **V Voevodsky, A Suslin, E M Friedlander**, *Cycles, transfers, and motivic homology theories*, Annals of Mathematics Studies 143, Princeton Univ. Press (2000) MR

FB, PAØ: Dipartimento di Matematica “Federigo Enriques”, Università degli Studi di Milano
Milan, Italy

DP: Department of Mathematics and Informatics, University of Wuppertal
Wuppertal, Germany

PAØ: Department of Mathematics, University of Oslo
Oslo, Norway

federico.binda@unimi.it, dpark@uni-wuppertal.de, paul.oestvaer@unimi.it

Proposed: Arend Bayer

Seconded: Stefan Schwede, Marc Levine

Received: 7 April 2024

Revised: 30 October 2024

Joyce structures on spaces of quadratic differentials

TOM BRIDGELAND

Consider the space parametrising complex projective curves of genus g equipped with a quadratic differential with simple zeroes. We use the geometry of isomonodromic deformations to construct a complex hyperkähler structure on the total space of its tangent bundle. This provides nontrivial examples of the Joyce structures introduced by the author in relation to Donaldson–Thomas theory.

14H60, 14H70

1 Introduction

Since Hitchin’s classic papers [1987a; 1987b], moduli spaces of Higgs bundles on algebraic curves have appeared in many areas of pure mathematics and mathematical physics. Consider for definiteness the space $\mathcal{M}_C(E, \Phi)$ parametrising $\mathrm{SL}_2(\mathbb{C})$ Higgs bundles (E, Φ) on a smooth complex projective curve C of genus $g > 1$. Two of its most important features are

- (i) a hyperkähler metric, defined using the non-Abelian Hodge correspondence, and
- (ii) a proper map $\mathcal{M}_C(E, \Phi) \rightarrow H^0(C, \omega_C^{\otimes 2})$, whose target parametrises quadratic differentials on C , and whose general fibres are abelian varieties, homeomorphic to $(S^1)^{6g-6}$.

In this paper we study a moduli space $\mathcal{M}(C, E, \nabla, \Phi)$ which in some respects resembles a complexification of $\mathcal{M}_C(E, \Phi)$. For a given integer $g > 1$ it parametrises the data of a curve C of genus g , together with an $\mathrm{SL}_2(\mathbb{C})$ bundle E on C , equipped with both a flat connection ∇ and a Higgs field Φ . We construct

- (i) a meromorphic complex hyperkähler metric, defined using isomonodromic flows, and
- (ii) a map $\mathcal{M}(C, E, \nabla, \Phi) \rightarrow \mathcal{M}(C, Q)$ whose fibres are algebraic tori $(\mathbb{C}^*)^{6g-6}$. The target $\mathcal{M}(C, Q)$ of this map is the “generic Hitchin base” parametrising curves C of genus g equipped with a quadratic differential $Q \in H^0(C, \omega_C^{\otimes 2})$ with simple zeroes.

It is important to note that the complex hyperkähler metric we construct on $\mathcal{M}(C, E, \nabla, \Phi)$ is a much simpler object than the hyperkähler metric on $\mathcal{M}_C(E, \Phi)$. In particular it is algebraic, in contrast to the Hitchin metric, which is highly transcendental. A striking demonstration of this difference appears on generalising to the setting where the Higgs fields and connections have poles of fixed orders [Zikidis \geq 2025]. One can then take the curve C to have genus $g = 0$, and in simple examples the resulting complex hyperkähler structures can be written explicitly in terms of rational functions [Bridgeland and Masoero 2023]. No such explicit formulae are expected for the Hitchin metric.

We expect our construction to generalise to gauge groups other than $\mathrm{SL}_2(\mathbb{C})$ but this will require new ideas. A key point in our construction is that the space of quadratic differentials $H^0(C, \omega_C^{\otimes 2})$ is the cotangent fibre to the moduli space of curves. To generalise to the gauge group $G = \mathrm{SL}_m(\mathbb{C})$, for example, would require a space of “higher complex structures” whose cotangent fibre is the Hitchin base $\bigoplus_{k=2}^m H^0(C, \omega^{\otimes k})$. A natural candidate is provided by the work of Fock and Thomas [2021]. The expectation is then that the total space of the cotangent bundle of this space should carry a meromorphic Joyce structure generalising the one constructed here.

1.1 DT invariants and [GMN]

In a celebrated development, Gaiotto, Moore and Neitzke [Gaiotto et al. 2010; 2013] uncovered a deep relation between the hyperkähler geometry of the moduli spaces $\mathcal{M}_C(E, \Phi)$ and the BPS invariants of a class of four-dimensional $N = 2$ supersymmetric gauge theories known as theories of class $S[A_1]$. More precisely, they introduced a class of nonlinear Riemann–Hilbert (RH) problems defined by the BPS invariants, and showed that their solutions describe twistor lines in the twistor space of $\mathcal{M}_C(E, \Phi)$. In mathematical terms these BPS invariants can be understood as the Donaldson–Thomas (DT) invariants of a certain three-dimensional Calabi–Yau (CY_3) triangulated category [Bridgeland and Smith 2015].

Our interest in the moduli spaces $\mathcal{M}(C, E, \nabla, \Phi)$ stems from a general programme [Bridgeland 2019; 2021] which attempts to encode the DT invariants of a CY_3 triangulated category in a geometric structure on its space of stability conditions. This procedure is currently highly conjectural, and involves a class of RH problems closely related to those considered by Gaiotto, Moore and Neitzke, and obtained from them by a procedure known in physics as the conformal limit [Gaiotto 2014]. In this limit it appears that the geometry of the Hitchin space considered in [Gaiotto et al. 2013] should be replaced by the simpler geometry of the space $\mathcal{M}(C, E, \nabla, \Phi)$ considered here.

The geometric structure on spaces of stability conditions envisaged in [Bridgeland 2021] is a kind of nonlinear Frobenius structure, and was christened a Joyce structure in honour of the paper [Joyce 2007], where the main ingredients were first identified. In later work [Bridgeland and Strachan 2021] it was shown that a Joyce structure on a complex manifold induces a complex hyperkähler structure on the total space of its tangent bundle. The main result of this paper is a construction of a meromorphic Joyce structure on the space $\mathcal{M}(C, Q)$.

The space $\mathcal{M}(C, Q)$ can be identified with a space of stability conditions on a CY_3 category by the work of Haiden [2024]. We leave for future research the problem of using the Joyce structure constructed here to solve the RH problems of [Bridgeland 2019] defined by the DT invariants of Haiden’s category. This would probably be more easily accomplished in the setting of meromorphic quadratic differentials [Bridgeland and Smith 2015; Zikidis \geq 2025], using the Fock–Goncharov cluster structure on the wild character variety [Fock and Goncharov 2006] and results from exact WKB analysis as in [Gaiotto et al. 2013]. One particular example was treated in detail in this way by Bridgeland and Masoero [2023], and other partial results have been obtained by Allegretti [2019; 2021].

1.2 Summary of the construction

We fix a genus $g > 1$ throughout the paper. The space $\mathcal{M}(C, Q)$ as defined above is a smooth Deligne–Mumford stack or, if we work in the analytic category, a complex orbifold. Since this may be uncomfortable for some readers, we shall also fix an integer $\ell > 0$, and insist that all curves C are equipped with a level ℓ structure. This extra data plays no essential role in our constructions so we will omit it from the notation. We always assume $\ell > 2$, since this has the pleasant consequence that all moduli spaces appearing are smooth quasiprojective varieties. But the reader happy with stacks can eliminate the level structures by taking $\ell = 1$ and working instead with smooth Deligne–Mumford stacks.

Let us then introduce the smooth quasiprojective variety $M = \mathcal{M}(C, Q)$ parametrising pairs (C, Q) consisting of a smooth projective complex curve C , equipped as always with a level ℓ structure, and a quadratic differential $Q \in H^0(C, \omega_C^{\otimes 2})$ with simple zeroes. Associated to a point $(C, Q) \in M$ is a smooth spectral curve Σ cut out in the cotangent bundle T_C^* by the equation $y^2 = Q$. The projection $p: \Sigma \rightarrow C$ is a branched double cover with a covering involution σ . The tangent space to M at the point (C, Q) can then be identified with the anti-invariant cohomology group $H^1(\Sigma, \mathbb{C})^-$. The dual of the integral anti-invariant homology defines an integral affine structure $T_M^{\mathbb{Z}} \subset T_M$ and we consider the quotient $X^{\#} = T_M / T_M^{\mathbb{Z}}$. The fibre of the induced projection $\pi: X^{\#} \rightarrow M$ over the point (C, Q) is the quotient of the group $H^1(\Sigma, \mathbb{C}^*)^-$ by the finite subgroup $p^*(H^1(C, \{\pm 1\}))$ and is isomorphic to $(\mathbb{C}^*)^{6g-6}$.

A Joyce structure on M is essentially the data of a pencil of flat symplectic nonlinear connections h_{ϵ} on the bundle $\pi: X^{\#} \rightarrow M$ parametrised by $\epsilon \in \mathbb{C}^*$. The associated complex hyperkähler structure on $X^{\#}$ is then defined by taking the eigenspaces of the operators I, J, K to be the horizontal subbundles of $T_{X^{\#}}$ defined by certain elements of this pencil. We construct the nonlinear connections h_{ϵ} as follows. We can realise elements of $H^1(\Sigma, \mathbb{C}^*)^-$ as the holonomy of anti-invariant line bundles with connection (L, ∂) on Σ . The usual spectral correspondence associates to L a Higgs bundle (E, Φ) on the curve C . Using an extension of this correspondence to connections, valid under a genericity assumption on L , we can use ∂ to induce a connection ∇ on E . The required family of nonlinear connections h_{ϵ} is then given by the isomonodromy flows for the connections $\nabla - \epsilon^{-1}\Phi$.

The complex hyperkähler structure we construct on M has poles; these arise from two interesting issues. Firstly, the extension of the spectral correspondence to connections requires a genericity assumption on the line bundle L . This relates to the theta divisor in the generalised Prym variety of the double cover $p: \Sigma \rightarrow C$. Secondly, given a fixed curve C equipped with a Higgs bundle (E, Φ) , we need to lift deformations of the quadratic differential $Q = \text{tr}(\Phi^2)$ to deformations of the Higgs field Φ . This relates to the wobbly locus in the space of Higgs bundles; see [Donagi and Pantev 2009].

The extended spectral correspondence in our construction can be viewed as an abelianization procedure for flat connections in the presence of a quadratic differential. In the case of meromorphic quadratic differentials, this de Rham abelianization can be compared with the Betti abelianization of [Hollands and Neitzke 2016; Nikolaev 2021], which depends on the choice of a spectral network on C . Their

relationship is highly nontrivial, and in fact, if we take the spectral network to be the WKB triangulation of the quadratic differential, one can view the solutions to the RH problems discussed above as intertwining these two abelianisation procedures.

Plan of the paper The aim of the paper is to construct a meromorphic Joyce structure on the space $M = \mathcal{M}(C, Q)$. A Joyce structure on a complex manifold M is a combination of two ingredients: a period structure and a pencil of nonlinear connections on the tangent bundle. The definitions of all these terms can be found in Section 2.

The required period structure on M is well-known and is described in Section 3. In Section 4 we recall the standard correspondence between Higgs bundles on C and line bundles on the spectral curve Σ , and explain how it can be extended to bundles with connection. Section 5 introduces the essential diagrams of moduli spaces which will be used to construct the pencil of nonlinear connections. We also prove two crucial generic finiteness results.

In Section 6 we recall the Atiyah–Bott symplectic form on the moduli space of flat connections and prove that our extended spectral correspondence preserves it. The meromorphic Joyce structure on M is finally constructed in Section 7 using isomonodromic flows. In Section 8, we describe an interesting compatibility relation between this Joyce structure and the Lagrangian submanifolds in M obtained by fixing the curve C . We include in the appendix a summary of the scheme-theoretic definitions and constructions of the various moduli spaces used in the main text.

Conventions and notation We use rather unconventional conventions for labelling moduli spaces. In general a symbol $\mathcal{M}_A(B)$ denotes the moduli space of objects of type B on a fixed object A . So for example $\mathcal{M}_C(E, \Phi)$ denotes the moduli space of $\mathrm{SL}_2(\mathbb{C})$ Higgs bundles (E, Φ) on a fixed curve C , whereas $\mathcal{M}(C, E, \Phi)$ denotes the moduli space where C is also allowed to vary. We can only apologise for the initially nonsensical appearance of statements such as “Take a point $(C, Q) \in \mathcal{M}(C, Q)$ ”, and hope that this proves less of an inconvenience than having to constantly consult a dictionary of the large number of moduli spaces that appear.

The paper contains many connections, both linear and nonlinear. Linear connections on a vector bundle E are specified by their covariant derivative $E \rightarrow E \otimes \Omega^1$ and are usually denoted by the symbols ∇ or ∂ . Nonlinear connections on a map $\pi: X \rightarrow M$ are specified by a bundle map $\pi^*(T_M) \rightarrow T_X$ and are denoted by small latin letters h, j , etc. Throughout the paper we encounter families of connections parametrised by $\epsilon \in \mathbb{C}^*$, which we refer to as pencils. Often the inverse $\zeta = \epsilon^{-1}$ would seem to be a more natural parameter, but we will nonetheless use ϵ since in the relations with mathematical physics this is the most natural variable, relating variously to the string coupling, Planck’s constant, etc.

We work with both complex manifolds and algebraic varieties. Except in the appendix, all algebraic varieties appearing are smooth and quasiprojective over \mathbb{C} . We view them as a subcategory of the category of complex manifolds. The derivative of a map of complex manifolds $f: X \rightarrow Y$ is denoted $f_*: T_X \rightarrow f^*(T_Y)$. The map f is called étale if f_* is an isomorphism.

Acknowledgements I am very grateful for correspondence and conversations with Dylan Allegretti, Anna Barbieri, Maciej Dunajski, Dima Korotkin, Andy Neitzke, Nikita Nikolaev, Ivan Smith, Ian Strachan, Jörg Teschner and Menelaos Zikidis. Extended discussions with Zikidis were particularly important in getting the results into the final form presented here. I am supported by a Royal Society Research Professorship at the University of Sheffield.

2 Joyce structures

This section introduces the notion of a Joyce structure on a complex manifold M . The definition arose from a line of work in Donaldson–Thomas theory [Bridgeland 2021] which originated with a paper of Joyce [2007]. We first define the notion of a pre-Joyce structure, which consists of a pencil of flat symplectic nonlinear connections on the tangent bundle $\pi: X = T_M \rightarrow M$. Following [Bridgeland and Strachan 2021], we show that a pre-Joyce structure on M induces a complex hyperkähler structure on X . This construction is well known in the twistor-theory literature, see for example [Dunajski and Mason 2000], and goes back to the work of Plebański [1975]. A Joyce structure is then defined to be a pre-Joyce structure with certain extra symmetries. The description of these symmetries involves a strengthening of the notion of an integral affine structure which we call a period structure.

2.1 Nonlinear connections

We begin by briefly summarising some basic facts about nonlinear connections in the sense of Ehresmann. We work with complex manifolds and holomorphic maps, but everything in this section holds also in the smooth setting.

Let $\pi: X \rightarrow M$ be a holomorphic submersion of complex manifolds. Denote the fibres by $X_m = \pi^{-1}(m)$. The derivative of π gives rise to a short exact sequence of vector bundles

$$(1) \quad 0 \rightarrow T_{X/M} \xrightarrow{i} T_X \xrightarrow{\pi_*} \pi^*(T_M) \rightarrow 0.$$

Definition 2.1 A nonlinear connection on the map π is a bundle map $h: \pi^*(T_M) \rightarrow T_X$ satisfying $\pi_* \circ h = 1$.

Writing $H = \text{im}(h)$ and $V = T_{X/M}$, the tangent bundle decomposes as a direct sum $T_X = H \oplus V$. We call tangent vectors and vector fields horizontal or vertical if they lie in H or V , respectively. Note that a vector field $u \in H^0(M, T_M)$ can be lifted to a horizontal vector field $h(u) \in H^0(X, T_X)$ by composing the pullback $\pi^*(u) \in H^0(X, \pi^*(T_M))$ with the map h .

Consider a smooth path $\gamma: [0, 1] \rightarrow M$. Given a point $x \in X_{\gamma(0)}$ we can look for a lifted path $\alpha: [0, \delta] \rightarrow X$ satisfying $\alpha_*(d/dt) = h(\gamma_*(d/dt))$ and $\alpha(0) = x$. Such a lift will exist for small enough $\delta > 0$. For $t \in [0, \delta]$ we call $\alpha(t) \in X_{\gamma(t)}$ the time t parallel transport of the point x along the path γ . Given a point $x_0 \in X_{\gamma(0)}$ we can find a $\delta > 0$ and open subsets $U_t \subset X_{\gamma(t)}$ with $x_0 \in U_0$, such that time t parallel transport along γ defines an isomorphism $\text{PT}_\gamma(t): U_0 \rightarrow U_t$ for each $t \in [0, \delta]$.

Given complex manifolds M and N there is a connection on the projection map $\pi_M: M \times N \rightarrow M$ induced by the canonical splitting $T_{M \times N} = \pi_M^*(T_M) \oplus \pi_N^*(T_N)$. A connection h on $\pi: X \rightarrow M$ is called flat if it is locally isomorphic to a connection of this form. More precisely:

Definition 2.2 *The connection h is flat if the following equivalent conditions hold:*

- (i) *For every $x \in X$ there are local coordinates (x_1, \dots, x_n) on X at x , and (y_1, \dots, y_d) on M at $\pi(x)$, such that $x_i = \pi^*(y_i)$ and $h(\partial/\partial y_i) = \partial/\partial x_i$ for $1 \leq i \leq d$.*
- (ii) *The subbundle $H = \text{im}(h) \subset T_X$ is involutive: $[H, H] \subset H$.*

Suppose given a relative symplectic form $\Omega_\pi \in H^0(X, \wedge^2 T_{X/M}^*)$ on the map π . It restricts to a symplectic form $\Omega_m \in H^0(X_m, \wedge^2 T_{X_m}^*)$ on each fibre X_m . Note that since $T_X/\text{im}(h) = T_{X/M}$, the relative form Ω_π can be lifted uniquely to a form $\Omega \in H^0(X, \wedge^2 T_X^*)$ satisfying $\ker(\Omega) = \text{im}(h)$. We say that the connection h preserves Ω_π if, for any path $\gamma: [0, 1] \rightarrow M$, the partially defined parallel transport maps $\text{PT}_\gamma(t): X_{\gamma(0)} \rightarrow X_{\gamma(t)}$ take $\Omega_{\gamma(0)}$ to $\Omega_{\gamma(t)}$.

Lemma 2.3 (i) *The connection h preserves Ω_π precisely if $i_{v_1}i_{v_2}(d\Omega) = 0$ for any two vertical vector fields $v_1, v_2 \in H^0(X, T_{X/M})$.*
 (ii) *If the connection h is flat, then it preserves Ω_π precisely if $d\Omega = 0$.*

Proof The first statement is [Gotay et al. 1983, Theorem 4]. For the second, take three vector fields u_1, u_2, u_3 on X and consider the expression defining $d\Omega(u_1, u_2, u_3)$. We can assume that each u_i is either horizontal or vertical. Since $i_h(\Omega) = 0$ for any horizontal vector field h , and horizontal vector fields are closed under the Lie bracket, we have $d\Omega(u_1, u_2, u_3) = 0$ as soon as two of the u_i are horizontal. The claim then follows from (i). □

Suppose that a discrete group G acts freely and properly on X preserving the map π . Then $Y = X/G$ is a complex manifold and the quotient map $q: X \rightarrow Y$ is étale. There is an induced submersion $\eta: Y \rightarrow M$ and a factorisation $\pi = \eta \circ q$. A connection $h: \pi^*(T_M) \rightarrow T_X$ will be called G -invariant if $g_* \circ h = h$ for all $g \in G$. There is then an induced connection $j: \eta^*(T_M) \rightarrow T_Y$ on η , uniquely defined by the condition that $q_* \circ h = q^*(j)$. We say that the connection h descends along the quotient map q .

2.2 Pre-Joyce structures

Let M be a complex manifold and let $\pi: X = T_M \rightarrow M$ be the total space of the tangent bundle of M . There is a canonical isomorphism $\nu: \pi^*(T_M) \rightarrow T_{X/M}$ obtained by composing the chain of identifications

$$(2) \quad \pi^*(T_M)_x = T_{M, \pi(x)} = T_{T_{M, \pi(x)}, x} = T_{X_{\pi(x)}, x} = T_{X/M, x},$$

and we set $v = i \circ \nu$. A connection $h: \pi^*(T_M) \rightarrow T_X$ on π then defines a family of such connections $h_\epsilon = h + \epsilon^{-1}v$ parametrised by $\epsilon \in \mathbb{C}^*$. We call such a family a ν -pencil of connections.

$$\begin{array}{ccccccc}
 0 & \longrightarrow & T_{X/M} & \xrightarrow{i} & T_X & \xrightarrow{\pi^*} & \pi^*(T_M) \longrightarrow 0 \\
 & & & & & \curvearrowright^{h_\epsilon} & \\
 & & & & & \curvearrowleft_v &
 \end{array}$$

Suppose that M is equipped with a holomorphic symplectic form $\omega \in H^0(M, \wedge^2 T_M^*)$. Via the isomorphism ν we obtain a relative symplectic form $\Omega_\pi \in H^0(X, \wedge^2 T_{X/M}^*)$. We say that a connection on π is symplectic if it preserves Ω_π .

Definition 2.4 A pre-Joyce structure on a complex manifold M consists of

- (i) a holomorphic symplectic form ω on M , and
- (ii) a nonlinear connection h on the tangent bundle $\pi: X = T_M \rightarrow M$,

such that for each $\epsilon \in \mathbb{C}^*$, the connection $h_\epsilon = h + \epsilon^{-1}v$ is flat and symplectic.

To clarify this definition we now describe it in local coordinates, although the resulting expressions will play no role in what follows. Given a local coordinate system (z_1, \dots, z_n) on M there are associated linear coordinates $(\theta_1, \dots, \theta_n)$ on the tangent spaces $T_{M,m}$ obtained by writing a tangent vector in the form $\sum_i \theta_i \cdot (\partial/\partial z_i)$. We thus get induced local coordinates (z_i, θ_j) on the space $X = T_M$. In these coordinates, $\nu(\partial/\partial z_i) = \partial/\partial \theta_i$.

We always assume that the coordinates z_i are Darboux, in the sense that

$$\omega = \frac{1}{2} \sum_{p,q} \omega_{pq} \cdot dz_p \wedge dz_q,$$

with ω_{pq} a constant skew-symmetric matrix. We denote by η_{pq} the inverse matrix.

The fact that the connection h is flat and symplectic ensures that we can write it in Hamiltonian form

$$(3) \quad h\left(\frac{\partial}{\partial z_i}\right) = \frac{\partial}{\partial z_i} + \sum_{p,q} \eta_{pq} \cdot \frac{\partial W_i}{\partial \theta_p} \cdot \frac{\partial}{\partial \theta_q},$$

for functions $W_i: X \rightarrow \mathbb{C}$. The connection h_ϵ is then flat for all $\epsilon \in \mathbb{C}^*$ if we can take $W_i = \partial W / \partial \theta_i$ for a single function $W: X \rightarrow \mathbb{C}$, which moreover satisfies

$$(4) \quad \frac{\partial^2 W}{\partial \theta_i \partial z_j} - \frac{\partial^2 W}{\partial \theta_j \partial z_i} = \sum_{p,q} \eta_{pq} \cdot \frac{\partial^2 W}{\partial \theta_i \partial \theta_p} \cdot \frac{\partial^2 W}{\partial \theta_j \partial \theta_q}.$$

The function W is called the Plebański function, and the partial differential equations (4) are known as Plebański’s second heavenly equations [Dunajski and Mason 2000].

2.3 Complex hyperkähler structures

By a complex hyperkähler structure on a complex manifold X we mean the data of a holomorphic metric $g: T_X \otimes T_X \rightarrow \mathcal{O}_X$, together with endomorphisms $I, J, K \in \text{End}_X(T_X)$ satisfying the quaternion relations

$$I^2 = J^2 = K^2 = IJK = -1,$$

which preserve g , and which are parallel for the holomorphic Levi-Civita connection ∇ :

$$(5) \quad g(R(u_1), R(u_2)) = g(u, v), \quad \nabla(R = 0, \quad R \in \{I, J, K\}.$$

Such structures have appeared before in the literature, often under different names.

Let M be a complex manifold with a holomorphic symplectic form ω . A nonlinear connection h on the tangent bundle $\pi: X = T_M \rightarrow M$ gives a decomposition

$$(6) \quad T_X = \text{im}(v) \oplus \text{im}(h) \cong \pi^*(T_M) \otimes_{\mathbb{C}} \mathbb{C}^2.$$

We can define a metric $g: T_X \otimes T_X \rightarrow \mathcal{O}_X$ by taking the tensor product of $\pi^*(\omega)$ with the standard symplectic form on \mathbb{C}^2 , and an action of the quaternions on T_X by identifying the complexification of the quaternions $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{C}$ with the algebra $\text{End}_{\mathbb{C}}(\mathbb{C}^2)$. With appropriate conventions this leads to the formulae

$$(7) \quad \begin{aligned} I \circ h &= i \cdot h, & J \circ h &= -v, & K \circ h &= i \cdot v, \\ I \circ v &= -i \cdot v, & J \circ v &= h, & K \circ v &= i \cdot h, \end{aligned}$$

which should be interpreted as equalities of maps $\pi^*(T_M) \rightarrow T_X$, and

$$(8) \quad g(h(u_1), v(u_2)) = \frac{1}{2}\omega(u_1, u_2), \quad g(h(u_1), h(u_2)) = 0 = g(v(u_1), v(u_2)).$$

It is easily checked that g is preserved by the endomorphisms I, J, K .

The following result implies in particular that a pre-Joyce structure on a complex manifold M induces a complex hyperkähler structure on the total space $X = T_M$.

Theorem 2.5 *The endomorphisms I, J, K are parallel for the Levi-Civita connection ∇ associated to g precisely if the connection h_ϵ is flat and symplectic for all $\epsilon \in \mathbb{C}^*$.*

Proof We begin with a general remark. Let $g: T_X \times T_X \rightarrow \mathcal{O}_X$ be a metric on a complex manifold X with associated Levi-Civita connection ∇ . Let $R \in \text{End}_X(T_X)$ be an endomorphism which is compatible with g and satisfies $R^2 = -1$. We can then define a 2-form Ω on X by setting $\Omega_R(u_1, u_2) = g(R(u_1), u_2)$. Let $H \subset T_X$ denote the $+i$ eigenbundle of R . Then standard proofs from Kähler geometry apply unchanged in this holomorphic context to give implications

$$(9) \quad \nabla(R) = 0 \implies [H, H] \subset H, \quad \nabla(R) = 0 \iff d\Omega_R = 0.$$

Return now to the setting above. For $\epsilon \in \mathbb{C}^*$ we introduce the endomorphism

$$(10) \quad J_\epsilon = I - i\epsilon^{-1}(J + iK).$$

A simple calculation using the definitions (7) shows that $J_\epsilon^2 = -1$, and that the $+i$ eigenbundle of J_ϵ coincides with $H_\epsilon = \text{im}(h_\epsilon)$.

As in Section 2.2, the symplectic form ω on M induces a relative symplectic form Ω_π on the projection $\pi : X \rightarrow M$. Moreover, as explained before Lemma 2.3, there is then a unique 2-form Ω_ϵ on X satisfying the conditions

$$(11) \quad \ker(\Omega_\epsilon) = H_\epsilon \quad \text{and} \quad \Omega_\epsilon(v(u_1), v(u_2)) = \omega(u_1, u_2),$$

where u_1, u_2 are arbitrary vector fields on M . Another calculation using (7) and (8) shows that this form is given explicitly by the formula

$$(12) \quad \Omega_\epsilon = \epsilon^{-2} \cdot \Omega_+ + 2i\epsilon^{-1} \cdot \Omega_I + \Omega_-, \quad \text{where } \Omega_\pm = \Omega_{J \pm iK}.$$

We can now prove the theorem. Suppose first that I, J, K are parallel. Then J_ϵ is parallel for all $\epsilon \in \mathbb{C}^*$, and applying (9) with $R = J_\epsilon$ we find that $[H_\epsilon, H_\epsilon] \subset H_\epsilon$ and hence that h_ϵ is flat. Since Ω_ϵ is also parallel and hence closed, applying Lemma 2.3 shows that h_ϵ is symplectic. Conversely suppose that for all $\epsilon \in \mathbb{C}^*$ the connection h_ϵ is flat and symplectic. Then by Lemma 2.3 again, $d\Omega_\epsilon = 0$ for all $\epsilon \in \mathbb{C}^*$, and this easily implies that $d\Omega_R = 0$ for $R \in \{I, J, K\}$. By (9) we conclude that I, J, K are parallel. \square

2.4 Period structures

Let M be a complex manifold and \mathcal{H} a holomorphic vector bundle on M . By a lattice in \mathcal{H} we mean a locally constant subsheaf of abelian groups $\mathcal{H}^{\mathbb{Z}} \subset \mathcal{H}$ such that the multiplication map $\mathcal{H}^{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathcal{O}_M \rightarrow \mathcal{H}$ is an isomorphism. There is an induced flat linear connection ∇ on \mathcal{H} whose flat sections are \mathbb{C} -linear combinations of the sections of $\mathcal{H}^{\mathbb{Z}}$.

Definition 2.6 *A period structure on a complex manifold M consists of*

- (P1) *a lattice $T_M^{\mathbb{Z}} \subset T_M$, whose associated flat connection we denote by ∇ , and*
- (P2) *a vector field $Z \in \Gamma(M, T_M)$ satisfying $\nabla(Z) = \text{id}$.*

Let $(T_M^{\mathbb{Z}}, \nabla, Z)$ be a period structure on a complex manifold M and take a basepoint $p \in M$. A basis of the free abelian group $T_{M,p}^{\mathbb{Z}}$ extends uniquely to a basis of ∇ -flat sections ϕ_1, \dots, ϕ_n of T_M over a contractible open neighbourhood $p \in U \subset M$. Writing the vector field Z in the form $Z = \sum_i z_i \cdot \phi_i$ then defines holomorphic functions $z_i : U \rightarrow \mathbb{C}$, and condition (P2) implies that (z_1, \dots, z_n) is a local coordinate system on M , and that $\phi_i = \partial/\partial z_i$. Note in particular that the connection ∇ is necessarily torsion-free.

Recall that an integral affine structure on a complex manifold M consists of a lattice $T_M^{\mathbb{Z}} \subset T_M$ whose associated flat connection ∇ is torsion-free [Kontsevich and Soibelman 2006]. A local coordinate system (z_1, \dots, z_n) is then called integral affine if the tangent vectors $\partial/\partial z_i$ lie in the lattice $T_M^{\mathbb{Z}}$. Such coordinate systems are uniquely defined up to affine transformations of the form $z_i \mapsto \sum_j a_{ij} z_j + v_i$ with $(a_{ij}) \in \text{GL}_n(\mathbb{Z})$ and $(v_i) \in \mathbb{C}^n$.

Given a period structure on a complex manifold M we obtain an integral affine structure by forgetting the vector field Z . A system of integral affine coordinates (z_1, \dots, z_n) will be called integral linear if $Z = \sum_i z_i \cdot (\partial/\partial z_i)$. Such coordinate systems are uniquely defined up to linear transformations of the form $z_i \mapsto \sum_j a_{ij} z_j$ with $(a_{ij}) \in \text{GL}_n(\mathbb{Z})$. Thus a period structure can be thought of as an integral *linear* structure.

Using the connection ∇ , we can lift the vector field $Z \in \Gamma(M, T_M)$ to a horizontal vector field $E \in \Gamma(X, T_X)$. Let us consider the case when there is a \mathbb{C}^* action on the manifold M whose generating vector field is Z . There is a \mathbb{C}^* action on $X = T_M$ obtained by combining the induced action of \mathbb{C}^* on $X = T_M$ with the rescaling action on the fibres of $\pi: X = T_M \rightarrow M$ of weight -1 . If $m_t: M \rightarrow M$ denotes the action of $t \in \mathbb{C}^*$ on M , this is the action for which $t \in \mathbb{C}^*$ sends $v \in T_{M,m}$ to $(m_t)_*(t^{-1}v) \in T_{M,m_t(v)}$.

Lemma 2.7 *The generating vector field for this \mathbb{C}^* action on X is the horizontal lift E .*

Proof If we take a system of integral linear coordinates (z_1, \dots, z_n) on M then by definition $Z = \sum_i z_i \cdot (\partial/\partial z_i)$. Taking associated coordinates (z_i, θ_j) on $X = T_M$ as before the ∇ -horizontal lift of Z is the vector field $E = \sum_i z_i \cdot (\partial/\partial z_i)$ on X . The \mathbb{C}^* action on X induced by that on M is given by $(z_i, \theta_j) \mapsto (tz_i, t\theta_j)$. Composing with the contraction in the fibres we obtain the action $(z_i, \theta_j) \mapsto (tz_i, \theta_j)$, whose generating vector field is E . □

Definition 2.8 *A period structure with skew form on a complex manifold M consists of a period structure $(T_M^{\mathbb{Z}}, \nabla, Z)$, together with a skew-symmetric form*

$$(13) \quad \eta: T_M^* \times T_M^* \rightarrow \mathcal{O}_M,$$

such that $\eta/2\pi i$ takes integral values on the lattices $(T_M^{\mathbb{Z}})^* \subset T_M^*$.

The pairing η is necessarily parallel for the flat connection ∇ , and it follows that it defines a holomorphic Poisson structure on M . We will be particularly interested in the case when the kernel of η is zero. Viewing η as a linear map $T_M^* \rightarrow T_M$, its inverse defines a complex symplectic form $\omega \in H^0(M, \wedge^2 T_M^*)$.

2.5 Joyce structures

Let M be a complex manifold with a period structure $(T_M^{\mathbb{Z}}, \nabla, Z)$. The rescaled lattice $(2\pi i)T_M^{\mathbb{Z}} \subset T_M$ acts on $X = T_M$ by translations in the fibres. We introduce the quotient

$$(14) \quad X^\# = T_M^\# = T_M / (2\pi i) T_M^{\mathbb{Z}}.$$

We also consider the involution $\iota: X \rightarrow X$ which acts by -1 on the fibres of $\pi: X \rightarrow M$.

Definition 2.9 Let M be a complex manifold, and let $\pi : X = T_M \rightarrow M$ denote the total space of the holomorphic tangent bundle. A Joyce structure on M consists of

- (a) a period structure with skew form $(T_M^{\mathbb{Z}}, Z, \nabla, \eta)$ on M , and
- (b) a pre-Joyce structure (ω, h) on M ,

satisfying the following conditions:

- (J1) The symplectic form ω is the inverse of the skew form η .
 - (J2) The connection h is invariant under the action of the lattice $(2\pi i) T_M^{\mathbb{Z}} \subset T_M$.
 - (J3) If E is the ∇ -horizontal lift of the vector field Z , then for any vector field v on M
- (15)
$$h([Z, v]) = [E, h(v)].$$
- (J4) The connection h is invariant under the action of the involution $\iota : X \rightarrow X$.

Note that once the period structure with skew form $(T_M^{\mathbb{Z}}, Z, \nabla, \eta)$ on M is fixed, the Joyce structure involves only one further piece of data, namely the nonlinear connection h . For the Joyce structures appearing in this paper the period structure is elementary and well-known, so all our work will go into defining the nonlinear connection h .

Let us express the conditions of Definition 2.9 in terms of a local coordinates as in Section 2.2. If we take a system of integral linear coordinates (z_1, \dots, z_n) on M then by definition $Z = \sum_i z_i \cdot (\partial/\partial z_i)$. Taking associated coordinates (z_i, θ_j) on $X = T_M$ as before, the ∇ -horizontal lift of Z is the vector field $E = \sum_i z_i \cdot (\partial/\partial z_i)$ on X . The symmetries (J2)-(J4) then translate into the following conditions on the Plebański function:

(16)
$$\frac{\partial^2 W}{\partial \theta_p \partial \theta_q}(z_i, \theta_j + 2\pi i k_j) = \frac{\partial^2 W}{\partial \theta_p \partial \theta_q}(z_i, \theta_j),$$

(17)
$$W(\lambda z_i, \theta_j) = \lambda^{-1} W(z_i, \theta_j),$$

(18)
$$W(z_i, -\theta_j) = -W(z_i, \theta_j),$$

where $(k_1, \dots, k_n) \in \mathbb{Z}^n$ and $\lambda \in \mathbb{C}^*$.

The construction we describe in this paper produces what we shall call a meromorphic Joyce structure. This means that the connection h has poles on certain subsets of X . More precisely, there is an effective divisor $D \subset X$, and h is defined by a bundle map $h : \pi^*(T_M) \rightarrow T_X(D)$ satisfying

(19)
$$(\pi_* \otimes \mathcal{O}_X(D)) \circ h = 1_{\pi^*(T_M)} \otimes s_D,$$

where $s_D : \mathcal{O}_X \rightarrow \mathcal{O}_X(D)$ is the canonical inclusion. This means that when expressed in terms of local coordinates as above, the Plebański function $W(z_i, \theta_j)$ is a meromorphic function.

3 The period structure on the space of quadratic differentials

For the rest of the paper we fix an integer $g > 1$ and a level $\ell > 2$. As discussed in Section 1.2 we insist that all curves C are equipped with a level ℓ structure, although we suppress this from the notation. In this section we introduce the space $M = \mathcal{M}(C, Q)$, which will form the base of our Joyce structure. It parametrises pairs (C, Q) consisting of a smooth projective curve C of genus g equipped with a level ℓ structure, and a quadratic differential $Q \in H^0(C, \omega_C^{\otimes 2})$ with simple zeroes. Any such pair (C, Q) determines a branched double cover $p: \Sigma \rightarrow C$, which we call the spectral curve. We construct a period structure with skew form on M , and give a moduli-theoretic description of the fibres of the map (14) in terms of line bundles with connection on Σ .

3.1 Moduli space of quadratic differentials

Let us begin by recalling the definition of a level structure. Given a smooth complex projective curve C of genus g , the homology group $H_1(C, \mathbb{Z}/\ell)$ is a free \mathbb{Z}/ℓ -module of rank $2g$. The intersection form defines a nondegenerate skew-symmetric form

$$(20) \quad \langle -, - \rangle: H_1(C, \mathbb{Z}/\ell) \times H_1(C, \mathbb{Z}/\ell) \rightarrow \mathbb{Z}/\ell.$$

A level ℓ structure on C is a basis $(\alpha_1, \dots, \alpha_g, \beta_1, \dots, \beta_g)$ for $H_1(C, \mathbb{Z}/\ell)$ which is symplectic, in the sense that $\langle \alpha_i, \alpha_j \rangle = 0 = \langle \beta_i, \beta_j \rangle$ and $\langle \alpha_i, \beta_j \rangle = \delta_{ij}$.

Let $\mathcal{M}(C)$ denote the moduli space of smooth complex projective curves of genus g equipped with a level ℓ structure. As we recall in the appendix, given our assumptions $g > 1$ and $\ell > 2$ this is a smooth quasiprojective complex variety of dimension $3g - 3$.

The tangent space to $\mathcal{M}(C)$ at a curve C is the cohomology group $H^1(C, T_C)$, and Serre duality gives $H^0(C, \omega_C^{\otimes 2}) = H^1(C, T_C)^*$, so the cotangent bundle $T_{\mathcal{M}(C)}^*$ parametrises pairs (C, Q) consisting of a smooth complex projective curve C equipped with a level ℓ structure, together with an element $Q \in H^0(C, \omega_C^{\otimes 2})$. We define $\mathcal{M}(C, Q) \subset T_{\mathcal{M}(C)}^*$ to be the open subset of pairs (C, Q) for which Q has simple zeroes. Then $M = \mathcal{M}(C, Q)$ is a smooth quasiprojective complex variety of dimension $6g - 6$.

3.2 Spectral curve

Let C be a smooth complex projective curve of genus g , and $Q \in H^0(C, \omega_C^{\otimes 2})$ a quadratic differential with simple zeroes. The spectral curve Σ associated to the pair (C, Q) is the smooth projective curve cut out inside the total space of the cotangent bundle T_C^* by the equation $y^2 = Q$. The projection $(x, y) \mapsto x$ defines a branched double cover $p: \Sigma \rightarrow C$, and there is a covering involution $\sigma: \Sigma \rightarrow \Sigma$ defined by $(x, y) \mapsto (x, -y)$. The assumption that Q has simple zeroes ensures that Σ is smooth, and the fact that Q has at least one zero implies that Σ is connected.

Denote by $R \subset \Sigma$ the branch divisor of the map $p: \Sigma \rightarrow C$. It has degree $4g - 4$, so by the Riemann–Hurwitz formula the spectral curve Σ has genus $4g - 3$. The dual of the derivative of p defines a canonical section $s: p^*(\omega_C) \rightarrow \omega_\Sigma$ fitting into a short exact sequence

$$(21) \quad 0 \rightarrow p^*(\omega_C) \xrightarrow{s} \omega_\Sigma \rightarrow \mathcal{O}_R \rightarrow 0.$$

On the other hand the square-root of $p^*(Q)$ defines a section of $p^*(\omega_C)$ with simple zeroes on R , and hence a short exact sequence

$$(22) \quad 0 \rightarrow \mathcal{O}_\Sigma \xrightarrow{\phi} p^*(\omega_C) \rightarrow \mathcal{O}_R \rightarrow 0.$$

We define the invariant and anti-invariant homology groups

$$H_1(\Sigma, \mathbb{Z})^\pm = \{\gamma \in H_1(\Sigma, \mathbb{Z}) : \sigma^*(\gamma) = \pm\gamma\},$$

and similarly for the cohomology groups $H^1(\Sigma, \mathbb{Z})^\pm, H^1(\Sigma, \mathbb{C})^\pm$, etc. There is a short exact sequence of free abelian groups

$$(23) \quad 0 \rightarrow H_1(\Sigma, \mathbb{Z})^- \rightarrow H_1(\Sigma, \mathbb{Z}) \xrightarrow{p_*} H_1(C, \mathbb{Z}) \rightarrow 0,$$

the map p_* being surjective because p is ramified.

Taking maps of (23) into \mathbb{Z} shows that the image of $p^*: H^1(C, \mathbb{Z}) \rightarrow H^1(\Sigma, \mathbb{Z})$ coincides with the subgroup $H^1(\Sigma, \mathbb{Z})^+$. The anti-invariant homology group $H_1(\Sigma, \mathbb{Z})^-$ is therefore free of rank $6g - 6$. We also consider the extended group

$$\tilde{H}^1(\Sigma, \mathbb{Z})^- := \text{Hom}_{\mathbb{Z}}(H_1(\Sigma, \mathbb{Z})^-, \mathbb{Z}) = H^1(\Sigma, \mathbb{Z})/H^1(C, \mathbb{Z}).$$

3.3 Period structure

Introduce the vector bundle $\mathcal{H} \rightarrow M$ whose fibre over a point (C, Q) is the vector space $H^1(\Sigma, \mathbb{C})^-$. It contains a lattice $\mathcal{H}^{\mathbb{Z}} \subset \mathcal{H}$ whose fibres are the groups $\tilde{H}^1(\Sigma, \mathbb{Z})^-$. The associated flat connection ∇^{GM} on \mathcal{H} is the Gauss–Manin connection. The dual bundle \mathcal{H}^* has fibres $H_1(\Sigma, \mathbb{C})^-$ and contains the dual lattice $(\mathcal{H}^{\mathbb{Z}})^*$ with fibres $H_1(\Sigma, \mathbb{Z})^-$. The intersection form defines a parallel skew-symmetric form on \mathcal{H}^* , which takes integral values on $(\mathcal{H}^{\mathbb{Z}})^*$.

For each point $(C, Q) \in M$, the tautological 1-form $y dx$ on T^*C restricts to a 1-form $\lambda \in H^0(\Sigma, \omega_\Sigma)$ satisfying $\lambda^{\otimes 2} = p^*(Q)$ and $\sigma^*(\lambda) = -\lambda$. This should not be confused with the section ϕ appearing in (22): there is a relation $\lambda = s \circ \phi$. The de Rham cohomology class associated to λ is an element of $H^1(\Sigma, \mathbb{C})^-$, and the resulting map

$$(24) \quad \delta: M \rightarrow \mathcal{H}, \quad (C, Q) \mapsto [\lambda] \in H^1(\Sigma, \mathbb{C})^-,$$

is a holomorphic section of the bundle \mathcal{H} .

Theorem 3.1 [Veech 1990] *The covariant derivative of δ with respect to the Gauss–Manin connection defines an isomorphism $\nabla^{\text{GM}}(\delta): T_M \rightarrow \mathcal{H}$.*

Taking a basis $(\gamma_1, \dots, \gamma_n) \subset H_1(\Sigma, \mathbb{Z})^-$ at some point $(C, Q) \in M$ and extending to nearby points using the Gauss–Manin connection gives locally defined functions on M

$$(25) \quad z_i = Z(\gamma_i) = ([\lambda], \gamma_i) = \int_{\gamma_i} \sqrt{Q} \quad \text{for } 1 \leq i \leq n.$$

Theorem 3.1 is then the statement that these functions are local coordinates on M . Note that the associated linear coordinates $\theta_i = (dz_i, -)$ on the fibres of the bundle $T_M \rightarrow M$ considered in Section 2.2 correspond, under the isomorphism of Theorem 3.1, to the functions on the fibres of the bundle $\mathcal{H} \rightarrow M$ given by pairing with the classes γ_i .

We can use the isomorphism of Theorem 3.1 to transfer the data $(\mathcal{H}^{\mathbb{Z}}, \nabla^{\text{GM}}, \delta)$ from the bundle \mathcal{H} to the tangent bundle T_M . This defines a period structure $(T_M^{\mathbb{Z}}, \nabla, Z)$ on M for which the periods (25) are integral linear coordinates. The required identity $\nabla(Z) = \text{id}$ holds by definition. The intersection form on \mathcal{H}^* induces a skew-symmetric form $T_M^* \times T_M^* \rightarrow \mathcal{O}_M$ which takes integral values on the lattice $(T_M^{\mathbb{Z}})^* \subset T_M^*$. We obtain a period structure with skew form by taking η to be this form multiplied by $2\pi i$. Since the intersection form is nondegenerate, the inverse to η defines a symplectic form $\omega \in H^0(M, \wedge^2 T_M^*)$.

3.4 Prym variety and abelian connections

We denote by $J(C)$ and $J(\Sigma)$ the Jacobians of the curves C and Σ . Set

$$J(\Sigma)^- = \{M \in J(\Sigma) : M \otimes \sigma^*(M) \cong \mathcal{O}_\Sigma\}, \quad J^2(C) = \{P \in J(C) : P^{\otimes 2} \cong \mathcal{O}_C\}.$$

The pullback map $p^*: J(C) \rightarrow J(\Sigma)$ is injective [Mumford 1974, Section 3], and we identify $J(C)$ with its image. The Prym variety is defined by either of the two quotients

$$(26) \quad P(\Sigma) = J(\Sigma)/J(C) = J(\Sigma)^-/J^2(C).$$

To see that the two quotients in (26) are indeed the same, consider the map $j: J(\Sigma)^- \rightarrow J(\Sigma)/J(C)$ induced by the inclusion $J(\Sigma)^- \subset J(\Sigma)$. Then j is surjective because for any $M \in J(\Sigma)$ we can write $M = N^{\otimes 2}$, and then

$$M = (N \otimes \sigma^*(N^*)) \otimes (N \otimes \sigma^*(N)).$$

The first factor clearly lies in $J(\Sigma)^-$, and it is proved in [Mumford 1974, Section 3] that the second lies in $J(C)$. The kernel of j is the intersection $J(C) \cap J(\Sigma)^- \subset J(\Sigma)$, and since any element $M \in J(C)$ satisfies $\sigma^*(M) = M$, this coincides with $J^2(C)$.

We also consider the spaces $J^\#(C)$ and $J^\#(\Sigma)$ of line bundles with connection. We can again identify $J^\#(C)$ with the image of the pullback map $p^*: J^\#(C) \rightarrow J^\#(\Sigma)$. We set

$$J^\#(\Sigma)^- = \{(M, \partial) \in J^\#(\Sigma) : (M, \partial) \otimes \sigma^*(M, \partial) \cong (\mathcal{O}_\Sigma, d)\}.$$

Note that if (N, ∂_N) is a line bundle with connection on C , and $P^{\otimes 2} \cong N$, then there is a unique connection ∂_P such that $(P, \partial_P)^{\otimes 2} \cong (N, \partial_N)$. In particular, each $P \in J^2(C)$ has a unique connection ∂_P satisfying $(P, \partial_P)^{\otimes 2} = (\mathcal{O}_C, d)$, and we can therefore identify $J^2(C)$ with a subgroup of $J^\#(\Sigma)$. We define

$$(27) \quad P^\#(\Sigma) = J^\#(\Sigma)/J^\#(C) = J^\#(\Sigma)^-/J^2(C),$$

with a similar argument as before showing that these two quotients are equal.

The exact sequence (23) shows that

$$\tilde{H}^1(\Sigma, \mathbb{C}^*)^- := \text{Hom}_{\mathbb{Z}}(H_1(\Sigma, \mathbb{Z})^-, \mathbb{C}^*) \cong H^1(\Sigma, \mathbb{C}^*)/H^1(C, \mathbb{C}^*).$$

The Riemann–Hilbert isomorphism $J^\#(\Sigma) \cong H^1(\Sigma, \mathbb{C}^*)$ then induces an isomorphism

$$(28) \quad P^\#(\Sigma) \cong \tilde{H}^1(\Sigma, \mathbb{C}^*)^-.$$

3.5 Anti-invariant branched connections

Let F be a vector bundle on the spectral curve Σ . By a branched connection on F we mean a meromorphic connection $\partial: F \rightarrow F \otimes \omega_\Sigma(R)$ with simple poles on the branch divisor $R \subset \Sigma$. The line bundle $\mathcal{O}_\Sigma(R)$ has a canonical branched connection $d: \mathcal{O}_\Sigma(R) \rightarrow \mathcal{O}_\Sigma(R) \otimes \omega_\Sigma(R)$ induced by the de Rham differential applied to functions on Σ with poles on R . This connection has a simple pole with residue -1 at each point of R .

The short exact sequence (22) induces an isomorphism $p^*(\omega_C) \cong \mathcal{O}_\Sigma(R)$. We therefore obtain a canonical branched connection d_ϕ on $p^*(\omega_C)$. It is uniquely defined by the condition that ϕ is a flat map of bundles with meromorphic connection $(\mathcal{O}_\Sigma, d) \rightarrow (p^*(\omega_C), d_\phi)$. We say that a line bundle with branched connection (L, ∂) on Σ is anti-invariant if

$$(29) \quad (L, \partial_L) \otimes \sigma^*(L, \partial_L) \cong (p^*(\omega_C), d_\phi).$$

It follows that ∂ has a simple pole with residue $-\frac{1}{2}$ at each point of R .

Let $J_{\text{br}}^\#(\Sigma)$ denote the space of line bundles L on Σ equipped with anti-invariant branched connections ∂ . The group $J^2(C) \subset J^\#(\Sigma)$ acts on this space by tensor product, and in analogy with (27) we define

$$P_{\text{br}}^\#(\Sigma) = J_{\text{br}}^\#(\Sigma)/J^2(C).$$

Lemma 3.2 *There is a canonical isomorphism $P^\#(\Sigma) \cong P_{\text{br}}^\#(\Sigma)$.*

Proof Tensor product gives $J_{\text{br}}^\#(\Sigma)$ the structure of a torsor over $J^\#(\Sigma)^-$, so choosing a point $(L_0, \partial_0) \in J_{\text{br}}^\#(\Sigma)$ gives a noncanonical identification of the two spaces

$$(30) \quad (L, \partial) \in J^\#(\Sigma) \mapsto (L, \partial) \otimes (L_0, \partial_0) \in J_{\text{br}}^\#(\Sigma)^-,$$

which descends to the quotients by $J^2(C)$. To obtain a canonical bijection, take a spin structure $\omega_C^{1/2}$ on C and let ∂_0 be the unique branched connection on $L_0 = p^*(\omega_C^{1/2})$ satisfying $(L_0, \partial_0)^{\otimes 2} = (p^*(\omega_C), d_\phi)$. Since $\omega_C^{1/2}$ is uniquely defined up to the action of $J^2(C)$, the resulting isomorphism $P^\#(\Sigma) \cong P_{\text{br}}^\#(\Sigma)$ is canonically defined. \square

4 The spectral correspondence

Let us fix a smooth projective curve C of genus g and a quadratic differential $Q \in H^0(C, \omega_C^{\otimes 2})$ with simple zeroes. Let $p: \Sigma \rightarrow C$ be the associated spectral curve with its covering involution σ . There is a very well-known correspondence relating $\text{SL}_2(\mathbb{C})$ Higgs bundles on C to anti-invariant line bundles on Σ . In this section we show how to extend this construction to include connections. The existence of this extension seems to be little known, although it is discussed by Donagi and Pantev [2009, Section 3.2] and also appears in a paper of Arinkin [2005].

4.1 Definition

The $\text{SL}_2(\mathbb{C})$ spectral correspondence [Hitchin 1987a, Section 8; Beauville et al. 1989, Section 3] defines a bijection between

- (i) rank 2 vector bundles E on C , with $\det(E) \cong \mathcal{O}_C$, equipped with a Higgs field $\Phi: E \rightarrow E \otimes \omega_C$ with $\text{tr}(\Phi) = 0$ and $\frac{1}{2} \text{tr}(\Phi^2) = Q$, and
- (ii) line bundles L on Σ satisfying $L \otimes \sigma^*(L) \cong p^*(\omega_C)$.

The line bundle L is obtained from the eigendecomposition of the pullback of the Higgs field $p^*(\Phi)$. In the reverse direction, the line bundle L is sent to the bundle $E = p_*(L)$ equipped with the Higgs field Φ which is the pushforward of the map $1 \otimes \phi: L \rightarrow L \otimes p^*(\omega_C)$.

We shall need an extension of this correspondence involving connections on the bundles L and $E = p_*(L)$. More precisely, the extension relates

- (i) connections ∇ on E inducing the trivial connection on $\det(E) \cong \mathcal{O}_C$,
- (ii) anti-invariant branched connections ∂ on the line bundle L .

We do not claim that this correspondence is a bijection in general, but it does define a birational map of the relevant moduli spaces: see Theorem 5.1 below.

The extended correspondence is defined as follows. Take a bundle $E = p_*(L)$ on C and a connection ∇ on E as in (i). The natural transformation $p^* p_*(L) \rightarrow L$ gives rise to a short exact sequence

$$(31) \quad 0 \rightarrow p^*(E) \xrightarrow{f} L \oplus \sigma^*(L) \rightarrow \mathcal{O}_R \rightarrow 0.$$

Since the map f is an isomorphism away from the divisor $R \subset \Sigma$, there is a unique meromorphic connection $\tilde{\nabla}$ on $L \oplus \sigma^*(L)$ with poles along R such that the map f becomes a map of bundles with meromorphic connection $p^*(E, \nabla) \rightarrow (L \oplus \sigma^*(L), \tilde{\nabla})$. The local calculation in the next subsection shows that the poles of $\tilde{\nabla}$ at the points of R are simple. Taking the component of $\tilde{\nabla}$ along L then gives the required branched connection ∂ .

To prove that (L, ∂) satisfies (29), take determinants of (31) to get a map $\det(f): \mathcal{O}_\Sigma \rightarrow L \otimes \sigma^*(L)$. The support of the cokernel is precisely R and it follows for degree reasons that $\text{coker}(\det(f)) = \mathcal{O}_R$. The sequence (22) then shows that we can identify $L \otimes \sigma^*(L)$ with $p^*(\omega_C)$ in such a way that the map $\det(f)$ coincides with the map ϕ . But then $\det(\tilde{\nabla}) = \partial \otimes 1 + 1 \otimes \sigma^*(\partial)$ is related to the trivial connection $\det(p^*(\nabla))$ by the meromorphic gauge change ϕ , and hence coincides with $d\phi$.

4.2 Local computation at branchpoint

Consider a σ -invariant neighbourhood $U \subset \Sigma$ of a branchpoint $p \in R$. Choose a local coordinate $w: U \rightarrow \mathbb{C}$ satisfying $\sigma^*(w) = -w$, and hence $w(p) = 0$. Choose also a local nonvanishing section $s \in H^0(U, L)$. In terms of the basis of sections $(s, \sigma^*(s))$ we can write the induced connection $\tilde{\nabla}$ on $L \oplus \sigma^*(L)$ in the form

$$\tilde{\nabla} = d + \begin{pmatrix} \alpha(w) & \beta(w) \\ \gamma(w) & \delta(w) \end{pmatrix} dw,$$

where $\alpha, \beta, \gamma, \delta: U \rightarrow \mathbb{C}$ are meromorphic functions defined in a neighbourhood of $0 \in \mathbb{C}$, and regular away from 0. The invariance of the pullback connection on $p^*(E)$ implies that $\gamma(w) = -\beta(-w)$ and $\delta(w) = -\alpha(-w)$.

The sequence (31) shows that a section of $p^*(E)$ over U is determined by sections of L and $\sigma^*(L)$ which agree at the branchpoint p . Since derivatives of regular sections of $p^*(E)$ are also regular sections of $p^*(E)$ we can write

$$(32) \quad \tilde{\nabla}_{\frac{\partial}{\partial w}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha(w) + \beta(w) \\ -\alpha(-w) - \beta(-w) \end{pmatrix} = \begin{pmatrix} c_+ \\ c_+ \end{pmatrix} + O(w),$$

$$(33) \quad \tilde{\nabla}_{\frac{\partial}{\partial w}} \begin{pmatrix} w \\ -w \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + w \begin{pmatrix} \alpha(w) - \beta(w) \\ \alpha(-w) - \beta(-w) \end{pmatrix} = \begin{pmatrix} c_- \\ c_- \end{pmatrix} + O(w),$$

with $c_{\pm} \in \mathbb{C}$. It follows that $\alpha(w)$ and $\beta(w)$ have at worst simple poles at $w = 0$ and we can therefore write

$$\alpha(w) = -\frac{a}{2w} + c + O(w), \quad \beta(w) = \frac{b}{2w} - d + O(w),$$

with $a, b, c, d \in \mathbb{C}$. Equation (32) then implies that $b = a$ and $d = c$, and (33) implies that $a + b = 2$.

Thus

$$\alpha(w) = -\frac{1}{2w} + c + O(w), \quad \beta(w) = \frac{1}{2w} - c + O(w),$$

for some element $c \in \mathbb{C}$. In particular, $\tilde{\nabla}$ has simple poles on R . The induced branched connection on L is given by $\partial = d + \alpha(w) dw$.

4.3 Explicit description

It will be useful later to have a more detailed description of the extended spectral correspondence described in Section 4.1.

The functor $p_* : \text{Coh}(\Sigma) \rightarrow \text{Coh}(C)$ has a left adjoint p^* and a right adjoint $p^!$. They are related by $p^!(-) = p^*(-) \otimes \omega_{\Sigma/C}$, where $\omega_{\Sigma/C} = \omega_{\Sigma} \otimes p^*(\omega_C^\vee)$ is the relative dualising sheaf. The short exact sequence (21) induces an identification of $\omega_{\Sigma/C}$ with $\mathcal{O}_{\Sigma}(R)$. We can then identify $p^!$ with the functor $p^*(-) \otimes \mathcal{O}_{\Sigma}(R)$. There are natural transformations

$$(34) \quad \eta : p^* \circ p_* \rightarrow 1, \quad \chi : 1 \rightarrow p^! \circ p_*.$$

Let L be a line bundle on Σ . Setting $f = (\eta_L, \eta_{\sigma^*(L)})$ and $g = (\chi_L, \chi_{\sigma^*(L)})$ gives maps

$$(35) \quad p^*(E) \xrightarrow{f} L \oplus \sigma^*(L) \xrightarrow{g} p^*(E) \otimes \mathcal{O}_{\Sigma}(R).$$

We claim that they are mutually inverse away from the ramification divisor:

Lemma 4.1 *Let $s_R : \mathcal{O}_{\Sigma} \rightarrow \mathcal{O}_{\Sigma}(R)$ denote the canonical inclusion. Then*

$$(36) \quad g \circ f = 1_{p^*(E)} \otimes s_R, \quad f \otimes \omega_{\Sigma/C} \circ g = 1_{L \oplus \sigma^*(L)} \otimes s_R.$$

Proof As in the previous section we can view local sections of $p^*(E)$ as consisting of pairs of local sections u of L and v of $\sigma^*(L)$ whose restrictions to the branch divisor R coincide. Then $\eta_L : p^*(E) \rightarrow L$ sends such a pair to the local section u , and $\chi_L(-R) : L(-R) \rightarrow p^*(E)$ sends a local section u of L which vanishes on R to the pair $(u, 0)$. The result follows. \square

The extended correspondence is defined by viewing the maps f and g as meromorphic gauge transformations, and using them to transfer the connection $p^*(\nabla)$ from $p^*(E)$ to $L \oplus \sigma^*(L)$. We then take the first component to obtain a meromorphic connection ∂ on L with poles along R . The connection ∂ is therefore given by the composite map

$$(37) \quad L \xrightarrow{\chi_L} p^*(E) \otimes \mathcal{O}_{\Sigma}(R) \xrightarrow{p^*(\nabla) \otimes 1 + 1 \otimes d} p^*(E) \otimes \mathcal{O}_{\Sigma}(R) \otimes \omega_{\Sigma}(R) \xrightarrow{\eta_L \otimes \omega_{\Sigma}(2R)} L \otimes \omega_{\Sigma}(2R).$$

Here d denotes the canonical branched connection on $\mathcal{O}_{\Sigma}(R)$ induced by the de Rham differential, as in Section 3.5. Note that although the resulting connection ∂ a priori has double poles along R , the local calculation in Section 4.2 shows that these poles are in fact of order one.

5 Two diagrams of moduli spaces

In this section we introduce a diagram of moduli spaces which will play a key role in our construction of the Joyce structure on the space $\mathcal{M}(C, Q)$. These moduli spaces parametrise smooth projective curves C of genus g equipped with various extra structures involving vector bundles, Higgs fields and connections.

We defer discussion of the existence and basic properties of the moduli spaces to the appendix. The main result of this section focuses on the two most interesting maps in our diagram: α and β_ϵ . We show that the first is birational and the second is generically étale.

5.1 The diagrams

We use the following notation:

- C is a complex projective curve of genus g , equipped with a level ℓ structure.
- $Q \in H^0(C, \omega_C^{\otimes 2})$ is a quadratic differential on C with simple zeroes.
- $p: \Sigma \rightarrow C$ is the spectral curve defined by the quadratic differential Q .
- E is a stable rank 2 vector bundle on C with trivial determinant.
- Φ is a trace-free Higgs field on E such that $\frac{1}{2} \text{tr}(\Phi^2)$ has simple zeroes.
- ∇ is a connection on E inducing the trivial connection on $\det(E)$.
- L is a line bundle on Σ such that $L \otimes \sigma^*(L) \cong p^*(\omega_C)$ and $p_*(L)$ is stable.
- ∂ is an anti-invariant branched connection on L .

Let us fix a parameter $\epsilon \in \mathbb{C}^*$ and contemplate the diagram

$$\begin{array}{ccccc}
 & & \mathcal{M}(C, E, \nabla, \Phi) & & \\
 & \swarrow \alpha & & \searrow \beta_\epsilon & \\
 (38) \quad \mathcal{M}(C, Q, L, \partial) & & & & \mathcal{M}(C, Q, E, \nabla) \xrightarrow{\rho'} \mathcal{M}(C, E, \nabla) \\
 \pi_3 \downarrow & & & & \pi_2 \downarrow \qquad \qquad \qquad \pi_1 \downarrow \\
 \mathcal{M}(C, Q) \xleftarrow{=} \mathcal{M}(C, Q) & \xrightarrow{\rho} & & & \mathcal{M}(C)
 \end{array}$$

Each moduli space parametrises the indicated objects, and the maps ρ, ρ' and π_i are the obvious projections. Note that $\mathcal{M}(C)$ and $\mathcal{M}(C, Q) = M$ are the moduli spaces appearing in Section 3.1. The map α is the extended spectral correspondence discussed in the previous subsection, and $\beta(\epsilon)$ is defined by the rule

$$\beta_\epsilon(C, E, \nabla, \Phi) = (C, \frac{1}{2} \text{tr}(\Phi^2), E, \nabla - \epsilon^{-1} \Phi).$$

We refer the reader to the appendix for further details on the moduli spaces appearing in (38). Given our standing assumptions $g > 1$ and $\ell > 2$, all the spaces appearing are smooth quasiprojective complex varieties which co-represent the relevant moduli functors.

Consider the bundle $\mathcal{H}^\# = \mathcal{H}/(2\pi i)\mathcal{H}^{\mathbb{Z}}$ over M with fibres $\tilde{H}^1(\Sigma, \mathbb{C}^*)^-$. Under the isomorphism of Theorem 3.1 it corresponds to the quotient (14). We now consider a second diagram of spaces, which can

be attached to the left-hand side of (38):

$$(39) \quad \begin{array}{ccccc} T_M^\# & \xrightarrow{\nabla^{\text{GM}}(\delta)} & \mathcal{H}^\# & \xleftarrow{\tau} & \mathcal{M}(C, Q, L, \partial) \\ \pi_5 \downarrow & & \pi_4 \downarrow & & \pi_3 \downarrow \\ M & \xleftarrow{=} & M & \xleftarrow{=} & \mathcal{M}(C, Q) \end{array}$$

Note that the fibre of the map π_3 over a pair (C, Q) is the open subset of the space $J_{\text{br}}^\#(\Sigma)$ defined by the condition that $p_*(L)$ is stable. The map τ is then given on fibres by the composite of the quotient map $J_{\text{br}}^\#(\Sigma) \rightarrow P_{\text{br}}^\#(\Sigma)$ with the isomorphism of Lemma 3.2 and the Riemann–Hilbert isomorphism (28). It is an étale map of complex manifolds.

5.2 Generic finiteness results

The following result contains the two main nontrivial facts we will need for the construction of our Joyce structure.

Theorem 5.1 *Fix a genus $g > 1$ and a level $\ell > 2$. Then*

- (i) *the map α is birational, and*
- (ii) *the map β_ϵ is generically étale.*

Proof Note first that the sources and targets of the maps α and β_ϵ are smooth quasiprojective varieties of the same dimension. By generic smoothness and the dimension theorem, for (i) it will be enough to prove that the general fibre of α is a single point, and for (ii) that the general fibre of β_ϵ is finite, or equivalently, that β_ϵ is dominant.

For (i), suppose we have two connections on E giving rise to the same connection on L . In terms of the local computation of Section 4.2 this means that we have two possible β_i with the same α , and in particular, the same $c \in \mathbb{C}$. Then the difference $\beta_2(w) - \beta_1(w)$ is regular and vanishes at the branchpoint $w = 0$. Globally the difference $(\beta_2(w) - \beta_1(w)) dw$ corresponds to a section $\sigma^*(L) \rightarrow L \otimes \omega_\Sigma$. But since Σ has genus $4g - 3$, and R has degree $4g - 4$

$$(40) \quad \chi(\sigma^*(L), L \otimes \omega_\Sigma(-R)) = \chi(\omega_\Sigma(-R)) = (4g - 3) - 1 - (4g - 4) = 0,$$

and it follows that for generic L any such section is zero.

For (ii), note first that by the relation $\nabla_\epsilon = \nabla - \epsilon^{-1}\Phi$, the source of β_ϵ may be equivalently viewed as parametrising C, E, Φ and ∇_ϵ , and therefore β_ϵ is a base-change of the map

$$(41) \quad \gamma: \mathcal{M}(C, E, \Phi) \rightarrow \mathcal{M}(C, E, Q)$$

defined by $Q = \frac{1}{2} \text{tr}(\Phi^2)$. That this map is dominant follows from the proof of [Beauville et al. 1989, Theorem 1]. Namely, we first extend γ by dropping the condition that Q has simple zeroes, and then show that the resulting map is dominant. But by the results of [Laumon 1988], for each curve C there exists a bundle E for which the fibre of γ over the point $(C, E, 0)$ is a single point. □

Note that the proof of Theorem 5.1 actually gives more: the open locus over which the map α is birational intersects each fibre $\pi_3^{-1}(m)$, and similarly, the open locus over which the map β_ϵ is étale intersects each fibre $\pi_2^{-1}(m)$. We record this as:

Corollary 5.2 *Let C be a smooth projective curve of genus $g > 1$, and $Q \in H^0(C, \omega_C^{\otimes 2})$ a quadratic differential with simple zeroes. Let $p: \Sigma \rightarrow C$ be the associated spectral curve. Then for each $\epsilon \in \mathbb{C}^*$, the composite $\beta_\epsilon \circ \alpha^{-1}$ defines a dominant rational map*

$$(42) \quad \gamma_\epsilon: \mathcal{M}_\Sigma(L, \partial) \dashrightarrow \mathcal{M}(E, \nabla)$$

between the moduli space of anti-invariant branched connections on Σ , and the moduli space of rank 2 connections on C with trivial determinant. □

6 Symplectic forms and their preservation

In this section we introduce the relevant symplectic forms on the moduli spaces and show that our maps preserve them.

6.1 The Atiyah–Bott symplectic form

Let E be a bundle on C and $\nabla: E \rightarrow E \otimes \omega_C$ a connection. There is an associated de Rham complex

$$(43) \quad 0 \rightarrow \mathcal{E}nd_{\mathcal{O}_C}(E) \xrightarrow{\nabla} \mathcal{E}nd_{\mathcal{O}_C}(E) \otimes \omega_C \rightarrow 0.$$

We denote by $\mathbb{H}_{\text{dR}}^i(C, \nabla)$ the i^{th} hypercohomology of this complex. The long exact sequence in hypercohomology gives relations

$$(44) \quad \mathbb{H}_{\text{dR}}^0(C, \nabla) = H^0(C, \mathcal{E}nd_{\mathcal{O}_C}(E)), \quad \mathbb{H}_{\text{dR}}^2(C, \nabla) = H^1(C, \mathcal{E}nd_{\mathcal{O}_C}(E) \otimes \omega_C),$$

and a short exact sequence

$$(45) \quad 0 \rightarrow H^0(C, \mathcal{E}nd_{\mathcal{O}_C}(E) \otimes \omega_C) \rightarrow \mathbb{H}_{\text{dR}}^1(C, \nabla) \rightarrow H^1(C, \mathcal{E}nd_{\mathcal{O}_C}(E)) \rightarrow 0.$$

Consider the moduli space $\text{Flat}_C(r)$ of rank r stable bundles on C equipped with a flat connection. It is a smooth, quasiprojective scheme whose tangent space at a point (E, ∇) is the group $\mathbb{H}_{\text{dR}}^1(C, \nabla)$. There is a canonical isomorphism $\int_C: H^1(C, \omega_C) \rightarrow \mathbb{C}$. The Atiyah–Bott form is the composite of the wedge product

$$(46) \quad \mathbb{H}_{\text{dR}}^1(C, \nabla) \times \mathbb{H}_{\text{dR}}^1(C, \nabla) \xrightarrow{\wedge} \mathbb{H}_{\text{dR}}^2(C, \nabla)$$

with the canonical maps

$$(47) \quad \mathbb{H}_{\text{dR}}^2(C, \nabla) \xrightarrow{\cong} H^1(C, \mathcal{E}nd_{\mathcal{O}_C}(E) \otimes \omega_C) \xrightarrow{\text{tr}} H^1(C, \omega_C) \xrightarrow{\int_C} \mathbb{C}.$$

There is a forgetful map

$$(48) \quad \pi : \text{Flat}_C(r) \rightarrow \text{Bun}_C(r)$$

to the moduli space of rank r , degree 0 stable bundles on C . The tangent space at a point of $\text{Bun}_C(r)$ is the cohomology group $\text{Ext}_C^1(E, E)$. The map π is an affine bundle for the vector bundle over $\text{Bun}_C(r)$ with fibres $\text{Hom}_C(E, E \otimes \omega_C)$, which can be identified with the cotangent bundle of $\text{Bun}_C(r)$ using the Serre duality pairing

$$(49) \quad \text{Hom}_C(E, E \otimes \omega_C) \times \text{Ext}_C^1(E, E) \rightarrow \text{Ext}^1(E, E \otimes \omega_C) \xrightarrow{\text{tr}} H^1(C, \omega_C) \xrightarrow{\int_C} \mathbb{C}.$$

The short exact sequence (1) defined by the derivative of the map π can be identified with (45), and the following result is then immediate.

Lemma 6.1 *The Atiyah–Bott form on $\text{Flat}_C(r)$ is uniquely characterised by the following two properties:*

- (i) *The fibres of the map (48) are Lagrangian.*
- (ii) *At any point (E, ∇) the induced pairing between the vertical tangent space for the map (48) and the cotangent space $T_E^* \text{Bun}_C(r)$ is the Serre duality pairing (49).*

Note that in the case of bundles of rank $r = 1$ the de Rham complex for (E, ∇) becomes the usual de Rham complex of C and hence $\mathbb{H}_{\text{dR}}^1(C, \nabla) = H^1(C, \mathbb{C})$. The Atiyah–Bott pairing can then be identified with the intersection form.

6.2 Properties of traces

Let X be a variety, and E a vector bundle on X . Taking the trace of endomorphisms defines a map $\text{tr} : \text{End}_{\mathcal{O}_X}(E) \rightarrow \mathcal{O}_X$. Tensoring by a line bundle L and taking cohomology gives linear maps

$$(50) \quad \text{tr}_E : \text{Ext}_X^p(E, E \otimes L) \rightarrow H^p(X, L).$$

In the proof of Theorem 6.2 we shall need the following properties of these maps, whose proofs we leave to the reader:

- (T1) Given bundles E, F and a line bundle L , and elements $g \in \text{Ext}_X^p(E, F)$ and $h \in \text{Ext}_X^q(F, E \otimes L)$, there is an identity

$$(51) \quad \text{tr}_E(h \circ g) = \text{tr}_F((g \otimes L) \circ h) \in H^{p+q}(X, L).$$

- (T2) Given a bundle E and line bundles L, M and elements $f \in \text{Ext}_X^p(E, E \otimes L)$ and $g \in \text{Ext}_X^q(L, M)$, there is an identity

$$(52) \quad g \circ \text{tr}_E(f) = \text{tr}_E((1_E \otimes g) \circ f).$$

(T3) Suppose given two bundles E_1, E_2 and a line bundle L , and an element $f \in \text{Ext}_X^p(G, G \otimes L)$, where $G = E_1 \oplus E_2$. Let $s_i: E_i \rightarrow G$ be the canonical inclusions, and $\pi_i: G \rightarrow E_i$ the canonical inclusions. Then

$$(53) \quad \text{tr}_G(f) = \text{tr}_{E_1}(\pi_1 \circ f \circ s_1) + \text{tr}_{E_2}(\pi_2 \circ f \circ s_2).$$

(T4) Given a map of varieties $f: X \rightarrow Y$, a bundle E on Y , a line bundle L on Y and an element $s \in \text{Ext}^p(E, E \otimes L)$, there is an identity

$$(54) \quad \text{tr}_{f^*(E)}(f^*(s)) = f^*(\text{tr}_E(s)) \in H^p(X, f^*(L)).$$

6.3 Preservation of symplectic forms

Each of the maps π_1, \dots, π_5 appearing in the diagram (38) and (39) carries a natural relative symplectic form, and we claim that the horizontal maps in these diagrams preserve these forms. For the most part these statements are rather obvious, but Theorem 6.2 below is quite nontrivial.

Let us consider each vertical map in turn. For the map π_3 we take the Atiyah–Bott form for rank 1 bundles with connection, restricted to the subset of anti-invariant connections. The tangent space to the fibres is $H^1(\Sigma, \mathbb{C})^-$ and the symplectic form is just the usual intersection form. We then take the same form on π_4 and by definition of the symplectic form ω on M this induces the required relative symplectic form on the map π_5 .

The map π_1 is equipped with a relative form whose restriction to each fibre is the Atiyah–Bott form discussed above. Note that since we are in the $\text{SL}_2(\mathbb{C})$ setting we should impose trace-free conditions appropriately. Since the right-hand square in (38) is Cartesian this induces a relative symplectic form on π_2 . In terms of the map $\gamma_\epsilon = \beta_\epsilon \circ \alpha^{-1}$ of Corollary 5.2 what is then left to prove is

Theorem 6.2 *The pullback of the Atiyah–Bott form on $\mathcal{M}_C(E, \nabla)$ by the rational map γ_ϵ is twice the Atiyah–Bott form on $\mathcal{M}_\Sigma(L, \partial)$.*

Proof We freely use notation from Section 4.3. At the level of bundles γ_ϵ is defined by $E = p_*(L)$. We have a map of short exact sequences in which all vertical arrows are isomorphisms

$$\begin{array}{ccccc} \text{Hom}_\Sigma(L, L \otimes \omega_\Sigma)_0 & \longrightarrow & T_{(L, \partial)}\mathcal{M}_\Sigma(L, \partial) & \longrightarrow & \text{Ext}_\Sigma^1(L, L)_0 \\ \gamma_{\epsilon,*} \uparrow & & \gamma_{\epsilon,*} \uparrow & & \downarrow p_* \\ \text{Hom}_C(E, E \otimes \omega_C)_0 & \longrightarrow & T_{(E, \nabla)}\mathcal{M}_C(E, \nabla) & \longrightarrow & \text{Ext}_C^1(E, E)_0 \end{array}$$

Take L an anti-invariant line bundle on Σ and set $E = p_*(L)$. Take elements

$$(55) \quad v \in \text{Hom}_C(E, E \otimes \omega_C), \quad w \in \text{Hom}_\Sigma^1(L, L).$$

We assume that $\text{tr}(v) = 0$ and w is anti-invariant, meaning that $w \otimes 1_{\sigma^*(L)} + 1_L \otimes \sigma^*(w) = 0$. What we must show is that

$$(56) \quad \int_C \text{tr}_E(v \circ p_*(w)) = \int_\Sigma \text{tr}_L(\alpha_*(v) \circ w).$$

By construction, $\alpha_*(v)$ is given by the composite

$$(57) \quad L \xrightarrow{\chi_L} p^*(E) \otimes \omega_{\Sigma/C} \xrightarrow{p^*(v) \otimes \omega_{\Sigma/C}} p^*(E) \otimes \omega_\Sigma \xrightarrow{\eta_L \otimes \omega_\Sigma} L \otimes \omega_\Sigma.$$

Thus we have

$$(58) \quad \int_\Sigma \text{tr}_L(\alpha_*(v) \circ w) = \int_\Sigma \text{tr}_L(\eta_L \otimes \omega_{\Sigma/C} \circ p^!(v) \circ \chi_L \circ w).$$

On the other hand, using the relation $\int_C \psi = \int_\Sigma (s \circ p^*(\psi))$, valid for all elements $\psi \in H^1(C, \omega_C)$, we can write

$$(59) \quad \int_C \text{tr}_E(v \circ p_*(w)) = \int_\Sigma s \circ p^* \text{tr}_E(v \circ p_*(w))$$

$$(60) \quad \stackrel{(T4)}{=} \int_\Sigma s \circ \text{tr}_{p^*(E)}(p^*(v) \circ p^* p_*(w))$$

$$(61) \quad \stackrel{(T2)}{=} \int_\Sigma \text{tr}_{p^*(E)}((1_{p^*(E)} \otimes s) \circ p^*(v) \circ p^* p_*(w))$$

$$(62) \quad = \int_\Sigma \text{tr}_{p^*(E)}(g \circ f \circ p^*(v) \circ p^* p_*(w))$$

$$(63) \quad \stackrel{(T1)}{=} \int_\Sigma \text{tr}_{L \oplus \sigma^*(L)}(f \otimes \omega_{\Sigma/C} \circ p^!(v) \circ p^! p_*(w) \circ g)$$

$$(64) \quad \stackrel{(T3)}{=} \int_\Sigma \text{tr}_L(\eta_L \otimes \omega_{\Sigma/C} \circ p^!(v) \circ \chi_L \circ w) + \int_\Sigma \text{tr}_{\sigma^*(L)}(\eta_{\sigma^*(L)} \otimes \omega_{\Sigma/C} \circ p^!(v) \circ \chi_{\sigma^*(L)} \circ \sigma^*(w)),$$

where we used the naturality of χ as well as (T3) in the final step. The same argument together with the assumption $\text{tr}_E(v) = 0$ shows that

$$(65) \quad \int_\Sigma \text{tr}_L(\eta_L \otimes \omega_{\Sigma/C} \circ p^!(v) \circ \chi_L) + \int_\Sigma \text{tr}_{\sigma^*(L)}(\eta_{\sigma^*(L)} \otimes \omega_{\Sigma/C} \circ p^!(v) \circ \chi_{\sigma^*(L)}) = 0.$$

For line bundles, (T2) shows that the trace of the composite is the composite of the traces. Using the assumption that w is anti-invariant we see that the two terms in (62) are equal, which proves the claim. \square

7 Construction of the Joyce structure

In this section we finally construct the required meromorphic Joyce structure on the complex manifold $M = \mathcal{M}(C, Q)$. The key ingredient is the isomonodromy connection on the map π_1 , which is both flat and symplectic. We pull this back across the diagrams (38) and (39) using the generic finiteness results of Theorem 5.1. This then gives the required nonlinear connections h_ϵ on the tangent bundle of M .

7.1 Nonlinear connections

Let $\pi : X \rightarrow Y$ be a smooth map of varieties. We can define the notion of a nonlinear connection on π exactly as before. Namely, the derivative of π gives a short exact sequence

$$0 \rightarrow T_{X/M} \xrightarrow{i} T_X \xrightarrow{\pi_*} \pi^*(T_Y) \rightarrow 0,$$

and we define a nonlinear connection on π to be a map of bundles $h : \pi^*(T_Y) \rightarrow T_X$ satisfying $\pi_* \circ h = \text{id}$. Let $D \subset X$ be an effective Cartier divisor with corresponding section $s_D : \mathcal{O}_X \rightarrow \mathcal{O}_X(D)$. We define a meromorphic connection on π with poles along D to be a map of bundles $h : \pi^*(T_Y) \rightarrow T_X(D)$ satisfying $(\pi_* \otimes \mathcal{O}_X(D)) \circ h = 1_{\pi^*(T_Y)} \otimes s_D$.

We shall need the following simple facts, whose proofs we leave to the reader:

- (C1) Suppose given smooth maps $\pi : X \rightarrow Y$ and $\eta : Y \rightarrow Z$ and connections $h : \pi^*(T_Y) \rightarrow T_X$ on π and $j : \eta^*(T_Z) \rightarrow T_Y$ on η . Then the composite $h \circ \pi^*(j) : \pi^*\eta^*(T_Z) \rightarrow T_X$ is a connection on the map $\eta \circ \pi : X \rightarrow Z$.
- (C2) If $\pi : X \rightarrow Y$ is étale then $\pi_*^{-1} : \pi^*(T_Y) \rightarrow T_X$ is the unique connection on π .
- (C3) Given a Cartesian square

$$\begin{array}{ccc} W & \xrightarrow{g} & X \\ \eta \downarrow & & \downarrow \pi \\ Z & \xrightarrow{f} & Y \end{array}$$

with π smooth, and a connection $h : \pi^*(T_Y) \rightarrow T_X$ on π , there exists a unique connection $j : \eta^*(T_Z) \rightarrow T_W$ on η such that

$$(64) \quad g_* \circ j = g^*(h) \circ \eta^*(f_*) : \eta^*(T_Z) \rightarrow g^*(T_X).$$

- (C4) Suppose given a smooth map $\pi : X \rightarrow Y$ and an open subset $U \subset X$. Denote the inclusion map by $i : U \rightarrow X$, and set $\pi_0 = \pi \circ i$. Then any connection $h_0 : \pi_0^*(T_Y) \rightarrow T_U$ on π_0 extends to a meromorphic connection on π . More precisely, there is an effective divisor $D \subset X$ and a meromorphic connection $h : \pi^*(T_Y) \rightarrow T_X(D)$ on π with poles along D such that $i^*(h) = \text{id}_{T_U} \otimes i^*(s_D) \circ h_0$.

7.2 Isomonodromy connection

Return to the diagram (38), and recall that we are imposing the condition that the bundles E are stable. In particular, the map

$$(65) \quad \pi_1 : \mathcal{M}(C, E, \nabla) \rightarrow \mathcal{M}(C)$$

is smooth. There is a flat nonlinear connection on this map known as the isomonodromy connection and constructed as follows. There is a map

$$(66) \quad \pi_0 : \mathcal{M}(C, \mathcal{E}) \rightarrow \mathcal{M}(C)$$

whose fibre over a curve C is the space of $\mathrm{SL}_2(\mathbb{C})$ local systems on C . The relative Riemann–Hilbert correspondence [Deligne 1970, Theorem 2.23] gives an open embedding $\mu: \mathcal{M}(C, E, \nabla) \rightarrow \mathcal{M}(C, \mathcal{E})$ sending a bundle with connection (E, ∇) to its monodromy local system. The fact that the universal family of curves over $\mathcal{M}(C)$ is locally trivial as a family of smooth surfaces defines a Gauss–Manin connection on π_0 , which by pullback along μ then induces the isomonodromy connection on π_1 .

We shall need two properties of the isomonodromy connection. Firstly, despite the fact that the monodromy map μ is highly transcendental, the isomonodromy connection is nonetheless an algebraic object. The basic reason is that the condition for a relative connection (\mathcal{E}, ∇) on a family of curves $f: \mathcal{C} \rightarrow T$ to be isomonodromic can be rephrased as the lifting of the relative flat connection on the bundle \mathcal{E} over \mathcal{C} to an actual flat connection, and the isomonodromy connection then arises from a zero curvature condition. A more abstract approach using crystals was explained by Simpson [1994b, Section 8].

The second property of the isomonodromy connection we need is that it is symplectic. Goldman [1984] proved that the Riemann–Hilbert map μ takes the Atiyah–Bott symplectic form on $\mathcal{M}_C(E, \nabla)$ to a natural symplectic form on the character variety defined using group cohomology. The only important point for us is that this second symplectic form is defined topologically, and is therefore independent of the complex structure on C . It follows that the parallel transport maps for the isomonodromy connection preserve the Atiyah–Bott symplectic form on the fibres of π_1 .

The right-hand square in (38) is Cartesian so we can pull back the isomonodromy connection using (C3) to obtain a connection on π_2 . Using (C1), (C2) and Theorem 5.1 we obtain a connection on an open subset of π_3 . Note that the bundle of groups $J^2(C)$ over $\mathcal{M}(C)$ acts by tensor product on the upper row of the diagram (38). Tensoring (E, ∇) by an element $(P, \partial_P) \in J^2(C)$ multiplies the monodromy by a homomorphism $H_1(C, \mathbb{Z}) \rightarrow \{\pm 1\}$. This action clearly preserves the isomonodromy connection, and the connection on π_3 therefore descends along the map τ appearing in (39). Using (C4) we can extend it to a meromorphic connection. Continuing across this diagram, we finally obtain a meromorphic connection h_ϵ on a dense open subset of $\pi_5: T_M^\# \rightarrow M$.

7.3 Joyce structure

In the last section we showed how to construct, for each $\epsilon \in \mathbb{C}^*$, a nonlinear connection h_ϵ on a dense open subset of π_5 . This connection is flat and symplectic because it is a pullback of the isomonodromy connection. To produce a pre-Joyce structure it remains to prove that as $\epsilon \in \mathbb{C}^*$ varies these connections form a v -pencil, ie that $h_\epsilon = h + \epsilon^{-1}v$, where $h = h_\infty$.

Consider the automorphism

$$r_\epsilon: \mathcal{M}(C, E, \nabla, \Phi) \rightarrow \mathcal{M}(C, E, \nabla, \Phi), \quad (C, E, \nabla, \Phi) \mapsto (C, E, \nabla + \epsilon^{-1}\Phi, \Phi).$$

Then $\beta_\epsilon = \beta_\infty \circ r_{-\epsilon}$, and it follows that $h_\epsilon = (r_\epsilon)_* \circ h$ as maps of bundles $\pi^*(T_M) \rightarrow T_X$. It will be enough to show that $(r_\epsilon)_* = \mathrm{id} + \epsilon^{-1}(v \circ \pi_*)$.

Recall the tautological differential $\lambda \in H^0(\Sigma, \omega_\Sigma)$ of Section 3.1 whose periods (25) around a basis of cycles $(\gamma_1, \dots, \gamma_n) \subset H_1(\Sigma, \mathbb{Z})^-$ define local flat integral coordinates z_i on M . After transferring along the birational map α , the map r_ϵ becomes

$$r_\epsilon: \mathcal{M}(C, Q, L, \partial) \rightarrow \mathcal{M}(C, Q, L, \partial), \quad (C, Q, L, \partial) \mapsto (C, Q, L, \partial + \epsilon^{-1}\lambda).$$

Thus r_ϵ is the operation of tensoring $(L, \partial_L) \in J_{\text{br}}^\#(\Sigma)$ with $(\mathcal{O}_\Sigma, d + \epsilon^{-1}\lambda) \in J^\#(\Sigma)^-$. Note that the monodromy of this second connection around a cycle γ_i is just $\exp(\epsilon^{-1}z_i)$. We now transfer the automorphism r_ϵ across the diagram (39). The map σ takes a point of the space $\mathcal{M}(C, Q, L, \partial)$ to the monodromy of the product (30). Thus taking fibre coordinates θ_i as in Section 2.2, we finally arrive at the automorphism of $X = T_M$ given in local coordinates by $\theta_i \mapsto \theta_i + \epsilon^{-1}z_i$, and the claim follows.

The final step is to check the compatibility between the period structure and the pre-Joyce structure. Conditions (J1) and (J2) of Definition 2.9 hold by construction. It remains to consider (J3) and (J4).

For (J3) we consider the \mathbb{C}^* action on $M = \mathcal{M}(C, Q)$ for which $t \in \mathbb{C}^*$ acts by $t \cdot (C, Q) = (C, t^2 \cdot Q)$. Combining the induced action on $X = T_M$ with the rescaling action on the fibres as in the paragraph before Lemma 2.7 gives a \mathbb{C}^* action on $X = T_M$. It is not hard to see that this descends to $X^\#$, and that when transferred across the diagrams (38) and (39) it becomes the \mathbb{C}^* action on $\mathcal{M}(C, E, \nabla, \Phi)$ which sends $\Phi \mapsto t\Phi$ and leaves (C, E, ∇) fixed. We denote by $m_t: M \rightarrow M$ and $n_t: X \rightarrow X$ the resulting actions of $t \in \mathbb{C}^*$. Note that the involution $\iota: X \rightarrow X$ coincides with n_{-1} .

After Lemma 2.7, to check (J3) it will be enough to show that for any vector field v on M we have $(n_t)_*(h(v)) = h((m_t)_*(v))$. Taking $t = -1$ this will also imply (J4). To prove this identity it is enough to show that $(n_t)_*(h(v))$ is a horizontal vector field for h . But this is clear by construction of h since in the diagram (38), $\rho' \circ n_t = \rho'$.

7.4 Restriction to the zero-section

It was explained in [Bridgeland and Strachan 2021, Section 3.2] that the involution property (iii) above implies that, when restricted to the zero-section $M \subset X = T_M$, the holomorphic Levi-Civita connection of the complex hyperkähler structure on X induces a flat, torsion-free connection ∇^J on the tangent bundle of M . This connection was referred to in [Bridgeland 2021, Section 7] as the linear Joyce connection, and is given in coordinates by

$$(67) \quad \nabla_{\frac{\partial}{\partial z_i}}^J \left(\frac{\partial}{\partial z_j} \right) = \sum_{p,q} \eta_{qp} \cdot \frac{\partial^3 W}{\partial \theta_i \partial \theta_j \partial \theta_p} \Big|_{\theta=0} \cdot \frac{\partial}{\partial z_q}.$$

Note, however, that locating the poles of the Joyce structure constructed above is a subtle problem, and in particular it is not clear whether the structure is regular along the zero-section $M \subset X = T_M$. This submanifold $M \subset X$ is the fixed locus of the involution ι , and when transferred across the diagram (39) corresponds to the multisection of π_3 consisting of points satisfying $(L, \partial_L)^{\otimes 2} = (p^*(\omega_C), \partial_{\text{can}})$. Understanding the properties of the Joyce structure at these points is difficult however, because they lie in

the exceptional locus of the birational map α . Note that in the one example that has been computed in detail [Bridgeland and Masoero 2023], the connection ∇^J is indeed well-defined, and turns out to be quite natural.

8 Good Lagrangian submanifolds

The spaces of complex and Kähler parameters on a compact Calabi–Yau threefold are expected to appear as complex Lagrangian submanifolds in the stability space of the associated CY_3 triangulated category. It has been a longstanding question to try to abstractly characterise these submanifolds in stability space; see eg [Bridgeland 2009, Section 7]. In this section we give a general definition of a good Lagrangian submanifold $B \subset M$ in the base of a Joyce structure. We then prove that for the Joyce structures constructed in Section 7, the submanifolds in $M = \mathcal{M}(C, Q)$ obtained by fixing the curve C and varying the quadratic differential Q are good Lagrangians in this sense.

8.1 General definition

Consider a Joyce structure on a complex manifold M and a complex Lagrangian submanifold $B \subset M$. Consider the normal bundle $\pi : N_B \rightarrow B$ fitting into the sequence

$$(68) \quad 0 \rightarrow T_B \xrightarrow{i} T_M|_B \xrightarrow{k} N_B \rightarrow 0.$$

Recall the pencil of connections $h_\epsilon = h + \epsilon^{-1}v$ on the bundle $\pi : X = T_M \rightarrow M$. For any complex submanifold $B \subset M$, and any $\epsilon^{-1} \in \mathbb{C}$, the connection h_ϵ restricts to a connection $h_\epsilon|_B$ on the bundle $X_B = T_M|_B \rightarrow B$.

Definition 8.1 *A complex Lagrangian submanifold $B \subset M$ will be called good if the restricted connection $h_\epsilon|_B$ descends via the map $k : X_B \rightarrow N_B$ to a connection n on the normal bundle $\pi : N_B \rightarrow B$.*

To explain this condition in more detail, take $x \in X_B$ with $\pi(x) = b \in B$. The bundle map k defines a map of complex manifolds $k : X_B \rightarrow N_B$, and we set $y = k(x) \in N_B$.

$$(69) \quad \begin{array}{ccccc} X & \hookrightarrow & X_B & \xrightarrow{k} & N_B \\ \pi \downarrow & & \downarrow & & \downarrow \pi \\ M & \hookrightarrow & B & \xrightarrow{=} & B \end{array}$$

Given a vector $w \in T_b B \subset T_b M$ the connection h_ϵ defines a lift $h_\epsilon(w) \in T_x X_B \subset T_x X$, and we define $n(w) = k_*(h_\epsilon(w)) \in T_y N_B$. Note that $n(w)$ is independent of ϵ , since $k_*(v(w)) = 0$. The condition of Definition 8.1 is that $n(w)$ depends only on $y \in N_B$, not on the element $x \in X_B$ satisfying $k(x) = y$. When this condition holds, the map n defines a connection on $\pi : N_B \rightarrow B$.

Lemma 8.2 *If $B \subset M$ is a good Lagrangian then the induced connection n on the normal bundle $\pi : N_B \rightarrow B$ is flat.*

Proof Recall that if $f : M \rightarrow N$ is a map of complex manifolds, and u, w are vector fields on M, N , respectively, then u, w are said to be f -related if $f_*(u_m) = w_{f(m)}$ for all $m \in M$. Given vector fields u_1, u_2 on M which are f -related to vector fields w_1, w_2 on N , it is easily checked that $[u_1, u_2]$ is f -related to $[w_1, w_2]$. We will apply this to the map $k : X_B \rightarrow N_B$.

Given a vector field u on B , we can extend it to a vector field on M which we also denote by u . We can then use the connection h_ϵ to lift it to the vector field $h_\epsilon(u)$ on X . The restriction of this vector field to X_B is a vector field on X_B , and is independent of the chosen extension. The good Lagrangian condition states that this vector field on X_B is k -related to a vector field on N_B , which by definition is $n(u)$.

The connection h_ϵ being flat is the condition that for any vector fields u_1, u_2 on M we have $h_\epsilon([u_1, u_2]) = [h_\epsilon(u_1), h_\epsilon(u_2)]$. But then it follows that $h_\epsilon([u_1, u_2])$ is k -related to $[n(u_1), n(u_2)]$, which by definition of n implies that $n([u_1, u_2]) = [n(u_1), n(u_2)]$ and hence that the connection n is flat. \square

To express the good Lagrangian condition more concretely, take local Darboux coordinates (z_1, \dots, z_{2d}) on M and assume that $B \subset M$ is given by the equations $z_{d+1} = \dots = z_{2d} = 0$, and that $\omega_{pq} = \pm 1$ if $q-p = \pm d$ and is otherwise zero. Lifting the vector fields $\partial/\partial z_i$ with $1 \leq i \leq d$ from M to X as in (3) gives

$$(70) \quad v_i = \frac{\partial}{\partial \theta_i}, \quad h_i = \frac{\partial}{\partial z_i} + \sum_{j=1}^d \left(\frac{\partial^2 W}{\partial \theta_i \partial \theta_{j+d}} \cdot \frac{\partial}{\partial \theta_j} - \frac{\partial^2 W}{\partial \theta_i \partial \theta_j} \cdot \frac{\partial}{\partial \theta_{j+d}} \right).$$

Applying the projection $k : T_M|_B \rightarrow N_B$ amounts to setting $\partial/\partial \theta_i = 0$ for $1 \leq i \leq d$. The condition of Definition 8.1 is then that for $1 \leq i \leq d$ the result of this projection should be independent of the coordinates θ_i for $1 \leq i \leq d$. This is equivalent to

$$(71) \quad \frac{\partial^3 W}{\partial \theta_i \partial \theta_j \partial \theta_k} = 0 \quad \text{for } 1 \leq i, j, k \leq d$$

along the locus $z_{d+1} = \dots = z_{2d} = 0$.

There is a canonical real structure on the tangent bundle T_M whose fixed locus $T_M^{\mathbb{R}} \subset T_M$ is the real span of the integral affine structure $T_M^{\mathbb{Z}} \subset T_M$. We call a complex Lagrangian $B \subset M$ nondegenerate if $T_B \cap T_M^{\mathbb{R}}|_B = (0) \subset T_M|_B$. When this holds, the restriction of the map k to the lattice $T_M^{\mathbb{Z}}|_B \subset T_M|_B$ is injective, and we denote its image by $N_B^{\mathbb{Z}} \subset N_B$. When the Lagrangian B is both good and nondegenerate, the connection n of Lemma 8.2 descends to a connection on the projection $\pi : N_B/N_B^{\mathbb{Z}} \rightarrow B$ whose fibres are compact tori $\mathbb{C}^d/\mathbb{Z}^{2d} \cong (S^1)^{2d}$.

8.2 Class $S[A_1]$ examples

Consider the Joyce structure on the space $M = \mathcal{M}(C, Q)$ constructed in this paper. Let us fix a curve $C \in \mathcal{M}(C)$ and consider the Lagrangian submanifold $B = \mathcal{M}_C(Q) \subset M$ which is the corresponding fibre of the projection $\rho : \mathcal{M}(C, Q) \rightarrow \mathcal{M}(C)$. Thus $B \subset H^0(C, \omega_C^{\otimes 2})$ parametrises quadratic differentials on C with simple zeroes.

Note that if $Q' \in H^0(C, \omega_C^{\otimes 2})$ is a quadratic differential on C , then $p^*(Q')$ vanishes to order two along the branch divisor $R \subset \Sigma$. It follows that the tangent space to B at a point (C, Q) can be identified with $H^0(\Sigma, \omega_\Sigma)^-$ via the map

$$(72) \quad T_b B = H^0(C, \omega_C^{\otimes 2}) \rightarrow H^0(\Sigma, \omega_\Sigma)^-, \quad Q' \mapsto p^*(Q')/2\lambda,$$

where λ is the tautological 1-form on Σ appearing in (24). Under the isomorphism of Theorem 3.1 the sequence (68) then corresponds to the Hodge filtration

$$(73) \quad 0 \rightarrow H^0(\Sigma, \omega_\Sigma)^- \xrightarrow{i} H^1(\Sigma, \mathbb{C})^- \xrightarrow{k} H^1(\Sigma, \theta_\Sigma)^- \rightarrow 0.$$

It follows from the isomorphism (28) that the fibres

$$(74) \quad H^1(\Sigma, \theta_\Sigma)^- / \tilde{H}^1(\Sigma, \mathbb{Z}) = P^\#(\Sigma) / H^0(\Sigma, \omega_\Sigma)^-$$

of the map $\pi : N_B / N_B^{\mathbb{Z}} \rightarrow B$ are the Prym varieties $P(\Sigma)$ appearing in Section 3.4.

Lemma 8.3 *For each curve C the submanifold $B = \mathcal{M}_C(Q) \subset M = \mathcal{M}(C, Q)$ is a good Lagrangian. The horizontal leaves of the induced meromorphic flat connection on $\pi : N_B / N_B^{\mathbb{Z}} \rightarrow B$ are defined by the condition that $E = p_*(L)$ is constant.*

Proof We use the notation $\mathcal{M}_C(Q, L, \partial)$ to denote the space parametrising data (Q, L, ∂) on the fixed curve C , and similarly for $\mathcal{M}_C(Q, E, \nabla_\epsilon)$, etc. Let w be a vector field on $B \subset M$ and let $u = h_\epsilon(w)$ be the lift to a vector field on $X_B \subset X$. Transferring across the diagram (39) we can consider u to be a vector field on $\mathcal{M}_C(Q, L, \partial)$, and we must show that it descends to the space $\mathcal{M}_C(Q, L)$. That is, the flow of the line bundle L under u should be independent of the connection ∂ . Passing through the diagram (38) we can view u as a vector field on $\mathcal{M}_C(Q, E, \nabla_\epsilon)$, and we must show that it descends to the space $\mathcal{M}_C(Q, E)$.

By definition, the connection h_ϵ on the projection π_2 is pulled back from the isomonodromy connection on π_1 . Since $\rho_*(w) = 0$, it follows that $\rho'_*(u) = 0$. That is, u is obtained by keeping the pair (E, ∇_ϵ) on C fixed as Q varies with w . It is then clear that u descends to $\mathcal{M}_C(Q, E)$, and the result follows. \square

Consider the diagram of moduli spaces

$$(75) \quad \begin{array}{ccccc} N_B / N_B^{\mathbb{Z}} & \xleftarrow{\tau} & \mathcal{M}_C(Q, L) & \xrightarrow{\kappa} & \mathcal{M}_C(E, \Phi) \\ \pi \downarrow & & \downarrow & & \downarrow \\ B & \xleftarrow{=} & \mathcal{M}_C(Q) & \xleftarrow{=} & \mathcal{M}_C(Q) \end{array}$$

Here κ is the isomorphism defined by the usual spectral construction sending a line bundle L on Σ to the Higgs bundle (E, Φ) on C , and τ is induced by the corresponding map from (39). The forgetful map $\mathcal{M}_C(E, \Phi)_0 \rightarrow \mathcal{M}_C(E)$ can be identified with the cotangent bundle of $\mathcal{M}_C(E)$, and according to Lemma 8.3, when transferred across the diagram (75), the fibres of this map become the horizontal leaves of the connection n .

Appendix Definition of the moduli spaces

In this section we give detailed constructions of the moduli spaces appearing in the diagram (38). All schemes are over $\text{Spec}(\mathbb{C})$. We fix a genus $g > 1$ and a level $\ell > 2$ throughout. We use the terminology bundle for locally free sheaf of finite rank, and line bundle for invertible sheaf.

A.1 Curves with level structure

A family of genus g curves is a smooth proper map of schemes $f: \mathcal{C} \rightarrow S$ of relative dimension 1 whose geometric fibres are connected and of genus g . Given a map of schemes $s: S' \rightarrow S$ we can pull back the family by forming the Cartesian diagram

$$(76) \quad \begin{array}{ccc} \mathcal{C}' & \xrightarrow{t} & \mathcal{C} \\ f' \downarrow & & \downarrow f \\ S' & \xrightarrow{s} & S \end{array}$$

Given a family of genus g curves $f: \mathcal{C} \rightarrow S$ as above, there is a locally constant sheaf of free \mathbb{Z}/ℓ -modules $\mathcal{V}_f = \mathbb{R}^1 f_*(\mathbb{Z}/\ell)$ on S . This construction commutes with base-change: given a diagram (76) there is a canonical isomorphism $\mathcal{V}_{f'} \cong s^*(\mathcal{V}_f)$. In particular, the pullback of \mathcal{V}_f to a \mathbb{C} -valued point of S is the cohomology group $H^1(C, \mathbb{Z}/\ell) \cong (\mathbb{Z}/\ell)^{\oplus 2g}$ of the corresponding genus g curve C . The intersection form on these cohomology groups defines a skew-symmetric \mathbb{Z}/ℓ -bilinear form

$$(77) \quad \varpi_f: \mathcal{V}_f \times \mathcal{V}_f \rightarrow \mathbb{Z}/\ell.$$

Let V be the free \mathbb{Z}/ℓ -module on the symbols $a_1, b_1, \dots, a_g, b_g$ equipped with the standard skew-symmetric form defined by

$$(78) \quad \varpi(a_i, a_j) = 0 = \varpi(b_i, b_j) \quad \text{and} \quad \varpi(a_i, b_j) = \delta_{ij} \quad \text{for } 1 \leq i, j \leq g.$$

A level ℓ structure on the family f is defined to be an isomorphism of \mathbb{Z}/ℓ -modules $\theta: V \rightarrow H^0(S, \mathcal{V}_f)$ relating the forms ϖ and ϖ_f . Given a diagram (76), a level ℓ structure on the family f defines a pulled-back level structure on the family f' in the obvious way.

Two families of curves $f_i: \mathcal{C}_i \rightarrow S$ with level ℓ structure θ_i are isomorphic if there is an isomorphism $g: \mathcal{C}_1 \rightarrow \mathcal{C}_2$ satisfying $f_2 \circ g = f_1$ and preserving the level structures in the obvious way. There is a functor $M(g, \ell): (\text{Sch}/\mathbb{C})^{\text{op}} \rightarrow \text{Sets}$ by sending a scheme S to the set of isomorphism classes of families of genus g curves over S equipped with level ℓ structure.

Theorem A.1 *When $g > 1$ and $\ell > 2$, the functor $M(g, \ell)$ is represented by a smooth quasiprojective scheme $\mathcal{M}(g, \ell)$.*

Proof This appears to be standard, although it is hard to find a complete proof in the literature. Grothendieck [1962, Section 2] shows that $M(g, \ell)$ is representable by an algebraic space for $\ell \gg 0$, and

attributes to Serre the statement that $\ell > 2$ is sufficient. Mumford, Fogarty and Kirwan [Mumford et al. 1994, Theorem 7.9] prove the analogous result on moduli spaces of abelian varieties, which implies that $\mathcal{M}(g, \ell)$ is representable by a quasiprojective scheme for $\ell \gg 0$, and they again attribute to Serre the statement that $\ell > 2$ is sufficient. \square

The closed points of $\mathcal{M}(g, \ell)$ parametrise smooth projective genus g curves C equipped with a choice of symplectic basis in the homology group $H_1(C, \mathbb{Z}/\ell)$.

A.2 Quadratic differentials

Given a family of genus g curves $f: \mathcal{C} \rightarrow S$ we denote by $\omega_{\mathcal{C}/S}$ the relative cotangent bundle. Given a Cartesian diagram (76) there is a canonical isomorphism $\omega_{\mathcal{C}'/S'} \cong g^*(\omega_{\mathcal{C}/S})$.

If C is any fibre of f then Serre duality gives $H^1(C, \omega_C^{\otimes 2}) = H^0(C, T_C)^* = 0$. Using cohomology and base-change it follows that $f_*(\omega_{\mathcal{C}/S}^{\otimes 2})$ is a vector bundle on S . By a quadratic differential on the family of curves f we mean a section of this vector bundle. Note that

$$(79) \quad H^0(S, f_*(\omega_{\mathcal{C}/S}^{\otimes 2})) = H^0(\mathcal{C}, \omega_{\mathcal{C}/S}^{\otimes 2}).$$

Applying this construction to the universal family of curves defines a vector bundle \mathcal{E} on the space $\mathcal{M}(g, \ell)$.

Define a functor $Quad(g, \ell): \text{Sch}/\mathbb{C}^{\text{op}} \rightarrow \text{Sets}$ by sending a scheme S to the set of isomorphism classes of families of genus g curves equipped with level structures and quadratic differentials.

Lemma A.2 *The functor $Quad(g, \ell)$ is represented by smooth quasiprojective variety $Quad(g, \ell)$ which is the total space of the vector bundle \mathcal{E} over $\mathcal{M}(g, \ell)$.*

The closed points of $Quad(g, \ell)$ parametrise smooth projective genus g curves C equipped with a choice of symplectic basis in the homology group $H_1(C, \mathbb{Z}/\ell)$ and a quadratic differential $Q \in H^0(C, \omega_C^{\otimes 2})$.

Consider a family of genus g curves $f: \mathcal{C} \rightarrow S$ and a section $Q \in H^0(S, \omega_{\mathcal{C}/S}^{\otimes 2})$. The relative critical locus of Q is a closed subscheme of \mathcal{C} , and since f is proper, its image is a closed subscheme in S . This construction commutes with base-change. Applying it to the universal family defines a closed subscheme of $Quad(g, \ell)$. We define $Quad_0(g, \ell) \subset Quad(g, \ell)$ to be the complementary open subscheme. By definition, a closed point of $Quad(g, \ell)$ lies in this open subscheme precisely if the corresponding quadratic differential $Q \in H^0(C, \omega_C^{\otimes 2})$ has simple zeroes.

A.3 Bundles, Higgs fields and flat connections

Let $f: \mathcal{C} \rightarrow S$ be a family of genus g curves over a scheme S . We now give the definitions of the relative moduli spaces of bundles, Higgs bundles and flat connections we will need.

A family of rank r bundles on f is simply a rank r bundle E on \mathcal{C} . Two such families E_i are equivalent if there is a line bundle L on S and an isomorphism $\theta: E_1 \rightarrow E_2 \otimes f^*(L)$. A family E is said to have trivial determinant if the associated family of rank 1 bundles $\wedge^r(E)$ is equivalent to the trivial family $\mathcal{O}_{\mathcal{C}}$.

A family of rank r Higgs bundles is a bundle E on \mathcal{C} equipped with a relative Higgs field $\Phi: E \rightarrow E \otimes \omega_{\mathcal{C}/S}$. Two such families (E_i, Φ_i) are equivalent if there is an isomorphism $\theta: E_1 \rightarrow E_2 \otimes f^*(L)$ which intertwines Φ_1 and $\Phi_2 \otimes 1_{f^*(L)}$. A family (E, Φ) has trivial determinant if the associated family of rank 1 Higgs bundles $(\wedge^r(E), \wedge^r(\Phi))$ is equivalent to the trivial family $(\mathcal{O}_{\mathcal{C}}, 0)$. (This is the usual condition that the Higgs bundle has zero trace.)

A family of rank r flat connections on f is a bundle E on \mathcal{C} equipped with a relative connection $\nabla: E \rightarrow E \otimes \omega_{\mathcal{C}/S}$. Two such families (E_i, ∇_i) are equivalent if there is an isomorphism $\theta: E_1 \rightarrow E_2 \otimes f^*(L)$ which is flat for the induced relative connection on $\mathcal{H}om_{\mathcal{O}_{\mathcal{C}}}(E_1, E_2 \otimes f^*(L))$. A family (E, ∇) has trivial determinant if the corresponding family of rank 1 flat connections $(\wedge^r(E), \wedge^r(\nabla))$ is equivalent to the trivial family $(\mathcal{O}_{\mathcal{C}}, d)$.

In all cases we say that a family on \mathcal{C} is stable if the restriction of the bundle E to each geometric fibre of f is stable. Note that for Higgs bundles and flat connections this is strictly stronger than the usual notion of stability. We use the stronger notion to ensure that the forgetful maps appearing in the diagram (38) are well-defined. Since stability is an open condition it corresponds to passing to open subsets of the usual moduli spaces.

Fix again a family of genus g curves $f: \mathcal{C} \rightarrow S$ and assume S to be of finite type over \mathbb{C} . There is a functor

$$(80) \quad \text{Bun}(\mathcal{C}/S, r): (\text{Sch}/S)^{\text{op}} \rightarrow \text{Sets}$$

which sends a map $m: T \rightarrow S$ to the set of equivalence classes of stable families of rank r bundles with trivial determinant on the pulled back family $f_T: \mathcal{C} \times_S T \rightarrow T$. We define moduli functors $\text{Higgs}(\mathcal{C}/S, r)$ and $\text{Flat}(\mathcal{C}/S, r)$ in the same way.

Theorem A.3 *The functors $\text{Bun}(\mathcal{C}/S, r)$, $\text{Higgs}(\mathcal{C}/S, r)$ and $\text{Flat}(\mathcal{C}/S, r)$ are co-representable by schemes $\text{Bun}(\mathcal{C}/S, r)$, $\text{Higgs}(\mathcal{C}/S, r)$ and $\text{Flat}(\mathcal{C}/S, r)$ respectively. Each of these schemes is smooth and quasiprojective over S . The obvious forgetful maps $\text{Higgs}(\mathcal{C}/S, r) \rightarrow \text{Bun}(\mathcal{C}/S, r)$ and $\text{Flat}(\mathcal{C}/S, r) \rightarrow \text{Bun}(\mathcal{C}/S, r)$ are smooth.*

Proof The co-representability follows from the results of Simpson [1994a]. The other statements are easy and well-known. □

We apply these results to the universal family of curves over the moduli space $\mathcal{M}(g, \ell)$. We denote the resulting moduli spaces as $\text{Bun}(g, \ell, r)$, $\text{Higgs}(g, \ell, r)$ and $\text{Flat}(g, \ell, r)$. They are smooth quasiprojective schemes. Note for example that $\text{Bun}(g, \ell, r)$ co-represents the functor which sends a scheme S to the set

of equivalence classes of pairs (f, E) consisting of a family of genus g curves $f: \mathcal{C} \rightarrow S$ and a stable family of rank r bundles with trivial determinant E over f . Similar remarks apply to $\text{Higgs}(g, \ell, r)$ and $\text{Flat}(g, \ell, r)$.

Given a family of Higgs fields $\Phi: E \rightarrow E \otimes \omega_{\mathcal{C}/S}$ on a family of genus g curves $f: \mathcal{C} \rightarrow S$ we can define a quadratic differential $Q \in H^0(\mathcal{C}, \omega_{\mathcal{C}/S}^{\otimes 2})$ by setting $Q = \frac{1}{2} \text{tr}(\Phi^2)$. This defines a map

$$(81) \quad \text{Higgs}(g, \ell, r) \rightarrow \text{Quad}(g, \ell),$$

and we define $\text{Higgs}_0(g, \ell, r)$ to be the inverse image of the open subset $\text{Quad}_0(g, \ell)$ of quadratic differentials with simple zeroes.

A.4 Anti-invariant branched connections

Consider a family of genus g curves $f: \mathcal{C} \rightarrow S$ equipped with a quadratic differential $Q \in H^0(S, \omega_{\mathcal{C}/S}^{\otimes 2})$ with no relative critical points. We can form a double cover $p: \Sigma \rightarrow \mathcal{C}$ by writing the equation $y^2 = Q$ inside the total space of the bundle $\omega_{\mathcal{C}/S}$. This construction commutes with base-change in the obvious way. There is a covering involution $\sigma: \Sigma \rightarrow \Sigma$ and a branch divisor $R \subset \Sigma$ which is flat over S . The composite $g = f \circ p: \Sigma \rightarrow S$ is a family of smooth genus $4g - 3$ curves.

A family of branched connections on $g: \Sigma \rightarrow S$ is a line bundle L on Σ equipped with a relative meromorphic connection $\partial: L \rightarrow L \otimes \omega_{\Sigma/S}(R)$. Two such families (L_i, ∂_i) are equivalent if there is a line bundle N on S and an isomorphism $\theta: L_1 \rightarrow L_2 \otimes g^*(N)$ which is flat for the induced relative meromorphic connection on $\mathcal{H}om_{\mathcal{O}_\Sigma}(L_1, L_2 \otimes g^*(N))$. A family of branched connections (L, ∂) is anti-invariant if the branched connection $(L, \partial) \otimes \sigma^*(L, \partial)$ is equivalent to the family of branched connections $(p^*(\omega_{\mathcal{C}}), \partial_{\text{can}})$.

There is a functor

$$(82) \quad \text{Flat}^{\text{br}}(\Sigma/S): (\text{Sch}/S)^{\text{op}} \rightarrow \text{Sets}$$

which sends a map $m: T \rightarrow S$ to the set of equivalence classes of anti-invariant families of branched connections on the pulled back family $g_T: \Sigma \times_S T \rightarrow T$.

Theorem A.4 *The functor $\text{Flat}^{\text{br}}(\Sigma/S)$ is representable by a scheme $\text{Flat}^{\text{br}}(\Sigma/S)$ which is smooth and quasiprojective over S .*

Proof For moduli spaces of flat connections with logarithmic singularities we can refer to Nitsure [1993], but since we are dealing with rank 1 connections this is really over-kill. For a more elementary approach we can pass to an étale cover of the functor $\text{Flat}^{\text{br}}(\Sigma/S)$ by adding the data of a square-root of the line bundle $\omega_{\mathcal{C}/S}$. Then, as in the proof of Lemma 3.2, we can replace the branched connections (L, ∂_L) with regular connections (M, ∂_M) . \square

We can apply the above construction to the universal family over $S = \text{Quad}_0(g, \ell)$. We denote the resulting moduli space by $\text{Flat}^{\text{br}}(g, \ell)$. It is a smooth quasiprojective variety. It represents the functor which sends a scheme S to the set of equivalence classes of quadruples (f, Q, L, ϑ) consisting of a family of genus g curves $f: \mathcal{C} \rightarrow S$ equipped with a quadratic differential $Q \in H^0(\mathcal{C}, \omega_{\mathcal{C}/S}^{\otimes 2})$ with simple zeroes, and a family of anti-invariant branched connections (L, ϑ) on the associated family of spectral curves $g: \Sigma \rightarrow S$.

A.5 Moduli spaces and maps

We can now define the moduli spaces and maps appearing in (38). Firstly we set $\mathcal{M}(C) = \mathcal{M}(g, \ell)$ and $\mathcal{M}(C, Q) = \text{Quad}_0(g, \ell)$. Then we take rank $r = 2$ and set $\mathcal{M}(C, E) = \text{Bun}(g, \ell, 2)$ and

$$(83) \quad \mathcal{M}(C, E, \nabla) = \text{Flat}(g, \ell, 2), \quad \mathcal{M}(C, E, \Phi) = \text{Higgs}_0(g, \ell, 2).$$

We define $\mathcal{M}(C, E, \nabla, \Phi)$ to be the fibre product of the obvious forgetful maps

$$\begin{array}{ccc} \mathcal{M}(C, E, \nabla, \Phi) & \longrightarrow & \mathcal{M}(C, E, \Phi) \\ \downarrow & & \downarrow \\ \mathcal{M}(C, E, \nabla) & \longrightarrow & \mathcal{M}(C, E) \end{array}$$

Similarly we define $\mathcal{M}(C, Q, E, \nabla)$ as the fibre product

$$\begin{array}{ccc} \mathcal{M}(C, Q, E, \nabla) & \longrightarrow & \mathcal{M}(C, Q) \\ \downarrow & & \downarrow \\ \mathcal{M}(C, E, \nabla) & \longrightarrow & \mathcal{M}(C) \end{array}$$

Finally we take $\mathcal{M}(C, Q, L, \vartheta)$ to be the open subvariety of $\text{Flat}^{\text{br}}(g, \ell)$ defined by the condition that $E = p_*(L)$ is stable. Each of these spaces are smooth quasiprojective varieties.

The maps in the diagram (38) are just the obvious forgetful maps, with the exception of α and β_ϵ . To define α we follow the same procedure in the text in the relative setting. To define β_ϵ note that $\beta_\epsilon = \beta_\infty \circ r_\epsilon$, where r_ϵ is the automorphism of the space $\mathcal{M}(C, E, \nabla, \Phi)$ defined by $\nabla \mapsto \nabla + \epsilon^{-1}\Phi$. The map β_∞ is given by the rule $(C, E, \nabla, \Phi) \mapsto (C, Q, E, \nabla)$ and is induced using the above fibre-product diagrams from the map (81).

References

[Allegretti 2019] **D G L Allegretti**, *Voros symbols as cluster coordinates*, J. Topol. 12 (2019) 1031–1068 MR Zbl
 [Allegretti 2021] **D G L Allegretti**, *Stability conditions, cluster varieties, and Riemann–Hilbert problems from surfaces*, Adv. Math. 380 (2021) art. id. 107610 MR Zbl
 [Arinkin 2005] **D Arinkin**, *On λ -connections on a curve where λ is a formal parameter*, Math. Res. Lett. 12 (2005) 551–565 MR Zbl

- [Beauville et al. 1989] **A Beauville, M S Narasimhan, S Ramanan**, *Spectral curves and the generalised theta divisor*, *J. Reine Angew. Math.* 398 (1989) 169–179 MR Zbl
- [Bridgeland 2009] **T Bridgeland**, *Spaces of stability conditions*, from “Algebraic geometry, I” (D Abramovich, A Bertram, L Katzarkov, R Pandharipande, M Thaddeus, editors), *Proc. Sympos. Pure Math.* 80, Part 1, Amer. Math. Soc., Providence, RI (2009) 1–21 MR Zbl
- [Bridgeland 2019] **T Bridgeland**, *Riemann–Hilbert problems from Donaldson–Thomas theory*, *Invent. Math.* 216 (2019) 69–124 MR Zbl
- [Bridgeland 2021] **T Bridgeland**, *Geometry from Donaldson–Thomas invariants*, from “Integrability, quantization, and geometry, II: Quantum theories and algebraic geometry” (S Novikov, I Krichever, O Ogievetsky, S Shlosman, editors), *Proc. Sympos. Pure Math.* 103.2, Amer. Math. Soc., Providence, RI (2021) 1–66 MR Zbl
- [Bridgeland and Masoero 2023] **T Bridgeland, D Masoero**, *On the monodromy of the deformed cubic oscillator*, *Math. Ann.* 385 (2023) 193–258 MR Zbl
- [Bridgeland and Smith 2015] **T Bridgeland, I Smith**, *Quadratic differentials as stability conditions*, *Publ. Math. Inst. Hautes Études Sci.* 121 (2015) 155–278 MR Zbl
- [Bridgeland and Strachan 2021] **T Bridgeland, I A B Strachan**, *Complex hyperkähler structures defined by Donaldson–Thomas invariants*, *Lett. Math. Phys.* 111 (2021) art. id. 54 MR Zbl
- [Deligne 1970] **P Deligne**, *Équations différentielles à points singuliers réguliers*, *Lecture Notes in Math.* 163, Springer (1970) MR Zbl
- [Donagi and Pantev 2009] **R Donagi, T Pantev**, *Geometric Langlands and non-abelian Hodge theory*, from “Geometry, analysis, and algebraic geometry: forty years of the Journal of Differential Geometry” (H-D Cao, S-T Yau, editors), *Surv. Differ. Geom.* 13, International, Somerville, MA (2009) 85–116 MR Zbl
- [Dunajski and Mason 2000] **M Dunajski, L J Mason**, *Hyper-Kähler hierarchies and their twistor theory*, *Comm. Math. Phys.* 213 (2000) 641–672 MR Zbl
- [Fock and Goncharov 2006] **V Fock, A Goncharov**, *Moduli spaces of local systems and higher Teichmüller theory*, *Publ. Math. Inst. Hautes Études Sci.* 103 (2006) 1–211 MR Zbl
- [Fock and Thomas 2021] **V Fock, A Thomas**, *Higher complex structures*, *Int. Math. Res. Not.* 2021 (2021) 15873–15893 MR Zbl
- [Gaiotto 2014] **D Gaiotto**, *Opers and TBA*, preprint (2014) arXiv 1403.6137
- [Gaiotto et al. 2010] **D Gaiotto, G W Moore, A Neitzke**, *Four-dimensional wall-crossing via three-dimensional field theory*, *Comm. Math. Phys.* 299 (2010) 163–224 MR Zbl
- [Gaiotto et al. 2013] **D Gaiotto, G W Moore, A Neitzke**, *Wall-crossing, Hitchin systems, and the WKB approximation*, *Adv. Math.* 234 (2013) 239–403 MR Zbl
- [Goldman 1984] **W M Goldman**, *The symplectic nature of fundamental groups of surfaces*, *Adv. in Math.* 54 (1984) 200–225 MR Zbl
- [Gotay et al. 1983] **M J Gotay, R Lashof, J Śniatycki, A Weinstein**, *Closed forms on symplectic fibre bundles*, *Comment. Math. Helv.* 58 (1983) 617–621 MR Zbl
- [Grothendieck 1962] **A Grothendieck**, *Techniques de construction en géométrie analytique, I: Description axiomatique de l’espace de Teichmüller et de ses variantes*, from “Familles d’espaces complexes et fondements de la géométrie analytique”, *Sém. H Cartan* 13 (1960/61), Secrétariat mathématique, Paris (1962) Exposé 7–8 Zbl

- [Haiden 2024] **F Haiden**, *3-D Calabi–Yau categories for Teichmüller theory*, *Duke Math. J.* 173 (2024) 277–346 MR Zbl
- [Hitchin 1987a] **N J Hitchin**, *The self-duality equations on a Riemann surface*, *Proc. London Math. Soc.* 55 (1987) 59–126 MR Zbl
- [Hitchin 1987b] **N Hitchin**, *Stable bundles and integrable systems*, *Duke Math. J.* 54 (1987) 91–114 MR Zbl
- [Hollands and Neitzke 2016] **L Hollands, A Neitzke**, *Spectral networks and Fenchel–Nielsen coordinates*, *Lett. Math. Phys.* 106 (2016) 811–877 MR Zbl
- [Joyce 2007] **D Joyce**, *Holomorphic generating functions for invariants counting coherent sheaves on Calabi–Yau 3-folds*, *Geom. Topol.* 11 (2007) 667–725 MR Zbl
- [Kontsevich and Soibelman 2006] **M Kontsevich, Y Soibelman**, *Affine structures and non-Archimedean analytic spaces*, from “The unity of mathematics” (P Etingof, V Retakh, IM Singer, editors), *Progr. Math.* 244, Birkhäuser, Boston, MA (2006) 321–385 MR Zbl
- [Laumon 1988] **G Laumon**, *Un analogue global du cône nilpotent*, *Duke Math. J.* 57 (1988) 647–671 MR Zbl
- [Mumford 1974] **D Mumford**, *Prym varieties, I*, from “Contributions to analysis (a collection of papers dedicated to Lipman Bers)” (L V Ahlfors, I Kra, B Maskit, L Nirenberg, editors), Academic, New York (1974) 325–350 MR Zbl
- [Mumford et al. 1994] **D Mumford, J Fogarty, F Kirwan**, *Geometric invariant theory*, 3rd edition, *Ergebnisse der Math.* (2) 34, Springer (1994) MR Zbl
- [Nikolaev 2021] **N Nikolaev**, *Abelianisation of logarithmic \mathfrak{sl}_2 -connections*, *Selecta Math.* 27 (2021) art. id. 78 MR Zbl
- [Nitsure 1993] **N Nitsure**, *Moduli of semistable logarithmic connections*, *J. Amer. Math. Soc.* 6 (1993) 597–609 MR Zbl
- [Plebański 1975] **J F Plebański**, *Some solutions of complex Einstein equations*, *J. Mathematical Phys.* 16 (1975) 2395–2402 MR
- [Simpson 1994a] **C T Simpson**, *Moduli of representations of the fundamental group of a smooth projective variety, I*, *Inst. Hautes Études Sci. Publ. Math.* 79 (1994) 47–129 MR Zbl
- [Simpson 1994b] **C T Simpson**, *Moduli of representations of the fundamental group of a smooth projective variety, II*, *Inst. Hautes Études Sci. Publ. Math.* 80 (1994) 5–79 MR Zbl
- [Veech 1990] **W A Veech**, *Moduli spaces of quadratic differentials*, *J. Analyse Math.* 55 (1990) 117–171 MR Zbl
- [Zikidis \geq 2025] **M Zikidis**, *Joyce structures on spaces of meromorphic quadratic differentials*, in preparation

Department of Pure Mathematics, University of Sheffield
Sheffield, United Kingdom

t.bridgeland@sheffield.ac.uk

Proposed: Mark Gross

Seconded: Simon Donaldson, Arend Bayer

Received: 17 April 2024

Revised: 7 October 2024

Big monodromy for higher Prym representations

AARON LANDESMAN

DANIEL LITT

WILL SAWIN

Let $\Sigma_{g'} \rightarrow \Sigma_g$ be a cover of an orientable surface of genus g by an orientable surface of genus g' , branched at n points, with Galois group H . Such a cover induces a virtual action of the mapping class group $\text{Mod}_{g,n+1}$ of a genus g surface with $n+1$ marked points on $H^1(\Sigma_{g'}, \mathbb{C})$. When g is large in terms of the group H , we calculate precisely the connected monodromy group of this action. The methods are Hodge-theoretic and rely on a “generic Torelli theorem with coefficients”.

14H30; 14D07, 14H10, 14H40, 14H60, 57K20

1. Introduction	2733
2. Notation and preliminaries on moduli	2742
3. Review of parabolic bundles and period maps	2745
4. Global generation of vector bundles on generic curves	2747
5. Preliminaries on variations of Hodge structure	2752
6. Generic Torelli theorems for unitary local systems	2756
7. Big monodromy for g large	2760
8. Big monodromy for n large	2771
9. Big monodromy for Kodaira fibrations	2773
10. Questions	2778
References	2779

1 Introduction

A classical result in geometric topology states that the action of the mapping class group Mod_g of a surface Σ_g of genus g on its first cohomology, $H^1(\Sigma_g, \mathbb{Z})$, is via the full group of automorphisms preserving the cup product, namely $\text{Sp}_{2g}(\mathbb{Z})$. Let the hyperelliptic mapping class group denote the subgroup of mapping classes commuting with an involution having genus-zero quotient. If one restricts to the hyperelliptic mapping class group the image is still of finite index in $\text{Sp}_{2g}(\mathbb{Z})$; see [A’Campo 1979].

Finally, one may consider the monodromy representation on the cohomology of Prym varieties arising from connected étale double covers of genus g curves. It follows from [Looijenga 1997] that the image of this representation is finite index in $\mathrm{Sp}_{2g-2}(\mathbb{Z})$.

In order to generalize the above cases, fix an *arbitrary* finite group H and consider a maximal family of (possibly branched) Galois H -covers of curves of genus g . What is the monodromy representation on the first cohomology of these covering curves? The three cases considered above correspond to the very special cases $H = \{\mathrm{id}\}$, the case $g = 0$ and $H = \mathbb{Z}/2\mathbb{Z}$, and the case where $g \geq 2$, $H = \mathbb{Z}/2\mathbb{Z}$, and the covers are unramified. Our main result asserts that once the genus g of the base curve is sufficiently large, the connected monodromy group of this family of H -covers is as large as possible. Namely, just as the action of the mapping class group on $H^1(\Sigma_g, \mathbb{Z})$ cannot be via all of $\mathrm{GL}_{2g}(\mathbb{Z})$ because it must preserve the cup product, a symplectic form, the monodromy of families of H -covers cannot be the full general linear group, but must preserve the symplectic form and respect the H -action. Moreover, as local systems of geometric origin are semisimple, it must be semisimple. These considerations show that the identity component of the Zariski closure of the image of the monodromy representation is contained in the derived subgroup of the centralizer of H in the symplectic group. We will show that it is in fact equal to this group, once the genus of the base curve is sufficiently large.

1.1 Statement of results We next set up notation to state our main results more precisely. Let $\Sigma_{g,n}$ be an orientable surface of genus g , with n punctures, and let $\mathrm{Mod}_{g,n}$ be the pure mapping class group of $\Sigma_{g,n}$. That is, $\mathrm{Mod}_{g,n} = \pi_0(\mathrm{Homeo}^+(\Sigma_{g,n}))$, where $\mathrm{Homeo}^+(\Sigma_{g,n})$ is the space of orientation-preserving homeomorphisms of $\Sigma_g = \Sigma_{g,0}$ fixing each of the n punctures. The goal of this paper is to study certain natural representations of these mapping class groups, arising from finite unramified covers of $\Sigma_{g,n}$. We call these representations *higher Prym representations*, following the terminology of Putman and Wieland [2013].

Let H be a finite group and let $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$ be a finite unramified Galois H -cover. We will next construct a homomorphism from a finite-index subgroup of $\mathrm{Mod}_{g,n+1}$ to the centralizer of H in $\mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))$. Fix a point x of $\Sigma_{g,n}$. All finite unramified H -torsors $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$ arise from surjection $\varphi: \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$, with isomorphism of torsors corresponding to conjugation (by H) of surjections. For any $x' \in \Sigma_{g',n'}$ mapping to x , we can identify the kernel of φ with $K := \pi_1(\Sigma_{g',n'}, x')$. The mapping class group $\mathrm{Mod}_{g,n+1}$ acts (up to isotopy) on $(\Sigma_{g,n}, x)$, where we view x as an $(n+1)^{\mathrm{st}}$ marked point of Σ_g . Thus, $\mathrm{Mod}_{g,n+1}$ acts on $\pi_1(\Sigma_{g,n}, x)$, and hence on the finite set of homomorphisms $\pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$. The stabilizer $\mathrm{Mod}_\varphi \subset \mathrm{Mod}_{g,n+1}$ of φ acts on the kernel K of φ . The induced action on $K^{\mathrm{ab}} = H_1(\Sigma_{g',n'}, \mathbb{Z})$ preserves the kernel of the natural morphism

$$H_1(\Sigma_{g',n'}, \mathbb{Z}) \rightarrow H_1(\Sigma_{g'}, \mathbb{Z}),$$

and thus φ gives rise to a virtual action of $\mathrm{Mod}_{g,n+1}$ on $H_1(\Sigma_{g'}, \mathbb{Z})$, ie an action of Mod_φ on $H_1(\Sigma_{g'}, \mathbb{Z})$.

This action manifestly commutes with the action of H and thus defines a homomorphism

$$(1-1) \quad R_\varphi : \text{Mod}_\varphi \rightarrow \text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H,$$

where Mod_φ is the stabilizer of φ as defined above, and $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$ denotes the centralizer of the action of H in $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))$. We may equivalently think of this as a virtual homomorphism from $\text{Mod}_{g,n+1}$ to $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$, where a virtual homomorphism is a homomorphism from a finite-index subgroup.

We are interested in the *connected monodromy group* of this virtual action of $\text{Mod}_{g,n+1}$, or, in other words, the identity component of the Zariski closure of the image of R_φ inside $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$. The slogan for this work is:

1.2 Slogan Monodromy groups should be as big as possible.

Based on this, we expect that, outside of a few exceptional cases, this Zariski closure should contain the commutator subgroup of $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$. (After passing to a further finite-index subgroup, we may assume the monodromy group is connected, and connected components of monodromy groups of local systems of geometric origin are semisimple, so the commutator subgroup is the largest possible.) Our main result is that this expectation holds under a lower bound on the genus:

1.3 Theorem *Let H be a finite group and let $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$ be an H -cover associated to a surjection $\varphi : \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$, where x is a basepoint of $\Sigma_{g,n}$. Let \bar{r} be the maximal dimension of an irreducible representation of H . Suppose that either*

- (1) $n = 0$ and $g \geq 2\bar{r} + 2$, or
- (2) n is arbitrary and $g > \max(2\bar{r} + 1, \bar{r}^2)$.

Let Mod_φ be the stabilizer of φ inside $\text{Mod}_{g,n+1}$. Then, using notation as in (1-1), the identity component of the Zariski closure of the image of

$$R_\varphi : \text{Mod}_\varphi \rightarrow \text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$$

is the commutator subgroup of $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$.

We prove Theorem 1.3 in Section 7.15. We refer to representations as in Theorem 1.3 as *higher Prym representations*.

1.4 Remark The constant \bar{r} reflects the group-theoretic properties of H . For example, $\bar{r} \leq \sqrt{\#H}$, and \bar{r} divides the index of any normal abelian subgroup of H . The case that the covering group H is abelian was considered by Looijenga [1997]. In this case, the representation was called a Prym representation. When H is abelian, $\bar{r} = 1$, and so it suffices to take $g \geq 4$ in the statement of Theorem 1.3. If H is dihedral, then $\bar{r} = 2$.

1.5 Remark One may reformulate Theorem 1.3 as follows: the Zariski-closure of the virtual image of $\text{Mod}_{g,n+1}$ under R_φ in $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$ is the commutator subgroup of $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$. Here the *Zariski-closure of the virtual image* is the intersection of the Zariski-closures of the images of all finite-index subgroups on which R_φ is defined.

1.6 Remark The Putman–Wieland conjecture [2013] is heavily influenced by Slogan 1.2. It predicts that if $g > 2$, the virtual action of $\text{Mod}_{g,n+1}$ on $H^1(\Sigma_{g'}, \mathbb{C})$ has no nonzero finite orbits. (In [Putman and Wieland 2013], this conjecture was made for all $g \geq 2$. However, Marković [2022, Theorem 1.3] gave a counterexample when $g = 2$.) The virtual action of $\text{Mod}_{g,n+1}$ on $H^1(\Sigma_{g'}, \mathbb{C})$ has no nonzero finite orbits if and only if the Zariski closure of $\text{Mod}_{g,n+1}$ in $\text{Aut}(H^1(\Sigma_{g'}, \mathbb{C}))$ has no nonzero finite orbits. Also the commutator subgroup of $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$ has no nonzero finite orbits on $H^1(\Sigma_{g'}, \mathbb{C})$, using (7-1). Therefore, the Putman–Wieland conjecture for a given covering $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$ is implied by the statement that the Zariski closure of $\text{Mod}_{g,n+1}$ in $\text{Aut}(H^1(\Sigma_{g'}, \mathbb{C}))$ is the commutator subgroup of $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$.

In particular, since the analogue of the Putman–Wieland conjecture does not hold for $g = 2$ due to Marković’s counterexample [2022, Theorem 1.3], as mentioned above, it follows that no analogue of Theorem 1.3 can hold for $g = 2$. That is, some lower bound on the genus is necessary. We prove Theorem 1.3 when the genus g is bounded below by a function depending on the maximal dimension of any irreducible H -representation. It remains possible, however, that the conclusion of Theorem 1.3 holds whenever $g > 2$, independent of H . As explained in the previous paragraph, if the conclusion of Theorem 1.3 were to hold whenever $g > 2$, one would obtain the Putman–Wieland conjecture as a consequence.

It is also unsurprising that the proof of Theorem 1.3 builds on the techniques used by the first two authors to prove the Putman–Wieland conjecture for g sufficiently large in terms of H ; see [Landesman and Litt 2024a, Theorem 7.2.1].

1.7 Remark It is reasonable to hope that an even stronger statement, describing the exact image of a finite-index subgroup of $\text{Mod}_{g,n+1}$ and not its Zariski closure, might be true, but a proof of this would require additional ideas. In particular, it is natural to ask if the image is an arithmetic subgroup of its Zariski closure.

In proving Theorem 1.3, it is natural to decompose $H^1(\Sigma_{g'}, \mathbb{C})$ into isotypic components corresponding to different irreducible representations of H . We introduce some notation to describe these:

Let G be a group and $\rho: G \rightarrow \text{GL}_r(\mathbb{C})$ an irreducible representation of G . If ρ is self-dual, then by Schur’s lemma, $(\rho \otimes \rho)^G$ is one-dimensional. As $\rho \otimes \rho = \text{Sym}^2(\rho) \oplus \wedge^2 \rho$, exactly one of $(\text{Sym}^2(\rho))^G$, $(\wedge^2 \rho)^G$ is nonzero.

1.8 Definition Let ρ be an irreducible self-dual finite-dimensional representation of a group G . If $(\text{Sym}^2(\rho))^G \neq 0$, we say ρ is *orthogonally self-dual*. If $(\wedge^2 \rho)^G \neq 0$ we say ρ is *symplectically self-dual*.

If \mathbb{V}^ρ is a unitary local system on $\Sigma_{g,n}$ corresponding to a representation ρ of $\pi_1(\Sigma_{g,n}, x)$, we define the *weight-one piece* $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho) \subset H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ to be

$$W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho) := H^1(\Sigma_g, j_* \mathbb{V}^\rho),$$

where $j: \Sigma_{g,n} \hookrightarrow \Sigma_g$ is the natural inclusion. This agrees with the usual notion of weights in algebraic geometry. The groups $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ and $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho^\vee})$ are naturally dual. In particular, if ρ is self-dual, $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ carries a natural perfect pairing with itself. By the graded-commutativity of the cup product, $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ is symplectically self-dual if ρ is orthogonally self-dual, and orthogonally self-dual if ρ is symplectically self-dual.

If ρ factors through a surjection $\varphi: \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$ with H a finite group, there is a natural isomorphism $\rho^\sigma \xrightarrow{\sim} \rho$ for each $\sigma \in \text{Mod}_\varphi$, the stabilizer of φ in $\text{Mod}_{g,n+1}$. Hence Mod_φ acts naturally on $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$, via a homomorphism we name $R_{\varphi,\rho}$. A large part of the proof of Theorem 1.3 consists of checking:

1.9 Theorem Let H be a finite group and let $\varphi: \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$ be a surjective homomorphism, where x is a basepoint of $\Sigma_{g,n}$. Suppose $\rho: H \rightarrow \text{GL}_r(\mathbb{C})$ is an irreducible representation of dimension r , with r satisfying either

- (1) $n = 0$ and $g \geq 2r + 2$, or
- (2) n is arbitrary and $g > \max(2r + 1, r^2)$.

Let \mathbb{V}^ρ denote the local system on $\Sigma_{g,n}$ associated to $\varphi \circ \rho$. Let Mod_φ be the stabilizer of φ inside $\text{Mod}_{g,n+1}$ and

$$R_{\varphi,\rho}: \text{Mod}_\varphi \rightarrow \text{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$$

the natural homomorphism.

Then the image of $R_{\varphi,\rho}$ is Zariski-dense in

- (1) $\text{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is symplectically self-dual,
- (2) $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is orthogonally self-dual, and
- (3) the product of $\text{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ with a finite subgroup of the center of $\text{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is not self-dual.

We prove Theorem 1.9 in Section 7.15.

One may draw a number of concrete corollaries of Theorems 1.3 and 1.9.

1.10 Corollary *Let H be a finite group and X a very general H -curve. Let \bar{r} be the maximal dimension of an irreducible representation of H . Suppose either*

- (1) *H acts freely on X and the genus of X/H is at least $2\bar{r} + 2$, or*
- (2) *the genus of X/H is greater than $\max(2\bar{r} + 1, \bar{r}^2)$.*

Then the Mumford–Tate group of $H^1(X, \mathbb{Q})$ contains the commutator subgroup of $\mathrm{Sp}(H^1(X, \mathbb{Q}))^H$ and is contained in $\mathrm{GSp}(H^1(X, \mathbb{Q}))^H$.

1.11 Corollary *Let X be as in Corollary 1.10. Then the endomorphism algebra of $\mathrm{Jac}(X)$ is $\mathbb{Q}[H]$.*

These corollaries are proven in Section 7.17. We expect Corollary 1.11 will likely have applications in equivariant birational geometry, generalizing the applications in [Hassett and Tschinkel 2022, Section 7] of results of [Grunewald et al. 2015].

We conclude the paper with a result on the monodromy of certain special Kodaira fibrations, that is, surfaces with a smooth projective map to a smooth curve. We refer to these as *Kodaira–Parshin fibrations* (see Definition 9.1); loosely speaking, these Kodaira–Parshin fibrations parametrize families of covers of a fixed curve, with a moving branch point. See Theorem 9.8 for a precise algebraic statement. These results have a purely topological interpretation — namely, they say that for the representations considered in Theorem 1.3, their restrictions to certain “point-pushing” subgroups have large image; see Corollary 9.9 for a precise statement.

1.12 The large n regime Another regime in which we expect big monodromy is the case that we fix g and let n grow large. Specifically, we conjecture the analogue of Theorem 1.3 in this context.

1.13 Conjecture Let H be a finite group and let $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$ be the H -cover associated to a homomorphism $\varphi: \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$, where x is a basepoint of $\Sigma_{g,n}$. Let $\rho: H \rightarrow \mathrm{GL}_r(\mathbb{C})$ be an irreducible H -representation and \mathbb{V}^ρ the local system associated to $\rho \circ \varphi$. We conjecture there is a function $c(g, \dim \rho)$ with the following property. Suppose there are $\Delta > c(g, \dim \rho)$ points of $\Sigma_g - \Sigma_{g,n}$ such that a small loop around each of these points is sent to a nonidentity matrix under the composition $\pi_1(\Sigma_{g,n}) \xrightarrow{\varphi} H \xrightarrow{\rho} \mathrm{GL}_r(\mathbb{C})$. Then, the image of the stabilizer Mod_φ of φ in $\mathrm{Mod}_{g,n+1}$ inside $\mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ is Zariski-dense in

- (1) $\mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is symplectically self-dual,
- (2) $\mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is orthogonally self-dual, and
- (3) the product of $\mathrm{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ with a finite subgroup of the center of $\mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is not self-dual.

1.14 Remark (motivation from arithmetic statistics) A number of works in arithmetic statistics over function fields have proven results in a large q limit setting by computing relevant monodromy groups with finite coefficients associated to spaces of H -covers. See [Achter 2008; Ellenberg et al. 2016; Feng et al. 2023; Park and Wang 2024; Ellenberg and Landesman 2023] for a few examples of this; the last three references are connected to H -covers for a particular group H via [Ellenberg and Landesman 2023, Proposition 6.4.5].

Verifying Conjecture 1.13 would give some evidence for analogous conjectures in number theory, as it would suggest a similar big monodromy result should be true with finite coefficients.

As evidence for Conjecture 1.13, we prove the following implication of the conjecture. If the monodromy is Zariski dense in the subgroups listed in Conjecture 1.13, then it is not contained in any nontrivial parabolic, and so in particular it does not fix any vectors.

1.15 Theorem *Let H be a finite group, $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$ an H -cover, and $\rho: H \rightarrow \mathrm{GL}_r(\mathbb{C})$ an irreducible H -representation. Suppose there are*

$$\Delta > \frac{3r^2}{\sqrt{g+1}} + 8r$$

points of $\Sigma_g - \Sigma_{g,n}$ such that a small loop around each of these points is sent to a nonidentity matrix under the composition $\pi_1(\Sigma_{g,n}) \xrightarrow{\varphi} H \xrightarrow{\rho} \mathrm{GL}_r(\mathbb{C})$. Then, setting $\mathrm{Mod}_\varphi \subset \mathrm{Mod}_{g,n+1}$ to be the stabilizer of φ , there are no nonzero vectors with finite orbit under the image of Mod_φ in $\mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$, for \mathbb{V}^ρ the local system associated to $\rho \circ \varphi$.

We prove Theorem 1.15 in Section 8.5.

1.16 Remark Theorem 1.15 verifies new cases of the Putman–Wieland conjecture [2013, Conjecture 1.2]. See also [Landesman and Litt 2024a, Section 1.8.4] for a summary of other known cases of the Putman–Wieland conjecture. Landesman and Litt [2024a, Theorem 1.4.2] give a variant of Theorem 1.15 when g is large relative to r , in comparison to Theorem 1.15, where we think of Δ as being large relative to g and r .

1.17 Remark We have opted to write Theorem 1.15 with a bound on Δ depending only on the rank r of our given representation ρ and on the genus g of $\Sigma_{g,n}$. We expect, however, that our methods could give stronger bounds in terms of the eigenvalues of the local monodromy of $\rho \circ \varphi$.

1.18 Previous work There is a great deal of past work related to big monodromy for higher Prym representations. In the case that H is abelian, Theorem 1.3 follows from [Looijenga 1997, Corollary 2.6]. Higher Prym representations corresponding to certain nonabelian covering groups H were considered in [Grunewald et al. 2015], and shown to give a rich class of representations of mapping class groups under certain conditions, namely when the cover is “ ϕ -redundant” [Grunewald et al. 2015, Theorem 1.2 and 1.6].

A variant for free groups was previously considered by Grunewald and Lubotzky [2009]. See also the recent paper by Looijenga [2025] for a criterion for big monodromy along somewhat different lines. In the four papers above, the representations above were in fact shown to have *arithmetic image*, meaning that they have finite index in the integral points of their Zariski closure. Our techniques seem not to be able to establish anything towards arithmeticity. There are known examples of families of cyclic branched covers of genus-zero curves with nonarithmetic monodromy [Deligne and Mostow 1986].

Further arithmeticity results associated to Prym representations, primarily in the case that H is abelian, were given by McMullen [2013], Venkataramana [2014a], and Venkataramana [2014b]. Salter and Tshishiku [2020] prove arithmeticity of monodromy groups of certain Kodaira fibrations, corresponding to H -covers with H a finite Heisenberg group. In a more arithmetic direction, big monodromy associated to certain Prym representations played a crucial role in the recent proof of Faltings' theorem given by Lawrence and Venkatesh [2020, Theorem 8.1].

In the setting of covers of projective lines, ie in a $g = 0$ analogue of the setting of Theorem 1.3, a big mod ℓ monodromy result was proven by Jain [2016, Theorem 5.4.2] when H has trivial Schur multiplier for $\ell \nmid 2|H|$. Big mod ℓ monodromy can often be used to prove that the ℓ -adic closure of the image of the mapping class group is large, which implies big Zariski closure. A key tool is a result of Conway and Parker (see eg [Fried and Völklein 1991, Appendix], as well as [Ellenberg et al. 2016, Proposition 3.4] and [Wood 2021]) which gives a stabilization of the braid group action on finite quotients of $\pi_1(\Sigma_{g,n})$. An analogue of this tool in the higher-genus case is work of Dunfield and Thurston [2006, Proposition 6.16], which shows a stabilization in g of the mapping class group action on finite quotients of $\pi_1(\Sigma_g)$ as g grows. Samperton [2020] gives a partial analogue of these results for the mapping class group action on $\pi_1(\Sigma_{g,n})$ with $n > 0$.

It may be possible to use [Dunfield and Thurston 2006, Proposition 6.16] to prove a higher-genus analogue of [Jain 2016, Theorem 5.4.2]. Such a result would have the advantage over Theorem 1.3 that it would give more precise information on the ℓ -adic image of the mapping class group, but the disadvantage that the required lower bound on the genus would be ineffective. Partial results in this direction were obtained in recent work of Sawin and Wood [2024] in the course of studying Heegaard splittings of 3-manifolds. This work uses [Dunfield and Thurston 2006, Proposition 6.16] to compute the intersections of maximal isotropic subspaces with their translates by random elements drawn from the monodromy group of Prym representations, though falls short of computing the actual monodromy group.

Finally, recent work of Landesman and Litt [2024a; 2023a; 2023b] shows that the monodromy group for higher Prym representations is not too small, in the sense that it has no nonzero finite orbit vectors. Here, we build on the methods developed in those papers to prove the monodromy group is as big as possible.

We next describe the primary new ideas of this work that do not appear in [Landesman and Litt 2024a; 2023a; 2023b].

1.19 Innovations of the proof To explain the main new ideas going into our paper, we begin with a sketch of the proof of Theorem 1.9.

Setup for the proof Consider a family of n -pointed curves $\pi: \mathcal{C} \rightarrow \mathcal{M}$, with associated family of punctured curves $\pi^\circ: \mathcal{C}^\circ \rightarrow \mathcal{M}$, such that the induced map $\mathcal{M} \rightarrow \mathcal{M}_{g,n}$ is dominant étale. Let \mathbb{V} be a complex local system on \mathcal{C}° with finite monodromy, whose restriction to a fiber of π° has monodromy given by ρ . It suffices to compute the connected monodromy group of $W_1 R^1 \pi_*^\circ \mathbb{V}$, as the monodromy representation on this local system factors through the representation we are interested in. The main idea of the proof is to analyze the derivative of the period map associated to $W_1 R^1 \pi_*^\circ \mathbb{V}$.

A novel technique: functorial reconstruction Using techniques building on those developed by Landesman and Litt [2024b; 2024a] we show in Theorem 6.2 and Proposition 6.4 that (given our assumptions on g and $\dim \mathbb{V}$) the monodromy representation ρ can be *functorially reconstructed* from the derivative of the period map associated to $W_1 R^1 \pi_*^\circ \mathbb{V}$ at a generic point of \mathcal{M} along so-called Schiffer variations. We think of this reconstruction as a new kind of “generic Torelli theorem with coefficients”—this is our main technical tool.

This enables a novel strategy to obtain information about the local system $W_1 R^1 \pi_*^\circ \mathbb{V}$. We first assume for the sake of contradiction that the monodromy group of $W_1 R^1 \pi_*^\circ \mathbb{V}$ has some undesirable property. Using the general theory of variations of Hodge structures, we describe the consequences of this property for the variation of Hodge structures, and in particular for the derivative of its period map. We then examine the consequences those properties have on the local system obtained by applying our functorial reconstruction algorithm, and finally show that these contradict known properties of \mathbb{V} .

Proving Theorem 1.9 by repeatedly applying functorial reconstruction For example, to show $W_1 R^1 \pi_*^\circ \mathbb{V}$ is irreducible as a representation of the monodromy group, we assume for the sake of contradiction that it is reducible. It would be convenient if this implied that $W_1 R^1 \pi_*^\circ \mathbb{V}$ is reducible as a representation of Hodge structures, but this is not the case—it could instead be, for example, the tensor product of a fixed irreducible Hodge structure with an irreducible variation of Hodge structure. However, in this case the derivative of the period map is still reducible in a suitable sense, and in fact one can check that this holds for any variation of Hodge structures whose underlying local system is reducible. Applying the reconstruction algorithm to a reducible derivative of the period map, we obtain a reducible local system, contradicting the irreducibility of \mathbb{V} .

A slight enhancement of this argument, namely Theorem 6.7, shows that the connected monodromy group of $W_1 R^1 \pi_*^\circ \mathbb{V}$ is in fact a *simple* group, acting irreducibly. By the classification of simple factors of Mumford–Tate groups of Abelian varieties, this group necessarily acts through a minuscule representation; see [Zarkhin 1984, Theorem 0.5.1(b)].

We rule out nonstandard representations in Section 7 via the same strategy. We show that, for these representations, the rank of the derivative of the period map along a Schiffer variation is necessarily large.

Applying the reconstruction algorithm gives a local system of large rank — in fact, larger than the rank of \mathbb{V} , leading to a contradiction.

The last case remaining to rule out is when the monodromy group of $W_1 R^1 \pi_*^\circ \mathbb{V}$ is a classical group but \mathbb{V} is not self-dual. In this case we show that $W_1 R^1 \pi_*^\circ \mathbb{V}$ is self-dual as a variation of Hodge structures, and deduce from the reconstruction algorithm that \mathbb{V} is self-dual, obtaining a contradiction. This concludes our sketch of the idea of the proof of Theorem 1.9.

Proving Theorem 1.3 To deduce Theorem 1.3, one may explicitly describe the commutator subgroup of $\mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$ as a product of the groups appearing in Theorem 1.9 — that theorem implies that the connected monodromy group of $H^1(\Sigma_{g'}, \mathbb{C})$ surjects onto each of the simple factors of $\mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$. It then suffices by the Goursat–Kolchin–Ribet criterion of Katz [1990, Proposition 1.8.2] to show that for ρ_1, ρ_2 irreducible H -representations, with associated local systems $\mathbb{U}_1, \mathbb{U}_2$ on \mathcal{C}° , an isomorphism $W_1 R^1 \pi_*^\circ \mathbb{U}_1 \simeq W_1 R^1 \pi_*^\circ \mathbb{U}_2$ necessarily comes from an isomorphism $\rho_1 \simeq \rho_2$. We deduce this from our functorial reconstruction results.

A new technical ingredient: global generation Although the heart of our functorial reconstruction results rest on understanding the derivative of a certain period map, following similar techniques introduced by Landesman and Litt [2023a; 2024a], there are a number of substantial innovations. In particular, the key to the proof of Theorem 6.2 is Proposition 4.9, which analyzes the global generation properties of flat vector bundles under isomonodromic deformation. In past work, the first two authors were only able to prove that the relevant vector bundles were *generically* globally generated, as opposed to actually being globally generated. By analyzing the obstruction for generically globally generated bundles to be globally generated, we are able to push the methods developed previously farther.

The proof of our main results underlies a new connection between *global generation* of certain vector bundles on curves and questions about big monodromy; see Section 10 for further details. As far as we are aware this connection is totally novel.

Acknowledgements Landesman was supported by the National Science Foundation under award DMS 2102955. Litt was supported by the NSERC Discovery grant “*Anabelian methods in arithmetic and algebraic geometry*”. Sawin was supported by NSF grant DMS-2101491 and a Sloan Research Fellowship. The authors are grateful for useful discussions with Kevin Chang, Josh Lam, Eduard Looijenga, Alex Lubotzky, Andrew Putman, Kasra Rafi, Andy Ramirez-Cote and Bena Tshishiku.

2 Notation and preliminaries on moduli

2.1 Notation Throughout this paper, we work over the complex numbers. Suppose we are given a smooth proper family of curves $\pi: \mathcal{C} \rightarrow \mathcal{M}$ with geometrically connected fibers and n sections $s_1, \dots, s_n: \mathcal{M} \rightarrow \mathcal{C}$ with disjoint images $\mathcal{D}_1, \dots, \mathcal{D}_n \subset \mathcal{C}$. Let $\mathcal{C}^\circ = \mathcal{C} - \{\mathcal{D}_1, \dots, \mathcal{D}_n\}$. If the induced map $\mathcal{M} \rightarrow \mathcal{M}_{g,n}$ is

dominant étale, we say π is a *versal family of n -pointed curves of genus g* , and refer to $\pi^\circ := \pi|_{\mathcal{C}^\circ}$ as the associated *versal family of n -punctured curves*.

Suppose, moreover, we have a diagram

$$(2-1) \quad \begin{array}{ccc} \mathcal{X} & \xrightarrow{f} & \mathcal{C} \\ & \searrow \pi' & \downarrow \pi \\ & & \mathcal{M} \end{array} \quad \left. \vphantom{\begin{array}{ccc} \mathcal{X} & \xrightarrow{f} & \mathcal{C} \\ & \searrow \pi' & \downarrow \pi \\ & & \mathcal{M} \end{array}} \right\} s_1, \dots, s_n$$

where π is a versal family of n -pointed curves of genus g , and π' is a smooth proper curve with geometrically connected fibers of genus g' . Suppose f is finite, Galois and unramified away from $\bigcup_{i=1}^n \mathcal{D}_i$. Let $\mathcal{X}^\circ = f^{-1}(\mathcal{C}^\circ)$, let $\pi^\circ := \pi|_{\mathcal{C}^\circ}$, and let $f^\circ = f|_{\mathcal{X}^\circ}$. For $m \in \mathcal{M}$ a point, we set $C = \mathcal{C}_m$, $C^\circ = \mathcal{C}_m^\circ$ and $D = C - C^\circ$; let $c \in C^\circ$ be a point. As f° is finite Galois, it induces a surjection $\pi_1(\mathcal{C}^\circ, c) \twoheadrightarrow H$ for some finite group H . As π' has geometrically connected fibers and f is Galois, the composition

$$\psi : \pi_1(C^\circ, c) \rightarrow \pi_1(\mathcal{C}^\circ, c) \rightarrow H$$

is surjective.

Let $x \in \Sigma_{g,n}$ be a basepoint and $\varphi : \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$ a surjection. A *versal family of φ -covers* is the data of a diagram as in (2-1) above, together with

- (1) a point $c \in \mathcal{C}^\circ$, $m = \pi^\circ(c)$, $C^\circ = \mathcal{C}_m^\circ$, and an identification $i : (\Sigma_{g,n}, x) \simeq (C^\circ, c)$, such that
- (2) under this identification the map ψ above identifies with φ .

If $\rho : H \rightarrow \text{GL}_r(\mathbb{C})$ is a representation of a finite group H , and $\varphi : \pi_1(\Sigma_{g,n}, x) \rightarrow H$ is a map, we often use \mathbb{V}^ρ to denote the local system on $\Sigma_{g,n}$ associated to $\rho \circ \varphi$.

2.2 Remark Versal families of φ -covers exist by [Wewers 1998, Theorem 4]. One may also construct such a versal family by taking an open substack of the stack $\mathcal{B}_{g,n}(H)$ as constructed in [Abramovich et al. 2003, Section 2.2] (where the group H here is called G there). The above constructions give versal families of Deligne–Mumford stacks. Hence, if one wishes, one may pass to a dominant étale cover of the \mathcal{M} thus constructed, making all the objects in question schemes.

2.3 Preliminaries on moduli In this section we explain how to interpret the main theorems of this paper as being about the monodromy of (summands of) $R^1\pi'_*\mathbb{C}$, for π' as in (2-1), arising from a versal family of H -covers.

Recall that, given a cover $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$, we are studying the action of finite-index subgroups of $\text{Mod}_{g,n+1}$ on the abelianization of $\pi_1(\Sigma_{g',n'}, x')$, a finite-index subgroup of $\pi_1(\Sigma_{g,n}, x)$. One may interpret the action of $\text{Mod}_{g,n+1}$ on $\pi_1(\Sigma_{g,n}, x)$ as follows. Let $\mathcal{M}_{g,n}$ be the Deligne–Mumford moduli stack of

genus g curves with n marked points, $\mathcal{C}_{g,n}/\mathcal{M}_{g,n}$ be the universal family, and $\mathcal{C}_{g,n}^\circ$ the associated family of n -punctured curves. Note that $\mathcal{C}_{g,n}^\circ$ is canonically isomorphic to $\mathcal{M}_{g,n+1}$.

Now consider the map

$$p_1: \mathcal{C}_{g,n}^\circ \times_{\mathcal{M}_{g,n}} \mathcal{C}_{g,n}^\circ \rightarrow \mathcal{C}_{g,n}^\circ$$

given by projection onto the first factor. This map has a canonical section given by the diagonal Δ . Thus there is a short exact sequence of fundamental groups

$$1 \rightarrow \pi_1(\Sigma_{g,n}) \rightarrow \pi_1(\mathcal{C}_{g,n}^\circ \times_{\mathcal{M}_{g,n}} \mathcal{C}_{g,n}^\circ) \xrightarrow{p_{1*}} \pi_1(\mathcal{C}_{g,n}^\circ) \rightarrow 1,$$

split by Δ_* , inducing a natural action of $\pi_1(\mathcal{C}_{g,n}^\circ)$ on $\pi_1(\Sigma_{g,n})$. It follows from eg [Farb and Margalit 2012, Section 10.6.3 and page 353] that there is a natural identification of $\pi_1(\mathcal{C}_{g,n}^\circ) = \pi_1(\mathcal{M}_{g,n+1})$ with $\text{Mod}_{g,n+1}$, under which this action identifies with the one we are studying.

Fix a surjection $\varphi: \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$ and a versal family of φ -covers with notation as in (2-1). In analogy with the fiber product construction above, we may consider the map

$$q_1: \mathcal{C}^\circ \times_{\mathcal{M}} \mathcal{X} \rightarrow \mathcal{C}^\circ$$

given by projection onto the first coordinate. We claim that the monodromy of $R^1q_{1*}\mathbb{C}$ factors through the representation R_φ studied in Theorem 1.3, and indeed factors through a finite-index subgroup of Mod_φ .

We first construct a map from $\pi_1(\mathcal{C}^\circ)$ to Mod_φ . The family of $(n+1)$ -pointed curves

$$\mathcal{C}^\circ \times_{\mathcal{M}} \mathcal{C} \rightarrow \mathcal{C}^\circ$$

(with sections given by the pullbacks of s_1, \dots, s_n and the tautological section induced by the diagonal) induces a map $\mathcal{C}^\circ \rightarrow \mathcal{M}_{g,n+1}$. By the discussion above and [Landesman and Litt 2024a, Lemma 2.1.4], the image of the induced map on fundamental groups $\pi_1(\mathcal{C}^\circ) \rightarrow \pi_1(\mathcal{M}_{g,n+1})$ has finite index in $\pi_1(\mathcal{M}_{g,n+1}) = \text{Mod}_{g,n+1}$. By the definition of \mathcal{M} , the image of the induced map is contained in Mod_φ , as defined as in Theorem 1.3, and hence the image has finite index in Mod_φ .

Unwinding this discussion, the monodromy representation on $R^1q_{1*}\mathbb{C}$ factors through the representation R_φ , with image a finite-index subgroup of the image of R_φ . Thus to understand the image of R_φ , it suffices to understand the image of this monodromy representation associated to $R^1q_{1*}\mathbb{C}$.

Moreover q_1 is itself the pullback of π' along π° , which induces a surjection on fundamental groups. Thus the image of the monodromy representation associated to $R^1q_{1*}\mathbb{C}$ is the same as that of the monodromy representation associated to $R^1\pi'_*\mathbb{C}$.

Now let $\rho: H \rightarrow \text{GL}_r(\mathbb{C})$ be a representation and \mathbb{V}^ρ the associated local system on \mathcal{C}° . A completely analogous argument shows that the image of the monodromy representation associated to $W_1R^1\pi_*^\circ\mathbb{V}^\rho$ is the same up to finite index as the representation $R_{\varphi,\rho}$ considered in Theorem 1.9.

3 Review of parabolic bundles and period maps

Let C be a smooth projective curve over \mathbb{C} , and $D = x_1 + \dots + x_n$ a reduced effective divisor on C .

3.1 Definition A *parabolic vector bundle* E_\star on (C, D) is a vector bundle E on C , a decreasing filtration $E_{x_j} = E_j^1 \supseteq E_j^2 \supseteq \dots \supseteq E_j^{n_j+1} = 0$ for each $1 \leq j \leq n$, and an increasing sequence of real numbers $0 \leq \alpha_j^1 < \alpha_j^2 < \dots < \alpha_j^{n_j} < 1$ for each $1 \leq j \leq n$, referred to as *weights*. Here, E_{x_j} refers to the fiber of E at x_j , and hence the filtration is merely a filtration of vector spaces on the fiber of E at x_j , not a filtration of the vector bundle E . We use $E_\star = (E, \{E_j^i\}, \{\alpha_j^i\})$ to denote the data of a parabolic bundle. Given a parabolic bundle E_\star , we will often write E_0 for the underlying vector bundle E .

3.2 Parabolic bundles admit a notion of *parabolic stability*, analogous to the usual notion of stability for vector bundles, which we next recall. First, the *parabolic degree* of a parabolic bundle E_\star is

$$\text{deg}(E_\star) := \text{deg}(E) + \sum_{j=1}^n \sum_{i=1}^{n_j} \alpha_j^i \dim(E_j^i / E_j^{i+1}).$$

Then, the *parabolic slope* is defined by $\mu_\star(E_\star) := \text{deg}(E_\star) / \text{rk}(E_\star)$. Any subbundle $F \subset E$ has an induced parabolic structure $F_\star \subset E_\star$ defined as follows: we take the filtration over x_j on F is obtained from the filtration

$$F_{x_j} = E_j^1 \cap F_{x_j} \supseteq E_j^2 \cap F_{x_j} \supseteq \dots \supseteq E_j^{n_j+1} \cap F_{x_j} = 0$$

by removing redundancies. For the weight associated to $F_j^i \subset F_{x_j}$ one takes

$$\max_{k, 1 \leq k \leq n_j} \{\alpha_j^k : F_j^i = E_j^k \cap F_{x_j}\}.$$

A parabolic bundle E_\star is *parabolically semistable* if for every nonzero subbundle $F \subset E$ with induced parabolic structure F_\star , we have $\mu_\star(F_\star) \leq \mu_\star(E_\star)$. Similarly, a parabolic bundle E_\star is *parabolically stable* if for every nonzero subbundle $F \subset E$ with induced parabolic structure F_\star , we have $\mu_\star(F_\star) < \mu_\star(E_\star)$.

We next introduce the notation E_\star^ρ for the parabolic bundle corresponding to a unitary representation ρ .

3.3 Notation Let C be a curve and $D \subset C$ a divisor. Under the Mehta–Seshadri correspondence from [Mehta and Seshadri 1980], there is a bijection between irreducible unitary representations of $\pi_1(C - D)$ and parabolic degree 0 stable parabolic vector bundles on C , with parabolic structure along D . Given an irreducible unitary representation

$$\rho: \pi_1(C - D) \rightarrow U(r)$$

of dimension r , we use E_\star^ρ to denote the parabolic bundle associated to ρ . The underlying vector bundle E_0^ρ of E_\star^ρ is the Deligne canonical extension of the flat bundle on $C - D$ associated to ρ , ie E_0^ρ carries a flat connection

$$\nabla: E_0^\rho \rightarrow E_0^\rho \otimes \Omega_C^1(\log D)$$

with monodromy ρ and whose residues have eigenvalues with real parts in $[0, 1)$. See [Landesman and Litt 2024b, Definition 3.3.1] for details of how to associate a parabolic structure to a connection.

Such unitary representations will often arise as follows: we will start with a surjection $\varphi: \pi_1(C - D) \twoheadrightarrow H$ for some finite group H . Then for each irreducible representation ρ of H , the representation $\rho \circ \varphi$ is unitary, and we will abuse notation to denote it by ρ as well.

Let $D_{\text{nontriv}}^\rho \subset D$ denote the subset of points $p \in D$ such that the local inertia at p under ρ is not the identity. Set Δ to be the number of points in D_{nontriv}^ρ .

3.4 Definition Given a parabolic bundle $E_\star = (E, \{E_j^i\}, \{\alpha_j^i\})$, let $J \subset \{1, \dots, n\}$ denote the set of integers $j \in \{1, \dots, n\}$ for which $\alpha_j^1 = 0$, and define

$$(3-1) \quad \widehat{E}_0 := \ker \left(E \rightarrow \bigoplus_{j \in J} E_{x_j} / E_j^2 \right).$$

(This is a special case of more general notation used for coparabolic bundles as in [Landesman and Litt 2024b, 2.2.8] or the equivalent [Boden and Yokogawa 1996, Definition 2.3], but is all we will need for this paper.) In particular, $\widehat{E}_0 \subset E$ is a subsheaf. Note that if E_\star is the parabolic bundle associated to a unitary representation of $\pi_1(C - D)$, the natural logarithmic connection on E_0 descends to a logarithmic connection on \widehat{E}_0 (albeit with different residues), by a local calculation.

3.5 Proposition Let \mathbb{V} be a unitary local system on $C - D$, and let E_\star be the associated parabolic bundle on C . Then there is a natural mixed Hodge structure on $H^1(C - D, \mathbb{V})$, and natural isomorphisms

$$(F^1 \cap W_1)H^1(C - D, \mathbb{V}) = H^0(C, \widehat{E}_0 \otimes \omega_C(D)),$$

$$W_1 H^1(C - D, \mathbb{V}) / (F^1 \cap W_1)H^1(C - D, \mathbb{V}) = H^1(C, E_0).$$

Proof Everything except the last line is [Landesman and Litt 2023a, Lemma 3.2]. The last line is just unwinding definitions; see eg [Landesman and Litt 2023a, Section 3] or [Landesman and Litt 2024a, Theorem 4.1.1]. \square

Now suppose $\pi: \mathcal{C} \rightarrow \mathcal{M}$ is a versal family of punctured n -pointed curves, $\pi^\circ: \mathcal{C}^\circ \rightarrow \mathcal{M}$ is the associated punctured family, and \mathbb{V} a unitary local system on \mathcal{C}° . It is well-known (see [Landesman and Litt 2024a, Theorem 4.1.1]) that $R^1 \pi_*^\circ \mathbb{V}$ carries an admissible complex variation of Hodge structures, with the fibers of $W_1 R^1 \pi_*^\circ \mathbb{V}$ as described in Proposition 3.5. Let $m \in \mathcal{M}$ be a point, $C = \pi^{-1}(m)$, $C^\circ = (\pi^\circ)^{-1}(m)$, and $D = C - C^\circ$. Let ρ be the monodromy representation associated to $\mathbb{V}|_{C^\circ}$. The derivative of the period map associated to $W_1 R^1 \pi_*^\circ \mathbb{V}$ is, by Proposition 3.5, a map

$$dP_m^\rho: T_m \mathcal{M} \rightarrow \text{Hom}(H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)), H^1(C, E_0^\rho)).$$

As π° is a versal family, for each $m \in \mathcal{M}$ we have that $T_m^* \mathcal{M} = H^0(\omega_C^{\otimes 2}(D))$, and so by Serre duality we may adjointly obtain a map

$$H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) \otimes H^0(C, (E_0^\rho)^\vee \otimes \omega_C) \rightarrow H^0(C, \omega_C^{\otimes 2}(D)).$$

There is another natural map between these vector spaces, which we denote by B_m^ρ and describe next.

There is a bilinear form

$$(3-2) \quad \widehat{E}_0^\rho \otimes (E_0^\rho)^\vee \rightarrow \mathcal{O}_C$$

arising from the inclusion $\widehat{E}_0^\rho \rightarrow E_0^\rho$ and the pairing $E_0^\rho \otimes (E_0^\rho)^\vee \rightarrow \mathcal{O}_C$. Twisting this bilinear form by $\omega_C^{\otimes 2}(D)$ yields the map of sheaves

$$\widehat{E}_0^\rho \otimes \omega_C(D) \otimes (E_0^\rho)^\vee \otimes \omega_C \rightarrow \omega_C^{\otimes 2}(D).$$

Taking global sections yields the map

$$(3-3) \quad H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D) \otimes (E_0^\rho)^\vee \otimes \omega_C) \rightarrow H^0(C, \omega_C^{\otimes 2}(D)).$$

Finally, we define

$$(3-4) \quad B_m^\rho: H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) \otimes H^0(C, (E_0^\rho)^\vee \otimes \omega_C) \rightarrow H^0(C, \omega_C^{\otimes 2}(D))$$

to be the composition of the multiplication map

$$H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) \otimes H^0(C, (E_0^\rho)^\vee \otimes \omega_C) \rightarrow H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D) \otimes (E_0^\rho)^\vee \otimes \omega_C)$$

with (3-3).

By combining Proposition 3.5 above with [Landesman and Litt 2024a, Theorem 5.1.6], we obtain the following description of the derivative of the period map.

3.6 Proposition *Let $\pi: \mathcal{C} \rightarrow \mathcal{M}$ be a versal family of marked curves of genus g as in Notation 2.1, and let $\pi^\circ: \mathcal{C}^\circ \rightarrow \mathcal{M}$ be the associated punctured family. Fix $m \in \mathcal{M}$ and set $C = \pi^{-1}(m)$, $C^\circ = (\pi^\circ)^{-1}(m)$ and $D = C - C^\circ$. Let \mathbb{V} be a unitary local system on \mathcal{C}° , and let ρ be the monodromy representation of $\mathbb{V}|_{C^\circ}$. Then the derivative of the period map associated to $W_1 R^1 \pi_* \mathbb{V}$ is identified with the map B_m^ρ defined in (3-4).*

The map dP_m^ρ described above gives rise to a number of other maps by adjointness; we will find it convenient to name some of them. We have denoted the bilinear pairing of (3-4) by B_m^ρ . We will denote by θ_m^ρ the adjoint map

$$(3-5) \quad \theta_m^\rho: H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) \rightarrow H^1(C, E_0^\rho) \otimes H^0(C, \omega_C^{\otimes 2}(D)),$$

where we identify $H^1(C, E_0^\rho)$ with $H^0(C, (E_0^\rho)^\vee \otimes \omega_C)^\vee$ by Serre duality.

4 Global generation of vector bundles on generic curves

4.1 A generalization of the non-ggg lemma We will need a generalization of the non-ggg (generically globally generated) lemma from [Landesman and Litt 2024b, Proposition 6.3.6]; see also [loc. cit., Proposition 6.3.1]. In order to prove this generalization, we will make use of the following lemma, bounding the rank of the global sections of a subbundle of a vector bundle in terms of various numerical invariants.

4.2 Lemma *Let F be a vector bundle on a smooth proper curve C of genus g , and let $V \subset F$ be a subbundle with $c := \text{rk}(F) - \text{rk}(V)$. Suppose that $0 = N^0 \subset N^1 \subset \dots \subset N^k = V$ is the Harder–Narasimhan filtration of V . Let μ be a rational number. If*

- (1) $h^0(C, F) - h^0(C, V) = \delta$,
- (2) $0 \leq \mu(N^i/N^{i-1}) \leq 2g - 2$ for $1 \leq i \leq k$, and
- (3) $\mu(F) \geq \mu$,

then

$$(2g - 1 - \mu) \text{rk}(F) \geq cg - \delta.$$

Proof Combining [Landesman and Litt 2024b, Lemma 6.2.1] with Riemann–Roch for F gives

$$\frac{1}{2} \deg(V) + \text{rk}(V) \geq h^0(C, V) = h^0(C, F) - \delta \geq \deg(F) - (g - 1) \text{rk}(F) - \delta.$$

By assumption (2) we obtain $\mu(V) \leq 2g - 2$, ie $\deg(V) \leq (2g - 2) \text{rk}(V)$, giving

$$g \text{rk}(V) \geq \deg(F) - (g - 1) \text{rk}(F) - \delta.$$

Then, $\text{rk}(V) = \text{rk}(F) - c$ and $\mu(F) \geq \mu$ gives

$$\begin{aligned} g(\text{rk}(F) - c) &= g(\text{rk}(V)) \geq \deg(F) - (g - 1) \text{rk}(F) - \delta \\ &\geq \mu(F) \text{rk}(F) - (g - 1) \text{rk}(F) - \delta \\ &\geq \mu \text{rk}(F) - (g - 1) \text{rk}(F) - \delta. \end{aligned}$$

Simplifying this gives

$$(4-1) \quad (2g - 1 - \mu) \text{rk}(F) \geq cg - \delta. \quad \square$$

4.3 Proposition *Let μ be a rational number. Suppose E is a vector bundle on a smooth proper connected genus g curve and any nonzero quotient bundle $E \twoheadrightarrow Q$ satisfies $\mu(Q) \geq \mu$. Let $U \subset E$ be a proper subbundle with $c := \text{rk}(E) - \text{rk}(U) > 0$ and $\delta := h^0(C, E) - h^0(C, U)$. If $\mu < 2g - 1$ then we have $\text{rk}(E) \geq (cg - \delta)/(2g - 1 - \mu)$, while if $\mu \geq 2g - 1$ then we have $\delta \geq cg + \mu - (2g - 1)$.*

Proof As a first step, we may reduce to the case U is generically globally generated by replacing U with the saturation of the image of $H^0(C, U) \otimes \mathcal{O}_C \rightarrow U \rightarrow V$. Next, let N^i denote the largest filtered part of the Harder–Narasimhan filtration of U , whose associated graded sequence of vector bundles all have slopes $> 2g - 2$. Because $U \subset E$ was assumed saturated, and $N^i \subset U$ is saturated, the quotient $U/N^i \subset E/N^i$ is also a saturated inclusion of vector bundles.

Now, let $F := E/N^i$ and let $V := U/N^i$. We now wish to apply Lemma 4.2 to the subbundle $V \subset F$, which will tell us

$$(4-2) \quad (2g - 1 - \mu) \text{rk}(F) \geq cg - \delta.$$

If we also assume $2g - 1 > \mu$, dividing both sides of (4-2) by $2g - 1 - \mu$ gives our desired inequality

$$\text{rk}(E) \geq \text{rk}(E/N^i) \geq \frac{cg - \delta}{2g - 1 - \mu}.$$

On the other hand, if $\mu \geq 2g - 1$, then rearranging (4-2), and using that $\text{rk}(F) \geq 1$ (since $V \subset F$ is a proper subbundle), gives the claimed inequality

$$\delta \geq cg - (2g - 1 - \mu) \text{rk}(F) = cg + (\mu - (2g - 1)) \text{rk}(F) \geq cg + \mu - (2g - 1).$$

To conclude, it remains to verify conditions (1)–(3) of Lemma 4.2.

For condition (1), we will show

$$\delta = h^0(C, E) - h^0(C, U) = h^0(C, E/N^i) - h^0(C, U/N^i).$$

Indeed, since $H^1(C, N^i) = 0$ by [Landesman and Litt 2024b, Lemma 6.3.5], we find $h^0(C, E/N^i) = h^0(C, E) - h^0(C, N^i)$ and $h^0(C, U/N^i) = h^0(C, U) - h^0(C, N^i)$. This implies $h^0(C, E) - h^0(C, U) = h^0(C, E/N^i) - h^0(C, U/N^i)$, and the former is δ by definition.

Condition (3) holds because F is a quotient of E , and so $\mu(F) \geq \mu$ by assumption.

Finally, we verify condition (2). We wish to show each associated graded piece of the Harder–Narasimhan filtration of V has slope between 0 and $2g - 2$. The upper bound follows from the construction of V as U/N^i . We conclude by verifying the lower bound. Recall we assumed above that U is generically globally generated. Therefore, V is generically globally generated as it is a quotient of U , and hence any quotient of V is itself generically globally generated, and thus has positive slope. Applying this to the smallest slope associated graded piece of the Harder–Narasimhan filtration verifies (2). \square

4.4 Lemma Suppose $E_\star = (E, \{E_j^i\}, \{\alpha_j^i\})$ is a nonzero parabolic bundle on (C, D) , where C is a smooth proper connected genus g curve and $D = x_1 + \dots + x_n$ is a reduced effective divisor. Assume E_\star is parabolically semistable of slope $\mu + n$ with $\mu < 2g - 1$. Suppose \hat{E}_0 has a proper subbundle $U \subset \hat{E}_0$ with $c := \text{rk}(\hat{E}_0) - \text{rk}(U)$ and $\delta := h^0(C, \hat{E}_0) - h^0(C, U)$. Then

$$\text{rk}(E) = \text{rk}(\hat{E}_0) \geq \frac{cg - \delta}{2g - 1 - \mu}.$$

Proof Note that [Landesman and Litt 2024b, Lemma 6.3.4] implies any quotient of vector bundles $\hat{E}_0 \rightarrow Q$ satisfies $\mu(Q) \geq \mu$. We may therefore conclude by applying Proposition 4.3 to the vector bundle \hat{E}_0 . \square

4.5 Lemma The bundle $E_0^{\rho \vee}$ is semistable if and only if \hat{E}_0^ρ is semistable. The bundle $E_0^{\rho \vee}$ is stable if and only if \hat{E}_0^ρ is stable.

Proof Stability of \hat{E}_0^ρ is equivalent to stability of $\hat{E}_0^\rho(D)$. In turn, this is equivalent to stability of $(E_0^{\rho \vee})^\vee$ by [Yokogawa 1995, (3.1)], which shows $\hat{E}_0^\rho(D) \simeq (E_0^{\rho \vee})^\vee$. Finally, stability of $(E_0^{\rho \vee})^\vee$ is equivalent to stability of $E_0^{\rho \vee}$. The same holds with semistability in place of stability. \square

4.6 Global generation and deformation theory In this section, we deduce Proposition 4.9 from Lemma 4.4 by passing to a generic curve. We note that Proposition 4.9 is a result about *global generation* of vector bundles, as opposed to just generic global generation.

Suppose C is a curve, D is a reduced effective divisor in C , and E is a vector bundle on C . Below we denote by $\text{At}_{(C,D)}(E)$ the preimage of $T_C(-D)$ under the natural map $\text{At}_C(E) \rightarrow T_C$, where $\text{At}_C(E)$ is the Atiyah bundle of E . For some background on Atiyah bundles relevant to the context of this paper, see [Landesman and Litt 2024b, Section 3]. We begin by recalling a couple of general facts from deformation theory relating to the Atiyah bundle.

4.7 Lemma *Let C be a smooth projective curve over a field k , and D a reduced effective divisor on C . Let E be a vector bundle on C . Then the space of first-order deformations of the triple (C, D, E) is naturally in bijection with the first cohomology of the Atiyah bundle, $H^1(C, \text{At}_{(C,D)}(E))$.*

Proof See [Landesman and Litt 2024b, Proposition 3.5.5], or see [Sernesi 2006, Theorem 3.3.11] for a proof in the case where E is a line bundle and D is empty; the general case is identical. \square

4.8 Lemma *With notation as in Lemma 4.7, let $s \in \Gamma(C, E)$ be a global section. Let*

$$\psi: \text{At}_{(C,D)}(E) \rightarrow E$$

be the map sending a differential operator ξ to $\xi(s)$. Then given $v \in H^1(C, \text{At}_{(C,D)}(E))$, corresponding to some first-order deformation $(\mathcal{C}, \mathcal{D}, \mathcal{E})$ over $k[\epsilon]/\epsilon^2$, the section s extends to a global section of \mathcal{E} if and only if

$$\psi(v) \in H^1(C, E)$$

vanishes.

In particular, if E carries a flat connection ∇ with logarithmic singularities (equivalently, the natural map $\text{At}_{(C,D)}(E) \rightarrow T_C(-D)$ is equipped with a section q^∇), then s extends to a first-order neighborhood in the universal isomonodromic deformation of E if and only if the map

$$\psi \circ q^\nabla: T_C(-D) \rightarrow E$$

induces the zero map

$$H^1(C, T_C(-D)) \rightarrow H^1(C, E).$$

Proof For the first paragraph, see [Sernesi 2006, Proposition 3.3.14] for the case where E is a line bundle and D is empty; the general case is identical. For the second, fix $v \in H^1(C, T_C(-D))$, corresponding to a first-order deformation of (C, D) . The element $q^\nabla(v) \in H^1(\text{At}_{(C,D)}(E))$ corresponds to, by [Landesman and Litt 2024b, Proposition 3.5.7], the first-order isomonodromic deformation of E in the direction v . Thus by the first paragraph, $\psi \circ q^\nabla(v) = 0$; as v was arbitrary, this completes the proof. \square

4.9 Proposition Fix a unitary representation $\rho: \pi_1(\Sigma_{g,n}) \rightarrow \text{GL}_r(\mathbb{C})$. Let (C, D) be a general n -pointed curve of genus $g \geq 2 + 2r$. Then $\widehat{E}_0^\rho \otimes \omega_C(D)$ is not only generically globally generated, but even globally generated.

Proof It suffices to consider the case that ρ is irreducible and nontrivial (as ω_C is globally generated). Letting \mathbb{V}^ρ denote the local system on C° associated to ρ , we may assume that $H^1(C, \widehat{E}_0^\rho \otimes \omega_C(D)) = 0$, as this is dual to $H^0(C, E_0^{\rho^\vee}) = H^0(C^\circ, \mathbb{V}^{\rho^\vee})$.

Suppose that the theorem is false, so that for a general n -pointed curve (C, D) , $\widehat{E}_0^\rho \otimes \omega_C(D)$ is not globally generated. Equivalently, there exists $p \in C$ such that $H^1(C, \widehat{E}_0^\rho \otimes \omega_C(D - p)) \neq 0$. Serre dually, $H^0(C, E_0^{\rho^\vee}(p))$ is nonzero. As (C, D) is general, we may assume that there exists a section $s \in H^0(C, E_0^{\rho^\vee}(p))$ such that for every first-order deformation (\tilde{C}, \tilde{D}) of (C, D) , there exists a first-order deformation \tilde{p} of p to \tilde{C} such that s survives to $H^0(\tilde{C}, \tilde{E}_0^{\rho^\vee}(\tilde{p}))$, where $\tilde{E}_0^{\rho^\vee}$ is the isomonodromic deformation of $E_0^{\rho^\vee}$ to (\tilde{C}, \tilde{D}) . If $p \in D$ but $\tilde{p} \notin \tilde{D}$, then we may, by passing to a different general curve (C', D') , assume that $p \notin D$. Thus we can and do assume in what follows that if $p \in D$, then $\tilde{p} \subset \tilde{D}$ for all first-order deformations (\tilde{C}, \tilde{D}) of (C, D) .

There are two cases, depending on whether or not $p \in D$; we set up notation to handle them simultaneously. If $p \in D$, we set $D' = D$; otherwise, we set $D' = D + p$. The bundle $\mathcal{O}(p)$ has a natural connection on it with regular singularities at p and trivial monodromy, and with residue -1 at p . We give $E_0^{\rho^\vee}(p)$ the tensor product connection, which has regular singularities along D' .

First-order deformations of (C, D) are parametrized by $H^1(C, T_C(-D))$, and first-order deformations of (C, D') are parametrized by $H^1(C, T_C(-D'))$. By Lemma 4.8, applied to (C, D') and the vector bundle $E_0^{\rho^\vee}(p)$, a section $s \in H^0(C, E_0^{\rho^\vee}(p))$ extends to the first-order deformation parametrized by $\eta \in H^1(C, T_C(-D'))$ if and only if the image of η under the map

$$H^1(C, T_C(-D')) \xrightarrow{H^1(\nabla s)} H^1(E_0^{\rho^\vee}(p))$$

is zero, where $\nabla s: T_C(-D') \rightarrow E_0^{\rho^\vee}(p)$ is the map sending a vector field X to $\nabla_X s$, the derivative of s with respect to the natural connection on $E_0^{\rho^\vee}(p)$ with regular singularities along D' , described in the previous paragraph. Note that ∇s is nonzero, since if it were zero, s would be flat, which is impossible as the monodromy representation ρ^\vee is nontrivial.

Thus it suffices to show that

$$\ker(H^1(C, T_C(-D')) \xrightarrow{H^1(\nabla s)} H^1(E_0^{\rho^\vee}(p)))$$

does not surject onto $H^1(C, T_C(-D))$ under the natural map induced by the inclusion $T_C(-D') \rightarrow T_C(-D)$.

Since $\text{deg } D' \geq \text{deg } D - 1$, it is enough to show that the rank of the map induced by ∇s on H^1 is at least 2, or equivalently that the Serre dual map

$$\widehat{E}_0^\rho \otimes \omega_C(D - p) \rightarrow \omega_C^{\otimes 2}(D')$$

induces a map of rank at least 2 on H^0 . Setting U to be the kernel of this map, $\mu := 2g - 3$, and δ to be the rank of the map

$$H^0(\widehat{E}_0^\rho \otimes \omega_C(D - p)) \rightarrow H^0(\omega_C^{\otimes 2}(D')),$$

we have by Lemma 4.4 that

$$r \geq \frac{1}{2}(g - \delta),$$

and hence that $\delta \geq g - 2r$. Hence $\delta \geq 2$ as long as $g - 2r \geq 2$, which holds by assumption. \square

5 Preliminaries on variations of Hodge structure

In this section, we freely use the terminology of Tannakian categories. For background, we suggest the reader consult [Deligne et al. 1982, II].

Let Y be a smooth variety over \mathbb{C} and x a point of Y .

5.1 Definition (integral K -VHS) Let K be a number field. A K -variation of Hodge structures on Y is a K -local system \mathbb{V} on Y equipped with a decreasing filtration F^\bullet on $W_{K/\mathbb{Q}}\mathbb{V} \otimes \mathcal{O}_Y$, where $W_{K/\mathbb{Q}}$ is the Weil restriction, turning $W_{K/\mathbb{Q}}\mathbb{V}$ into a polarizable \mathbb{Q} -variation of Hodge structure so that the natural action of K is via morphisms of variations of Hodge structure. We say that a K -variation \mathbb{V} is *integral* if there exists a locally constant sheaf of locally free \mathcal{O}_K -modules \mathbb{W} on Y and an isomorphism $\mathbb{W} \otimes_{\mathcal{O}_K} K \simeq \mathbb{V}$.

We denote by $\text{VHS}(Y, K)$ the neutral Tannakian category of semisimple K -variations of mixed Hodge structure (ie direct sums of irreducible pure K -variations of Hodge structure), and by $\text{VHS}(Y, \mathcal{O}_K)$ the full (neutral Tannakian) subcategory consisting of direct sums of integral pure K -variations of Hodge structure. Note that $\text{VHS}(Y, \mathcal{O}_K)$ is a K -linear category, not an \mathcal{O}_K -linear category. We equip this category with the fiber functor sending a local system to its fiber at x .

If \mathbb{V} is a \mathbb{Q} -VHS, the *generic Mumford–Tate group* of \mathbb{V} is the identity component of the Tannakian group associated to the full subcategory of $\text{VHS}(Y, \mathbb{Q})$ generated by \mathbb{V} (see eg [Moonen 2017, penultimate paragraph of Section 4.1] for more discussion and a comparison to other definitions of the generic Mumford–Tate group; in particular, if x is very general, there is a canonical isomorphism between the generic Mumford–Tate group and the Mumford–Tate group of \mathbb{V}_x).

For any local system \mathbb{V} on a variety Y , with associated monodromy representation $\rho: \pi_1(Y, y) \rightarrow \text{GL}(\mathbb{V}_y)$, the *algebraic monodromy group* of \mathbb{V} is the identity component of the Zariski closure of the image of ρ .

5.2 Lemma *Let \mathbb{V} be an integral K -variation of Hodge structure on a smooth variety Y . Then the algebraic monodromy group of \mathbb{V} is a normal subgroup of the derived subgroup of the generic Mumford–Tate group of the Weil restriction $W_{K/\mathbb{Q}}\mathbb{V}$.*

Proof This is immediate from [André 1992, Theorem 1 on page 10]. \square

If \mathbb{V} is a K -variation of Hodge structure on Y and $\iota: K \hookrightarrow \mathbb{C}$ is an embedding, then $\mathbb{V} \otimes_{K,\iota} \mathbb{C}$ naturally obtains the structure of a \mathbb{C} -variation of Hodge structure.

5.3 Definition (infinitesimal VHS) A *weak infinitesimal variation of Hodge structure on Y at x* , or weak IVHS, is a finite-dimensional \mathbb{Z} -graded complex vector space V^\bullet equipped with an action by $T_x Y$ by commuting linear operators of degree -1 , ie with a map $\delta^i: T_x Y \rightarrow \text{Hom}(V^i, V^{i-1})$, such that $\delta^{i-1}(w_1) \circ \delta^i(w_2)(v) = \delta^{i-1}(w_2) \circ \delta^i(w_1)(v)$ for all $w_1, w_2 \in T_x Y, v \in V^i$. A morphism of weak IVHS is a graded map of vector spaces commuting with the action of $T_x Y$. We denote the category of weak IVHS on Y at x by $\text{IVHS}(Y, x)$.

5.4 Remark We call the above *weak* IVHS because the conditions we impose are weaker than the standard conditions on infinitesimal variations of Hodge structure; see for example [Carlson et al. 1983].

5.5 Proposition *The category $\text{IVHS}(Y, x)$ is a neutral Tannakian category, and the forgetful functor $\text{IVHS}(Y, x) \rightarrow \text{Vect}_{\mathbb{C}}$ sending V^\bullet to its underlying vector space is a fiber functor. The corresponding Tannakian group is $T_x Y \rtimes \mathbb{G}_m$, where \mathbb{G}_m acts on $T_x Y$ by inverse scaling.*

Proof It suffices to prove that $\text{IVHS}(Y, x)$ is equivalent to the category of representations of $T_x Y \rtimes \mathbb{G}_m$, with the equivalence respecting the tensor product and forgetful functor to vector spaces, as the category of representations of any pro-algebraic group is a neutral Tannakian category with fiber functor the forgetful functor and Tannakian group the initial pro-algebraic group [Deligne et al. 1982, II, Example 1.25].

Given a weak infinitesimal variation of Hodge structures, we obtain an action of the algebraic group $T_x Y$ on V^\bullet by exponentiating δ , ie for $v \in V^i$ and $w \in T_x Y$ we have

$$w \cdot v = v + \delta^i(w)(v) + \frac{\delta^{i-1}(w) \circ \delta^i(w)(v)}{2!} + \frac{\delta^{i-2}(w) \circ \delta^{i-1}(w) \circ \delta^i(w)(v)}{3!} + \dots$$

(The power series converges since $V^{i-k} = 0$ for all k sufficiently large.) The relation

$$\delta^{i-1}(w_1) \circ \delta^i(w_2)(v) = \delta^{i-1}(w_2) \circ \delta^i(w_1)(v)$$

implies that $w_1 \cdot (w_2 \cdot v) = (w_1 + w_2) \cdot v$ and thus this is an action of the additive group $T_x Y$ on V^\bullet .

We also obtain an action of \mathbb{G}_m on V^\bullet by, for $v \in V^i$ and $\lambda \in \mathbb{G}_m$, taking $\lambda \cdot v = \lambda^i v$. Then for $\lambda \in \mathbb{G}_m, w \in T_x Y$ and $v \in V^\bullet$ it is not hard to check that we have $\lambda \cdot (w \cdot (\lambda^{-1} \cdot v)) = (\lambda^{-1} w) \cdot v$. This relation implies that the two actions combine to give an action of $T_x Y \rtimes \mathbb{G}_m$ on V^\bullet .

Conversely, given a vector space V with an action of \mathbb{G}_m , we obtain a grading by taking V^i to be the subspace on which $\lambda \in \mathbb{G}_m$ acts with eigenvalue λ^i for all λ , and the derivative at the identity of the action of the additive group $T_x Y$ defines an action of $T_x Y$ by commuting linear operators.

These two constructions are inverse, as can be checked separately on the $T_x Y$ and \mathbb{G}_m parts, the first part being the standard equivalence between representations of a unipotent algebraic group and nilpotent

representations of its Lie algebra, and the second being the standard equivalence between graded vector spaces and representations of \mathbb{G}_m .

They are compatible with tensor product (for the natural notion of tensor product on weak infinitesimal variations of Hodge structures) and, trivially, with the forgetful functor to the underlying vector space. \square

Given a K -variation of Hodge structure \mathbb{V} on Y and an embedding $\iota: K \hookrightarrow \mathbb{C}$, we can form the graded vector space $\bigoplus_{i \in \mathbb{Z}} (F^i \mathbb{V}_x \otimes_{K, \iota} \mathbb{C}) / (F^{i+1} \mathbb{V}_x \otimes_{K, \iota} \mathbb{C})$, which admits an action of the tangent space $T_x Y$ by commuting linear maps of degree -1 (by Griffiths transversality). This formation is functorial, giving rise to a functor GH_x , where GH stands for “graded Hodge”,

$$(5-1) \quad \text{GH}_x: \text{VHS}(Y, K) \rightarrow \text{IVHS}(Y, x)$$

(which depends on ι , but we suppress this dependence from the notation).

Composing this functor with the Weil restriction functor $W_{K/\mathbb{Q}}$ induces a homomorphism from $T_x Y \rtimes \mathbb{G}_m$ to the generic Mumford–Tate group of $W_{K/\mathbb{Q}} \mathbb{V}$, since $T_x Y \rtimes \mathbb{G}_m$ is connected and hence any homomorphism from $T_x Y \rtimes \mathbb{G}_m$ to the Tannakian group lands in its identity component.

5.6 Lemma *Let K be a number field, Y a smooth complex variety, x a point of Y , and \mathbb{V} a pure integral K -VHS on Y . Let M be the monodromy group of \mathbb{V} , ie the Zariski closure of the monodromy representation $\pi_1(Y, x) \rightarrow \text{GL}(\mathbb{V}_x)$, and let \mathfrak{m} be the Lie algebra of the identity component of M .*

Suppose \mathbb{V} has at most $k+1$ nonvanishing Hodge numbers. Then either $\text{GH}_x(\mathbb{V})$ splits as a direct sum in $\text{IVHS}(Y, x)$ or \mathfrak{m} acts irreducibly on \mathbb{V} and has at most k simple factors.

In particular, if \mathbb{V} has at most 2 nonvanishing Hodge numbers, then either $\text{GH}_x(\mathbb{V})$ splits, or the identity component of M is a simple algebraic group acting irreducibly on \mathbb{V} .

Proof Since monodromy groups are preserved when a local system is extended to a larger coefficient field, we may assume K is Galois over \mathbb{Q} .

Let G be the generic Mumford–Tate group of $W_{K/\mathbb{Q}} \mathbb{V}$ and \mathfrak{g} the Lie algebra of G . The action of $T_x Y \rtimes \mathbb{G}_m$ on $\text{gr}^{F^\bullet}(W_{K/\mathbb{Q}} \mathbb{V}_x \otimes \mathbb{C})$ is given by a homomorphism $T_x Y \rtimes \mathbb{G}_m \rightarrow G_{\mathbb{C}}$.

Because the category $\text{VHS}(Y, \mathcal{O}_K)$ is semisimple, G is reductive. Thus if \mathbb{V} is reducible as a representation of G_K (that is, after passing to a finite étale cover, as a K -VHS), it splits as a direct sum of two representations of G , hence a direct sum of two representations of $T_x Y \rtimes \mathbb{G}_m$, so $\text{GH}_x(\mathbb{V})$ splits as a sum of two graded vector spaces with actions of $T_x Y$. Hence we may assume \mathbb{V} is irreducible as a representation of G_K , and thus irreducible as a representation of \mathfrak{g} .

Because G is reductive, \mathfrak{g} splits as a product of n simple Lie algebras $\mathfrak{g}_1, \dots, \mathfrak{g}_n$ times a trivial Lie algebra, and because \mathbb{V} is irreducible as a representation of \mathfrak{g} , it is a tensor product of n nontrivial irreducible representations V_1, \dots, V_n of the n simple Lie algebras with a one-dimensional representation of the trivial Lie algebra. Each V_i is a representation of the Lie algebra of $T_x Y \rtimes \mathbb{G}_m$, and thus admits a \mathbb{C} -grading and an action of $T_x Y$ by commuting linear maps of degree -1 .

If $T_x Y$ acts trivially on some V_i , then V_i splits as a direct sum since every graded vector space of dimension > 1 splits as a direct sum of graded vector spaces (and one-dimensional representations of simple Lie algebras are trivial), so \mathbb{V} splits as a direct sum of graded vector spaces with actions of $T_x Y$. So we may assume $T_x Y$ acts nontrivially on each V_i . It follows that each V_i has vectors of at least two different grades, so the tensor product \mathbb{V} of the V_i has vectors of at least $n + 1$ different grades, and thus $n \leq k$.

As M is a normal subgroup of G , the Lie algebra \mathfrak{m} of the identity component of M is an ideal of G . Therefore, \mathfrak{m} is a sum of some of the \mathfrak{g}_i , possibly with a trivial algebra. Since monodromy groups of \mathbb{Q} -VHS's are simple, \mathfrak{m} is a sum of some of the \mathfrak{g}_i . If some \mathfrak{g}_i does not appear in this sum, then its adjoint representation corresponds to a \mathbb{Q} -VHS (possibly on a cover of X) with trivial monodromy, hence a constant Hodge structure, so the derivative of its period map vanishes, and thus $T_x Y$ acts trivially on this adjoint representation. But this implies that the image of $T_x Y$ in \mathfrak{g}_i is zero, and thus $T_x Y$ acts trivially on V_i , contradicting our assumption that $T_x Y$ acts nontrivially on each V_i .

Thus \mathfrak{m} is the sum of all the \mathfrak{g}_i and thus acts irreducibly on \mathbb{V} and, in addition, has $n \leq k$ simple factors. \square

5.7 Lie algebras and weights We discuss some generalities on representations of the Lie algebra of the generic Mumford–Tate group.

For \mathbb{V} a \mathbb{Q} -VHS on a variety X and x a point of X , Proposition 5.5 gives a homomorphism $T_x Y \rtimes \mathbb{G}_m \rightarrow G$, where G is the generic Mumford–Tate group of \mathbb{V} . Thus we obtain a Lie algebra homomorphism $T_x Y \rtimes \mathbb{C} \rightarrow \mathfrak{g}$, where \mathfrak{g} is the Lie algebra of G , \mathbb{C} is the Lie algebra of \mathbb{G}_m , and \mathbb{C} acts on $T_x Y$ by $[1, v] = -v$ for $v \in T_x Y$, where the minus sign appears because \mathbb{G}_m acts on $T_x Y$ by inverse scaling, by Proposition 5.5.

For any representation of \mathfrak{g} , we refer to the eigenvalues of $\mathbb{C} \subseteq T_x Y \rtimes \mathbb{C}$ on that representation as the weights and the generalized eigenspace of a given eigenvalue as the weight space. For the representation arising from the action of G on \mathbb{V}_x , these weights agree with the Hodge weights of \mathbb{V} , and in particular are integers, but for arbitrary representations they will be complex numbers.

For any representation that factors through a finite covering of G , the action of \mathbb{C} factors through a finite covering of \mathbb{G}_m . This implies that the weights will be rational numbers, and the action of \mathbb{C} is semisimple so the generalized eigenspaces will be eigenspaces.

The identity $[1, v] = -v$ for $v \in T_x Y$ implies that the elements of $T_x Y$ send elements of weight w to elements of weight $w - 1$.

5.8 Example Let \mathbb{V} be an integral K -variation of Hodge structure on X , with Mumford–Tate (isogenous to) GL_ν acting through the $\wedge^k : \mathrm{GL}_\nu \rightarrow \mathrm{GL}(\binom{\nu}{k})$, the k^{th} wedge power of the standard representation, where $1 < k < \nu - 1$. Let $\mathrm{std} : \mathfrak{gl}_\nu \rightarrow \mathfrak{gl}_\nu$ be the standard representation. As in the discussion above, we

have a Lie algebra homomorphism $\iota: T_x Y \rtimes \mathbb{C} \rightarrow \mathfrak{g} = \mathfrak{gl}_\nu$, so that the weights of $\wedge^k \circ \iota(1)$ are the Hodge weights of \mathbb{V} .

We may in this case consider $\text{std} \circ \iota(1)$, which acts on \mathbb{C}^ν with weights in $\frac{1}{k}\mathbb{Z}$. By definition, $\wedge^k \circ \iota(1) = \wedge^k(\text{std} \circ \iota(1))$. Thus, letting a_1, \dots, a_ν be the weights of $\text{std} \circ \iota(1)$, we have that the weights of $\wedge^k \circ \iota(1)$ are precisely the sums $a_{i_1} + \dots + a_{i_k}$, where $1 \leq i_1, \dots, i_k \leq \nu$ are distinct integers.

The following case will be used in the proof Lemma 7.9. Suppose every weight of $\wedge^k \circ \iota(1)$ is either 0 or 1. Then the weights of $\text{std} \circ \iota(1)$ must be either

$$\left\{ \frac{1}{k}, \dots, \frac{1}{k}, -\frac{k-1}{k} \right\} \quad \text{or} \quad \{0, \dots, 0, 1\},$$

by an elementary combinatorial analysis.

6 Generic Torelli theorems for unitary local systems

Let $\pi: \mathcal{C} \rightarrow \mathcal{M}$ be a versal family of n -punctured curves of genus g , and let $\pi^\circ: \mathcal{C}^\circ \rightarrow \mathcal{M}$ be the associated punctured versal family, defined as in Notation 2.1. Let $m \in \mathcal{M}$ be a general point and $C = \mathcal{C}_m$, $C^\circ = \mathcal{C}_m^\circ$ and $D = C - C^\circ$. Let \mathbb{U} be a unitary local system on \mathcal{C}° . The goal of this section is to show that, in many cases, $\mathbb{U}|_{C^\circ}$ can be functorially recovered from the local system $W_1 R^1 \pi_{\star}^\circ \mathbb{U}$. In fact, it will be recoverable from the associated weak IVHS at a general point of m . We view this as a generic Torelli theorem for curves with a unitary local system, and indeed the proof is closely related to classical proofs of the generic Torelli theorem, as in [Harris 1985].

6.1 Reconstructing unitary bundles In this section we reconstruct from $W_1 R^1 \pi_{\star}^\circ \mathbb{U}$ the vector bundle \hat{E}_0 corresponding to the parabolic bundle on (C, D) associated to $\mathbb{U}|_{C^\circ}$. In fact we will reconstruct \hat{E}_0 from $\text{GH}_m(W_1 R^1 \pi_{\star}^\circ \mathbb{U})$, with GH as defined in (5-1). In the case $D = \emptyset$, this in fact recovers $\mathbb{U}|_C$ by the Narasimhan–Seshadri correspondence [Narasimhan and Seshadri 1965]. When D is nonempty, some additional work will be required to recover $\mathbb{U}|_C$.

6.2 Theorem *With notation as above, let \mathbb{U} be a unitary local system on C° , with monodromy representation ρ and associated parabolic bundle E_\star on (C, D) . Let $r = \text{rk}(\mathbb{U})$, and assume $g \geq 2 + 2r$. Let ψ_m^ρ be the map*

$$H^0(C, \hat{E}_0 \otimes \omega_C(D)) \otimes \mathcal{O}_C \xrightarrow{\theta_m^\rho \otimes \text{id}} H^1(C, E_0) \otimes H^0(C, \omega_C^{\otimes 2}(D)) \otimes \mathcal{O}_C \xrightarrow{\text{id} \otimes \text{ev}} H^1(C, E_0) \otimes \omega_C^{\otimes 2}(D)$$

obtained as the composition of the map induced by θ_m^ρ , defined in (3-5), with the map induced by the evaluation map $\text{ev}: H^0(C, \omega_C^{\otimes 2}(D)) \otimes \mathcal{O}_C \rightarrow \omega_C^{\otimes 2}(D)$. Then ψ_m^ρ factors through the evaluation map

$$\text{ev}: H^0(C, \hat{E}_0 \otimes \omega_C(D)) \otimes \mathcal{O}_C \rightarrow \hat{E}_0 \otimes \omega_C(D),$$

and induces an isomorphism

$$\hat{E}_0 \otimes \omega_C(D) \xrightarrow{\sim} \text{im}(\psi_m^\rho).$$

Proof It suffices to show that

$$(6-1) \quad \ker(\psi_m^\rho) = \ker(\text{ev}: H^0(C, \widehat{E}_0 \otimes \omega_C(D)) \otimes \mathcal{O}_C \rightarrow \widehat{E}_0 \otimes \omega_C(D)),$$

as $\widehat{E}_0 \otimes \omega_C(D)$ is globally generated by Proposition 4.9, using that m is a general point of \mathcal{M} .

We first observe that $H^1(C, E_0)$ is, by Serre duality, dual to $H^0(C, E_0^\vee \otimes \omega_C)$. As m is general, $E_0^\vee \otimes \omega_C$ is globally generated by Proposition 4.9 since, setting F_\star to be the parabolic dual to E_\star , we have $E_0^\vee \otimes \omega_C = \widehat{F}_0 \otimes \omega_C(D)$.

Let s be a local section to $H^0(C, \widehat{E}_0 \otimes \omega_C(D)) \otimes \mathcal{O}_C$. Now s is in the kernel of ψ_m^ρ if and only if, for all $t \in H^1(C, E_0)^\vee = H^0(C, E_0^\vee \otimes \omega_C)$, we have that $t(\psi_m^\rho(s)) = 0$ as a local section to $\omega_C^{\otimes 2}(D)$. Note that $t(\psi_m^\rho(s)) = B_m^\rho(s, t)$, for B_m^ρ as defined in (3-4). Now, $B_m^\rho(s, t)$ vanishes for all t if and only if $\text{ev}(s) = 0$, as $E_0^\vee \otimes \omega_C$ is (generically) globally generated and B_m^ρ induces a perfect pairing between the generic fibers of $\widehat{E}_0 \otimes \omega_C(D)$ and $E_0^\vee \otimes \omega_C$. This implies (6-1). \square

6.3 Functoriality The upshot of Theorem 6.2 is that $\widehat{E}_0 \otimes \omega_C(D)$ (and hence \widehat{E}_0 itself) may be constructed from $\text{GH}_m(W_1 R^1 \pi_\star \mathbb{U})$ for generic m . This construction is functorial: given \mathbb{U} and \mathbb{U}' unitary local systems on \mathcal{C}° , and given a map $\text{GH}_m(W_1 R^1 \pi_\star^\circ \mathbb{U}) \rightarrow \text{GH}_m(W_1 R^1 \pi_\star^\circ \mathbb{U}')$, one obtains a map $\widehat{E}_0 \rightarrow \widehat{E}'_0$, where E'_\star is the parabolic bundle on C corresponding to $\mathbb{U}'|_{C^\circ}$. The goal of this section is to show that in many cases this map is necessarily flat for the natural unitary connections on $\widehat{E}_0, \widehat{E}'_0$. This is automatic from the Narasimhan–Seshadri correspondence if $D = \emptyset$, but not in general.

6.4 Proposition *With notation as in Notation 2.1, let $m \in \mathcal{M}$ be general, set $C = \mathcal{C}_m, C^\circ = \mathcal{C}_m^\circ$ and $D = C - C^\circ$. Let \mathbb{U} and \mathbb{U}' be unitary local systems on \mathcal{C}° such that $\mathbb{U}|_{C^\circ}$ and $\mathbb{U}'|_{C^\circ}$ are irreducible, of dimensions r and r' , respectively. Assume that*

$$\text{GH}_m(W_1 R^1 \pi_\star^\circ \mathbb{U}) \quad \text{and} \quad \text{GH}_m(W_1 R^1 \pi_\star^\circ \mathbb{U}')$$

are isomorphic to one another. Suppose that $g \geq \max(2 + 2r, 2 + 2r')$ and either

- (1) $D = \emptyset$, or
- (2) $g > rr'$.

Then the natural map

$$(6-2) \quad \text{Hom}_{\text{LocSys}}(\mathbb{U}|_{C^\circ}, \mathbb{U}'|_{C^\circ}) \rightarrow \text{Hom}_{\text{IVHS}(\mathcal{M}, m)}(\text{GH}_m(W_1 R^1 \pi_\star^\circ \mathbb{U}), \text{GH}_m(W_1 R^1 \pi_\star^\circ \mathbb{U}'))$$

is a bijection. In particular, $\mathbb{U}|_{C^\circ}$ is isomorphic to $\mathbb{U}'|_{C^\circ}$.

Proof Let ρ and ρ' be the monodromy representations of \mathbb{U} and \mathbb{U}' , and let $E_\star^\rho, E_\star^{\rho'}$ be the parabolic bundles associated to these local systems. By Theorem 6.2, there exists an isomorphism $\widehat{E}_0^\rho \simeq \widehat{E}_0^{\rho'}$, and hence $r = r'$.

We now prove (6-2) is injective. Indeed, a map $\mathbb{U}|_{C^\circ} \rightarrow \mathbb{U}'|_{C^\circ}$ is determined by the induced map $\widehat{E}_0^\rho \rightarrow \widehat{E}_0^{\rho'}$, which is in turn determined by the induced map

$$H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) \rightarrow H^0(C, \widehat{E}_0^{\rho'} \otimes \omega_C(D)),$$

as the vector bundles in question are globally generated by Proposition 4.9.

To prove the surjectivity of (6-2), consider a map $\psi: \mathrm{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{U}) \rightarrow \mathrm{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{U}')$. By Theorem 6.2, it induces a map $\tilde{\psi}: \widehat{E}_0^\rho \rightarrow \widehat{E}_0^{\rho'}$. We must show this map is flat. Equivalently, we wish to show that the map

$$\nabla \tilde{\psi}: T_C(-D) \rightarrow \mathrm{Hom}(\widehat{E}_0^\rho, \widehat{E}_0^{\rho'})$$

sending a vector field X to $\nabla_X \tilde{\psi}$ is identically zero. Here ∇_X is differentiation along X , using the connection ∇ on $\mathrm{Hom}(\widehat{E}_0^\rho, \widehat{E}_0^{\rho'})$. When $D = \emptyset$ this is immediate by the functoriality of the Narasimhan–Seshadri correspondence, so we need only consider the case $D \neq \emptyset$.

By Lemma 4.8, we know that $\nabla \tilde{\psi}$ induces the zero map

$$H^1(C, T_C(-D)) \rightarrow H^1(\mathrm{Hom}(\widehat{E}_0^\rho, \widehat{E}_0^{\rho'})),$$

as m is general and hence ψ extends to a first-order neighborhood of m . Serre-dually, the map

$$\mathrm{Hom}(\widehat{E}_0^{\rho'}, \widehat{E}_0^\rho) \otimes \omega_C \rightarrow \omega_C^{\otimes 2}(D),$$

obtained by dualizing $\nabla \tilde{\psi}$ and tensoring with ω_C , induces zero on H^0 . Thus if $\nabla \tilde{\psi}$ is nonzero, the bundle $\mathrm{Hom}(\widehat{E}_0^{\rho'}, \widehat{E}_0^\rho) \otimes \omega_C$ is not generically globally generated.

Now as $g > rr' = r^2$, $E_0^{(\rho')^\vee}$ and $E_0^{\rho^\vee}$ are semistable, by [Landesman and Litt 2024b, Corollary 6.1.2]; hence by Lemma 4.5, the same is true for $\widehat{E}_0^{\rho'}$ and \widehat{E}_0^ρ . As these bundles are isomorphic by the first paragraph above, we have that $\mathrm{Hom}(\widehat{E}_0^{\rho'}, \widehat{E}_0^\rho) \otimes \omega_C$ is semistable of slope $2g - 2$ and rank $rr' = r^2$. As $g > rr'$ it is thus generically globally generated by [loc. cit., Proposition 6.3.1(b)], whence the proof is complete. \square

The following will not be used in what follows, but we felt it might be of independent interest.

6.5 Construction With notation as in Notation 2.1, let \mathbb{U} be a unitary local system on \mathcal{C}° of rank r . Let $g > r^2$. Let $W_1 R^1 \pi_*^\circ \mathbb{U}$ be the above defined \mathbb{C} -VHS and let $m \in \mathcal{M}$ be general, with $C^\circ = \mathcal{C}_m^\circ$ and $C = \mathcal{C}_m$. We next sketch how to directly reconstruct the connection on \widehat{E}_0 , where E_\star is the parabolic bundle associated to $\mathbb{U}|_{C^\circ}$.

We may recover the connection from the restriction of $W_1 R^1 \pi_*^\circ \mathbb{U}$ to a small neighborhood of m , though we do not know how to do so from $\mathrm{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{U})$.

Recall that the data of a logarithmic connection on a bundle F is the same as an \mathcal{O}_C -linear splitting s of the natural map

$$\mathrm{At}_{(C,D)}(F) \rightarrow T_C(-D)$$

(see eg [loc. cit., Proposition 3.1.6]). The induced map

$$H^1(C, T_C(-D)) \rightarrow H^1(C, \text{At}_{(C,D)}(F))$$

may be interpreted as the map sending a first-order deformation (\tilde{C}, \tilde{D}) of (C, D) to the triple $(\tilde{C}, \tilde{D}, \tilde{E})$, where \tilde{E} is the isomonodromic deformation of E ; see [Landesman and Litt 2024b, Proposition 3.5.7]. This latter map may be recovered (taking $F = \hat{E}_0$) from $W_1 R^1 \pi_*^\circ \mathbb{U}$ if $g \geq r^2$, by performing the construction of Theorem 6.2 in families. Serre-dually, we may recover the equivalent data of the Serre-dual map

$$H^0(C, (\text{At}_{(C,D)}(\hat{E}_0))^\vee \otimes \omega_C) \rightarrow H^0(C, \omega_C^{\otimes 2}(D)).$$

Now $\omega_C^{\otimes 2}(D)$ is globally generated, and $(\text{At}_{(C,D)}(\hat{E}_0))^\vee \otimes \omega_C$ is generically globally generated by similar arguments to those in the proof of Proposition 6.4.

Since a map of generically globally generated vector bundles is determined by the induced map on global sections, the argument is complete.

6.6 Simplicity of the monodromy representation We now prove that the monodromy representation is simple.

6.7 Theorem *With notation as in Notation 2.1, let $\rho: H \rightarrow \text{GL}_r(\mathbb{C})$ be an irreducible H -representation, and let \mathbb{V} be the corresponding local system on \mathcal{C}° . Suppose that either $n = 0$ and $g > 2r + 1$, or n is arbitrary and $g > \max(2r + 1, r^2)$. Then the variation of Hodge structure $W_1 R^1 \pi_*^\circ \mathbb{V}$ has simple monodromy group, acting irreducibly.*

Proof Note first that \mathbb{V} is an integral variation of Hodge structure because every representation of a finite group is defined over the ring of integers over some number field and is polarizable because representations of finite groups are unitary. Thus $W_1 R^1 \pi_*^\circ \mathbb{V}$ is an integral variation of Hodge structures because these are stable under derived pushforward and passing to subspaces in the weight filtration. (See eg [Landesman and Litt 2024a, Theorem 4.1.1] for a discussion without the integrality condition.)

Now suppose that the conclusion of the theorem is false, ie that either the monodromy group of $W_1 R^1 \pi_*^\circ \mathbb{V}$ is not simple, or it does not act irreducibly. Then by Lemma 5.6, $\text{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{V})$ splits as a direct sum for general $m \in \mathcal{M}$. Thus we have

$$\dim \text{Hom}_{\text{IVHS}(\mathcal{M}, m)}(\text{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{V}), \text{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{V})) \geq 2.$$

But taking $\mathbb{U} = \mathbb{U}' = \mathbb{V}$ in Proposition 6.4, we have

$$\dim \text{Hom}_H(\rho, \rho) \geq 2.$$

But this contradicts the irreducibility of ρ , by Schur's lemma. □

6.8 Remark We could have argued using Theorem 6.2, instead of Proposition 6.4, as follows. The splitting of $\mathrm{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{V})$ implies, by the construction of Theorem 6.2, that \widehat{E}_0^ρ itself splits as a direct sum. But forthcoming work of Ramirez-Cote, following ideas of Landesman and Litt [2024b], shows that \widehat{E}_0^ρ is stable for irreducible ρ when $g \geq r^2$, contradicting this splitting.

6.9 Corollary *With notation as in Notation 2.1, let $\rho_i: H \rightarrow \mathrm{GL}_{r_i}(\mathbb{C})$ for $i = 1, 2$ be irreducible representations, and let \mathbb{V}_1 and \mathbb{V}_2 be the corresponding local systems on \mathcal{C}° . Suppose that $g \geq \max(2 + 2r_1, 2 + 2r_2)$ and either $D = \emptyset$ or $g > r_1 r_2$. If the local systems $W_1 R^1 \pi_*^\circ \mathbb{V}_1$ and $W_1 R^1 \pi_*^\circ \mathbb{V}_2$ are isomorphic, then ρ_1 and ρ_2 are conjugate.*

Proof By Theorem 6.7, the local systems $W_1 R^1 \pi_*^\circ \mathbb{V}_1, W_1 R^1 \pi_*^\circ \mathbb{V}_2$ are irreducible. Hence, by the theorem of the fixed part applied to the tensor product $(W_1 R^1 \pi_*^\circ \mathbb{V}_1)^\vee \otimes W_1 R^1 \pi_*^\circ \mathbb{V}_2$, any isomorphism between them is necessarily an isomorphism of \mathbb{C} -VHS (up to a shift of the weight and Hodge filtrations, but these shifts must vanish by consideration of the weight and Hodge numbers of both sides), and in particular induces an isomorphism $\mathrm{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{V}_1) \simeq \mathrm{GH}_m(W_1 R^1 \pi_*^\circ \mathbb{V}_2)$ for any $m \in \mathbb{M}$. Now the result follows by Proposition 6.4. \square

7 Proofs of Theorems 1.3 and 1.9: big monodromy for g large

In this section, we prove our main results, Theorems 1.3 and 1.9. We next describe the possibilities for the connected monodromy groups of the variations of Hodge structure appearing in Theorem 6.7 by applying a result of Deligne (implicit in his classification of Shimura varieties of Abelian type [Deligne 1979, 1.3.6], and explicit in [Zarkhin 1984, Theorem 0.5.1(b)]), and proceed to rule out many of these possibilities through several lemmas. We conclude the proofs in Section 7.15.

7.1 Describing the possibilities for the monodromy representation With notation as in Theorem 6.7, the Hodge filtration of $W_1 R^1 \pi_*^\circ \mathbb{V}$ has only two parts, and the monodromy group of $W_1 R^1 \pi_*^\circ \mathbb{V}$ is simple by Theorem 6.7. As it is a normal subgroup of the generic Mumford–Tate group of $W_1 R^1 \pi_*^\circ \mathbb{V}$, by [André 1992, Theorem 1 on page 10], it follows that it must be a simple factor of this group. It follows from [Zarkhin 1984, Theorem 0.5.1(b)] (which Zarkhin attributes to Deligne) that the monodromy group is isogenous to $\mathrm{SL}_\nu, \mathrm{Sp}_{2\nu}, \mathrm{SO}_\nu$ for some ν , acting in one of the following ways:

- (1) The standard representation.
- (2) The trivial representation.
- (3) The spin or half-spin representation of the spin cover of the group SO_ν .
- (4) A wedge power of the standard representation of the group SL_ν .

We first rule out the trivial representation.

7.2 Lemma With notation as in Notation 2.1, let $\rho: H \rightarrow \mathrm{GL}_r(\mathbb{C})$ be an irreducible H -representation, and let \mathbb{V} be the corresponding local system on \mathcal{C}° . If $g > 2$, the connected monodromy group of $W_1 R^1 \pi_* \mathbb{V}$ is nontrivial.

Proof We are free to pass to finite étale covers of the base \mathcal{M} in our setup. Therefore, we may assume the Zariski closure of monodromy is already connected. We also know the monodromy representation is irreducible by Theorem 6.7. Further,

$$\dim W_1 R^1 \pi_* \mathbb{V}_m \geq (2g - 2) \mathrm{rk}(\mathbb{V}) > 1,$$

since we are assuming $g \geq 2$, so we conclude that the monodromy group is not acting via the trivial representation. \square

7.3 Rank estimates

7.4 Lemma With notation as in Notation 2.1, let \mathbb{U} be a unitary local system on \mathcal{C}° and $m \in \mathcal{M}$ a general point. Let ρ be the monodromy representation of $\mathbb{U}|_{\mathcal{C}^\circ}$ and let E_\star be the associated parabolic bundle on (C, D) . Let $r = \mathrm{rk}(\mathbb{U})$, and assume $g \geq 2 + 2r$. For p in C , consider the one-dimensional space of $H^0(C, \omega_C^{\otimes 2}(D))^\vee = T_m \mathcal{M}$ spanned by the functional sending a global section to its value at p , and then composing with a linear map $\omega_C^{\otimes 2}(D)|_p \rightarrow \mathbb{C}$.

For p a general point of C and α_p an element of the associated one-dimensional subspace of $T_m \mathcal{M}$, the rank of $dP_m^\rho(\alpha_p) \in \mathrm{Hom}(H^0(C, \hat{E}_0 \otimes \omega_C(D)), H^1(C, E_0))$ is equal to r .

Here dP_m^ρ is defined as in Section 3, immediately above Proposition 3.6.

Proof Theorem 6.2 shows that $\hat{E}_0 \otimes \omega_C(D)$, a sheaf of rank r , is the image of the map ψ_m^ρ , defined as in the statement of that theorem. For any map $f: V \rightarrow W$ of vector bundles, the rank of the image of f is equal to the rank of the fiber $f_q: V_q \rightarrow W_q$ of f at a general point q . So it suffices to check that the rank of the fiber of ψ_m^ρ at a general point p in C is equal to the rank of $dP_m(\alpha_p)$.

The fiber of the evaluation map $\mathrm{ev}: H^0(C, \omega_C^{\otimes 2}(D)) \otimes \mathcal{O}_C \rightarrow \omega_C^{\otimes 2}(D)$ at p is, up to scalars, the linear form $\alpha_p \in (H^0(C, \omega_C^{\otimes 2}(D)) \otimes \mathcal{O}_C)^\vee$.

Since ψ_m^ρ is the composition of θ_m^ρ with $\mathrm{id} \otimes \mathrm{ev}$, the fiber of ψ_m^ρ is the composition of θ_m^ρ with $\mathrm{id} \otimes \alpha_p$. Previously, in (3-5), we defined θ_m^ρ as the adjoint of

$$dP_m^\rho: H^0(C, \omega_C^{\otimes 2}(D))^\vee \rightarrow \mathrm{Hom}(H^0(C, \hat{E}_0 \otimes \omega_C(D)), H^1(C, E_0)),$$

namely by

$$\theta_m^\rho: H^0(C, \hat{E}_0^\rho \otimes \omega_C(D)) \rightarrow H^1(C, E_0) \otimes H^0(C, \omega_C^{\otimes 2}(D)).$$

The defining property of this adjoint is that composing with $\mathrm{id} \otimes \alpha$ for a linear form $\alpha \in H^0(C, \omega_C^{\otimes 2}(D))^\vee$ gives a map $H^0(C, \hat{E}_0^\rho \otimes \omega_C(D)) \rightarrow H^1(C, E_0)$ defined by $dP_m^\rho(\alpha)$, so the composition of θ_m^ρ with $\mathrm{id} \otimes \alpha_p$ is $dP_m^\rho(\alpha_p)$, as desired. \square

7.5 Remark The subspaces of $H^0(C, \omega_C^{\otimes 2}(D))^\vee = H^1(C, T_C(-D))$ spanned by the α_p of Lemma 7.4 are known as *Schiffer variations*, see eg [Collino and Pirola 1995, 1.2.4]. There is some history of using (variants of) Schiffer variations for Torelli-style results, eg in [Voisin 2022].

7.6 Corollary *With notation as above, let \mathbb{U} be a unitary local system on C° , with monodromy representation ρ and associated parabolic bundle E_\star on (C, D) . Let $r = \text{rk}(\mathbb{U})$, and assume $g \geq 2 + 2r$. We have*

$$r \geq \inf\{\text{rank}(dP_m^\rho(\alpha)) \mid \alpha \in T_m\mathcal{M}, dP_m^\rho(\alpha) \neq 0\}.$$

Proof This follows from Lemma 7.4 since any member of a set is at least its minimal value, and r is a member of this set because $r = \text{rank}(dP_m^\rho(\alpha_p))$ with $dP_m^\rho(\alpha_p) \neq 0$, as $r > 0$. □

7.7 Lemma *For ρ a unitary representation of $\pi_1(C - D)$ of rank r , we have*

$$\dim H^1(C, E_0^\rho) \geq (g - 1)r \quad \text{and} \quad \dim H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) \geq (g - 1)r.$$

Proof Since E_\star^ρ has parabolic degree 0, the degree of E_0^ρ is ≤ 0 , and so $H^1(C, E_0^\rho) \geq (g - 1)r$ by Riemann–Roch. By Serre duality, [Landesman and Litt 2024b, Proposition 2.6.6],

$$\dim H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) = \dim H^1(C, \widehat{E}_0^{\rho^\vee}).$$

The latter is $\geq (g - 1)r$ by the first part. □

Combining Corollary 7.6 and Lemma 7.7, we obtain an inequality expressed only in terms of the weak IVHS of $W_1 R^1 \pi_\star^\circ \mathbb{V}$, ie

$$\begin{aligned} \dim H^1(C, E_0^\rho) &\geq (g - 1) \inf\{\text{rank}(dP_m^\rho(\alpha)) \mid \alpha \in T_m\mathcal{M}, dP_m^\rho(\alpha) \neq 0\}, \\ \dim H^0(C, \widehat{E}_0^\rho \otimes \omega_C(D)) &\geq (g - 1) \inf\{\text{rank}(dP_m^\rho(\alpha)) \mid \alpha \in T_m\mathcal{M}, dP_m^\rho(\alpha) \neq 0\}. \end{aligned}$$

7.8 Ruling out nonstandard minuscule representations In this section we show that if the monodromy group of $W_1 R^1 \pi_\star^\circ \mathbb{V}$ is SL_ν or SO_ν , it must act via the standard representation. We will use the notation discussed in Section 5.7.

7.9 Lemma *With notation as in Notation 2.1, let $\rho: H \rightarrow \text{GL}_r(\mathbb{C})$ be an irreducible H -representation, and let \mathbb{V} be the corresponding local system on \mathcal{C}° . Suppose $g \geq 2r + 2$. If the connected monodromy group of $W_1 R^1 \pi_\star^\circ \mathbb{V}$ is isogenous to SL_ν acting via \wedge^k of the standard representation, then either $k = 1$ or $k = \nu - 1$.*

Proof The identity component of the normalizer of the monodromy group SL_ν is $\text{GL}_\nu / \mu_{\text{gcd}(k, \nu)}$, with $\mu_{\text{gcd}(k, \nu)}$ embedded as scalar matrices, so its Lie algebra is \mathfrak{gl}_ν ; in particular, we are provided with a natural Lie algebra homomorphism $T_x Y \rtimes \mathbb{C} \rightarrow \mathfrak{gl}_\nu$, as in Section 5.7. Let (a_1, \dots, a_ν) be the weights of the standard representation of \mathfrak{gl}_ν .

We first determine the possible values for the a_i ; see also Example 5.8. Since the k^{th} wedge power of the standard representation has only weights 0 and 1, as the Hodge structure $W_1 R^1 \pi_*^\circ \mathbb{V}$ has only two parts, we must have that $\sum_{i \in I} a_i \in \{0, 1\}$ for any subset $I \subset \{1, \dots, v\}$ of size k . By Lemma 7.2, $k \neq 0, v$. Thus Lemma 7.10 below shows that the only possibilities for (a_1, \dots, a_v) , up to reordering, are

$$a_1 = 1, \quad a_2 = \dots = a_v = 0 \quad \text{or} \quad a_1 = \frac{1-k}{k}, \quad a_2 = \dots = a_v = \frac{1}{k}.$$

We now consider how elements of $T_m \mathcal{M}$ act on the standard representation of \mathfrak{gl}_v . In the first case, any element of $T_m \mathcal{M}$ must send the weight 1 space to the weight 0 space and the weight 0 space to zero. Since the weight 1 space is one-dimensional, any element of $T_m \mathcal{M}$ must be zero or nilpotent of rank 1. In the second case, the element sends the weight $1/k$ space to the weight $(1-k)/k$ space, and the weight $(1-k)/k$ space to zero, and again must be zero or nilpotent of rank 1.

The action of a nilpotent element of rank one in \mathfrak{gl}_v on the representation \wedge^k has rank $\binom{v-2}{k-1}$ by Lemma 7.11 below. So the minimum nonzero rank of an element of $T_m \mathcal{M}$ acting on \mathbb{V} is $\binom{v-2}{k-1}$. It follows from Corollary 7.6 that $r \geq \binom{v-2}{k-1}$.

On the other hand, $\dim H^1(C, E_0^\rho)$ and $\dim H^0(C, \widehat{E}_0^\rho) \otimes \omega_C(D)$ are either respectively equal to $\binom{v-1}{k-1}$ and $\binom{v-1}{k}$, or respectively equal to $\binom{v-1}{k}$ and $\binom{v-1}{k-1}$. By Lemma 7.7, these are both at least $(g-1)r$, which gives

$$\binom{v-1}{k-1} \geq (g-1)r \geq (g-1)\binom{v-2}{k-1} \quad \text{and} \quad \binom{v-1}{k} \geq (g-1)r \geq (g-1)\binom{v-2}{k-1}.$$

Dividing both sides by $\binom{v-2}{k-1}$ we obtain

$$\frac{v-1}{v-k} \geq (g-1) \quad \text{and} \quad \frac{v-1}{k} \geq (g-1),$$

so

$$1 = \frac{k}{v} + \frac{v-k}{v} < \frac{k}{v-1} + \frac{v-k}{v-1} \leq \frac{1}{g-1} + \frac{1}{g-1},$$

which implies $g-1 < 2$, contradicting the assumption that $g \geq 2r+2 \geq 4$. □

7.10 Lemma *Let v and k be natural numbers with $1 < k < v$. Let a_1, \dots, a_v be a tuple of complex numbers such that each sum of exactly k of the a_1, \dots, a_v is either 0 or 1, with both possibilities occurring. Then, up to reordering, we have either*

$$a_1 = 1, \quad a_2 = \dots = a_v = 0 \quad \text{or} \quad a_1 = \frac{1-k}{k}, \quad a_2 = \dots = a_v = \frac{1}{k}.$$

Proof We first observe that there can be at most two distinct values appearing in $\{a_1, \dots, a_v\}$, as otherwise one could find size k subsets summing to three different values. Further, if a_1 is distinct from a_2 , then, after possibly switching a_1 and a_2 , all other values of a_i must agree with a_2 , using that $1 < k < v-1$, or else we could again obtain three distinct sums from subsets of size k . Hence, we must have $a_2 = \dots = a_v$, and then a subset of size k drawn from $\{a_2, \dots, a_v\}$ must sum to 0 or 1, yielding the two possibilities claimed above. □

7.11 Lemma Let $\nu \geq 2$ and k be natural numbers and Let $N \in \mathfrak{gl}_\nu$ be a nilpotent element of rank 1. Then the action of N on the representation \wedge^k of \mathfrak{gl}_ν is by a matrix of rank $\binom{\nu-2}{k-1}$.

Proof We may choose a basis e_1, \dots, e_ν where the element sends e_1 to e_2 and each other basis vector to 0, and then the image is generated by wedges of e_2 with $k-1$ of the remaining $\nu-2$ basis vectors e_3, \dots, e_ν . \square

7.12 Lemma With notation as in Notation 2.1, let $\rho: H \rightarrow \mathrm{GL}_r(\mathbb{C})$ be an irreducible H -representation, and let \mathbb{V} be the corresponding local system on \mathcal{C}° . Suppose $g \geq 2r+2$. The connected monodromy group of $W_1 R^1 \pi_* \mathbb{V}$ cannot be isogenous to Spin_ν acting via a spin or half-spin representation unless the monodromy group is also a classical group acting via its standard representation.

Note that (as discussed in the first paragraph of the proof of Lemma 7.12) the group Spin_ν acting via the spin representation is isogenous to a classical group via an isogeny sending the spin representation to the standard representation, if and only if $\nu = 3, 4, 5, 6, 8$.

Proof We first observe that for $\nu = 3, 4, 5, 6$, the spin group is respectively $\mathrm{SL}_2, \mathrm{SL}_2 \times \mathrm{SL}_2, \mathrm{Sp}_4, \mathrm{SL}_4$, and the spin representation in each case is the standard representation (or, in $\nu = 2$, the standard representation of one of the two factors). For $\nu = 8$, the spin group is not a classical group but each half-spin representation factors through the standard representation of a different quotient of Spin_8 isomorphic to SO_8 . (The existence of these three different SO_8 quotients is known as triality). So we may assume that $\nu > 8$ or $\nu = 7$.

Now, the identity component of the normalizer of the monodromy group is GSpin_ν , with Lie algebra \mathfrak{go}_ν . The weights of the standard representation of \mathfrak{go}_ν come in opposite pairs $b+a_1, b-a_1, \dots, b+a_k, b-a_k$ if $\nu = 2k$ is even, while if $\nu = 2k+1$ is odd, the weights have the form $b+a_1, b-a_1, \dots, b+a_k, b-a_k, b$. In either case we may assume all the values a_i are nonnegative.

First, we recall properties of the spin representation, as defined by Chevalley [1954, pages 119–122]. In the case that $\nu = 2k$ is even, there are two irreducible half-spin representation of dimension 2^{k-1} , and if $\nu = 2k+1$ is odd, then there is an irreducible spin representation of dimension 2^k .

The weights of the spin representation are then obtained as the 2^k sums $\frac{1}{2}(\pm a_1 \pm a_2 \pm \dots \pm a_k) + \frac{1}{2}kb$. One half-spin representation has weights which are sums as above with an odd number of minus signs, and the other consists of weights which are sums as above with an even number of minus signs. Applying Lemma 7.13, in case (1) if $\nu = 2k+1$ is odd (where we have $\nu \geq 7$ so that $k \geq 3$) and in case (2) if $\nu = 2k$ is even (where we have $\nu > 8$ so $k > 4$), we can assume $a_1 = 1$ and $a_2, \dots, a_k = 0$.

It follows that the action of a generator of the Lie algebra \mathbb{C} on the standard representation of \mathfrak{so}_ν has a one-dimensional $(1+b)$ -eigenspace, a one-dimensional $(b-1)$ -eigenspace, and a $(\nu-2)$ -dimensional b -eigenspace. Applying Lemma 7.14, where w is obtained by subtracting the scalar b from the generator 1

of the Lie algebra \mathbb{C} and the commutator relation comes from Section 5.7, we see that elements of $T_m\mathcal{M}$ acting on \mathbb{V} have rank either 0, 2^{k-3} or 2^{k-2} . In particular, the minimum nonzero rank is $\geq 2^{k-3}$, and hence $r \geq 2^{k-1}$. But the weight-0 and weight-1 spaces of the spin representation both have dimension 2^{k-2} , so that $\dim H^1(C, E_0^\rho)$ and $\dim H^0(C, \hat{E}_0^\rho \otimes \omega_C(D))$ are each equal to 2^{k-2} and hence $\leq 2r$. Since $g \geq 2r + 2 > 3$, this contradicts Lemma 7.7. \square

7.13 Lemma *Let k be a natural number. Let a_1, \dots, a_k and b be rational numbers. Suppose that either:*

- (1) $k \geq 1$ and all sums of the form $\frac{1}{2}(\pm a_1 \pm a_2 \pm \dots \pm a_k) + \frac{1}{2}kb$ are equal to 0 or 1, with both values attained.
- (2) $k > 4$ and either
 - (a) all sums of the form $\frac{1}{2}(\pm a_1 \pm a_2 \pm \dots \pm a_k) + \frac{1}{2}kb$ with an even number of minus signs are equal to 0 or 1, with both values attained, or
 - (b) all sums of the form $\frac{1}{2}(\pm a_1 \pm a_2 \pm \dots \pm a_k) + \frac{1}{2}kb$ with an odd minus sign are equal to 0 or 1, with both values attained.

Then, up to reordering and swapping signs, we have $a_1 = 1, a_2, \dots, a_k = 0$, and $b = 1/k$.

Proof By swapping signs, we may assume all the a_i are nonnegative.

In case (1), if two of the a_i s are nonzero, then without loss of generality we may assume a_1 and a_2 are both nonzero, and we obtain three distinct sums

$$\frac{1}{2}\left(-a_1 - a_2 + \sum_{i=3}^k a_k + kb\right), \quad \frac{1}{2}\left(-a_1 + a_2 + \sum_{i=3}^k a_k + kb\right), \quad \frac{1}{2}\left(a_1 + a_2 + \sum_{i=3}^k a_k + kb\right).$$

So all but one of the a_i are 0, and up to reordering only a_1 may be nonzero, so the only possible weights are $\frac{1}{2}a_1 + \frac{1}{2}kb$ and $-\frac{1}{2}a_1 + \frac{1}{2}kb$. Since the difference between these must be 1, we must have $a_1 = 1$, which implies $b = 1/k$.

In case (2), if two of the a_i are nonzero, without loss of generality a_1 and a_2 , the sums with an even number of minus signs will include the three distinct values

$$\begin{aligned} &\frac{1}{2}\left(a_1 + a_2 + a_3 + a_4 + \sum_{i=5}^k a_k + kb\right), \\ &\frac{1}{2}\left(-a_1 + a_2 - a_3 + a_4 + \sum_{i=5}^k a_k + kb\right), \\ &\frac{1}{2}\left(-a_1 - a_2 - a_3 - a_4 + \sum_{i=5}^k a_k + kb\right). \end{aligned}$$

Similarly, the sums with an odd number of minus signs will include the three distinct values

$$\begin{aligned} & \frac{1}{2} \left(a_1 + a_2 + a_3 + a_4 - a_5 + \sum_{i=6}^k a_i + kb \right), \\ & \frac{1}{2} \left(-a_1 + a_2 - a_3 + a_4 - a_5 + \sum_{i=6}^k a_i + kb \right), \\ & \frac{1}{2} \left(-a_1 - a_2 - a_3 - a_4 - a_5 + \sum_{i=6}^k a_i + kb \right). \end{aligned}$$

So only one of the a_i may be nonzero. We conclude the argument the same way as in case (1). \square

7.14 Lemma *Let $\nu \geq 7$ be a natural number. Let $w \in \mathfrak{so}_\nu$ be an element with a one-dimensional 1-eigenspace, a one-dimensional (-1) -eigenspace, and a $(\nu-2)$ -dimensional 0-eigenspace. Let $N \in \mathfrak{so}_\nu$ be an element such that $wN - Nw = -N$. Then if $\nu = 2k - 1$ is odd, the action of N on the 2^{k-1} -dimensional spin representation of \mathfrak{so}_ν has rank either 2^{k-2} , 2^{k-3} or 0. If $\nu = 2k$ is even, the action of N on either of the 2^{k-1} -dimensional half-spin representations of \mathfrak{so}_ν has rank either 2^{k-2} , 2^{k-3} or 0.*

Proof The action of N sends the one-dimensional weight 1-eigenspace to the $(\nu-2)$ -dimensional 0-eigenspace, and sends the 0-eigenspace to the (-1) -eigenspace. Thus N is determined by its action on a generator of the 1-eigenspace space as the condition that it is equal to minus its transpose will then determine its action on the 0-eigenspace. Fix a two-dimensional subspace W of the 0-eigenspace where the quadratic form is nondegenerate. Then W contains nontrivial elements where the quadratic form attains any fixed value, including zero, and thus W intersects each orbit of $\mathrm{SO}_{\nu-2}$ on the 0-eigenspace.

Hence any element of the 0-eigenspace is $\mathrm{SO}_{\nu-2}$ -conjugate to an element of W , and thus N is conjugate to an element that sends the generator of the 1-eigenspace to an element of W . It follows from $N = -N^T$ that N sends every element of the 0-eigenspace perpendicular to W to the zero element of the (-1) -eigenspace, ie N sends the orthogonal complement of W in the 0-eigenspace to 0. Elements of \mathfrak{so}_ν sending the orthogonal complement of W in the 0-eigenspace to 0 form a Lie algebra, isomorphic to \mathfrak{so}_4 , so we conclude that N is conjugate to a nilpotent element of \mathfrak{so}_4 .

The restriction to \mathfrak{so}_4 of the spin representation of \mathfrak{so}_ν for $\nu = 2k - 1$ is odd is isomorphic to the sum of 2^{k-3} copies of each half-spin representation of \mathfrak{so}_4 , and the same is true for the restriction to \mathfrak{so}_4 of the half-spin representation of \mathfrak{so}_ν for $\nu = 2k$. This may be checked by comparing weights, since a representation of a simple Lie group is uniquely determined by its weight multiplicities. Since \mathfrak{so}_4 is isomorphic to $\mathfrak{sl}_2 \times \mathfrak{sl}_2$, with each spin representation the standard representation of one factor, every nilpotent element is a pair of two nilpotent elements in \mathfrak{sl}_2 , each of which may be zero. The rank of the element is 0 if both elements of the pair vanish, 2^{k-3} if one element is nonvanishing, or 2^{k-2} if both are nonvanishing. \square

7.15 Proof of main theorems We next give the proof of Theorem 1.9. This theorem says that, under suitable hypotheses on n, g and r , the monodromy map $R_{\varphi, \rho}: \text{Mod}_{\varphi} \rightarrow \text{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ has image with Zariski closure $\text{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ when ρ is symplectically self-dual, $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ when ρ is orthogonally self-dual, and the product of $\text{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ with a finite central subgroup when ρ is not self-dual.

Proof of Theorem 1.9 As in Notation 2.1, let $\pi: \mathcal{C} \rightarrow \mathcal{M}$ be a versal family of n -pointed curves of genus g , with associated punctured versal family $\pi^{\circ}: \mathcal{C}^{\circ} \rightarrow \mathcal{M}$, and suppose $f: \mathcal{X} \rightarrow \mathcal{C}$ gives a versal family of H -covers. Let $\rho: H \rightarrow \text{GL}_r(\mathbb{C})$ be an irreducible representation of H , let \mathbb{V}^{ρ} denote the associated local system on $\Sigma_{g,n}$ and let \mathbb{U}^{ρ} be the associated local system on \mathcal{C}° ; we wish to analyze the connected monodromy group of $R^1 \pi_*^{\circ} \mathbb{U}^{\rho}$. By the discussion of Section 2.3, this suffices.

By Theorem 6.7, the connected monodromy group associated to $W_1 R^1 \pi_*^{\circ} \mathbb{U}^{\rho}$ is a simple group, acting irreducibly. As described in Section 7.1, there are several possibilities for the monodromy representation. Recall we are assuming $g \geq 2r + 2$, so the bounds on g in Lemmas 7.2, 7.9 and 7.12 are satisfied. By removing the possibilities excluded via Lemmas 7.2, 7.9 and 7.12, we see that the monodromy representation must act via the standard representation of SL_v, SO_v or Sp_v , ie must be $\text{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$, $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ or $\text{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$.

We next show that the identity component of the monodromy group is

- (1) $\text{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ if ρ is symplectically self-dual,
- (2) $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ if ρ is orthogonally self-dual, and
- (3) $\text{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ if ρ is not self-dual.

We first handle the case that ρ is self-dual.

If ρ is symplectically self-dual, the antisymmetric pairing on ρ induces a symmetric pairing on the weight-one piece $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho})$, which implies that the connected monodromy group must be contained in $\text{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$. This implies that the monodromy cannot be $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ or $\text{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ so must be $\text{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$. If ρ is orthogonally self-dual, the monodromy must be contained in $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$, and we similarly obtain that it must be equal to $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$.

To conclude the calculation of the identity component, it remains to show that the identity component of the Zariski closure of the image of monodromy is $\text{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ when ρ is not self dual. As explained above, there are only three possibilities, and hence it remains to show that the monodromy cannot be $\text{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$ or $\text{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho}))$. If the monodromy has this form, then there is an isomorphism of local systems between $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho})$ and $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho})^{\vee}$, which by Poincaré duality is $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho^{\vee}})$. It follows from Corollary 6.9 that ρ and ρ^{\vee} are conjugate, contradicting the assumption that ρ is not self-dual.

Having described the connected monodromy group, we now describe the monodromy group itself. If ρ is orthogonally self-dual, Poincaré duality gives a symplectic form on $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$, so the monodromy group is contained in $\mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ and contains $\mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$, and thus must equal $\mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$.

Similarly, if ρ is symplectically self-dual, Poincaré duality gives a symmetric form on $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$, so the monodromy group is contained in $O(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ and contains $\mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$, and thus must equal either $\mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ or $O(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$. However, we now check that the monodromy group is not $O(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$: The representation ρ , being symplectic and unitary, necessarily has the structure as a representation over the quaternions \mathbb{H} , ie has an \mathbb{R} -linear action of the quaternions compatible with the action of H . Hence the complex local system $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ has an \mathbb{R} -linear action of the quaternions compatible with the mapping class group action. Thus, for each element σ in the mapping class group, the eigenspace with eigenvalue 1 of σ , acting on $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$, has an action of the quaternions and hence is an even-dimensional complex vector space. However, any elements of $O_N - \mathrm{SO}_N$ for even N have an odd-dimensional 1-eigenspace, so the image of the mapping class group is contained in $\mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ and thus the monodromy group must be $\mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$.

Finally, if ρ is not self-dual, the monodromy group is a subgroup of $\mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ whose identity component is $\mathrm{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ and hence must be the product of $\mathrm{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ with a finite subgroup of the center of $\mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$. □

For the proof of Theorem 1.3, we will also need the following explicit description of the symplectic centralizer.

7.16 Lemma *Let $\rho_1, \dots, \rho_\alpha$ denote the orthogonally self-dual complex irreducible representations of H , let $\rho_{\alpha+1}, \dots, \rho_{\alpha+\beta}$ denote the symplectically self-dual irreducible complex representations of H , and let $(\rho_{\alpha+\beta+1}, \rho_{\alpha+\beta+\gamma+1}), \dots, (\rho_{\alpha+\beta+\gamma}, \rho_{\alpha+\beta+2\gamma})$ denote the dual pairs of complex irreducible representations of H . We use \mathbb{V}^ρ to denote the local system on $\Sigma_{g,n}$ corresponding to ρ . Then the isomorphism*

$$H^1(\Sigma_{g'}, \mathbb{C}) \xrightarrow{\sim} \prod_{i=1}^{\alpha+\beta+2\gamma} \rho_i^\vee \otimes W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})$$

induces an isomorphism

$$(7-1) \quad \mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H \xrightarrow{\sim} \prod_{i=1}^{\alpha} \mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})) \times \prod_{i=\alpha+1}^{\alpha+\beta} O(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})) \times \prod_{i=\alpha+\beta+1}^{\alpha+\beta+\gamma} \mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})).$$

Hence the commutator of $\mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$ is

$$(7-2) \quad \prod_{i=1}^{\alpha} \mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})) \times \prod_{i=\alpha+1}^{\alpha+\beta} \mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})) \times \prod_{i=\alpha+\beta+1}^{\alpha+\beta+\gamma} \mathrm{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})).$$

The above lemma follows from the explicit description of the symplectic centralizer given in [Jain 2016, Theorem 3.1.10] (which seems to implicitly work over finite fields, but the same proof works over the complex numbers).

We next prove our main result, Theorem 1.3, which states that under suitable hypotheses on g, n and the maximal dimension \bar{r} of an irreducible representation of H , the identity component of the Zariski closure of the monodromy map $R_\varphi : \mathrm{Mod}_\varphi \rightarrow \mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$ is the commutator subgroup of $\mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$.

Proof of Theorem 1.3 Let G be the identity component of the Zariski closure of the image of the mapping class group in (7-1). In other words, G is the connected monodromy group of $\bigoplus_{i=1}^{\alpha+\beta+\gamma} W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})$. Checking Theorem 1.3, ie that the virtual image of the mapping class group is Zariski dense in the commutator subgroup of this group, is equivalent to checking that G contains (7-2). By the discussion of Section 2.3, we may interpret all the representations in question as monodromy representations associated to local systems on the base \mathcal{M} of a versal family of φ -covers.

To check that G contains (7-2), we apply the Goursat–Kolchin–Ribet criterion of Katz [1990, Proposition 1.8.2]. We let V_i be the representation of G acting on $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})$. Let G_i be the image of G in $\mathrm{GL}(V_i)$, which we know from Theorem 1.9 is $\mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i}))$ for i from 1 to α , $\mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i}))$ for i from $\alpha + 1$ to $\alpha + \beta$, and $\mathrm{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i}))$ for i from $1 + \alpha + \beta$ to $\alpha + \beta + \gamma$. Then [loc. cit., Proposition 1.8.2] guarantees that $G^{0,\mathrm{der}} = \prod_{i=1}^{\alpha+\beta+\gamma} G_i^{0,\mathrm{der}}$, which is the desired (7-2), as long as four conditions are satisfied, which we verify next.

The first condition is that for each i , $G_i^{0,\mathrm{der}}$ operates irreducibly on V_i , and its Lie algebra is simple. Theorem 1.9 guarantees that the action is by the standard representation, which is irreducible except in the case of the two-dimensional standard representation of SO_2 , and the Lie algebra is simple except in the case of the four-dimensional standard representation of SO_4 . But our assumptions on r and g imply each representation has dimension $\geq 2r_i(g - 1) \geq 2r_i(2r_i + 1) \geq 6$ where $r_i = \dim \rho_i$, so this condition is always satisfied.

The second condition is that for any $i \neq j$, $(G_i^{0,\mathrm{der}}, V_i)$ and $(G_j^{0,\mathrm{der}}, V_j)$ are Goursat-adapted in the sense of [Katz 1990, Section 1.8], but this follows by [loc. cit., Example 1.8.1] from the fact that G_i is a classical group and V_i is its standard representation of dimension ≥ 6 with the possible exception that $G_i \cong G_j \cong \mathrm{SO}_8$. However, we know from Theorem 1.9 that G_i is SO_n only if ρ_i is symplectically self-dual, which implies ρ_i has rank $r_i \geq 2$ since all symplectically self-dual representations are even-dimensional and thus $\dim V_i \geq 2r_i(2r_i + 1) \geq 20 > 8$, so G_i cannot be SO_8 .

The third and fourth conditions say that for each $i \neq j$ and each character χ of G , neither the representation V_i nor its dual is isomorphic to $V_j \otimes \chi$. Since G is the connected monodromy group of a local system of geometric origin, it is necessarily simple, and so χ is finite-order. Thus by passing to a finite cover of \mathcal{M} , the character χ becomes trivial, showing that $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i})$ is isomorphic to $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_j})$ or its dual over this finite covering. This case is ruled out by Corollary 6.9 unless ρ_i is conjugate to ρ_j or its dual, which is impossible as $i \neq j$ and $i, j \leq \alpha + \beta + \gamma$ so they cannot be part of a dual pair.

Since all the conditions of the Goursat–Kolchin–Ribet criterion of Katz [1990, Proposition 1.8.2] are satisfied, $G^{0,\text{der}} = \prod_{i=1}^{\alpha+\beta+\gamma} G_i^{0,\text{der}}$ and thus $\prod_{i=1}^{\alpha+\beta+\gamma} G_i^{0,\text{der}}$ is a subgroup of G , as desired. \square

7.17 Proofs of corollaries We next prove Corollary 1.10, which states that, for X a very general H -curve, under suitable hypotheses on g and H , the Mumford–Tate group of $H^1(X, \mathbb{Q})$ contains the commutator subgroup of $\text{Sp}(H^1(X, \mathbb{Q}))^H$ and is contained in $\text{GSp}(H^1(X, \mathbb{Q}))^H$.

Proof of Corollary 1.10 This is immediate from André’s theorem of the fixed part [André 1992, Theorem 1 on page 10] and Theorem 1.3; André’s theorem implies the generic Mumford–Tate group contains the monodromy group, namely the commutator subgroup of $\text{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$. On the other hand, it is contained in the centralizer of H in $\text{GSp}(H^1(\Sigma_{g'}, \mathbb{C}))$, as it centralizes H and preserves the symplectic pairing on H^1 up to scaling (as the symplectic pairing corresponds to a Hodge class). \square

We next prove Corollary 1.11, which, under the same hypotheses as in the previous corollary, states that the endomorphism algebra of the Jacobian of a very general H -curve X is $\mathbb{Q}[H]$.

Proof of Corollary 1.11 Let G be the Mumford–Tate group of $H^1(X, \mathbb{Q})$. It suffices to show that the natural map $\mathbb{Q}[H] \rightarrow \text{End}_{\text{HS}}(H^1(X, \mathbb{Q})) = \text{End}_G(H^1(X, \mathbb{Q}))$ is a bijection, where HS is the category of \mathbb{Q} -polarizable variations of Hodge structure. As G contains the commutator subgroup S of $\text{Sp}(H^1(X, \mathbb{Q}))^H$ by Corollary 1.10, it moreover suffices to show that the map

$$\mathbb{Q}[H] \rightarrow \text{End}_S(H^1(X, \mathbb{Q}))$$

is a bijection; we may do so after tensoring with \mathbb{C} , whence $\mathbb{C}[H] = \prod_{\rho_i} \text{End}(\rho_i)$, where the ρ_i run over the irreducible complex representations of H . Similarly, for \mathbb{V}^{ρ_i} the local system on $X/H - D$ associated to ρ_i ,

$$H^1(X, \mathbb{C}) = \bigoplus \rho_i \otimes W_1 H^1(X/H - D, \mathbb{V}^{\rho_i}),$$

where $D \subset X/H$ is the branch locus of the natural map $X \rightarrow X/H$. By the description of the symplectic centralizer, Lemma 7.16, the $W_1 H^1(X/H - D, \rho_i)$ are simple and pairwise nonisomorphic as S -representations. Hence

$$\text{End}_S(H^1(X, \mathbb{C})) = \prod_{\rho_i} \text{End}_{\mathbb{C}}(\rho_i)$$

has dimension $|H|$. As the map we are studying is evidently injective, it is necessarily surjective as well by a dimension count. \square

8 Proof of Theorem 1.15: big monodromy for n large

We now prove Theorem 1.15. We will require the following slight strengthening of [Landesman and Litt 2024b, Theorem 1.3.4], which was not quite stated optimally. See [Landesman and Litt 2024b, Definitions 1.2.1 and 1.2.3] for the definitions of *hyperbolic curve* and *analytically very general*.

8.1 Theorem *Let (C, D) be hyperbolic of genus g and let (E, ∇) be a flat vector bundle on C with regular singularities along D , and irreducible monodromy. Suppose that (E', ∇') is an isomonodromic deformation of (E, ∇) to an analytically general nearby curve, with Harder–Narasimhan filtration*

$$0 = (F')^0 \subset (F')^1 \subset \dots \subset (F')^m = E'.$$

For $1 \leq i \leq m$, let μ_i denote the slope of $\text{gr}_{\text{HN}}^i E' := (F')^i / (F')^{i-1}$. Suppose that E' is not semistable. Then for every $0 < i < m$, there exists $j < i < k$ with

$$\text{rk}(\text{gr}_{\text{HN}}^{j+1} E') \cdot \text{rk}(\text{gr}_{\text{HN}}^k E') \geq g + 1 \quad \text{and} \quad 0 < \mu_{j+1} - \mu_k \leq 1.$$

Proof In fact this is precisely the output of the proof of [Landesman and Litt 2024b, Theorem 1.3.4]. \square

8.2 Corollary *With notation as in Theorem 8.1, let $\mu_{\text{diff}} = \mu_1 - \mu_m$. Then*

$$\mu_{\text{diff}} \leq \frac{3 \text{rk}(E)}{2\sqrt{g+1}} + 3.$$

Proof Set s to be the greatest integer which is strictly less than $\mu_{\text{diff}}/3$. We first show that

$$\text{rk}(E) \geq 2\sqrt{g+1} \cdot s.$$

Fix disjoint closed intervals I_1, \dots, I_s in $[\mu_m, \mu_1]$, each of length 3. For each interval, $I_t = [a_t, a_t + 3]$, there exists i_t with $\mu_{i_t} \in [a_t + 1, a_t + 2]$, by Theorem 8.1. By the same theorem, there exists $j_t < i_t < k_t$ with

$$(8-1) \quad \text{rk}(\text{gr}_{\text{HN}}^{j_t+1} E') \cdot \text{rk}(\text{gr}_{\text{HN}}^{k_t} E') \geq g + 1$$

and

$$0 < \mu_{j_t+1} - \mu_{k_t} \leq 1.$$

In particular, $\mu_{j_t+1}, \mu_{k_t} \in I_t$, and hence all the integers $j_t + 1, k_t$ are distinct.

Now we have

$$\text{rk}(E) = \text{rk}(E') \geq \sum_{t=1}^s (\text{rk}(\text{gr}_{\text{HN}}^{j_t+1} E') + \text{rk}(\text{gr}_{\text{HN}}^{k_t} E')),$$

which is bounded below by $(2\sqrt{g+1}) \cdot s$ by the AM–GM inequality and (8-1). So $s \leq \text{rk}(E) / (2\sqrt{g+1})$, which implies

$$\mu_{\text{diff}} \leq 3s + 3 \leq \frac{3 \text{rk}(E)}{2\sqrt{g+1}} + 3. \quad \square$$

8.3 Lemma *With notation as in Notation 3.3, one of E_0^ρ and $E_0^{\rho^\vee}$ has slope less than or equal to $-\Delta/2r$.*

Proof It suffices to show that $E_0^\rho \oplus E_0^{\rho^\vee} = E_0^{(\rho \oplus \rho^\vee)}$ has degree less than or equal to $-\Delta$. But if λ is an eigenvalue of local monodromy of ρ at a point x of D_{nontriv}^ρ , then λ^{-1} is an eigenvalue of local monodromy of ρ^\vee at x . Hence if $\alpha \neq 0$ is a parabolic weight of E_\star^ρ at x , then $1 - \alpha$ is a parabolic weight of $E_\star^{\rho^\vee}$ at x . In particular, the sum of the parabolic weights of $E_\star^{(\rho \oplus \rho^\vee)}$ at x is at least 1 for each $x \in D_{\text{nontriv}}^\rho$.

Hence,

$$\deg E_0^{(\rho \oplus \rho^\vee)} = - \sum_{x_j \in D_{\text{nontriv}}^\rho} \sum_{i=1}^{n_j} \alpha_j^i \dim(E_j^i/E_j^{i+1}) \leq - \sum_{x \in D_{\text{nontriv}}^\rho} 1 \leq -\Delta. \quad \square$$

8.4 Lemma *With notation as in Notation 3.3, suppose (C, D) is a general n -pointed curve. If*

$$(8-2) \quad \Delta > \frac{3r^2}{\sqrt{g+1}} + 8r,$$

then at least one of $(E_0^\rho)^\vee \otimes \omega_C$ and $(E_0^{\rho^\vee})^\vee \otimes \omega_C$ is globally generated.

Proof Without loss of generality we may (by replacing ρ with ρ^\vee if necessary) assume

$$\mu(E_0^\rho) \leq -\frac{\Delta}{2r},$$

by Lemma 8.3. We will show that in this case $(E_0^\rho)^\vee \otimes \omega_C$ is globally generated.

Let $\mu_1 > \dots > \mu_m$ be the set of slopes of the graded pieces of the Harder–Narasimhan filtration of E_0^ρ , as in Theorem 8.1, and let $\mu_{\text{diff}} = \mu_1 - \mu_m$. By Corollary 8.2, we have

$$\mu_{\text{diff}} \leq \frac{3r}{2\sqrt{g+1}} + 3.$$

Hence

$$\mu_1 \leq -\frac{\Delta}{2r} + \mu_{\text{diff}} \leq -\frac{\Delta}{2r} + \frac{3r}{2\sqrt{g+1}} + 3 < -1$$

by rearranging the assumption (8-2). Thus the Harder–Narasimhan slopes of $(E_0^\rho)^\vee \otimes \omega_C$ are all greater than $2g - 1$.

We claim any vector bundle V such that each graded piece of its Harder–Narasimhan filtration has slope more than $2g - 1$ is globally generated. This is well known, but we explain it for completeness. A vector bundle V is globally generated at p if $H^1(C, V(-p)) = 0$, as then $H^0(C, V) \rightarrow H^0(C, V|_p)$ is surjective. By [Landesman and Litt 2024b, Lemma 6.3.5], if W is a vector bundle such that each graded piece of its Harder–Narasimhan filtration has slope more than $2g - 2$, $H^1(C, W) = 0$. Therefore, $H^1(C, V(-p)) = 0$ for any point p , so V is globally generated. This shows $(E_0^\rho)^\vee \otimes \omega_C$ is globally generated. □

8.5 Proof of Theorem 1.15 We are aiming to prove that, once Δ is sufficiently large (larger than $(3r^2/\sqrt{g+1}) + 8r$), there are no nonzero vectors with finite orbit in $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ under the image of Mod_φ ; here, Mod_φ is the stabilizer of $\varphi: \pi_1(\Sigma_{g,n}) \rightarrow H$ in $\text{Mod}_{g,n+1}$.

As the virtual representations of $\text{Mod}_{g,n+1}$ on

$$W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho), W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho^\vee})$$

are semisimple and dual to one another, it suffices to prove this for one of ρ and ρ^\vee , so we may without loss of generality assume by Lemma 8.4 that $(E_0^\rho)^\vee \otimes \omega_C$ is globally generated for (C, D) a general n -pointed curve. A similar argument to that given in [Landesman and Litt 2023a, Proposition 3.4] shows that the virtual action of $\text{Mod}_{g,n+1}$ on $\text{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ has no nonzero finite-orbit vectors.

To make this proof slightly more self-contained, we recall briefly the idea of the proof of [Landesman and Litt 2023a, Proposition 3.4]. Namely, the derivative of the period map as described in Proposition 3.6 associated to the Hodge filtration of $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ can be identified with a map

$$H^0(C, \hat{E}_0^\rho \otimes \omega_C(D)) \otimes H^0(C, (E_0^\rho)^\vee \otimes \omega_C) \rightarrow H^0(C, \omega_C^{\otimes 2}(D)).$$

If there is a vector with finite orbit in $W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)$ then the adjoint map

$$H^0(C, \hat{E}_0^\rho \otimes \omega_C(D)) \rightarrow \text{Hom}(H^0(C, (E_0^\rho)^\vee \otimes \omega_C), H^0(C, \omega_C^{\otimes 2}(D)))$$

has a nonzero kernel. A vector in the kernel yields a nonzero map

$$\phi: (E_0^\rho)^\vee \otimes \omega_C \rightarrow \omega_C^{\otimes 2}(D)$$

inducing the 0 map on global sections. Then, $\ker \phi \subset (E_0^\rho)^\vee \otimes \omega_C$ would be a proper subbundle inducing an isomorphism on global sections, implying $(E_0^\rho)^\vee \otimes \omega_C$ is not generically globally generated, hence not globally generated, a contradiction. \square

9 Big monodromy for Kodaira fibrations

In this section we analyze the connected monodromy groups of certain smooth proper families

$$\pi: S \rightarrow Z$$

over \mathbb{C} , referred to as *Kodaira–Parshin fibrations*, where Z is a smooth curve and π is smooth and proper of relative dimension one, with connected fibers. Loosely speaking, these families will parametrize covers of a fixed curve C , branched at a moving point. We next give a few definitions to fix notation for Kodaira–Parshin fibrations.

9.1 Definition Let Z be a smooth, not necessarily proper, connected curve. A smooth proper morphism $\pi: S \rightarrow Z$ of relative dimension one with connected fibers is a *Kodaira–Parshin fibration* if there exists a smooth proper curve C , a reduced effective divisor $D \subset C$, a nonconstant map $f: Z \rightarrow C$, and a dominant finite map $q: S \rightarrow Z \times C$, branched only over the graph of f and $Z \times D$, such that π factors

as $\pi = \pi_1 \circ q$, where π_1 is projection onto the first coordinate, as in the diagram

$$(9-1) \quad \begin{array}{ccccc} S & \xrightarrow{q} & Z \times C & \xrightarrow{\pi_2} & C \\ & \searrow \pi & \downarrow \pi_1 & & \\ & & Z & & \end{array}$$

9.2 Definition Continuing with notation as in Definition 9.1, fix $z \in Z$, and consider the map $\pi^{-1}(z) \rightarrow C$ given as $\pi_2 \circ q|_{\pi^{-1}(z)}$, where $\pi_2: Z \times C \rightarrow C$ is projection on to the second coordinate. We refer to the underlying map on topological spaces (in the Euclidean topology) as the *topological type* of the family π .

9.3 Remark The topological type of π , as defined in Definition 9.2, is independent of z up to homeomorphism, by Ehresmann’s theorem. That is, every fiber of π is a cover of C of the same topological type. Hence it makes sense to speak of the topological type of π , and not just of π over z .

9.4 Definition Continuing with notation as in Definition 9.2, the topological type of a Kodaira–Parshin fibration is a ramified map of (topological) surfaces $t: \Sigma_{g'} \rightarrow \Sigma_g$; we refer to the Galois group of (the Galois closure of) this map as the *Galois group* of the Kodaira–Parshin fibration. Note that Σ_g has a *distinguished point*, which corresponds to $f(z)$ under the isomorphism $\Sigma_g \simeq C$. We refer to the conjugacy class of the monodromy about this point in the Galois group H of the cover as the *distinguished class* $[h] \subset H$.

We next recall the classical Kodaira–Parshin trick; see [Parshin 1968, Proposition 7] for the original construction, and also [Landesman and Litt 2024b, Proposition 5.1.1] for a more modern construction in families. See also [Atiyah 1969] and [Kodaira 1967] for closely related constructions.

9.5 Example (the Kodaira–Parshin trick) Let C be a smooth proper curve, and let $p: S \rightarrow C \times C$ be a dominant finite map branched only over the diagonal Δ . More precisely, S is the normalization of $C \times C - \Delta$ in the function field of a finite étale cover of $C \times C - \Delta$. Let

$$S \xrightarrow{\pi} Z \rightarrow C$$

be the Stein factorization of the composition $\pi_1 \circ p$, where $\pi_1: C \times C \rightarrow C$ is projection onto the first factor. Then π is a Kodaira–Parshin fibration.

9.6 Remark Define a *Kodaira fibration* to be a surjective smooth proper morphism from a smooth projective surface to a smooth *proper* curve. In the setting of Example 9.5, C itself is proper, so π is in fact a Kodaira fibration. The construction of Example 9.5 is known as the *Kodaira–Parshin trick*, and is used eg by Faltings in his proof of the Mordell conjecture.

The main theorem of this section is a computation of the connected monodromy group of a Kodaira–Parshin fibration of topological type $t: \Sigma_{g'} \rightarrow \Sigma_g$, when g is large compared to the dimensions of the irreducible representations of H . We give a statement for Kodaira–Parshin fibrations with Galois topological type,

but one may easily deduce an analogous statement for arbitrary Kodaira–Parshin fibrations by passing to Galois closures. To state the theorem, we use the following notation.

9.7 Notation Let $\pi : S \rightarrow Z$ be a Kodaira–Parshin fibration of topological type $t : \Sigma_{g'} \rightarrow \Sigma_g$, with t a Galois cover branched at n points, with Galois group H and distinguished class $[h] \subset H$, as defined in Definition 9.1. Let $\varphi : \pi_1(\Sigma_{g,n}, x) \twoheadrightarrow H$ denote the surjection associated to t , for x a chosen basepoint. Let $\rho_1^O, \dots, \rho_\alpha^O$ be the irreducible orthogonally self-dual complex H -representations with $\rho_i^O([h])$ nontrivial, $\rho_1^{Sp}, \dots, \rho_\beta^{Sp}$ the irreducible symplectically self-dual H irreducible complex representations with $\rho_i^{Sp}([h])$ nontrivial, and let $(\rho_1^{SL}, \rho_{\gamma+1}^{SL}), \dots, (\rho_\gamma^{SL}, \rho_{2\gamma}^{SL})$ be the set of dual pairs of complex irreducible H -representations with $\rho_i^{SL}([h])$ nontrivial. Let \bar{s} be the maximal dimension of an irreducible representation ρ of H with $\rho([h])$ nontrivial. We use \mathbb{V}^ρ to denote the local system on $\Sigma_{g,n}$ associated to the composition $\rho \circ \varphi$.

9.8 Theorem With notation as in Notation 9.7, suppose $g > \max(2\bar{s} + 1, \bar{s}^2)$. Then the identity component of the Zariski-closure of $\pi_1(Z, z)$ in $GL(H^1(\pi^{-1}(z), \mathbb{C}))$ is

$$(9-2) \quad \prod_{i=1}^{\alpha} Sp(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i^O})) \times \prod_{i=1}^{\beta} SO(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i^{Sp}})) \times \prod_{i=1}^{\gamma} SL(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i^{SL}})),$$

the subgroup of the derived subgroup of the centralizer of H in $Sp(H^1(\Sigma_{g'}, \mathbb{C}))$ corresponding to those H -irreps which are nontrivial on $[h]$.

The following purely topological corollary follows immediately from the definitions, as in Section 2.3—loosely speaking, it says that in the representations considered in Theorem 1.3, the restriction to the “point-pushing subgroup” still has large image.

9.9 Corollary With notation as in Notation 9.7, suppose that $n > 0$, and let $h \in \pi_1(\Sigma_{g,n}, x)$ be a loop around a fixed puncture z of $\Sigma_{g,n}$, such that $\varphi(h) = [h]$ is the distinguished class. Suppose that $g > \max(2\bar{s} + 1, \bar{s}^2)$. Let P_φ be the stabilizer of φ in the kernel of the map

$$\text{Mod}_{g,n+1} \rightarrow \text{Mod}_{g,n}$$

induced by forgetting z . (Here we view $\Sigma_{g,n}$ as an $(n+1)$ -marked surface, with x an additional marked point.) The identity component of the Zariski closure of

$$R_\varphi|_{P_\varphi} : P_\varphi \rightarrow Sp(H^1(\Sigma_{g'}, \mathbb{C}))^H$$

in $GL(H^1(\Sigma_{g'}, \mathbb{C}))$ is

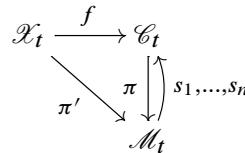
$$\prod_{i=1}^{\alpha} Sp(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i^O})) \times \prod_{i=1}^{\beta} SO(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i^{Sp}})) \times \prod_{i=1}^{\gamma} SL(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i^{SL}})),$$

the subgroup of the derived subgroup of the centralizer of H in $Sp(H^1(\Sigma_{g'}, \mathbb{C}))$ corresponding to those H -irreps nontrivial on $[h]$. Here R_φ is defined as in Theorem 1.3.

In order to prove Theorem 9.8, we recall some notation from Notation 2.1. Given a branched cover

$$t: \Sigma_{g'} \rightarrow \Sigma_g$$

of topological surfaces, with Galois group H and n branch points (here consisting of the points of D together with the $f(z)$), there is a versal family of covers over a variety \mathcal{M}_t , parametrizing branched covers of Riemann surfaces of topological type t (see eg [Bertin and Romagny 2011, Section 6]), carrying a family of covers $\mathcal{X}_t/\mathcal{C}_t$, as in the diagram



After replacing Z with an étale cover, our given Kodaira–Parshin fibration $S \rightarrow Z$ of topological type t is the pullback of π' by a map $\iota: Z \rightarrow \mathcal{M}_t$. The family of pointed curves $\mathcal{C}_t/\mathcal{M}_t$ induces a dominant map $\mathcal{M}_t \rightarrow \mathcal{M}_{g,n}$ with finitely many geometric points in each fiber. By definition, ι dominates a component of the fiber of the composition $\mathcal{M}_t \rightarrow \mathcal{M}_{g,n} \rightarrow \mathcal{M}_{g,n-1}$, given by forgetting the distinguished point as defined in Definition 9.4.

The following lemma is the main ingredient in the proof, and describes the monodromy of a Kodaira–Parshin family associated to a particular representation.

9.10 Lemma *With notation as above, let $[h] \in H$ be the distinguished class, and let $\rho: H \rightarrow \mathrm{GL}_r(\mathbb{C})$ be an irreducible H -representation. Suppose $g > \max(2r + 1, r^2)$. Then the identity component of the Zariski-closure of the monodromy representation*

$$(9-3) \quad \pi_1(Z, z) \rightarrow \mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$$

is trivial if $\rho([h])$ is trivial. If $\rho([h])$ is nontrivial, it is

- (1) $\mathrm{SO}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is symplectically self-dual,
- (2) $\mathrm{Sp}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is orthogonally self-dual, or
- (3) $\mathrm{SL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho))$ if ρ is not self-dual.

Proof We first observe that if $\rho([h])$ is trivial, the connected component of the monodromy group in question is also trivial. Indeed, in this case, let $K \subset H$ be the kernel of ρ ; the monodromy representation in question appears in the cohomology of $S/K \rightarrow Z$, which is isotrivial. Indeed, the fibers of this map are covers of C branched at a fixed divisor, and hence do not vary in moduli.

We now suppose $\rho([h])$ is nontrivial. We first claim that it suffices to show that the monodromy representation (9-3) has infinite image. Indeed, by the discussion above the statement of the lemma, (9-3) factors through a representation

$$(9-4) \quad \pi_1(\mathcal{M}_t) \rightarrow \mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^\rho)).$$

By Theorem 1.9, the representation (9-4) has image with Zariski-closure SO , Sp or SL , depending on the self-duality properties of ρ . The connected monodromy group of (9-3) is in fact a *normal* subgroup of this group, as the fundamental group of the fiber of the map $\mathcal{M}_t \rightarrow \mathcal{M}_{g,n-1}$ is normal in the fundamental group of \mathcal{M}_t . Thus if the image in question is infinite, it must be Zariski-dense in the Zariski-closure of the image of (9-4), as this group is simple by Theorem 6.7.

We prove the infinitude of the image of this representation by Hodge-theoretic methods. As the monodromy representation in question is a topological invariant, we may take Z to dominate a component of a general fiber of the map $\mathcal{M}_t \rightarrow \mathcal{M}_{g,n-1}$ and hence assume that C is general.

Fix general $z \in Z$, and let $X = \pi^{-1}(z)$, let $D \subset C$ be the branch locus of the cover $X \rightarrow C$, and let $p \in C$ be the image of z in C ; note that (C, D) is general, and $p \in D$ by the definition of a Kodaira–Parshin fibration. Recall that the theorem of the fixed part implies a VHS with finite monodromy has constant period map. Using this, it suffices to show that the derivative of the period map

$$T_z Z \rightarrow \mathrm{Hom}(H^0(C, \hat{E}_0^\rho \otimes \omega_X(D)), H^1(C, E_0^\rho))$$

is nonzero for generic $z \in Z$. There is a natural identification

$$T_z Z = \ker(H^1(C, T_C(-D)) \rightarrow H^1(C, T_C(-D + p))),$$

as the map $Z \rightarrow C$ is generically étale, so Serre-dually we wish to show that the pairing

$$\begin{aligned} H^0(C, \hat{E}_0^\rho \otimes \omega_X(D)) \otimes H^0(C, (E_0^\rho)^\vee \otimes \omega_C) &\rightarrow H^0(C, \omega_C^{\otimes 2}(D - p)) \\ &\rightarrow \mathrm{coker}(H^0(C, \omega_C^{\otimes 2}(D - p)) \rightarrow H^0(C, \omega_C^{\otimes 2}(D))) \end{aligned}$$

is nonzero, where the first map is induced by B_z^ρ , as in Proposition 3.6. Evaluation at p identifies

$$\mathrm{coker}(H^0(C, \omega_C^{\otimes 2}(D - p)) \rightarrow H^0(C, \omega_C^{\otimes 2}(D)))$$

with $\omega_C^{\otimes 2}(D)|_p$. Using that (C, D) is general, so $\hat{E}_0^\rho \otimes \omega_X(D)$ and $(E_0^\rho)^\vee \otimes \omega_C$ are globally generated by Proposition 4.9, it is enough to show that the pairing

$$(\hat{E}_0^\rho \otimes \omega_X(D)) \otimes (E_0^\rho)^\vee \otimes \omega_C \rightarrow \omega_X^{\otimes 2}(D)|_p$$

is nonzero. But this pairing is induced by the corresponding pairing of sheaves between $E_0^\rho \otimes \omega_X(D)$ and $(E_0^\rho)^\vee \otimes \omega_C$, which is perfect, via the inclusion $\hat{E}_0^\rho \hookrightarrow E_0^\rho$. So it is nonzero as long as this inclusion is nonzero at p , ie as long as $\rho([h])$ is nontrivial, as desired. \square

To complete the proof of Theorem 9.8, we will need to compute the total monodromy group, which is contained in a product of groups indexed by the different irreps of H . To prove our result we will use Goursat’s lemma, and the following lemma is a key input to verifying the hypotheses of Goursat’s lemma.

9.11 Lemma *Let $\rho_1, \rho_2: H \rightarrow \mathrm{GL}_r(\mathbb{C})$ be irreducible H -representations with $\rho_i([h])$ nontrivial for $i = 1, 2$. If $g > r^2$ and the monodromy representations*

$$(9-5) \quad \pi_1(Z, z) \rightarrow \mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_i}))$$

for $i = 1, 2$ are isomorphic to one another, then $\rho_1 \simeq \rho_2$.

Proof We can factor (9-5) as a composition

$$(9-6) \quad \pi_1(Z, z) \rightarrow \pi_1(\mathcal{M}_t) \xrightarrow{\phi_t} \mathrm{GL}(W_1 H^1(\Sigma_{g,n}, \mathbb{V}^{\rho_t})),$$

and by Proposition 6.4, it suffices to show that ϕ_1 is isomorphic to ϕ_2 , possibly after passing to a cover of \mathcal{M}_t . By [Landesman and Litt 2024a, Lemma 2.2.2] (using that our two representations of $\pi_1(Z, z)$ in question are irreducible by Lemma 9.10), the projectivizations of ϕ_1 and ϕ_2 are isomorphic. Hence, we may assume $\phi_1 \simeq \phi_2 \otimes \chi$, for χ a character. Observe that χ must be of finite order because its connected monodromy group is simple (as also argued in the penultimate paragraph of the proof of Theorem 1.3). Hence, we may trivialize χ by passing to a cover of \mathcal{M}_t , and therefore reduce to the case $\phi_1 \simeq \phi_2$, as desired. \square

Proof of Theorem 9.8 The proof follows from the Goursat–Kolchin–Ribet criterion of Katz [1990, Proposition 1.8.2], precisely following the argument in the proof of Theorem 1.3; we use Lemma 9.10 in place of Theorem 1.9, and Lemma 9.11 in place of Proposition 6.4. The final statement describing (9-2) follows from Lemma 7.16. \square

10 Questions

We conclude with a number of open questions motivated by the preceding results.

10.1 Improving the bounds in our results Theorem 1.3 gives a large monodromy result for the mapping class group action on the homology of an H -cover of a genus g curve, on the assumption that g is large compared to the dimensions of the irreducible representations of H . However, we don't know a counterexample to the statement of Theorem 1.3 as long as $g \geq 3$.

10.2 Question Let H be a finite group and let $\Sigma_{g',n'} \rightarrow \Sigma_{g,n}$ be an H -cover. Suppose that $g \geq 3$. Is the identity component of the Zariski closure of the virtual image $\mathrm{Mod}_{g,n+1}$ in $\mathrm{GL}(W_1 H^1(\Sigma_{g'}))$ the commutator subgroup of $\mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{C}))^H$?

It seems natural to conjecture that the answer is yes, strengthening the Putman–Wieland conjecture [2013] as explained in Remark 1.6. Moreover one might expect that as soon as $g \geq 3$, the image is arithmetic:

10.3 Question Does the virtual image of $\mathrm{Mod}_{g,n+1}$ in $\mathrm{Sp}(H^1(\Sigma_{g'}, \mathbb{Z}))^H$ have finite index?

While our methods say nothing about Question 10.3, strengthening our results on (generic) global generation would imply a positive answer to Question 10.2 in some cases, as we now explain.

10.4 Question Let $g \geq 3$, and let $\rho: \pi_1(\Sigma_{g,n}) \rightarrow \mathrm{GL}_r(\mathbb{C})$ be a representation with finite image. Fix a very general complex structure (C, D) on a pointed surface $(\Sigma_g, x_1, \dots, x_n)$ and let E_\star be the parabolic bundle corresponding to ρ . Is the vector bundle $\hat{E}_0 \otimes \omega_C(D)$

- (a) generically globally generated, or even
- (b) globally generated?

10.5 Remark When $n = 0$, a positive answer to Question 10.4(b) for arbitrary r would imply a positive answer to Question 10.2 by the methods of this paper. Similarly, for arbitrary n , a positive answer to Question 10.4(a) would imply a positive answer to the Putman–Wieland conjecture [2013].

10.6 Remark One might naturally pose Question 10.4 for arbitrary unitary representations ρ . We are grateful to Eric Larson and Isabel Vogt for explaining to us that in this case one must necessarily take the meaning of “very general” in the question to *depend* on ρ . That is, every smooth proper curve C of genus $g \geq 2$ admits a stable vector bundle E of degree zero (corresponding, by the Narasimhan–Seshadri correspondence, to some unitary representation) such that $E \otimes \omega_C$ is not generically globally generated.

10.7 Remark A positive answer to Question 10.4 for general unitary representations would provide some evidence for the well-known conjecture that mapping class groups have Kazhdan’s property T in genus $g \geq 3$, as it would imply that certain unitary representations of mapping class groups are rigid, following the methods of [Landesman and Litt 2024a]. See [Ivanov 2006, Section 8] for a discussion of this question.

10.8 Arithmetic statistics As mentioned in Remark 1.14, the results of this paper are closely related to a number of results in arithmetic statistics, which concern understanding the monodromy with $\mathbb{Z}/\ell\mathbb{Z}$ coefficients, instead of complex coefficients. As noted in the introduction, Jain [2016] is able to prove a big monodromy result over genus 0 bases with the number n of punctures having monodromy in every conjugacy class sufficiently large. However, his result does not say how large n has to be.

10.9 Question Is it possible to obtain a big monodromy result analogous to Theorem 1.3 which is effective in g , or a result analogous to Conjecture 1.13 which is effective in n , with $\mathbb{Z}/\ell\mathbb{Z}$ coefficients instead of complex coefficients?

If the above were possible, can one use it to deduce some large q limit versions of the Cohen–Lenstra heuristics, as alluded to in Remark 1.14?

10.10 Analogs for free groups Finally, it is natural to pose analogues of the questions here for representations of groups other than the mapping class group $\text{Mod}_{g,n}$. For example, the group $\text{Aut}(F_n)$ acts virtually on finite-index subgroups of the free group F_n , and hence on their abelianizations.

10.11 Question Let $n \geq 3$ and fix a finite-index subgroup $K \subset F_n$. Is the image in $\text{GL}(K^{\text{ab}})$ of the stabilizer of K inside $\text{Aut}(F_n)$ arithmetic? What is its Zariski-closure?

This last question is studied in many interesting special cases by Grunewald and Lubotzky [2009].

References

[Abramovich et al. 2003] **D Abramovich, A Corti, A Vistoli**, *Twisted bundles and admissible covers*, *Comm. Algebra* 31 (2003) 3547–3618 MR Zbl

- [A'Campo 1979] **N A'Campo**, *Tresses, monodromie et le groupe symplectique*, Comment. Math. Helv. 54 (1979) 318–327 MR Zbl
- [Achter 2008] **J D Achter**, *Results of Cohen–Lenstra type for quadratic function fields*, from “Computational arithmetic geometry” (K E Lauter, K A Ribet, editors), Contemp. Math. 463, Amer. Math. Soc., Providence, RI (2008) 1–7 MR Zbl
- [André 1992] **Y André**, *Mumford–Tate groups of mixed Hodge structures and the theorem of the fixed part*, Compos. Math. 82 (1992) 1–24 MR Zbl
- [Atiyah 1969] **M F Atiyah**, *The signature of fibre-bundles*, from “Global analysis” (D C Spencer, S Iyanaga, editors), Univ. Tokyo Press (1969) 73–84 MR Zbl
- [Bertin and Romagny 2011] **J Bertin, M Romagny**, *Champs de Hurwitz*, Mém. Soc. Math. France 125-126, Soc. Math. France, Paris (2011) MR Zbl
- [Boden and Yokogawa 1996] **H U Boden, K Yokogawa**, *Moduli spaces of parabolic Higgs bundles and parabolic $K(D)$ pairs over smooth curves, I*, Int. J. Math. 7 (1996) 573–598 MR Zbl
- [Carlson et al. 1983] **J Carlson, M Green, P Griffiths, J Harris**, *Infinitesimal variations of Hodge structure, I*, Compos. Math. 50 (1983) 109–205 MR Zbl
- [Chevalley 1954] **C C Chevalley**, *The algebraic theory of spinors*, Columbia Univ. Press, New York (1954) MR Zbl
- [Collino and Pirola 1995] **A Collino, G P Pirola**, *The Griffiths infinitesimal invariant for a curve in its Jacobian*, Duke Math. J. 78 (1995) 59–88 MR Zbl
- [Deligne 1979] **P Deligne**, *Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques*, from “Automorphic forms, representations and L -functions, II” (A Borel, W Casselman, editors), Proc. Sympos. Pure Math. 33, Amer. Math. Soc., Providence, RI (1979) 247–289 MR Zbl
- [Deligne and Mostow 1986] **P Deligne, G D Mostow**, *Monodromy of hypergeometric functions and nonlattice integral monodromy*, Inst. Hautes Études Sci. Publ. Math. 63 (1986) 5–89 MR Zbl
- [Deligne et al. 1982] **P Deligne, J S Milne, A Ogus, K-y Shih**, *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Math. 900, Springer (1982) MR Zbl
- [Dunfield and Thurston 2006] **N M Dunfield, W P Thurston**, *Finite covers of random 3-manifolds*, Invent. Math. 166 (2006) 457–521 MR Zbl
- [Ellenberg and Landesman 2023] **J S Ellenberg, A Landesman**, *Homological stability for generalized Hurwitz spaces and Selmer groups in quadratic twist families over function fields*, preprint (2023) arXiv 2310.16286
- [Ellenberg et al. 2016] **J S Ellenberg, A Venkatesh, C Westerland**, *Homological stability for Hurwitz spaces and the Cohen–Lenstra conjecture over function fields*, Ann. of Math. 183 (2016) 729–786 MR Zbl
- [Farb and Margalit 2012] **B Farb, D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl
- [Feng et al. 2023] **T Feng, A Landesman, E M Rains**, *The geometric distribution of Selmer groups of elliptic curves over function fields*, Math. Ann. 387 (2023) 615–687 MR Zbl
- [Fried and Völklein 1991] **M D Fried, H Völklein**, *The inverse Galois problem and rational points on moduli spaces*, Math. Ann. 290 (1991) 771–800 MR Zbl
- [Grunewald and Lubotzky 2009] **F Grunewald, A Lubotzky**, *Linear representations of the automorphism group of a free group*, Geom. Funct. Anal. 18 (2009) 1564–1608 MR Zbl

- [Grunewald et al. 2015] **F Grunewald, M Larsen, A Lubotzky, J Malestein**, *Arithmetic quotients of the mapping class group*, *Geom. Funct. Anal.* 25 (2015) 1493–1542 MR Zbl
- [Harris 1985] **J Harris**, *An introduction to infinitesimal variations of Hodge structures*, from “Workshop Bonn 1984” (F Hirzebruch, J Schwermer, S Suter, editors), *Lecture Notes in Math.* 1111, Springer (1985) 51–58 MR Zbl
- [Hassett and Tschinkel 2022] **B Hassett, Y Tschinkel**, *Equivariant geometry of odd-dimensional complete intersections of two quadrics*, *Pure Appl. Math. Q.* 18 (2022) 1555–1597 MR Zbl
- [Ivanov 2006] **N V Ivanov**, *Fifteen problems about the mapping class groups*, from “Problems on mapping class groups and related topics” (B Farb, editor), *Proc. Sympos. Pure Math.* 74, Amer. Math. Soc., Providence, RI (2006) 71–80 MR Zbl
- [Jain 2016] **L Jain**, *Big mod ℓ monodromy for families of G covers*, PhD thesis, University of Wisconsin–Madison (2016) Available at <https://www.proquest.com/docview/1808033223>
- [Katz 1990] **N M Katz**, *Exponential sums and differential equations*, *Ann. of Math. Stud.* 124, Princeton Univ. Press (1990) MR Zbl
- [Kodaira 1967] **K Kodaira**, *A certain type of irregular algebraic surfaces*, *J. Anal. Math.* 19 (1967) 207–215 MR Zbl
- [Landesman and Litt 2023a] **A Landesman, D Litt**, *Applications of the algebraic geometry of the Putman–Wieland conjecture*, *Proc. Lond. Math. Soc.* 127 (2023) 116–133 MR Zbl
- [Landesman and Litt 2023b] **A Landesman, D Litt**, *An introduction to the algebraic geometry of the Putman–Wieland conjecture*, *Eur. J. Math.* 9 (2023) art. id. 40 MR Zbl
- [Landesman and Litt 2024a] **A Landesman, D Litt**, *Canonical representations of surface groups*, *Ann. of Math.* 199 (2024) 823–897 MR Zbl
- [Landesman and Litt 2024b] **A Landesman, D Litt**, *Geometric local systems on very general curves and isomonodromy*, *J. Amer. Math. Soc.* 37 (2024) 683–729 MR Zbl
- [Lawrence and Venkatesh 2020] **B Lawrence, A Venkatesh**, *Diophantine problems and p -adic period mappings*, *Invent. Math.* 221 (2020) 893–999 MR Zbl
- [Looijenga 1997] **E Looijenga**, *Prym representations of mapping class groups*, *Geom. Dedicata* 64 (1997) 69–83 MR Zbl
- [Looijenga 2025] **E Looijenga**, *Arithmetic representations of mapping class groups*, *Algebr. Geom. Topol.* 25 (2025) 677–698
- [Marković 2022] **V Marković**, *Unramified correspondences and virtual properties of mapping class groups*, *Bull. Lond. Math. Soc.* 54 (2022) 2324–2337 MR Zbl
- [McMullen 2013] **C T McMullen**, *Braid groups and Hodge theory*, *Math. Ann.* 355 (2013) 893–946 MR Zbl
- [Mehta and Seshadri 1980] **V B Mehta, C S Seshadri**, *Moduli of vector bundles on curves with parabolic structures*, *Math. Ann.* 248 (1980) 205–239 MR Zbl
- [Moonen 2017] **B Moonen**, *Families of motives and the Mumford–Tate conjecture*, *Milan J. Math.* 85 (2017) 257–307 MR Zbl
- [Narasimhan and Seshadri 1965] **M S Narasimhan, C S Seshadri**, *Stable and unitary vector bundles on a compact Riemann surface*, *Ann. of Math.* 82 (1965) 540–567 MR Zbl

- [Park and Wang 2024] **S W Park, N Wang**, *On the average of p -Selmer ranks in quadratic twist families of elliptic curves over global function fields*, *Int. Math. Res. Not.* 2024 (2024) 4516–4540 MR Zbl
- [Parshin 1968] **A N Parshin**, *Algebraic curves over function fields, I*, *Izv. Akad. Nauk SSSR Ser. Mat.* 32 (1968) 1191–1219 MR Zbl In Russian; translated in *Math. USSR-Izv.* 2 (1968) 1145–1170
- [Putman and Wieland 2013] **A Putman, B Wieland**, *Abelian quotients of subgroups of the mappings class group and higher Prym representations*, *J. Lond. Math. Soc.* 88 (2013) 79–96 MR Zbl
- [Salter and Tshishiku 2020] **N Salter, B Tshishiku**, *Arithmeticity of the monodromy of some Kodaira fibrations*, *Compos. Math.* 156 (2020) 114–157 MR Zbl
- [Samperton 2020] **E Samperton**, *Schur-type invariants of branched G -covers of surfaces*, from “Topological phases of matter and quantum computation” (P Bruillard, C Ortiz Marrero, J Plavnik, editors), *Contemp. Math.* 747, Amer. Math. Soc., Providence, RI (2020) 173–197 MR Zbl
- [Sawin and Wood 2024] **W Sawin, M M Wood**, *Finite quotients of 3-manifold groups*, *Invent. Math.* 237 (2024) 349–440 MR Zbl
- [Sernesi 2006] **E Sernesi**, *Deformations of algebraic schemes*, *Grundle Math. Wissen.* 334, Springer (2006) MR Zbl
- [Venkataramana 2014a] **T N Venkataramana**, *Image of the Burau representation at d -th roots of unity*, *Ann. of Math.* 179 (2014) 1041–1083 MR Zbl
- [Venkataramana 2014b] **T N Venkataramana**, *Monodromy of cyclic coverings of the projective line*, *Invent. Math.* 197 (2014) 1–45 MR Zbl
- [Voisin 2022] **C Voisin**, *Schiffer variations and the generic Torelli theorem for hypersurfaces*, *Compos. Math.* 158 (2022) 89–122 MR Zbl
- [Wewers 1998] **S Wewers**, *Construction of Hurwitz spaces*, PhD thesis, University of Essen (1998) Zbl
- [Wood 2021] **M M Wood**, *An algebraic lifting invariant of Ellenberg, Venkatesh, and Westerland*, *Res. Math. Sci.* 8 (2021) art. id. 21 MR Zbl
- [Yokogawa 1995] **K Yokogawa**, *Infinitesimal deformation of parabolic Higgs sheaves*, *Int. J. Math.* 6 (1995) 125–148 MR Zbl
- [Zarkhin 1984] **Y G Zarkhin**, *Weights of simple Lie algebras in the cohomology of algebraic varieties*, *Izv. Akad. Nauk SSSR Ser. Mat.* 48 (1984) 264–304 MR Zbl In Russian; translated in *Math. USSR-Izv.* 24 (1985) 245–281

*Department of Mathematics FAS, Harvard University
Cambridge, MA, United States*

*Department of Mathematics, University of Toronto
Toronto, ON, Canada*

*Department of Mathematics, Princeton University
Princeton, NJ, United States*

aaronlandesman@gmail.com, daniel.litt@utoronto.ca, wsawin@math.princeton.edu

Proposed: Benson Farb
Seconded: Dan Abramovich, David Fisher

Received: 29 April 2024
Accepted: 19 October 2024

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 Bib \TeX 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 29 Issue 5 (pages 2251–2782) 2025

Quantitative Thomas–Yau uniqueness	2251
YANG LI	
Knot surgery formulae for instanton Floer homology, I: The main theorem	2269
ZHENKUN LI and FAN YE	
Morse actions of discrete groups on symmetric spaces: local-to-global principle	2343
MICHAEL KAPOVICH, BERNHARD LEEB and JOAN PORTI	
Nearly geodesic immersions and domains of discontinuity	2391
COLIN DAVALO	
Graded character sheaves, HOMFLY-PT homology, and Hilbert schemes of points on \mathbb{C}^2	2463
QUOC P HO and PENGHUI LI	
Asymptotically Calabi metrics and weak Fano manifolds	2547
HANS-JOACHIM HEIN, SONG SUN, JEFFREY VIACLOVSKY and RUOBING ZHANG	
The distribution of critical graphs of Jenkins–Strebel differentials	2571
FRANCISCO ARANA-HERRERA and AARON CALDERON	
Tian’s stabilization problem for toric Fanos	2609
CHENZI JIN and YANIR A RUBINSTEIN	
On the logarithmic slice filtration	2653
FEDERICO BINDA, DOOSUNG PARK and PAUL ARNE ØSTVÆR	
Joyce structures on spaces of quadratic differentials	2695
TOM BRIDGELAND	
Big monodromy for higher Prym representations	2733
AARON LANDESMAN, DANIEL LITT and WILL SAWIN	